CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 114, No. 4, Pt. 1 October 2003

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 114, NO. 4, PT. 1, OCTOBER 2003

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 114, NO. 4, PT. 1, OCTOBER 2003

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 114, NO. 4, PT. 1, OCTOBER 2003

CONTENTS—*Continued from preceding page*

ACOUSTICAL NEWS—USA

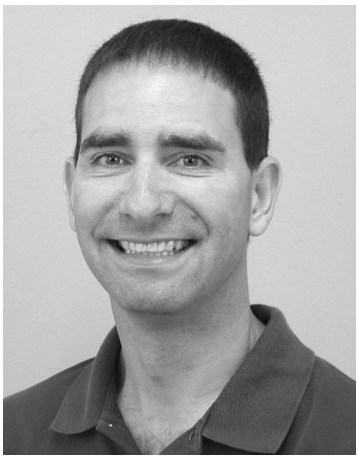
Elaine Moran
Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502

Editor's Note: *Readers of this Journal are encouraged to submit news items on awards, appointments, and other activities about themselves or their colleagues. Deadline dates for news items and notices are 2 months prior to publication.*

New Fellows of the Acoustical Society of America

Murray Campbell—For contributions to musical acoustics.

Peter Gerstoft—For contributions to geoacoustic inverse theory and methods.



Thomas J. Matula—For contributions to the understanding of nonlinear bubble dynamics.

Hee-Chun Song—For contributions to time-reversal acoustics and signal processing.

Report of the Auditor

Published herewith is a condensed version of our auditor's report for calendar year ended 31 December 2002.

Independent Auditors' Report

To the Executive Council
Acoustical Society of America

We have audited the accompanying statements of financial position of the Acoustical Society of America as of 31 December 2002 and 31 December 2001 and the related statements of activities and cash flows for the years then ended. These financial statements are the responsibility of the Society's management. Our responsibility is to express an opinion on the financial statements based on our audits.

We conducted our audits in accordance with auditing standards generally accepted in the United States of America. Those standards require that we plan and perform the audit to obtain reasonable assurance about whether the financial statements are free of material misstatement. An audit includes examining on a test basis, evidence supporting the amounts and disclosures in the financial statements. An audit also includes assessing the accounting principles used and significant estimates made by management, as well as evaluating the overall financial statement presentation. We believe that our audits provide a reasonable basis for our opinion.

In our opinion, the financial statements referred to above present fairly, in all material respects, the financial position of the Acoustical Society of America as of 31 December 2002 and 31 December 2001 and the changes in its net assets and its cash flows for the years then ended in conformity with accounting principles generally accepted in the United States of America.

CONROY, SMITH & CO.
Certified Public Accountants
4 June 2003
New York, NY

ACOUSTICAL SOCIETY OF AMERICA
STATEMENTS OF FINANCIAL POSITION
AS OF 31 DECEMBER 2002 AND 2001

	2002	2001
Assets:		
Cash and cash equivalents	$1,282,066	$1,326,755
Accounts receivable	303,948	136,551
Marketable securities, at market	5,171,458	5,917,761
Furniture, fixtures, and equipment—net	79,532	107,286
Other assets	362,264	306,765
Total assets	$7,199,268	$7,795,118
Liabilities:		
Accounts payable and accrued expenses	$ 313,297	$ 429,131
Deferred revenue	1,430,153	1,264,467
Total liabilities	$1,743,450	$1,693,598
Net assets:		
Unrestricted	$4,588,629	$5,055,355
Temporarily restricted	459,356	543,275
Permanently restricted	407,833	502,890
Total net assets	$5,455,818	$6,101,520
Total liabilities and net assets	$7,199,268	$7,795,118

ACOUSTICAL SOCIETY OF AMERICA
STATEMENTS OF ACTIVITIES
FOR THE YEARS ENDED 31 DECEMBER 2002 AND 2001

	2002	2001
Changes in Unrestricted Net Assets		
Revenues:		
Dues	$ 659,240	$ 637,075
Publishing—JASA	2,074,151	2,076,400
Standards	322,925	259,614
Spring Meeting	182,808	222,550
Fall Meeting	239,367	160,076
Other member services revenue	16,175	42,899
Other	93,071	98,619
Net assets released from restrictions	72,637	39,955
	$3,660,374	$3,537,188
Expenses:		
Publishing	$1,458,604	$1,563,726
Standards	460,782	514,061
Spring Meeting	232,650	266,513
Fall Meeting	289,208	277,514
Member Services	186,211	188,005
Other	471,317	373,350
Administration	462,961	432,758
	$3,561,733	$3,615,927
Net income from operations	$ 98,641	($ 78,739)
Nonoperating activities:		
Interest and dividends	$ 130,915	$ 232,550
Realized (loss) gain on investments	(89,461)	176,243
Unrealized (loss) on investments	(606,821)	(793,681)
	($ 565,367)	($ 384,888)
Decrease in unrestricted net assets	($ 466,726)	($ 463,627)
Changes in Temporarily Restricted Net Assets		
Contributions	$ 10,858	$ 25,000
Interest and dividends	14,082	23,824
Realized (loss) gain	(9,622)	18,055
Unrealized (loss)	(65,277)	(81,304)
Net assets released from restrictions	(33,960)	(7,955)
Decrease in temporarily restricted net assets	($ 83,919)	($ 22,380)
Changes in Permanently Restricted Net Assets		
Contributions	-	$ 361
Interest and dividends	$ 13,055	24,214
Realized (loss) gain	(8,921)	18,351
Unrealized (loss)	(60,514)	(82,642)
Net assets released from restrictions	(38,677)	(32,000)
Decrease in permanently restricted net assets	($ 95,057)	($ 71,716)
Decrease in net assets	($ 645,702)	($ 557,723)
Net assets at beginning of year	6,101,520	6,659,243
Net assets at end of year	$5,455,818	$6,101,520

ACOUSTICAL SOCIETY OF AMERICA
STATEMENTS OF CASH FLOWS
FOR THE YEARS ENDED 31 DECEMBER 2002 AND 2001

	Total All Funds 2002	2001
Operating Activities		
(Decrease) in net assets	($ 645,702)	($ 557,723)
Adjustments to reconcile (decrease) in net assets to net cash provided by operating activities:		
Depreciation	31,733	30,897
Loss on disposal of equipment and leasehold improvements	-	40,257
Net realized loss (gain) on marketable securities	108,004	(212,649)
Unrealized loss on marketable securities	732,612	957,627
Changes in operating assets and liabilities:		
(Increase) decrease in accounts receivable	(167,397)	187,859
(Increase) decrease in other assets	(55,499)	42,454
(Decrease) in accounts payable and accrued expenses	(115,834)	(24,413)
(Decrease) in deferred rent liability	-	(41,989)
Increase (decrease) in deferred revenue	165,686	(185,116)
Net cash flows provided by operating activities	$ 53,603	$ 237,204
Investing Activities		
Purchase of furniture, fixtures, and equipment	($ 3,979)	($ 73,148)
Proceeds from sale of marketable securities	1,681,182	4,270,712
Purchase of marketable securities	(1,775,495)	(4,409,384)
Net cash (used in) investing activities	($ 98,292)	($ 211,820)
(Decrease) increase in cash and cash equivalents	($ 44,689)	$ 25,384
Cash and cash equivalents, beginning of year	1,326,755	1,301,371
Cash and cash equivalents, end of year	$1,282,066	$1,326,755
Supplemental Information		
Cash paid for unrelated business income tax	$ 7,000	$ 4,906

ASA Awards Presented at the 2003 Intel International Science and Engineering Fair

ASA presented awards to four students with outstanding acoustics research projects at the 2003 Intel International Science and Engineering Fair (see Fig. 1). This year's ISEF was held 11–17 May in downtown Cleveland, Ohio, a stone's throw from the Rock and Roll Hall of Fame and the Great Lakes Science Center. Each of the participating 1099 projects and 1334 students came as a winner from one of the ISEF-affiliated science fairs in 38 countries. Approximately 30 projects had some connection to acoustics, spanning 6 of the fair's 14 subject categories, with most from Behavioral and Social Sciences, Engineering, Math, and Physics. Of these, 10 finalists were identified as offering the best contributions to acoustics research, and their works certainly are worthy of presentation at an ASA meeting. These students clearly have bright futures ahead of them!

From this excellent pool, the winner of the ASA Special Award of a $500 check was Sita Chandrika Palepu, for her project *Absolute vs Relative Pitch: It's Music to My Ears! Or Is It to My Mind?* A senior at James Madison High School in Vienna, Virginia, Ms. Palepu developed a method for assessing whether a subject used absolute or relative pitch-detection mechanisms in different contexts, and she was especially thorough in testing it. She reports that subjects possessing absolute pitch tend to have a decreased ability to use relative pitch mechanisms, even in contexts when relative pitch mechanisms would be normally favored.

Three Honorable Mentions were also awarded. Elisabeth Lee Crabtree, a junior at Bying High School in Ada, Oklahoma, presented *The McGurk Effect: Does Age Matter?* She explored the role that age and language experience played in subjects' interpolation between aural and visual cues to perception of consonants. Samrat A. Amin is a senior at Laramie Senior High School, in Laramie, Wyoming. In his project, *Sonoluminescence*, Mr. Amin attained sonoluminescence in an oscillating bubble in a fluid-filled flask, and explored the effects of dissolved air and of different fluids. Ashley Renee Woodall plays mallet percussion instruments and is in her freshman year at North Garland High School, in Garland, Texas. She was seeking a

FIG. 1. ASA award winners at 2003 Intel ISEF (1 to r): Ashley Renee Woodall, Samrat A. Amin, ASA judge Jack Mowry, Elisabeth Lee Crabtree, and Sita Chandrika Palepu.

replacement for the rosewood traditionally used in marimbas. In *Physics of Mallet Percussion*, Ms. Woodall constructed a series of marimba bar sets, tuning two resonances on each bar, then measuring decay times. [Editor's note: Ms. Woodall was the recipient of an ''Outstanding Acoustics Project Award'' from the ASA North Texas Regional Chapter at the 45th Dallas Morning News—Toyota Regional Science and Engineering Fair held in 2002 [J. Acoust. Soc. Am. **112**, 1 (2002)].

A veritable army of judges was required for the prodigious task of judging all the excellent projects at ISEF. More than 1000 judges reviewed projects for the Grand Awards, ranging from $500 up to $5,000, which were presented by ISEF. Another 500 judges represented the 70 individual organizations that sponsored Special Awards. Judging for the ASA were Osman Mawardi, Professor Emeritus of Case Western Reserve University, Jack Mowry, editor and publisher of *Sound and Vibration* magazine, David Proctor, Professor Emeritus of Baldwin-Wallace College, and Peter Hoekje, Associate Professor of Physics at Baldwin-Wallace.

Each of the four ASA winners was granted a 1-year subscription to JASA on CD-ROM. Each also received two books contributed by Jack Mowry and *Sound and Vibration*. We hope to hear much more of these talented scientists in the future. Take note of their names and keep an ear out for them!

PETER L. HOEKJE

USA Meetings Calendar

Listed below is a summary of meetings related to acoustics to be held in the U.S. in the near future. The month/year notation refers to the issue in which a complete meeting announcement appeared.

2003

10–14 Nov. 146th Meeting of the Acoustical Society of America, Austin, TX [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

2004

24–28 May 75th Anniversary Meeting (147th Meeting) of the Acoustical Society of America, New York, NY [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

3–7 August 8th International Conference on Music Perception and Cognition, Evanston, IL [School of Music, Northwestern University, Evanston, IL 60201; WWW: www.icmpc.org/conferences.html].

15–19 Nov. 148th Meeting of the Acoustical Society of America, San Diego, CA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

Cumulative Indexes to the *Journal of the Acoustical Society of America*

Ordering information: Orders must be paid by check or money order in U.S. funds drawn on a U.S. bank or by Mastercard, Visa, or American Express credit cards. Send orders to Circulation and Fulfillment Division, American Institute of Physics, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2270. Non-U.S. orders add $11 per index.

Some indexes are out of print as noted below.

Volumes 1–10, 1929–1938: JASA and Contemporary Literature, 1937–1939. Classified by subject and indexed by author. Pp. 131. Price: ASA members $5; nonmembers $10.

Volumes 11–20, 1939–1948: JASA, Contemporary Literature, and Patents. Classified by subject and indexed by author and inventor. Pp. 395. Out of print.

Volumes 21–30, 1949–1958: JASA, Contemporary Literature, and Patents. Classified by subject and indexed by author and inventor. Pp. 952. Price: ASA members $20; nonmembers $75.

Volumes 31–35, 1959–1963: JASA, Contemporary Literature, and Patents. Classified by subject and indexed by author and inventor. Pp. 1140. Price: ASA members $20; nonmembers $90.

Volumes 36–44, 1964–1968: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 485. Out of print.

Volumes 36–44, 1964–1968: Contemporary Literature. Classified by subject and indexed by author. Pp. 1060. Out of print.

Volumes 45–54, 1969–1973: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 540. Price: $20 (paperbound); ASA members $25 (clothbound); nonmembers $60 (clothbound).

Volumes 55–64, 1974–1978: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 816. Price: $20 (paperbound); ASA members $25 (clothbound); nonmembers $60 (clothbound).

Volumes 65–74, 1979–1983: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 624. Price: ASA members $25 (paperbound); nonmembers $75 (clothbound).

Volumes 75–84, 1984–1988: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 625. Price: ASA members $30 (paperbound); nonmembers $80 (clothbound).

Volumes 85–94, 1989–1993: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 736. Price: ASA members $30 (paperbound); nonmembers $80 (clothbound).

Volumes 95–104, 1994–1998: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 632. Price: ASA members $40 (paperbound); nonmembers $90 (clothbound).

Revision List

New Associates

Abraham, Bruce M., Applied Physical Sciences, 2 State St., Ste. 300, New London, CT 06371

Ahting, Peter A., Dynamic Test & Measurement Solutions, 726 Fox Creek Ln., Cincinnati, OH 45245

Beurskens, Kees L., Gebroeders van Doornelaan 3, Horst 5961 BA, The Netherlands

Catanzariti, Scott P., SARA, Inc., 6300 Gateway Dr., Cypress, CA 90630-4844

Chen, Julian C., Voice Systems, 8051 Congress Ave., Rm. 2079, Boca Raton, FL 33487

Choi, Jee Woong, Applied Physics Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698

Cosi, Piero, ISTC-CNR, SPFD, Via G. Anghinoni 10, Padova PD 35121, Italy

Costello, James B., 11081 East Raintree Dr., Scottsdale, AZ 85255

Cotter, Maxwell T., Mayo Hearing Clinic, 500 East Olive Ave., Ste. 500, Burbank, CA 91501

Coudron, Chad D., TRW Automotive Electronics, 5676 Industrial Park Rd., Winona, MN 55987

Council, Orin P., O.P. Council, MD, PC, 214 14th Ave., Ste. 104, Sidney, MT 59270

Denny, Stuart E., Veritas DGC, Inc., 10300 Town Park, Houston, TX 77072

DeVito, Felix A., 343 State Rd., Eliot, ME 03903

Deyesso, Joseph P., EMC Corp., 21 Coslin Dr., Southboro, MA 01772

Eddy, F. N., 40 Finn Rd., Harvard, MA 01451-1923

Ferat, Patrick A., Johns Hopkins Univ., Applied Physics Lab., 11100 Johns Hopkins Rd., Laurel, MD 20723

Fokin, Vladimir N., National Ctr. for Physical Acoustics, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677

Fokina, Margarita S., National Ctr. for Physical Acoustics, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677

Gainer, Dan L., Sound Progression, 9728 Dayton Ave. North, Seattle, WA 98103

Gerhard, David B., Computer Science, Univ. of Regina, 3737 Wascana Pkwy., Regina, SK S4S 02A, Canada

Haser, David C., RR #2, Box 2002, Carbondale, PA 18407
Holmes, Stephen D., Psychology, Univ. of Birmingham, Birmingham B15 2TT, U.K.
Houser, Dorian S., Biomimetica, 5750 Amaya Dr., #24, La Messa, CA 91942
Hsu, Chaur-Jian, Schlumberger-Doll Research, 36 Old Quarry Rd., Ridgefield, CT 06877
Hu, Fang, Inst. of Linguistics, Phonetics Lab., Chinese Academy of Social Sciences, No. 5 Jianguo, Mennei Dajie, Beijing 100732, P. R. China
Ikegami, Toru, Schlumberger-Doll Research, 36 Old Quarry Rd., Ridgefield, CT 06877
Johansson, Magnus, Peltor AB, Box 2341, Varnamo SE 331 02, Sweden
Jordan, Pedro M., Naval Research Lab., Code 7181, Bldg. 1005, Stennis Space Center, MS 39529
Kari, Leif, Aeronautical & Vehicle Eng., Royal Inst. of Technology, Teknikringen 8, Stockholm 10044, Sweden
Kimizuka, Ikuo, CE & RAS Design, IBM Japan, Ltd., 1623-14 Shimotsuruma, Yamamto-Shi, Kanagawa-Ken 242-8502, Japan
Lee, Sungbok, Linguistics, Univ. of Southern California, 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693
Lindwall, Dennis A., Naval Research Lab., SeaFl. Geosciences, Code 7432, Stennis Space Center, MS 39529
Marvit, Peter, Anatomy & Neurobiology, Univ. of Maryland at Baltimore, 685 West Baltimore St., HSF 222, Baltimore, MD 21201
Miyazaki, Hideo, 235 E 40th St., #16D, New York, NY 10016 Moulton, David R., Moulton Laboratories, 39 Ames Rd., Groton, MA 01450
Okuno, Hiroshi G., Graduate School of Informatics, Kyoto Univ., Yoshida-Honmachi, Sakyo, Kyoto 606-8501, Japan
Parent, Marc D., 4 Yorkshire Rd., Windham, NH 03087
Pascal, Jean-Claude, UMRS CNRS 6613, ENSIM, rue Aristote, Le Mans 72085, France
Peterson, Phillip P., 15 North Lincoln St., Redlands, CA 92374-4146
Porterfield, Randall L., Reactor Compartment Pump Design and Eng., Curtiss Wright Electro-Mechanical Corp., 1000 Cheswick Ave., Cheswick, PA 15024-1356
Pryshepa, John A., Technical Solutions, 18100 Railroad Ln., Gazelle, GA 96034
Sakurai, Masatsugu, Yoshimasa Electronic, Inc., Yoyogi 1-58-10, shibuya-ku, Tokyo 151-0053, Japan
Scribner, Kelvin, Sagetech Corp., 3901 Twilight Ave., Enid, OK 73703
Seo, Jong-Soo, Samsung Heavy Industries Co., Ltd., 103-28 Daeduk Science Town, Munji-Dong, Yooseong-Gu, Daejeon 305-380, South Korea
Skinner, Christian M., Ingersoll-Rand Company, Business Development Ctr., 1467 Route 31 South, P.O. Box 970, Annandale, NJ 08801
Stevens, Alexander A., Oregon Health & Science Univ., Psychiatry, 3181 SW Sam Jackson Park Rd., Mail Code CR139, Portland, OR 97239-3098
Tallapragada, Bhanuprakash, Plot 96, Sector 8, MVP Colony, Vishakapatnam 530017, India
Vento, Barbara A., Communication Sciences & Disorders, Univ. of Pittsburgh, 4033 Forbes Tower, Pittsburgh, PA 15260
Zhang, Xiaoming, Dept. of Physiology and Biophysics, Mayo Clinic Rochester, 200 First St. SW, Rochester, MN 55905

New Students

Bachand, Corey L., 26 Nye St., New Bedford, MA 02746
Billon, Alexis J., LEPTAB, Univ. de la Rochelle, Av. M. Crepeau Cedex 1, La Rochelle 17042, France
Blum, Nathan R., 2803 La Ventana, San Clemente, CA 92672
Breznik, Evelyn A., 433 Gillett Ave., Waukegan, IL 60085
Carballeira, Andrew C., Cavanaugh Tocci Associates, Inc., 327F Boston Post Rd., Sudbury, MA 01776
Damljanovic, Vesna, Theoretical and Applied Mechanics, Univ. of Illinois at Urbana-Champaign, 104 South Wright St., Urbana, IL 61801
Einarsson, Thorvaldur, 8500 New Hampshire Ave., Apt. 125, Silver Spring, MD 20903
Fan, Zhiliang, 1150 Collier Rd. NW, Apt. H6, Atlanta, GA 30318
Finney, Nathaniel R., Marstalsgade 21, 5, th., KBH 2100, Denmark
Fujimoto, Antonio K., Rua Ana Aurora Lisboa 104, Sao Paulo 05374-080, Brazil
Gao, Wen, 802 Conway Dr., Apt. 201, Williamsburg, VA 23185
Gorr, John W., Audio Arts and Acoustics, Columbia College Chicago, 676 North LaSalle, Ste. 300, Chicago, IL 60610
Goupell, Matthew, 4484 Karen Ann Dr., Apt. G203, Okemos, MI 48864
Gregoire, Jerry, ECE, MSU-B, Bozeman, MT 59718
Hawbaker, Joel, 324 Park St., Bensenville, IL 60106
Hawkins, Nathan C., 2641 East Brooks St., Gilbert, AZ 85296
Howell, Carl A., 17125 South Dooneen Ave., Tinley Park, IL 60477
Jeung, Changmo, Beodeune Apt. 112-1002 Taepyeong 2, Jung-gu, Daejon 301-780, Korea
Johnson, Christopher T., 3300 Poinsett Highway, #29471, Greenville, SC 39613
Johnson, Keith E., 801 East 10th St., Apt. 6, Tucson, AZ 85719
Jones, Tyrone L., 1798 North Atherton St., Apt. 7, State College, PA 16803
Kim, Bumjun, 2909 SW 13th St., #37, Gainesville, FL 32608
Lang, Daniel A., 4 Suellen Ct., Islip, NY 11751-4107
Lawrence, Cecile A., Dartmouth College, 6018 McNutt Hall, Hanover, NH 03755
LeDantec, Nicolas, Scripps Inst. of Oceanography, Marine Physical Lab., 9500 Gilman Dr., Code 0213, La Jolla, CA 92093-0213
Li, Guangyan, Dept. of Physics, Univ. of Mississippi, University, MS 38677
Light, Jason R., P.O. Box 15639, Boston, MA 02215
Lin, Hejie, 1502 Grandville, Apt. 305, Pontiac, MI 48340
Lin, Pei-Yu, Inst. of Biomedical Engineering, Yang-Ming Univ., 155, Li-Nong Steet, Sec. 2, Taipei 112, Taiwan
Linscomb, Sommer H., 2403 Liberty St., Beaumont, TX 77702
Liu, Yunbo, Mechanical Engineering, Duke Univ., 133 Hudson Hall, P.O. Box 90300, Durham, NC 27708
McClanahan, Richard D., 2600 East Idaho, #176-L, Las Cruces, NM 88011
McGuffey, Christopher S., 1192 Wellington, Duncanville, TX 75137
Mace, Emily K., MSC 2528, 623 Holly St., Nampa, ID 83686
Myers, Melodie M., 10 North Main St., #201A, Memphis, TN 38103
Namasivayam, Aravind K., Dept. of Speech Language Pathology, Univ. of Toronto, 500 University Ave., 10th Fl., Toronto, ON M5G 1V7, Canada
Nandur, Vuday, 2917 SW 13th St., Apt. 86, Gainesville, FL 32608
Ngan, Kwok Hung, A704, Tower A, HKUST, Clear Water Bay, Kowloon, Hong Kong SAR, 852, China
O'Donnell, Matthew L., Physics and Astronomy, CSULB, 1250 Bellflower Boulevard, Long Beach, CA 90840-0106
Olsen, Caroledith, 3346 Lyndale Ave. North, Minneapolis, MN 55412-2554
Roufes, Timothy, 7686 VanBuren St., Fridley, MN 55432
Roy, Johanna-Pascale, Inst. de Phonetique, Univ. Marc Bloch (UFR Lettres), 22 rue Rene Descartes, Strasbourg 67084, France
Saintval, Wendy, 1510 15th St., 2nd Fl., Troy, NY 12180
Schad, Martha R., P4 Regency Pl., Lawrence, KS 66049
Scimone, Lisa M., 19 Riverview Ave., Danvers, MA 01923
Scott, James N., 192 Macaw Dr., Spearsville, LA 71277
Stone, Michael E., Mechanical Engineering, Univ. of Texas at Austin, 1 Univ. Station C2200, Austin, TX 78712
Surendran, Dinoj R., Computer Science, Univ. of Chicago, 1100 East 58th St., Chicago, IL 60615
Szamatowicz, Brandon J., 2501 Kelly Pl., Dyer, IN 46311
Teddy, Sintiani D., Blk. 640 Jurong West St., 11-16, 640640 Singapore
Ulker Karbeyaz, Basak, Electrical and Computer Engineering, Northeastern Univ., Ste. 302, Stearns Ctr., 360 Huntington Ave., Boston, MA 02115-5000
van Wassenhove, Virginie, Cognitive Neuroscience of Language Lab., Univ. of Maryland, 1401 Marie Mount Hall, College Park, MD 20742
Viss, Barret, 719 East Main St., Ripon, CA 95366
Walker, Christopher D., Signal Research Ctr., Univ. of Southern Mississippi, Box 5173, Hattiesburg, MS 39406
Wang, Zhenhai, ECE Dept., Univ. of Massachusetts Dartmouth, 285 Old Westport Rd., North Dartmouth, MA 02740
Watts, Matthew K., 7334 Major Ave., Norfolk, VA 23505
Wienke, Eric T., Human Factors/Psychology, Univ. of South Dakota, 414 East Clark St., Vermillion, SD 57069
Wilson, Joshua D., Ocean Engineering, Massachusetts Inst. of Technology, 77 Massachusetts Ave., 5-435, Cambridge, MA 02139-4307

New Electronic Associates

Arrabito, Robert, Defense Research and Development Canada, Human Factors Res. and Eng., 1133 Sheppard Ave. West, Toronto, ON M3M 3B9, Canada

Douligeris, Konstantinos D., El. Venizelou 13, N. Faliron, Piraeuc 18547, Greece
Fripp, Deborah, 3826 Cemetery Hill Rd., Carrollton, TX 75007
Fujii, Kenji, Yoshimasa Electronic, Daiichi Nishiwaki Bldg., 1-58-10 Yoyogi, Shibuya, Tokyo 151-0053, Japan
Hawkins, Timothy, Applied Research Labs., Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029
Hazelrig, Donald T., Timescape Acoustics, 152 Dexter Circle, Madison, AL 35757-8005
Koessel, Manfred, Zoological Inst., Univ. of Frankfurt, Siesmayerstr. 70, Frankfurt 60323, Germany
Lessley, Ted D., 1719 East Vine Ave., West Covina, CA 91791
Mahon, Mark P., Pennsylvania State Univ./ARL, P.O Box 30, State College, PA 16804-0030
Miller, Maxwell H., 9902 Evergreen Ave., Columbia, MD 21046
Nieukirk, Sharon L., NOAA/CIMRS, 2115 SE OSU Dr., Newport, OR 97365
Paeng, Dong-Guk, Faculty of Marine Industrial Eng., Cheju National Univ., 1 Ara 1-dong, Jeju, Jujudo 690-756, Korea
Poinsett, David, R. M. Young Company, 2801 Aero-Park Dr., Traverse City, MI 49686
Qiang, Lin, Univ. of Bath, Claverton Down, Bath BA2 7AY, England
Rust, Wallace R., 523 Britton Rd., Rochester, NY 14616-3219
Stack, Janet W., 385 Claymont Dr., Earlysville, VA 22936
Wong, Yuen Kwan, Rm. 930, Oi Tak House, Yau Oi Estate, Tuen Mun, Hong Kong

New Corresponding Electronic Associates

Agarwal, Mahesh C., National Instruments, 113 Prestige Pinnacle, 7th Block, Koramangala, Bangalore, Karnataka 560095, India
Hyun, Gwoanchul, 1704-Ho, 103-Dong, Hakyeuul-Apart, Nowon-Ku, Seoul 139-734, Rep. of Korea
Paeng, Dong-Guk, Faculty of Marine Industrial Eng., Cheju National Univ., 1 Ara 1-dong, Jeju, Jujudo 690-756, Korea

Members Elected Fellows

G. L. Augspurger, M. D. Campbell, K. A. Cunefare, G. B. Deane, D. R. Dowling, T. J. Foulkes, R. R. Freiheit, P. Gerstoft, J. J. Guinan, Jr., J. M. Harrison, T. J. Matula, J. H. Miller, J. A. Nystuen, M. D. Richardson, J. O. Smith III, H. C. Song

Associates Elected Fellows

D. Chu, S. G. Kargl, J. F. Lingevitch, E. K. Westwood

Students to Associates

J. Braasch, K. Lee, M. M. Matlo, B. E. Noel, G. D. Yates

Associate to Student

M. Gordon

Associates to Electronic Associates

J. C. Bachenko, N. M. Gage, J. S. Madison, D. K. Peterson, G. A. Torio

Students to Electronic Associates

M. R. Kasam, S. C. Renaud

Associate to Corresponding Electronic Associate

M. Guirao

Resigned

W. L. Clearwaters, M. Haggard—*Fellows*
R. Wayne Gatehouse—*Member*
B. Nita—*Associate*
S. Alves, M. H. Thomas—*Students*

Deceased

W. C. Meecham, S. R. Silverman—*Fellows*
K. L. Landahl—*Associate*

Fellows	906
Members	2479
Associates	2688
Students	916
Electronic Associates	199
	7188

ACOUSTICAL NEWS—INTERNATIONAL

Walter G. Mayer
Physics Department, Georgetown University, Washington, DC 20057

International Meetings Calendar

Below are announcements of meetings and conferences to be held abroad. Entries preceded by an * are new or updated listings.

October 2003

15–17 **34th Spanish Congress on Acoustics**, Bilbao, Spain. (Fax: +34 91 411 7651; Web: www.ia.csic.es/sea/index.html)

15–17 **Acoustics Week in Canada**, Edmonton, AB, Canada. (Fax: +1 780 414 6376; Web: caa-aca.ca/edmonton-2003.html)

30–31 **Autumn Meeting of the Swiss Acoustical Society**, Basel, Switzerland. (Fax: +41 419 62 13; Web: www.sga-ssa.ch)

November 2003

5–6 **Institute of Acoustics (UK) Autumn Conference**, Oxford, UK. (Fax: +44 1727 850553; Web: www.ioa.org.uk)

7–9 **Reproduced Sound**, Oxford, UK. (Fax: +44 1727 850553; Web: www.ioa.org.uk)

12–14 **Tenth Asia-Pacific Vibration Conference (APVC 2003)**, Gold Coast, Queensland, Australia. (Web: www.apvc.net)

26–28 **Tenth Mexican International Congress on Acoustics**, Puebla, Pua, Mexico. (Instituto Mexicano de Acustica, P.O. Box 12-1022, Col. Narvarte 03001 Mexico, D.F., Mexico; Fax: +52 555523 4742; e-mail: sberista@hotmail.com)

December 2003

10–12 **3rd International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications**, Firenze, Italy. (Fax: +39 55 479 6767; Web: www.maveba.org)

March 2004

17–19 **Spring Meeting of the Acoustical Society of Japan**, Atsugi, Japan. (Fax: +81 3 5256 1022; Web: www.soc.nii.ac.jp/asj/index-e.html)

22–25 **Joint Congress of the French and German Acoustical Societies (SFA-DEGA)**, Strasbourg, France. (Fax: +33 1 48 88 90 60; Web: www.sfa.asso.fr/cfa-daga2004)

23–26 ***International Conference: Speech Prosody 2004**, Nara, Japan. (K. Hirose, School of Frontier Sciences, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan; Fax: +81 3 5841 6648; Web: www.gavo.t.u-tokyo.ac.jp/sp2004)

31–3 **International Symposium on Musical Acoustics (ISMA2004)**, Nara, Japan. (Fax: +81 77 495 2647; Web: www2.crl.go.jp/jt/a132/isma2004)

April 2004

5–9 **18th International Congress on Acoustics (ICA2004)**, Kyoto, Japan. (Fax: +81 66 879 8025; Web: www.ica2004.or.jp)

11–13 **International Symposium on Room Acoustics (ICA2004 Satellite Meeting)**, Hyogo, Japan. (Fax: +81 78 803 6043; Web: rad04.iis.u-tokyo.ac.jp)

May 2004

8–10 **116th AES Convention**, Berlin, Germany. (Web: aes.org/events/116)

17–21 **International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004)**, Montréal, Canada. (Web: www.icassp2004.com)

June 2004

8–10 **Joint Baltic-Nordic Acoustical Meeting**, Mariehamn, Åland, Finland. (Fax: +358 09 460 224; e-mail: asf@acoustics.hut.fi)

July 2004

5–8 **7th European Conference on Underwater Acoustics (ECUA 2004)**, Delft, The Netherlands. (Fax: +31 70 322 9901; Web: www.ecua2004.tno.nl)

11–16 **12th International Symposium on Acoustic Remote Sensing (ISARS)**, Cambridge, UK. (Fax: +44 161 295 3815; Web: www.isars.org.uk)

August 2004

23–27 **2004 IEEE International Ultrasonics, Ferroelectrics, and Frequency Control 50th Anniversary Conference**, Montréal, Canada. (Fax:+1 978 927 4099; Web: www.ieee-uffc.org/index2-asp)

22–25 **Inter-noise 2004**, Prague, Czech Republic. (Web: www.internoise2004.cz)

September 2004

13–17 **4th Iberoamerican Congress on Acoustics, 4th Iberian Congress on Acoustics, 35th Spanish Congress on Acoustics**, Guimarães, Portugal. (Fax: +351 21 844 3028; Web: www.spacustica.pt/novidades.htm)

November 2004

4–5 **Autumn Meeting of the Swiss Acoustical Society**, Rapperswil, Switzerland. (Fax: +41 419 62 13; Web: www.sga-ssa.ch)

August 2005

7–10 **Inter-Noise**, Rio de Janeiro, Brazil. (Details to be announced later)

28–2 ***EAA Forum Acusticum Budapest 2005**, Budapest, Hungary. (I. Bába, OPAKFI, Fö u. 68, Budapest 1027, Hungary; Fax: +36 1 202 0452; Web: www.fa2005.org)

September 2007

2–7 ***19th International Congress on Acoustics (ICA2007)**, Madrid, Spain. (SEA, Serrano 144, 28006 Madrid, Spain; Web: www.ia.csic.es/sea/index.html)

Preliminary Announcement

June 2008

23–27 **Joint Meeting of European Acoustical Association (EAA), Acoustical Society of America (ASA), and Acoustical Society of France (SFA)**, Paris, France. (Details to be announced later)

ADVANCED-DEGREE DISSERTATIONS IN ACOUSTICS

Editor's Note: *Abstracts of Doctoral and Master's theses will be welcomed at all times. Please note that they must be limited to 200 words, must include the appropriate PACS classification numbers, and formatted as shown below. If sent by postal mail, note that they must be double spaced. The address for obtaining a copy of the thesis is helpful. Submit abstracts to: Acoustical Society of America, Thesis Abstracts, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, e-mail: asa@aip.org*

Development of an acoustic levitation technique to obtain foam material properties [43.25.Uv, 43.25.Yw, 43.35.Yb]—Li Liu, *Department of Aerospace and Mechanical Engineering, Boston University, Boston, Massachusetts, January 2003 (MS).* Aqueous foam is an impermanent form of matter in which a kind of gas, often air, is dispersed as an agglomeration of bubbles that are separated from each other by films of liquid. Foams are of tremendous economical importance in industry. Foam material properties are sensitive functions of the void fraction. A ''wet foam'' is a bubbly liquid that cannot support shearing motion; inside the wet foam the individual bubbles are free to move around. A ''transitional'' or ''critical foam'' is composed of bubbles whose dynamics are strongly interacting and whose surfaces may be in mechanical contact with each other. Finally, a ''dry foam'' is composed of bubbles who have a fixed position in a lattice for low to moderate straining rates. An acoustic levitation technique is developed which provides a noncontact means of estimating the properties of the foam by acoustically levitating aqueous foam drops and exciting their spheroidal modes oscillation. Assuming linear oscillation of foam drops, experimental data for frequency and damping show good agreement with a bubble dynamics-based theoretical model.

Thesis advisor: R. Glynn Holt

Copies of this thesis may be obtained by contacting the advisor, Glynn Holt, Dept. of Aerospace and Mechanical Engineering, Boston University, 110 Cummington St., Boston, MA 02215. E-mail address: rgholt@bu.edu

Investigation of bubble dynamics and heating during focused ultrasound insonation in tissue-mimicking materials [43.25.Yw, 43.35.Wa, 43.80.Gx, 43.80.Sh]—Xinmai Yang, *Department of Aerospace and Mechanical Engineering, Boston University, Boston, Massachusetts, May 2003 (Ph.D.).* The deposition of ultrasonic energy in tissue can cause tissue damage due to local heating. For pressures above a critical threshold, cavitation will occur in tissue and bubbles will be created. These oscillating bubbles can induce a much larger thermal energy deposition in the local region. The present work is an attempt to control and utilize this bubble-enhanced heating. First, by applying appropriate bubble dynamic models, limits on the asymptotic bubble size distribution are obtained for different driving pressures at 1 MHz. The size distributions are bounded by two thresholds: the bubble shape instability threshold and the rectified diffusion threshold. The growth rate of bubbles in this region is also given, and the resulting time evolution of the heating in a given insonation scenario is modeled. Experimental results have been obtained to investigate the bubble-enhanced heating in an agar and graphite based tissue-mimicking material. By fitting appropriate bubble densities in the ultrasound field, the peak temperature changes observed in experiments are simulated. Finally, a simple bubbly liquid model is presented to estimate shielding effects which may be important even for low void fraction during high intensity focused ultrasound (HIFU) treatment.

Thesis advisor: R. Glynn Holt

Copies of this thesis may be obtained by contacting the advisor, Glynn Holt, Dept. of Aerospace and Mechanical Engineering, Boston University, 110 Cummington St., Boston, MA 02215. E-mail address: rgholt@bu.edu

Sonoluminescence in varying acceleration environments [43.25.Uv, 43.25.Yw]—Sean C. Wyatt, *Department of Aerospace and Mechanical Engineering, Boston University, Boston, Massachusetts, June 2000 (MS).* In order to prepare for experiments in variable-g environments, some of the multiple effects caused by the time-varying acceleration of acoustic resonators used in bubble levitation experiments are considered. The coupled effects of the induced changes in ambient pressure (due to a changing hydrostatic head) and bubble position (due to a change in buoyant body force) were modeled. Changing the ambient pressure, while holding the acoustic pressure amplitude constant, causes changes in the radial bubble response and diffusive equilibrium requirements. Changing the bubble levitation position causes both the local acoustic pressure amplitude and gradient to change, which will again impact bubble response. If the bubble remains in a stable diffusive equilibrium, both of these effects will force the bubble's equilibrium radius to change. By using an empirical relation for emitted sonoluminescence intensity versus bubble response, the variation of emitted light intensity as a function of the changing ambient acceleration can be predicted. The predicted results are shown to agree quantitatively with existing experimental data from other research groups. An experiment was designed and built to fly onboard a KC-135 aircraft that has been reinforced for parabolic flight maneuvers to simulate micro and hyper gravity acceleration.

Thesis advisor: R. Glynn Holt

Copies of this thesis may be obtained by contacting the advisor, Glynn Holt, Dept. of Aerospace and Mechanical Engineering, Boston University, 110 Cummington St., Boston, MA 02215. E-mail address: rgholt@bu.edu

Design, fabrication and use of a KC-135 experiment for studying sonoluminescence [43.25.Uv, 43.25.Yw]—Charles Robert Thomas, *Department of Aerospace and Mechanical Engineering, Boston University, Boston, Massachusetts, April 2003 (MS).* The effects of many external parameters on single-bubble sonoluminescence (SBSL) have been studied since its discovery 13 years ago in 1990. These studies include the effects due to host water temperature, dissolved gas, ambient pressure, and even magnetic fields. Because the stability of the bubble depends in part on gravity, through its buoyancy, it is natural to study SBSL in microgravity. This is the subject of this thesis: the design and fabrication of an experiment to study the effects of gravity on SBSL for use on the NASA KC-135 aircraft. The KC-135 aircraft flies in a parabolic trajectory to simulate weightlessness for about 20 s per parabolic maneuver, performing roughly 40 parabolas per flight. During those flights, the bubble's size, position in the imposed sound field, and emitted light intensity were measured. The apparatus and techniques used to make these measurements, along with the constraints imposed by the unique environment afforded by experimentation aboard the KC-135, are discussed in detail. Preliminary results are reported, and a full uncertainty analysis is presented for each measurement.

Thesis advisor: R. Glynn Holt

Copies of this thesis may be obtained by contacting the advisor, Glynn Holt, Dept. of Aerospace and Mechanical Engineering, Boston University, 110 Cummington St., Boston, MA 02215. E-mail address: rgholt@bu.edu

BOOK REVIEWS

P. L. Marston
Physics Department, Washington State University, Pullman, Washington 99164

These reviews of books and other forms of information express the opinions of the individual reviewers and are not necessarily endorsed by the Editorial Board of this Journal.

Editorial Policy: *If there is a negative review, the author of the book will be given a chance to respond to the review in this section of the Journal and the reviewer will be allowed to respond to the author's comments. [See "Book Reviews Editor's Note," J. Acoust. Soc. Am.* ***81****, 1651 (May 1987).]*

Mechanical Shock, Vol. 2 of Mechanical Vibration & Shock, Fatigue Damage, Vol. 4 of Mechanical Vibration & Shock

Christian Lalanne

Taylor & Francis, New York, 2002.
xv+320 pp. Price: $150 (hardcover) ISBN: 1560329866 (Vol. 2).
ix+350 pp. Price: $150 (hardcover) ISBN: 1560329890 (Vol. 4).

The author's approach is well stated on the books' dust jackets: ''This ...series of five volumes has been written with both the professional engineer and the academic in mind. Christian Lalanne explores...vibration and shock ...from both theoretical and refreshingly practical standpoints.'' The other three volumes of the series, which are not reviewed here, deal with Sinusoidal Vibration, Random Vibration, and Specification Development. Although the subject matters of all five volumes clearly are interrelated, each volume is complete on its own.

Volume 2, *Mechanical Shock*, begins in Chap. 1 with a series of definitions and a discussion of Fourier analysis, including digital calculation of the Fourier transforms. Chapter 2 deals with shock response spectra in general and presents dimensionless spectra corresponding to some idealized pulses. Chapter 3, which describes the characteristics of shock spectra, provides practical insights into the parameter dependences, discusses the relation of the shock spectrum to the Fourier spectrum, and provides guidance in the measurement and calculation of spectra. Chapter 4 addresses development of shock test specifications, including representation of measured data by idealized shocks and the related difficulties and consequences. Chapter 5 presents the characteristics of common idealized shock pulses, and Chap. 6 describes commonly used shock testing machines, together with their capabilities and limitations. Chapter 7 deals with the generation of shocks by means of shakers, Chap. 8 briefly discusses means for simulating pyrotechnic shocks, and Chap. 9 describes in considerable detail the control of shakers for shock testing. Two brief appendices summarize similitude in mechanics and the history of mechanical shock testing.

The first chapter of Vol. 4, *Fatigue Damage*, presents an extensive summary of the responses of a linear single-degree-of-freedom system to random vibration, and the second chapter proceeds with discussion of the characteristics of these responses that are relevant to fatigue estimation. Chapter 3 addresses the various material fatigue hypotheses, their uses, and their representations, and Chap. 4 deals similarly with the accumulation of fatigue damage. Chapter 5 presents and evaluates various counting methods for analyzing random time-histories. Chapter 6 discusses the calculation of fatigue damage induced in a single-degree-of-freedom system by a number of approaches and presents some comparisons of theoretical to experimental data. Chapter 7 presents a number of approaches to calculation of the standard deviation of fatigue damage and Chap. 8 discusses other fatigue damage calculations. An appendix deals rather comprehensively with the gamma function and the incomplete gamma function.

Both books contain extensive bibliographies. The fact that these contain no references that were published later than 1990 may be due to the relative maturity of the fields, resulting in limited research in these areas in recent years.

It is unfortunate that poor proofreading, some imprecise language, some awkward English, and illustrations of rather poor quality mar the treasure trove of information contained in these books. Nevertheless, readers with some background knowledge should be able to understand the author's intents fully and will find these books to be extremely valuable.

ERIC E. UNGAR
Acentech, Incorporated
33 Moulton Street
Cambridge, Massachusetts 02138-1118

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

The purpose of these acoustical patent reviews is to provide enough information for a Journal reader to decide whether to seek more information from the patent itself. Any opinions expressed here are those of reviewers as individuals and are not legal opinions. Printed copies of United States Patents may be ordered at $3.00 each from the Commissioner of Patents and Trademarks, Washington, DC 20231. Patents are available via the Internet at http://www.uspto.gov.

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
HASSAN NAMARVAR, *Department of BioMed Engineering, University of Southern California, Los Angeles, California 90089*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
CARL J. ROSENBERG, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, Univ. of Rochester, Rochester, New York 14627*

6,543,177

43.30.Gv ACOUSTICALLY VISIBLE FISHING NET

Donald P. King and Norman L. Holy, assignors to Atlantic Gillnet Supply, Incorporated
8 April 2003 (Class 43/4.5); filed 26 March 1999

Thermoplastic resin compositions are doped with additives such as barium sulfate (3% to 30% by weight) to produce a filament with enhanced acoustic reflectivity. This is then used to fabricate fishing nets that possess higher acoustic reflectivity which may reduce the unwanted capture of echolocating sea mammals such as dolphins, porpoises, and whales.—WT

6,552,336

43.30.Pc NON-INVASIVE, OPTO-ACOUSTIC WATER CURRENT MEASUREMENT SYSTEM AND METHOD

Jack Lloyd and Jeff Rish, assignors to The United States of America as represented by the Secretary of the Navy
22 April 2003 (Class 250/306); filed 11 October 2000

A system for measuring water current is described. A laser beam, modulated at some convenient frequency in the acoustic spectrum, is aimed obliquely at the surface of the water. This incident electromagnetic energy generates an acoustic wave in the water at the modulating frequency. However, this frequency is then Doppler shifted because of the current. A second laser beam, with a different wavelength than the initial beam, interrogates the water's surface at a number of different spots on the surface that surround the point of initial incidence. This second beam becomes frequency modulated at the Doppler-shifted frequency and therefore the frequency content of the echo returns of this beam is indicative of both the magnitude and direction of the water current.—WT

6,552,964

43.30.Tg STEERABLE BEAMFORMING SYSTEM

Alice M. Chiang and Steven R. Broadstone, assignors to TeraTech Corporation
22 April 2003 (Class 367/138); filed 6 April 2001

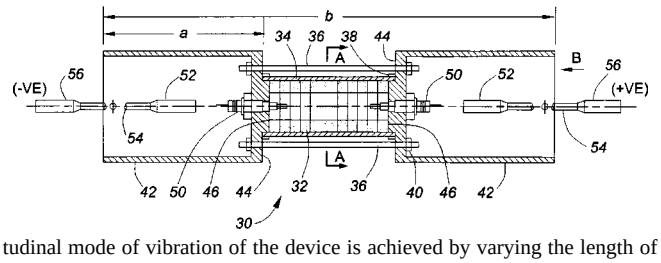
This patent discusses the concept of a small, diver-operated, 2-D, sparsely populated, sonar receiving array and associated beamformer for forming multiple, steered, narrow beams (less than 0.3° beamwidth) with relatively low side lobes.—WT

6,545,949

43.30.Yj AXIAL DRIVE RESONANT PIPE PROJECTOR (ADRPP)

J. Barrie Franklin, assignor to Her Majesty the Queen in right of Canada, as represented by the Minister of National Defence
8 April 2003 (Class 367/176); filed 21 September 2001

A low-frequency projector comprises a central motor section consisting of a coaxially aligned set of piezoceramic discs **32** waterproofed by a thin neoprene boot **34**, a set of prestress rods **36** which apply static compression to the stack of discs, and end caps **44** which are actually free-flooding cylindrical cups that function as waveguides. Tuning of the longi-



tudinal mode of vibration of the device is achieved by varying the length of the waveguides, the length of the motor section, and other dimensions and material parameters. To lower this resonant frequency even further, the waveguide cavities could be filled with a fluid having a lower sound speed than water and then sealed by a thin flexible membrane.—WT

6,555,947

43.30.Yj PRESSURE-BALANCED UNDERWATER ACOUSTIC TRANSDUCER

Kyung Soo Bahk *et al.*, assignors to Korea Ocean Research and Development Institute
29 April 2003 (Class 310/337); filed in the Republic of Korea 23 May 2001

An underwater acoustic transducer is pressure-compensated by allowing a free-flooding layer of water at the back face of the piezoceramic motor section. The thickness of this layer is adjustable so that the phase of the rear radiation, when subsequently reflected forward off a reflector unit in the back of the transducer's housing, matches that of the direct radiation, thereby resulting in constructive interference of these two signals in the forward direction, at one operating frequency.—WT

6,556,510

43.30.Yj INTEGRATED MARINE SEISMIC SOURCE AND METHOD

Loran D. Ambs, assignor to WesternGeco, L.L.C.
29 April 2003 (Class 367/20); filed 29 November 2000

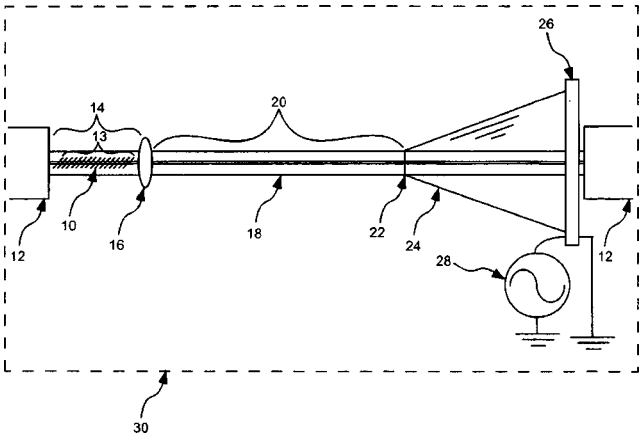
Seismic streamers are discussed that contain a number of sources so that the acoustic output of the line array can be steered in various directions or so that the temporal characteristics of the radiation can be varied. The preferred sources consist of slotted piezoceramic cylinders. Multiple hydrophones are also positioned along the streamers.—WT

6,556,729

43.35.Sx ACOUSTO-OPTIC BANDPASS FILTER

Duane Anthony Satorius, assignor to The United States of America as represented by the National Security Agency
29 April 2003 (Class 385/7); filed 27 July 2001

In this tunable acousto-optic filter, all light within the optical bandwidth of operation is first coupled from the core mode of an optical fiber to a specific cladding mode by a long-period photo-induced grating **10**. These cladding mode light waves then proceed to an acousto-optic interaction



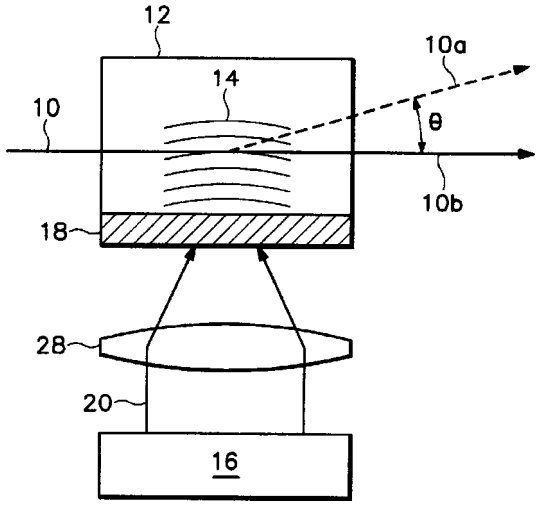
region **20** where selected flexure wave frequencies, induced by an acoustic wave amplifier **24**, recouple selected bands of wavelengths back into the core mode. An acoustic absorber **16** limits acoustic interaction in the interaction region through isolation from the broadband coupler.—DRR

6,560,005

43.35.Sx ACOUSTO-OPTIC DEVICES

Timothy N. Thomas, assignor to TKD, Incorporated
6 May 2003 (Class 359/305); filed 7 August 2001

This acousto-optic device uses a pulsed light source to control a steady light beam via acoustic waves. The beam **10** from a steady light source is directed through a transparent material **12**. The light beam **20** from a pulsed source is directed onto a thin layer of light-absorbing material **18** which is attached to the transparent material. The pulsed beam produces acoustic



waves within the light-absorbing material, which then propagate into the transparent material and generate a refractive index wave **14**, which behaves much like a sinusoidal grating, altering the path of the steady light beam. The diffracted angle is proportional to the acoustic frequency. A number of applications are suggested in the well-written patent.—DRR

6,559,389

43.35.Yb HIGH-DENSITY CABLE AND METHOD THEREFOR

William Paul Kornrumpf *et al.*, assignors to General Electric Company
6 May 2003 (Class 174/254); filed 25 August 2000

This is a cable for transmitting ultrasound signals from an ultrasonic probe to multiplexing circuitry during a medical ultrasound procedure. The cable contains one or more flexible circuits arranged within flexible sheaths. Each flexible circuit features an elongated flexible substrate with oppositely disposed surfaces and multiple conductors on at least one of these surfaces.—DRR

6,529,869

43.38.Hz PROCESS AND ELECTRIC APPLIANCE FOR OPTIMIZING ACOUSTIC SIGNAL RECEPTION

Joachim Wietzke and Rainer Cornelius, assignors to Robert Bosch GmbH
4 March 2003 (Class 704/240); filed 1 June 2000

Recent patents suggest that electronically steered microphone arrays can be used advantageously in teleconference systems, hearing aids, hands-free telephones, and communications centers. This invention is concerned mainly with processing the signals from such an array. A speech signal is first identified, one of several possible locations is determined, and the array pickup pattern is then steered to that location.—GLA

6,535,610

43.38.Hz DIRECTIONAL MICROPHONE UTILIZING SPACED APART OMNI-DIRECTIONAL MICROPHONES

Brett B. Stewart, assignor to Morgan Stanley & Company Incorporated
18 March 2003 (Class 381/92); filed 7 February 1996

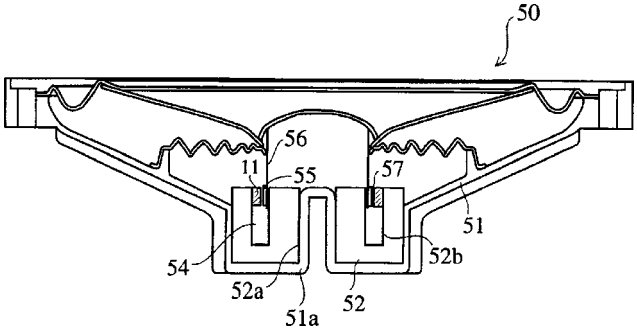
The current buzzword in audio pickup for teleconferencing is ''beamforming.'' If mere beamforming is not up to the task, then ''adaptive beamforming'' will surely do the trick. This latest invention mounts a few omnidirectional microphones around the periphery of a video display, digitizes their outputs, and then processes the signals through tapped delay lines under computer control.—GLA

6,529,107

43.38.Ja SPEAKER COMPRISING RING MAGNET

Motoharu Shimizu and Hiroyuki Daichoh, assignors to Hitachi Metals Limited
4 March 2003 (Class 335/302); filed in Japan 16 December 1999

Modern magnetic materials allow moving coil loudspeakers to be built with thin, lightweight ring magnets. In the illustration, voice coil **55** moves in magnetic gap **57** formed between ring magnet **11** and pole piece **52**. It seems obvious that the magnet must have one pole on its inner surface facing the voice coil and the other pole on its outer surface abutting the pole



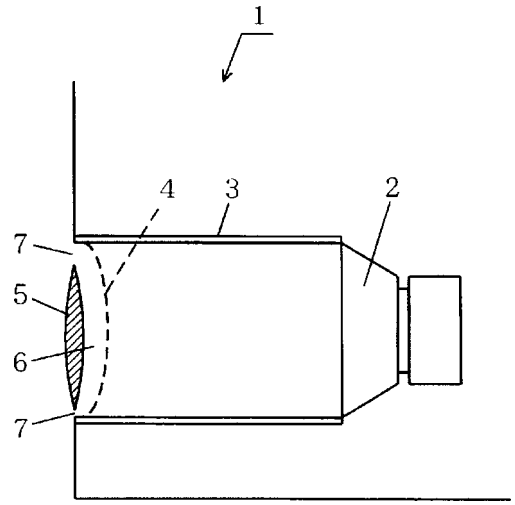
piece. According to this patent, a more linear magnetic field can be achieved if magnetization is divided into three areas. Only the central portion of the ring is magnetized at right angles. The upper and lower portions are magnetized at angles of 40° or less.—GLA

6,533,063

43.38.Ja VIDEO EQUIPMENT SPEAKER DEVICE WITH ACOUSTIC LENS

Kazuhiko Ikeuchi and Tomio Shiota, assignors to Matsushita Electric Industrial Company, Limited
18 March 2003 (Class 181/176); filed in Japan 6 July 1999

In the 1950s and 1960s a good percentage of loudspeaker patents were issued to amateur inventors who relied largely on voodoo acoustics. Today, the gauntlet seems to have been taken up by Japanese electronics manufacturers. Instead of mounting a TV loudspeaker behind a simple slatted grill,



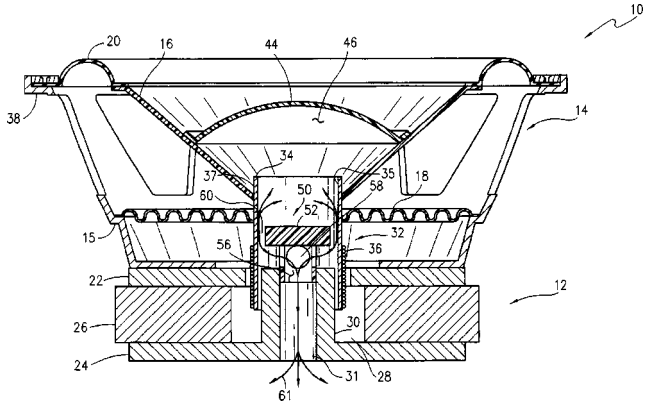
the inventors suggest that a larger fin-shaped obstruction **5** not only looks better but acts as a diverging acoustic lens. The latter assertion is not backed up by any test results. Why are we not surprised?—GLA

6,535,613

43.38.Ja AIR FLOW CONTROL DEVICE FOR LOUDSPEAKER

Jason A. Ssutu, assignor to JL Audio, Incorporated
18 March 2003 (Class 381/397); filed 28 December 1999

Airtight dust cap **44** pumps air in and out of cavity **46** through rear



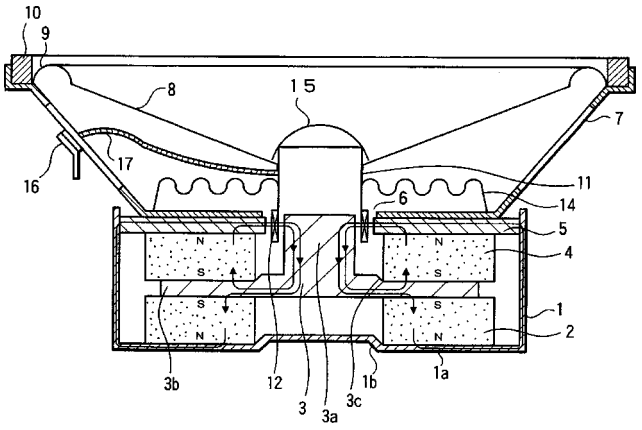
vent **31**. Plate **52** directs the air flow against the inner surface of bobbin **35** to cool voice coil **36**.—GLA

6,542,617

43.38.Ja SPEAKER

Masao Fujihira *et al.*, assignors to Sony Corporation
1 April 2003 (Class 381/402); filed in Japan 26 May 1999

What appears to be a conventional self-shielded loudspeaker is in fact a small unit designed to reproduce frequencies up to 70 kHz or so. Voice coil bobbin **11** is made of a conductive material such as aluminum. Coil **6** is attached to the bobbin by a soft bonding agent that decouples the coil from



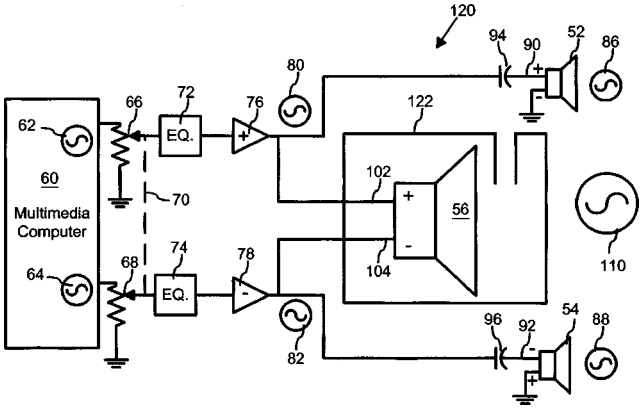
the bobbin at very high frequencies. The patent explains that in this very high range the bobbin is driven as an induction motor and continues to radiate sound, presumably from dust cap **15**.—GLA

6,529,787

43.38.Lc MULTIMEDIA COMPUTER SPEAKER SYSTEM WITH BRIDGE-COUPLED SUBWOOFER

Raymond K. Weikel and Jeffrey S. Anderson, assignors to Labtec Corporation
4 March 2003 (Class 700/94); filed 15 November 1999

The circuit shown is a clever way to power a center speaker or subwoofer without the complications of a dual voice coil or a third amplifier. Historically, it is a stepchild of telephone multiplexing in which differential and common mode signals can be handled independently. Interestingly, all



28 patent claims clearly identify the invention as a multimedia computer speaker system. Are other applications protected by other patents, known prior art, or just not worth bothering about?—GLA

6,535,269

43.38.Md VIDEO KARAOKE SYSTEM AND METHOD OF USE

Gary Sherman and Michael Chase, both of Los Angeles, California
18 March 2003 (Class 352/6); filed 29 June 2001

The sound track of a commercial motion picture is already created and stored in a multi-track format. Individual tracks may be rerecorded later to correct audio problems or to dub dialog into another language. Suppose that you could purchase a DVD for home viewing that allowed you to record and replace selected dialog tracks with your own overdubs. The patent describes an interactive system to facilitate such customized viewing.—GLA

6,535,467

43.38.Md APPARATUS AND METHOD FOR PROCESSING AUDIO SIGNALS RECORDED ON A MEDIUM

Han Jung, Bangbae-dong, Seoul and Ki Won Kim, Seryu 2-dong, both of the Republic of Korea
18 March 2003 (Class 369/47.16); filed in the Republic of Korea 17 September 1993

This patent describes a CD format for recording and a method of controlling audio-video karaoke entertainment during playback. Audio tracks are provided for vocals alone and for vocals with background instrumentals. These can be manipulated by the user/performer for increased enjoyment.—GLA

6,544,099

43.38.Md SUSPENDABLE TALKING APPARATUS

Joseph Shafik, Annandale, Virginia
8 April 2003 (Class 446/404); filed 22 January 2001

Trash talking is one way to rile your boxing opponent. But don't single practitioners of the pugilistic arts feel left out? Don't they need to be verbally abused to increase their output? Just put the ubiquitous sound chip in the punching bag, suspend it from the ceiling, add appropriate force sensors, and you're ready. Don't forget the battery.—MK

6,544,122

43.38.Md BACKGROUND-SOUND CONTROL SYSTEM FOR A VIDEO GAME APPARATUS

Shigeru Araki *et al.*, assignors to Konami Company, Limited; Kabushi Kaisha Konami Computer Entertainment Osaka
8 April 2003 (Class 463/35); filed in Japan 8 October 1998

A video game has multiple virtual sound background sources. This patent proposes precomputed fade functions. This is clearly obvious.—MK

6,545,595

43.38.Md CD QUALITY WIRELESS DOOR CHIME

Thomas G. Xydis, assignor to The Lamson & Sessions Company
8 April 2003 (Class 340/392.1); filed 22 September 2000

Ordinary door bells are, well, just too ordinary. So why not replace them with a high quality recording of your favorite greeting? The possibilities are endless.—MK

6,546,370

43.38.Md RECORDING MEDIUM WITH AUDIO DATA FROM CODER USING CONSTANT BITRATE REAL-TIME LOSSLESS ENCODING BY MOVING EXCESS DATA AMOUNTS

Jae-Hoon Heo, assignor to Samsung Electronics Company, Limited
8 April 2003 (Class 704/500); filed in the Republic of Korea 6 May 1998

The issue here is frame packing for a lossless encoder. Somehow the obvious idea of reallocating frames that exceed the constant bitrate limit has been granted. And how exactly is this different from any other coder with a bitrate ceiling?—MK

6,553,436

43.38.Md APPARATUS AND METHOD FOR PLAYBACK OF WAVEFORM SAMPLE DATA AND SEQUENCE PLAYBACK OF WAVEFORM SAMPLE DATA

Tokiharu Ando and Takashi Suzuki, assignors to Yamaha Corporation
22 April 2003 (Class 710/25); filed in Japan 9 January 1998

In digital multichannel playback and recording, samples must be buffered before being placed on the processor bus for transfer to memory. This is particularly important when the bus speeds are so high and burst mode transfers are de rigeur. The patent is highly specific to an unnamed Yamaha chip where the DSP and codecs are connected to the PCI bus.—MK

6,560,349

43.38.Md AUDIO MONITORING USING STEGANOGRAPHIC INFORMATION

Geoffrey B. Rhoads, assignor to Digimarc Corporation
6 May 2003 (Class 382/100); filed 28 December 1999

Steganography, from the Greek ''covered writing,'' is the process of concealing a secret message within ordinary material. In this extensive patent (there are more than two pages of patent references), the basic idea is to generate a noiselike signal that can be embedded in program material. Although audio encoding is mentioned, most of the patent is devoted to video encoding methods. However, in contrast to many patents granted these days, this patent writer knows what prior art is.—MK

6,542,614

43.38.Si BOOMLESS HEARING/SPEAKING CONFIGURATION FOR SOUND RECEIVING MEANS

Heinz Renner, assignor to Koninklijke Philips Electronics N.V.
1 April 2003 (Class 381/370); filed in the European Patent Office 21 March 2001

Certain hands-free communication applications require the user to hear local sounds as well as incoming signals. This requirement can be met by a single headphone having an attached boom microphone. Although the microphone is advantageously close to the user's mouth, its proximity results in unwanted pickup of pops and air noises as well. Moreover, the assembly is easily dislodged if the user is moving. This patent describes a lightweight, clip-on earpiece that contains not only a headphone but an embedded directional microphone.—GLA

6,516,669

43.40.Le VIBRATION-TYPE CONTACT DETECTION SENSOR

Kaoru Matsuki *et al.*, assignors to Mitutoyo Corporation
11 February 2003 (Class 73/651); filed in Japan 15 May 2000

The sensor described in this patent is intended for detecting contact of a stylus with a workpiece, as in three-dimensional coordinate measuring machines, for example. The present patent in essence involves a stylus that is driven by a piezoelectric element at the natural frequency the stylus exhibits when its tip just makes contact. The resulting resonant motion then provides a highly sensitive indication that contact has occurred.—EEU

6,545,390

43.40.Ng DEVICE FOR GENERATING HIGH-FREQUENCY MECHANICAL VIBRATIONS FOR A DENTAL HANDPIECE

Rainer Hahn and Uwe Grotz, assignors to Durr Dental GmbH & Company KG
8 April 2003 (Class 310/317); filed in Germany 11 April 1999

This is a dental drilling device designed to provide high-frequency mechanical vibration with constant output amplitude that is independent of the load applied to the vibration generator. The effect is achieved through a feedback circuit controlling the oscillator.—DRR

6,517,060

43.40.Tm MICRO VIBRATION ISOLATION DEVICE

Zoltan A. Kemeny, assignor to Vistek Incorporated
11 February 2003 (Class 267/136); filed 26 April 2000

This isolator is intended to attenuate the transmission of vibrations of small amplitudes. It consists of a stack of vertically acting nonlinear springs, such as Belleville washers, mounted in a housing atop an arrangement that resembles a ball bearing. The nonlinear springs are selected to have very low net vertical stiffness when they support the design load, so that they provide good vertical isolation. Radial displacement of the housing causes the balls to move up the slope of their raceway, with gravity then providing the restoring force. The horizontal stiffness is controlled by appropriate design of the raceway.—EEU

6,520,678

43.40.Tm VEHICLE CENTER BEARING ASSEMBLY INCLUDING PIEZO-BASED DEVICE FOR VIBRATION DAMPING

Albert Aiken and James A. Duggan, assignors to Spicer Driveshaft, Incorporated
18 February 2003 (Class 384/1); filed 27 March 2001

Piezo devices are attached around a bearing assembly so that mechanical vibrations of the assembly produce electrical current, which is dissipated in resistive elements as heat. The center frequency of operation of this arrangement may be varied by varying the magnitude of the resistance and/or by adjusting the stiffness of the piezo elements electronically. These variations may be effected by a controller in response to a signal from a vibration sensor.—EEU

6,523,816

43.40.Tm METHOD OF DAMPING VIBRATION, ACTIVE HYDRAULIC ANTI-VIBRATION MOUNT AND VEHICLE INCLUDING SUCH A MOUNT

Jean-Luc Gastineau *et al.*, assignors to Hutchinson
25 February 2003 (Class 267/140.14); filed 7 November 2000

This patent describes an isolator for motor vehicle applications that is intended to attenuate high-frequency vibrations of small amplitude, as well as low-frequency vibrations of large amplitude. The isolator consists of a hollow conical elastomeric element that is filled with a fluid and terminated by a rigid partition. On the other side of that partition there is a second fluid-filled chamber, which communicates with the first via two regulating nozzles. One of these nozzles is throttled by a movable device whose motion is computer-controlled and the other is designed to limit the volume variations of the fluid-filled chambers.—EEU

6,517,326

43.50.Gf BLOWING APPARATUS

Hiroyasu Fujinaka *et al.*, assignors to Matsushita Electric Industrial Company, Limited
11 February 2003 (Class 417/354); filed in Japan 14 April 2000

This patent describes the design of small electric axial-flow fans in which matching of the frequencies resulting from various noise sources is avoided, so that relatively quiet operation is obtained. The aforementioned matching is avoided by designing the fan/motor system so that the numbers of the following are not integer multiples of each other: salient poles on stator, magnetic poles on rotor, number of fan blades, sides of outer periphery of fan, spokes supporting motor from housing.—EEU

6,543,576

43.50.Gf MULTIPLE LAYER FIBER FILLED SOUND ABSORBER AND A METHOD OF MANUFACTURING THE SAME

Cameron Gorrell Cofer *et al.*, assignors to Owens-Corning Fiberglas Technology, Incorporated; Owens-Corning Sweden AB; Futaba Industrial Company, Limited
8 April 2003 (Class 181/252); filed 18 July 2000

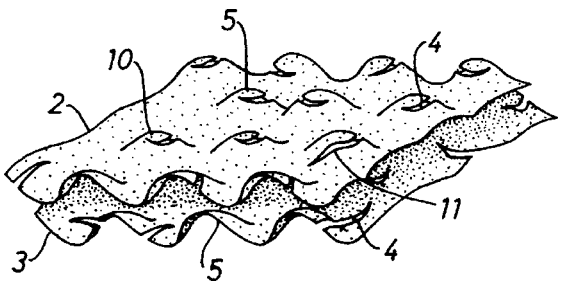
The patent covers the manner in which a vehicle muffler, with different densities of absorptive fill around the gaseous passage, can be manufactured more efficiently. The method involves a direct-fill process in which continuous fibers are injected into the sound absorber chambers.—CJR

6,555,246

43.50.Gf METHOD OF PRODUCING A SOUND-ABSORBENT INSULATING ELEMENT AND INSULATING ELEMENT PRODUCED ACCORDING TO THIS METHOD

Evelyn Zwick, assignor to Rieter Automotive (International) AG
29 April 2003 (Class 428/596); filed in Switzerland 2 February 1999

In the process of forming an insulating element or so-called heat-sink (for example, between a hot catalytic converter and the floor of a vehicle), sheets of metal are crimped and formed to have knobs or embossments that separate the sheets. In this invention, the knobs are further deformed so



there are fissures that act as perforations. These fissures improve the sound absorbing properties of the sheets and this is provided in a simplified one-step manufacturing process.—CJR

6,546,328

43.50.Jh KNOCK AND MISFIRE DETECTION SYSTEM

James M. Slicker, assignor to Eaton Corporation
8 April 2003 (Class 701/111); filed 3 January 2001

This system for detecting knocking in an internal combustion engine features a crank angle indicator, a vibration sensor, and a signal processor. The signal processor receives signals from the indicator and the sensor, performs a wavelet transform analysis of the signals from the sensor to produce a vibration frequency signal on a time scale, and compares the vibration frequency signal to a predetermined value to establish the existence of knocking in the combustion processes. It also compares the time scale of the vibration frequency signal to the signal from the indicator to establish which of the engine cylinders is knocking.—DRR

6,547,868

43.55.Ev SCRATCH RESISTANT DISCONTINUOUS ACOUSTICAL SURFACE COATING

Hector Belmares and Kenneth G. Caldwell, assignors to AWI Licensing Company
15 April 2003 (Class 106/217.3); filed 30 May 2000

A film is sprayed onto a sound-absorbing acoustical panel. The film is made with filler (typically clay), binder (typically starch), and mica, and when it dries, it forms a scratch resistant primer coating through which sound can pass. The coated panel is either air dried or heated in an oven. A finish and color can be added to this coating.—CJR

6,543,575

43.55.Rg DOUBLE-WALLED STRUCTURE AND CONNECTION ARRANGEMENT

Dwight A. Marcellus, assignor to Lindab AB
8 April 2003 (Class 181/224); filed 14 June 2000

This patent describes an improved connection between male and female segments of a double-wall ventilation duct. A double-wall duct comprises an inner wall spaced away from an outer wall, and the space between is filled with insulating material. The coupling connects just the outer walls of the duct, and the rest of the duct thickness is sealed off.—CJR

6,554,101

43.55.Rg STRUCTURE AND METHOD OF ABSORBING AND SHIELDING SOUND

Kyoichi Watanabe *et al.*, assignors to Nissan Motor Company, Limited
29 April 2003 (Class 181/290); filed in Japan 4 September 2000

This panel for isolation and shielding of engine noise into an automotive passenger compartment comprises four layers: a first panel (to block sound), a layer of absorptive material, a space for a layer of air, and a second panel layer. The absorptive material has ventilating properties and it is claimed that this improves the acoustical properties of the whole panel without adding undue weight.—CJR

6,555,979

43.58.Fm SYSTEM AND METHOD FOR CONTROLLING ELECTRICAL CURRENT FLOW AS A FUNCTION OF DETECTED SOUND VOLUME

L. Taylor Arnold and Richard Michelli, both of Raleigh, North Carolina
29 April 2003 (Class 318/272); filed 6 December 2000

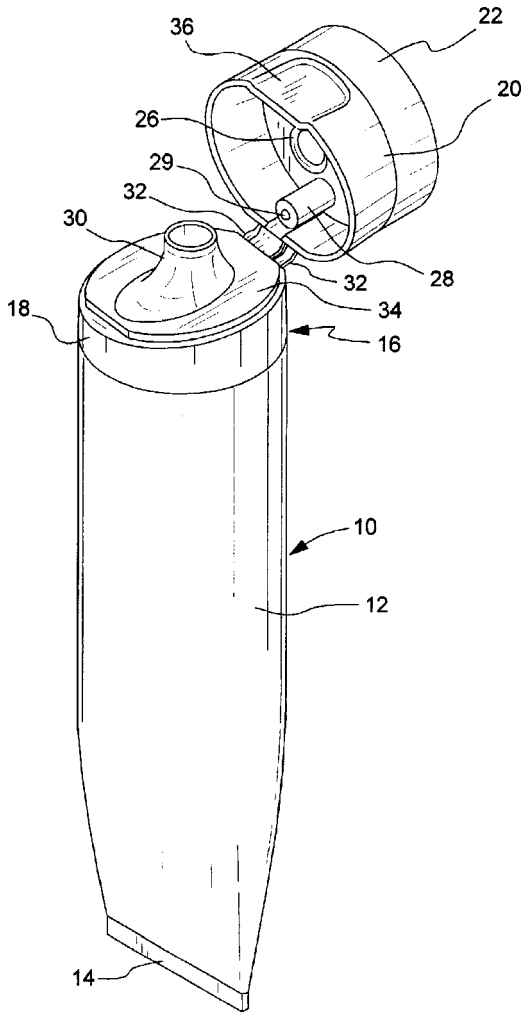
The idea is simple enough: amplify and filter the output of a microphone and use it to control a device, for example, an electric motor in a child's toy. The patent shows a very simple implementation using one op amp and a few discrete components.—MK

6,557,728

43.58.Wc MUSICAL TOOTHPASTE TUBE CLOSURE

Jeffrey Blake Anderson and Michael Barnett, assignors to Colgate-Palmolive Company
6 May 2003 (Class 222/39); filed 15 June 2000

Another invasion of your private auditory space comes from the



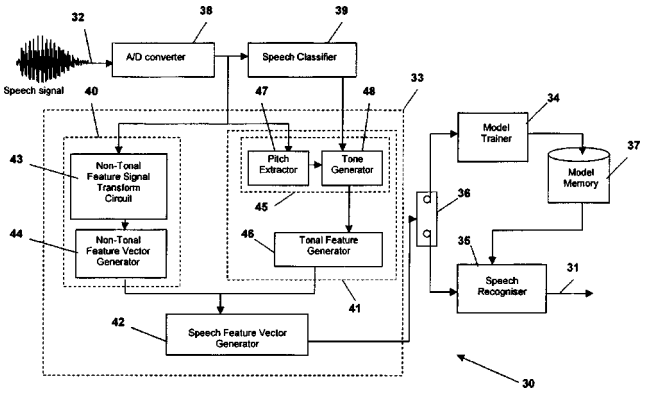
inventors at Colgate-Palmolive, who propose adding the ubiquitous sound chip to a toothpaste tube. The possibilities are endless.—MK

6,553,342

43.60.Lq TONE BASED SPEECH RECOGNITION

Yaxin Zhang *et al.*, assignors to Motorola, Incorporated
22 April 2003 (Class 704/255); filed 2 February 2000

Speech recognition specialized for many Asian languages, e.g., Mandarin Chinese, uses speech tonal information via tonal feature vectors during the voiced speech segments. Combining the tonal and nontonal features and



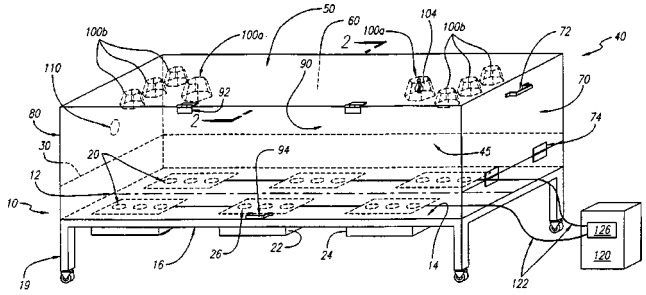
comparing these to prestored speech vector models complete the classification phase.—HHN

6,544,165

43.64.Ri METHOD AND APPARATUS FOR APPLYING FREQUENCY VIBRATIONS THERAPEUTICALLY

Barry McNew, Cottonwood, Arizona
8 April 2003 (Class 600/27); filed 23 July 1999

This far-fetched device is designed to apply sound and light into a dark, almost coffinlike box containing a person undergoing therapeutic treatment. The box contains at least two loudspeakers positioned just below the person's body, one underneath the right side and the other underneath the left side. In combination with a light display, the sound is supposed to affect



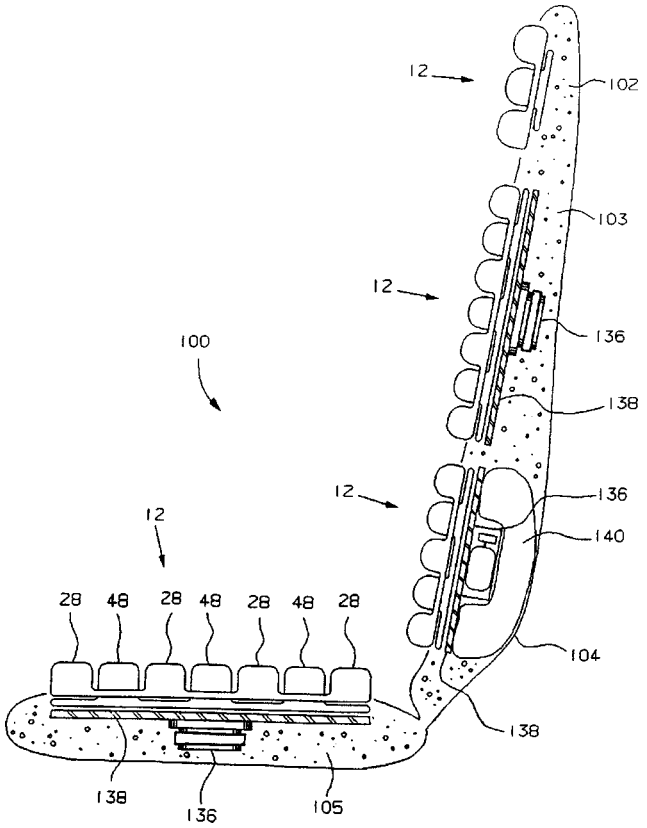
the user through the ears and through the body in a salutary fashion. As a reviewer of acoustical patents, I found the phrase ''...using acoustic vibrations, also known as sound'' to be rather grating. I would recommend listening to Mozart on a good quality stereo system for even better therapy without the necessity of being shut up in a dark box.—DRR

6,551,450

43.64.Ri UNIQUE AIR AND SONIC MASSAGING APPARATUS

Paul B. Thomas and Robert D. Leventhal, assignors to D2RM Corporation
22 April 2003 (Class 156/580.1); filed 7 September 2001

This apparatus, which can be formed into either a bed or a seat, functions as an air and sonic massaging system. The apparatus includes an air device and a sonic device located below the air device. A patterned inflation and deflation of the air-containing sacs provides the massaging effect, while



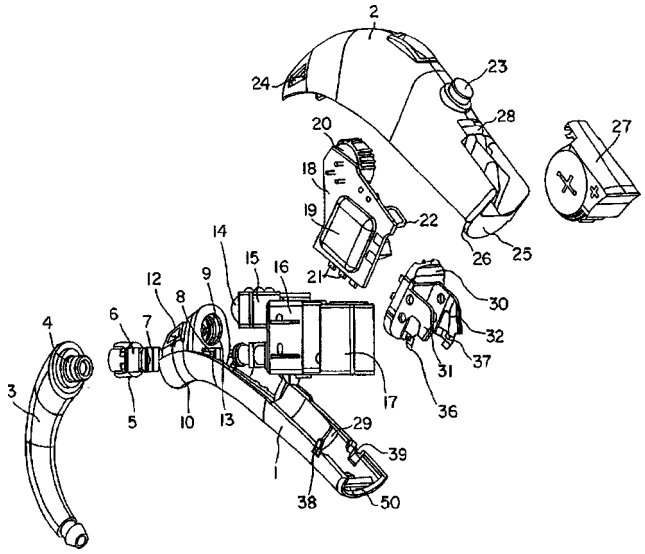
the sonic device operates through a sound transducer that generates acoustic waves from a flat plate. The acoustic energy may be directly coupled to the air device or directly to the user's body cavity.—DRR

6,522,764

43.66.Ts HEARING AID

Tom Bøgeskov-Jensen, assignor to Oticon A/S
18 February 2003 (Class 381/330); filed in Denmark 7 October 1998

Post-auricular hearing aids have traditionally been assembled using screws, which causes stress in the housing parts that may break the housing material if the hearing aid is dropped on a hard surface. Described is a



post-auricular hearing aid design in which the housing quickly snaps together without causing stress on the parts.—DAP

6,529,607

43.66.Ts METHOD FOR PRODUCING A CONSTANT SOUND PRESSURE LEVEL IN HEARING AIDS AND CORRESPONDING HEARING AID

Ullrich Sigwanz and Frank Wagner, assignors to Siemens Audiologische Technik GmbH
4 March 2003 (Class 381/107); filed in Germany 11 December 1998

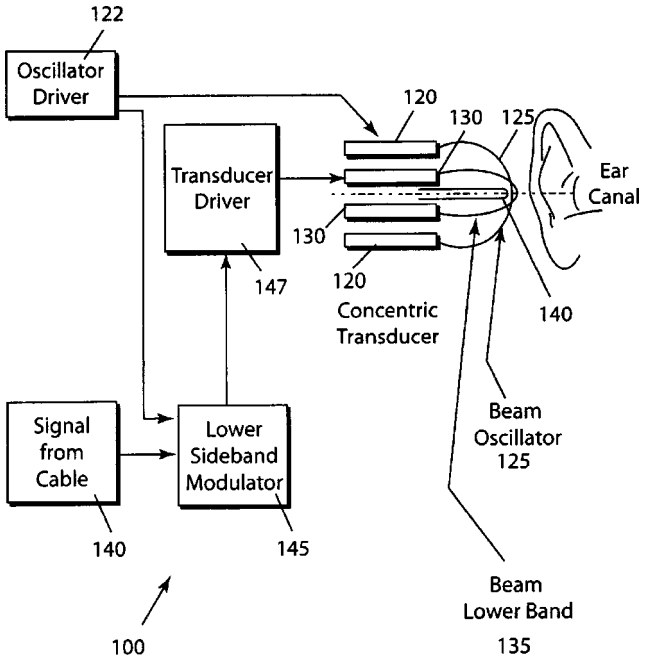
To prevent loudness discomfort, compression is frequently used to automatically control the gain of hearing aids. This patent describes a means for preventing loudness discomfort by digitally holding the output SPL constant via envelope generation. An envelope generator produces an envelope value of a signal value to be processed. A divider calculates the ratio of the signal value and its calculated envelope value. Ideally, these ratios lie in a restricted value range, e.g., -1 to $+1$. A comparator compares the envelope value with a gain parameter. Signal values whose envelope values exceed an adjustable threshold outside of the restricted range are divided by the envelope values, producing a constant envelope within the restricted range that is routed to the hearing aid receiver.—DAP

6,532,292

43.66.Ts METHOD AND APPARATUS TO TRANSMIT AUDIO INTO THE HUMAN EAR

Edward A. Lygas, assignor to Sony Corporation; Sony Electronics, Incorporated
11 March 2003 (Class 381/74); filed 3 March 1999

In hearing aids, cellular phones, or headsets containing both microphone and receiver, echoes or crosstalk resulting from the output signal returning to the microphone and being reamplified are often problematic. Described is a method of reducing the crosstalk while preventing the audio signal from being heard by others. Two ultrasonic waves having different frequencies are generated in the ear canal of the wearer by concentric cylindrical transducers having different diameters. The first transducer is driven by an oscillator producing the frequency f1. The second transducer is driven by a mixer that produces the frequency $f1 - f_{audio}$. The audio is said to



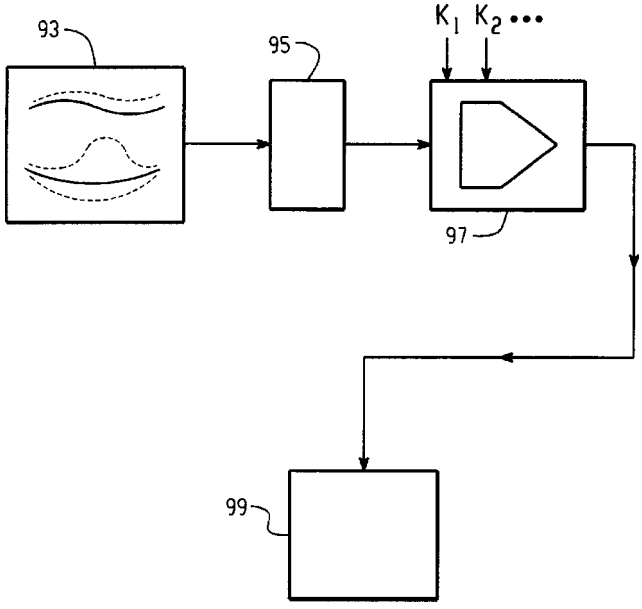
be produced within the ear canal by a difference frequency resulting from interference occurring in the location of a high-intensity overlap between the first and second signals.—DAP

6,533,062

43.66.Ts PRODUCTION PROCESS FOR CUSTOM-MOULDED EAR-PLUG DEVICES

Christoph Widmer *et al.*, assignors to Phonak AG
18 March 2003 (Class 181/129); filed 25 September 2000

The fit of a custom molded ear-plug or in-the-ear hearing aid is improved by accounting via computer modeling for the effect on ear canal shape during mastication of at least two different positions of the jaw. In production, custom molded ear-plug or in-the-ear hearing aids are produced under computer control by incrementally building up the surface via laser sintering or via stereo lithography of a laser-hardening photopolymer. Corrections to the outer contour surfaces are made via digital imaging to form



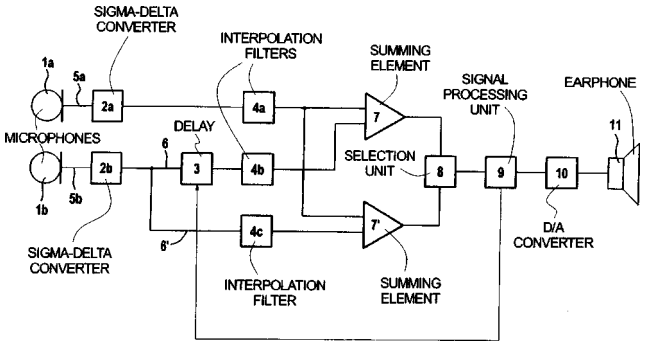
flexible or compressible sections so as to account for ear canal shape changes.—DAP

6,539,096

43.66.Ts METHOD FOR PRODUCING A VARIABLE DIRECTIONAL MICROPHONE CHARACTERISTIC AND DIGITAL HEARING AID OPERATING ACCORDING TO THE METHOD

Ullrich Sigwanz and Fred Zoels, assignors to Siemens Audiologische Technik GmbH
25 March 2003 (Class 381/313); filed in Germany 30 March 1998

A variable directional microphone characteristic is produced with greater adjustment resolution without increasing the processing sampling frequency and hence power consumption. Instead, finer adjustment graduations are provided by delaying the 1-bit output of a sigma-delta converter which is connected to the output signal path of one of at least two micro-



phones in the directional system. Delaying a bit stream of only 1-bit width, as compared to delaying a 16-bit byte, effectively multiplies the sampling rate by typically 20 to 40 times. An interpolation low-pass filter attenuates signal disturbances that may occur when switching between directional microphone characteristics.—DAP

6,546,082

43.66.Ts METHOD AND APPARATUS FOR ASSISTING SPEECH AND HEARING IMPAIRED SUBSCRIBERS USING THE TELEPHONE AND CENTRAL OFFICE

Tommy R. Alcendor *et al.*, assignors to International Business Machines Corporation
8 April 2003 (Class 379/52); filed 2 May 2000

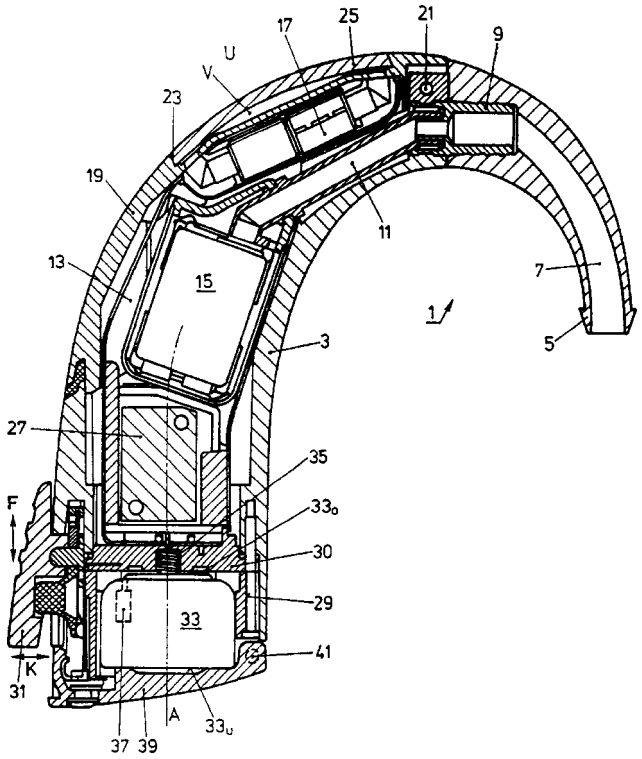
This is a system to enable speech- and hearing-impaired persons to use a telephone. The system consists of the following steps: (1) a central office receives a request to communicate from an impaired caller using a special unit, (2) the central office determines whether the caller is speech-impaired, and (3) if the caller is speech-impaired, then the central office prompts that caller for text input, receives the text input, and then converts the text input, using a text-to-speech resource, into an audio output that is forwarded to the receiving party.—DRR

6,546,110

43.66.Ts BEHIND-THE-EAR HEARING AID AND ATTACHMENT MODULE FOR SAME

Andi Vonlanthen, assignor to Phonak AG
8 April 2003 (Class 381/330); filed in the World IPO 16 June 1999

A post-auricular hearing aid design includes a detachable module that may contain a rechargeable battery, an audio input, wireless interface cir-



cuitry, or programming plug-in means. For the battery, its cylindrical axis is approximately coaxial with the longitudinal axis of the hearing aid, resulting in more efficient use of space at the bottom of the hearing aid. Electronic terminals are provided in the module that connect to those in the hearing aid. Several such modules can be stacked onto the base of the hearing aid.—DAP

6,549,611

43.66.Ts TEXT ENHANCED TELEPHONY

Robert M. Engelke and Kevin Colwell, assignors to Ultratec, Incorporated
15 April 2003 (Class 379/52); filed 15 October 2001

This telephone system for assisting the hard of hearing uses text enhanced telephone (TET). The hard of hearing user uses a TET device that permits the user to speak normally, but to read the words spoken by the other party. The TET device supplies to the user a stream of text of the words spoken by the other party using currently available voice-to-text relays to help the deaf community to communicate with hearing persons over the telephone. Both the TET relay and TET device are capable of separating voice and text-carrying digital communications frequencies that travel simultaneously over the telephone line.—DRR

6,549,633

43.66.Ts BINAURAL DIGITAL HEARING AID SYSTEM

Søren Erik Westermann, assignor to Widex A/S
15 April 2003 (Class 381/312); filed 23 February 2000

A binaural hearing aid system is described in which differences in hearing loss between the two ears and different compensation for each ear are taken into account. Two individual digital signal processors are utilized, one for the signal from the input transducer on each side of the head. Each of the individual ear signal processors simulates signal processing occurring in the opposite ear. A third digital signal processor processes the information derived from the two individual ear signal processors. A mix of individual and combined processing ranging from monaural to binaural is output to the hearing aid speaker. Communication between the ears may be via a bi-directional wireless link.—DAP

6,549,634

43.66.Ts BEHIND-THE-EAR HEARING AID

Andi Vonlanthen, assignor to Phonak AG
15 April 2003 (Class 381/322); filed 28 June 1999

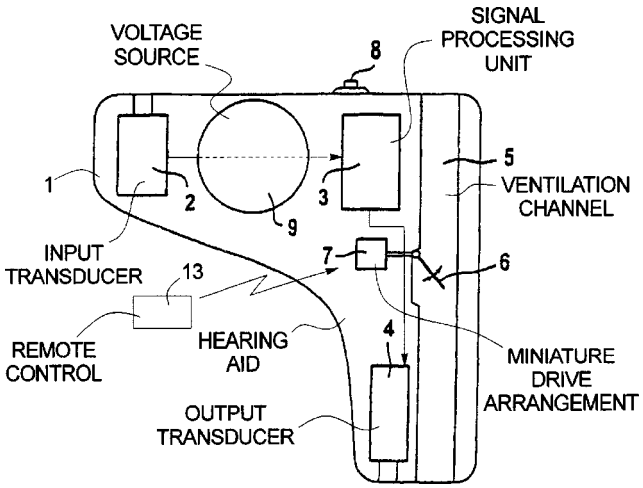
An intermediate space is created in a behind-the-ear hearing aid between the hearing aid housing and the loudspeaker housing to provide greater low-frequency output from the hearing aid. A removable capsule with metal cup construction surrounds the loudspeaker and the additional space and acts as a magnetic shield.—DAP

6,549,635

43.66.Ts HEARING AID WITH A VENTILATION CHANNEL THAT IS ADJUSTABLE IN CROSS-SECTION

Anton Gebert, assignor to Siemens Audiologische Technik GmbH
15 April 2003 (Class 381/324); filed in Germany 7 September 1999

A ventilation channel (vent) is utilized frequently in hearing aids to equalize the ear canal to atmospheric pressure or to reduce the occluding effect caused by the hearing aid earmold. Such a deliberate leak can create acoustic feedback problems and reduce the effectiveness of directional microphones in the hearing aid. Therefore, to make the vent as small as possible, a mechanism is described for adjusting the cross-sectional size of a



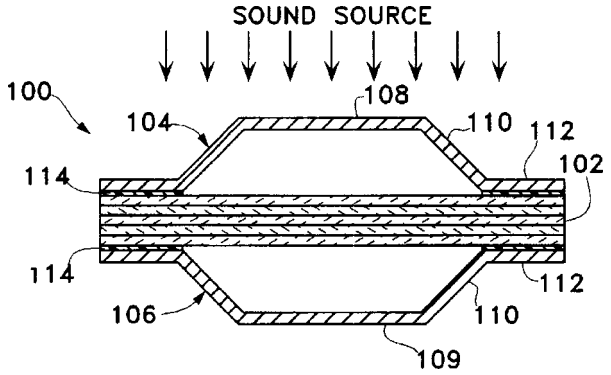
vent under control of the hearing aid signal processor. One or more adjusting elements such as valves, pins, slides, or flaps with holes in them are positioned within the vent by miniature electrical and/or magnetic drives. The size of the vent may be varied by the hearing aid wearer in different environmental listening conditions with a switch or rotional device on the hearing aid or via remote control.—DAP

6,554,761

43.66.Ts FLEXTENSIONAL MICROPHONES FOR IMPLANTABLE HEARING DEVICES

Sunil Puria and Rodney C. Perkins, assignors to Soundport Corporation
29 April 2003 (Class 600/25); filed 29 October 1999

The sensitivity of a surgically implanted piezoelectric microphone may be increased by sandwiching its sensor element between two endcaps. The endcaps, for which two shapes are recommended, amplify the forces



mechanically, producing a greater output voltage from the piezo. The result is that signal-to-noise ratio resulting from internal circuit noise of the piezo microphone is said to be increased.—DAP

6,554,762

43.66.Ts IMPLANTABLE HEARING SYSTEM WITH MEANS FOR MEASURING ITS COUPLING QUALITY

Hans Leysieffer, assignor to Cochlear Limited
29 April 2003 (Class 600/25); filed in Germany 25 August 2000

The quality of coupling of the output transducer of a middle ear or cochlear implant is assessed by measuring mechanical impedance. Postoperative monitoring of the coupling quality is enabled by circuitry within the implant triggering an impedance measurement automatically at periodic intervals or in response to a request from software. The impedance measurement may consist of electrical input impedance of the output transducer coupled to the biological load determined via a voltage measurement across a known resistor. The impedance value is transmitted via telemetry to the hearing professional for evaluation.—DAP

6,556,686

43.66.Ts PROGRAMMABLE HEARING AID DEVICE AND METHOD FOR OPERATING A PROGRAMMABLE HEARING AID DEVICE

Tom Weidner, assignor to Siemens Audiologische Technik GmbH
29 April 2003 (Class 381/312); filed in Germany 14 April 1999

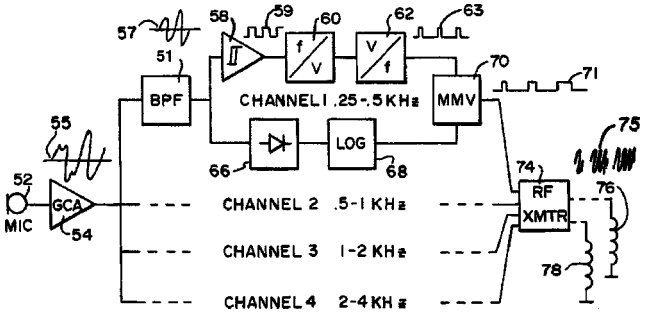
Hearing aids made with the same hardware may be adapted to individual fitting adjustments with a key that enables or disables adjustable properties. The keys, which may be hardware-based and/or software-based, permit the hearing aid wearer to modify performance parameters within a subset of maximum ranges defined by the hearing aid hardware. Hardware keys may be detachable and interchangeable and consist of mechanical, electrical, or magnetic switches, permitting the wearer to exchange performance parameter sets for different listening environments. Software keys are signals transmitted from a remote programming device to the hearing aid via wired or wireless methods.—DAP

6,556,870

43.66.Ts PARTIALLY INSERTED COCHLEAR IMPLANT

Clemens M. Zierhofer *et al.*, assignors to Med-El Elektromedizinische Geraete GmbH
29 April 2003 (Class 607/57); filed 31 January 2001

To make the external portion of a cochlear implant less visible, an ear canal module is described that contains a microphone, signal processor, and, in some embodiments, also a wireless transmitter. The signal processor



forms signals suitable for stimulating the cochlea using techniques such as continuous interleaved sampling, spectral peak processing, and channel specific sampling sequences. An implantable battery may be used to power the implant and be recharged from the wireless transmissions.—DAP

6,546,515

43.72.Gy METHOD OF ENCODING A SIGNAL

Peter Vary *et al.*, assignors to Alcatel
8 April 2003 (Class 714/746); filed in the European Patent Office 18 June 1999

Described is a method for encoding a speech signal in which source coding (reducing redundancy in the original signal and maximizing transmitted SNR for the available bit rate) and channel coding (use of error correcting codes) are both optimized. The excess bits after source coding are used to increase error robustness at the transmit end. Code redundancy is incorporated at the transmit end with means such as error correcting codes. These codes are decoded at the receive end indirectly using parameter value estimation rather than hard or soft decisions.—DAP

6,526,379

43.72.Ne DISCRIMINATIVE CLUSTERING METHODS FOR AUTOMATIC SPEECH RECOGNITION

Luca Rigazio *et al.*, assignors to Matsushita Electric Industrial Company, Limited
25 February 2003 (Class 704/245); filed 29 November 1999

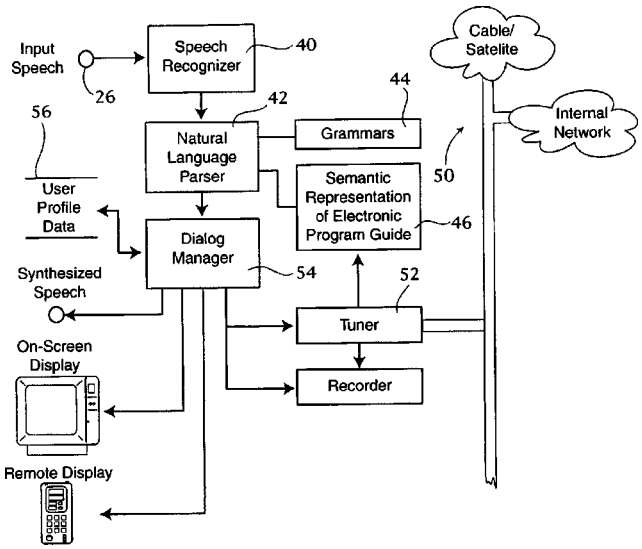
Hidden Markov models (HMMs) are commonly used to represent features extracted from speech. Probabilities associated with these HMMs are frequently represented as Gaussian distributions. As the number of states representing each speech unit is increased and as more complex Gaussian mixture density components are used, complexity increases, resulting in larger memory and power requirements. To reduce complexity, this patent describes a technique for defining clusters of Gaussians that are sufficiently close to each other to be merged. Cluster dispersion is reduced with the Lloyd-Max algorithm by moving Gaussians from one cluster to another so that the Bhattacharyya distance inside the clusters is minimized.—DAP

6,553,345

43.72.Ne UNIVERSAL REMOTE CONTROL ALLOWING NATURAL LANGUAGE MODALITY FOR TELEVISION AND MULTIMEDIA SEARCHES AND REQUESTS

Roland Kuhn *et al.*, assignors to Matsushita Electric Industrial Company, Limited
22 April 2003 (Class 704/275); filed 26 August 1999

An interactive remote control using natural speech is used to access a display of the available TV programs for viewing and recording. A touch



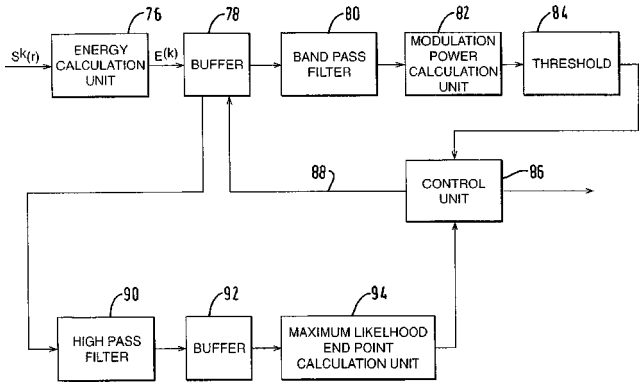
screen display allows a user to easily make a selection by tapping with a pen or a finger.—HHN

6,560,575

43.72.Ne SPEECH PROCESSING APPARATUS AND METHOD

Robert Alexander Keiller, assignor to Canon Kabushiki Kaisha
6 May 2003 (Class 704/241); filed in the United Kingdom 20 October 1998

To check the consistency between training words for speech recognition or verification, a dynamic programming alignment process is used to calculate an average frame score and a worst frame score in a number of



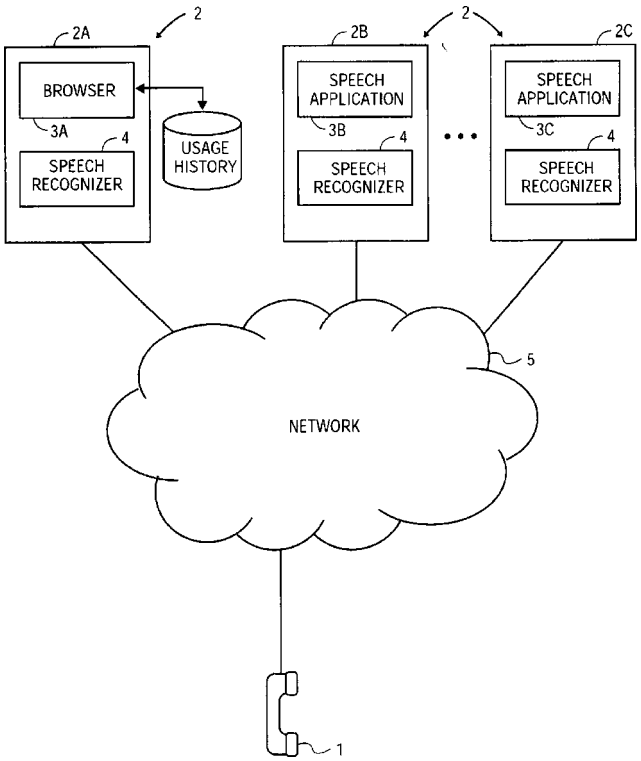
successive frames. These numbers are compared with prestored numbers for consistent examples.—HHN

6,560,576

43.72.Ne METHOD AND APPARATUS FOR PROVIDING ACTIVE HELP TO A USER OF A VOICE-ENABLED APPLICATION

Michael H. Cohen *et al.*, assignors to Nuance Communications
6 May 2003 (Class 704/270); filed 25 April 2000

An active voice-enabled help system uses a dialog history of each user to provide help to a user. The system may maintain a different set of



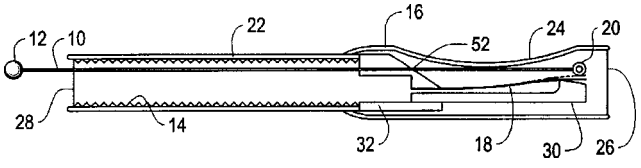
conditions and information for each specific help topic.—HHN

6,547,627

43.75.Ef GAME CALLING DEVICE HAVING ADJUSTABLE PITCH, INTENSITY, TONE AND INFLECTION

David E. Oathout, assignor to David E. Oathout
15 April 2003 (Class 446/208); filed 23 August 2001

While acoustic musical instruments are rather fixed in their principles of operation, apparatus for calling game to hunters is not. In this invention, the combination of a single reed **18**, an elastic skin **14**, and a control rod **12**



allows the instrumentalist (or hunter) to control various aspects of game vocalization timbre.—MK

6,555,736

43.75.Hi RHYTHM SHAKER

Daniel Delosreyes, assignor to Latin Percussion, Incorporated
29 April 2003 (Class 84/402); filed 16 January 2002

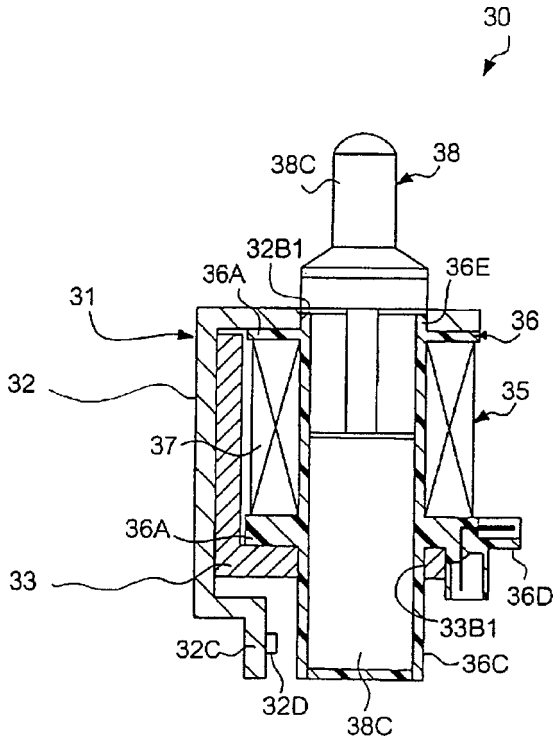
Simply put, the idea is to provide an enclosed volume whereby one surface is reactive to solid pellets and the others damped. This is, unfortunately, obvious, even to those not ''trained in the art.''—MK

6,552,251

43.75.Mn ACTUATING DEVICE EASILY ASSEMBLED AND KEYBOARD MUSICAL INSTRUMENT EQUIPPED THEREWITH

Shigeru Muramatsu and Syuzo Okabe, assignors to Yamaha Corporation
22 April 2003 (Class 84/3); filed in Japan 1 June 2000

This Yamaha patent improves on previous Disklavier™ patents [e.g., reviewed in J. Acoust. Soc. Am. **112**(6), 2521–2522 (2002)] by making the



unit easier to assemble and repair by placing the solenoids in a retaining frame.—MK

6,559,369

43.75.Mn APPARATUS AND METHOD FOR SELF-TUNING A PIANO

Donald A. Gilmore, Kansas City, Missouri
6 May 2003 (Class 84/455); filed 14 January 2002

As the inventor notes, the piano is virtually the only instrument not typically tuned by the owner. So, the proposal is to apply a voltage across the strings to use thermal expansion to tune the string which in turn is monitored by a microcontroller. Small complications arise, such as the windings on bass strings which decrease electrical resistance to a near short. The inventor's solution is to wind the strings with insulated magnetic wire. This was also mentioned in the ''New Scientist'' as reported by the Winter 2003 Society's ''Echoes.''—MK

6,545,210

43.75.Wx MUSICAL SOUND GENERATOR

Toru Morita, assignor to Sony Computer Entertainment Incorporated
8 April 2003 (Class 84/622); filed in Japan 3 March 2000

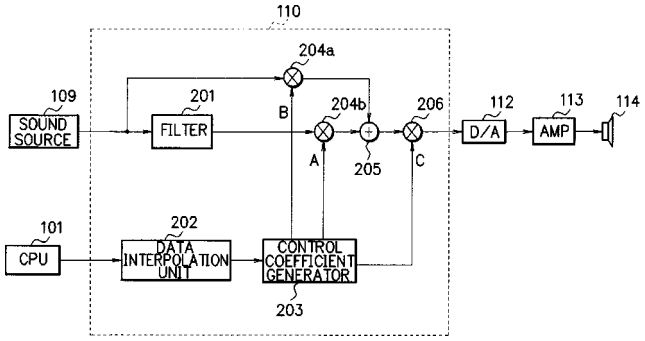
When the patent examiner supplies the references to all the past work you know the patent is in trouble. Here, Sony attempts to patent mixed musical note and sound effects with a particular attention to their implementation. But surely the notion of variable pointers to sound data is obvious, even to those not ''skilled in the art.''—MK

6,548,749

43.75.Wx ELECTRONIC MUSICAL INSTRUMENT AND TONE VOLUME CONTROL METHOD

Masayuki Suda, assignor to Kabushiki Kaisha Kawai Gakki Seisakusho
15 April 2003 (Class 84/660); filed in Japan 31 October 2000

The filter **201** has a inverse Fletcher-Munson response and can be mixed with the original by control of the multipliers **204a** and **204b** and



adder **205**. Ideally, this is all controlled by a CPU with a single potentiometer, which is digitized and used to generate the proper mixing curves.—MK

6,544,187

43.80.Qf PARAMETRIC IMAGING ULTRASOUND CATHETER

James B. Seward, assignor to Mayo Foundation for Medical Education and Research
8 April 2003 (Class 600/466); filed 5 March 2002

According to this patent, parametric imaging is defined as the imaging of quantifiable ''parameters,'' which fall into the categories of visible two-, three-, four-dimensional or nonvisible higher dimensional temporal physiologic events. Visible motion is considered to be a fourth-dimensional event and it may entail cardiac muscle contraction, heart wall motion, valve leaflet motion, etc. Nonvisible events, such as aging, healing, remodeling, transformation, etc., are deemed higher dimensional events. An ultrasonic catheter incorporating parametric imaging capability can obtain dynamic digital information from the surrounding environment and convert information contents into static or dynamic geometric figures from which discrete or gross quantifiable information can be derived.—DRR

6,544,189

43.80.Qf HANDHELD SENSOR FOR ACOUSTIC DATA ACQUISITION

Raymond Watrous, assignor to Zargis Medical Corporation
8 April 2003 (Class 600/528); filed 27 June 2001

The device is a hand-held sensor for acquiring acoustic data for the purpose of medical diagnosis. The module features three studs on one side, at least one of which serves as an electrode to provide a temporal reference signal. The modules also contain a sensing unit that contacts the patient, capturing an acoustic cardiovascular signal and the temporal reference signal. The device also incorporates a telemetry sensor that is connected to the sensing unit and communicates the acoustic signal, the temporal reference signal, and information on the positioning of the sensing unit to a data processor.—DRR

6,544,198

43.80.Qf STETHOSCOPE SYSTEM FOR SELF-EXAMINATION USING INTERNET

Hoon Chong and Jong Gie Kim, assignors to Hoseo University; Hoon Chong
8 April 2003 (Class 600/586); filed in the Republic of Korea 11 June 2001

The device is a stethoscope system for self-examination by which an individual can diagnose his or her health status by comparing the sound waves generated from various parts of the patient's body with sound waves characteristic of specific diseases. The sensor part of the system is placed over the appropriate portions of the body and the signals are relayed over the Internet to a server that analyzes the received information to provide the diagnosis.—DRR

6,546,276

43.80.Qf ULTRASONIC BASED DETECTION OF INTERVENTIONAL MEDICAL DEVICE CONTACT AND ALIGNMENT

Claudio I. Zanelli, Menlo Park, California
8 April 2003 (Class 600/424); filed 12 September 2000

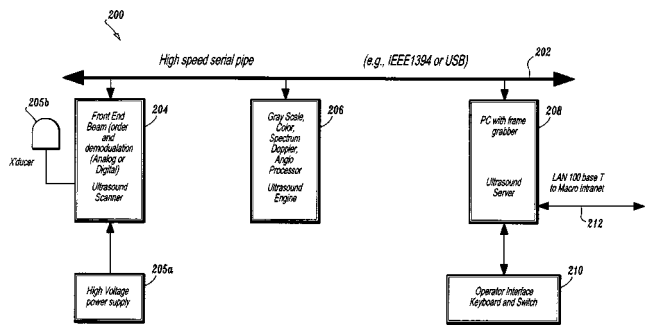
The patent covers a method and apparatus to monitor the location of an interventional medical device (IMD) through the use of ultrasonic signals. Both the proximity and alignment of the IMD are established from ultrasonic signals reflected off the tissue surface. The inclusion of an offset between the distal end of the IMD and the ultrasonic transducer is said to provide accurate position and alignment monitoring when the IMD is in contact with or very close to the tissue surface. A comparison of the strengths of multiple reflected signals is used to measure the alignment of the IMD in three-dimensional space or perpendicular to a given surface. It is also stated that the device can be used with a variety of IMDs in a variety of medical procedures.—DRR

6,547,730

43.80.Qf ULTRASOUND INFORMATION PROCESSING SYSTEM

Shengtz Lin and Zengpin Yu, assignors to U-Systems, Incorporated
15 April 2003 (Class 600/437); filed 31 December 1998

The use of ultrasonic medical imaging results in the need for efficient processing of echo mode algorithms, Doppler shift echo processing algorithms, color flow mode processing, etc. An architecture and protocol are provided to allow a flexible ultrasound information processing system whereby ultrasound information processing functions are performed by a plurality of ultrasound modules coupled to a high speed information bus. In another embodiment, a hospital ultrasound information network is formed by the ultrasound station modules and the ultrasound information bus at a given clinical site, each ultrasound module being assigned a unique TCP/IP address.—DRR



6,547,731

43.80.Qf METHOD FOR ASSESSING BLOOD FLOW AND APPARATUS THEREOF

D. Jackson Coleman *et al.*, assignors to Cornell Research Foundation, Incorporated; University of Virginia Patent Foundation
15 April 2003 (Class 600/437); filed 5 May 1999

This method for assessing blood flow in living tissue involves directing an ultrasonic beam through the tissue along overlapping lines of sight. Blood flow data is then generated from the overlaps to provide an evaluation of the flow. This stems from the fact that partially overlapping beams are generated at fixed temporal intervals. The spatial overlap allows the spatial distance between overlapping lines of sight to be ignored, while moving reflectors, within any overlapping lights of sight, will cause detectable changes in the range of the moving reflector from one line of sight to the next. The rate of motion can be ascertained from the measured change in range and the known time interval between vectors.—DRR

6,547,736

43.80.Qf DOPPLER ULTRASOUND METHOD AND APPARATUS FOR MONITORING BLOOD FLOW AND DETECTING EMBOLI

Mark A. Moehring and Timothy R. Myers, assignors to Spentech, Incorporated
15 April 2003 (Class 600/454); filed 18 May 2000

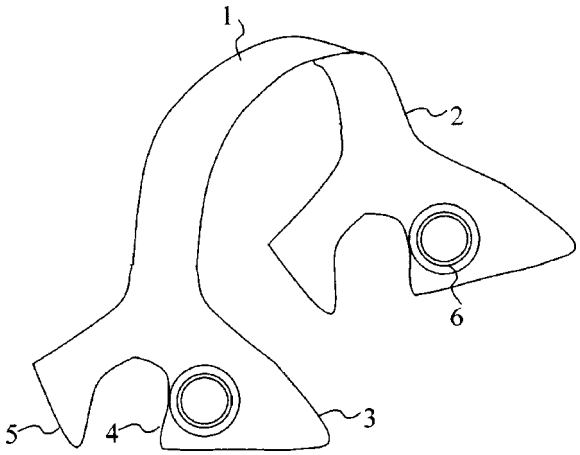
We have here another pulse Doppler ultrasound system for monitoring blood flow and detecting emboli. The system features a graphical information display that can provide simultaneous depth-mode and spectrogram displays. The depth-mode display indicates the various positions along the ultrasound beam axis at which the blood flow is detected. These positions are denoted by one or more colored regions, with the colors indicating direction of blood flow and varying in intensity as a function of the detected Doppler ultrasound signal amplitude or detected blood flow velocity. The depth-mode display also includes a pointer that can be positioned by the user. The spectrogram displayed then corresponds to the location selected by the pointer.—DRR

6,547,737

43.80.Qf INTELLIGENT TRANSCRANIAL DOPPLER PROBE

Philip Chidi Njemanze, Owerri, Imo, Nigeria
15 April 2003 (Class 600/454); filed 15 March 2002

The device, designed to be mounted on a patient's head, makes use of transcranial Doppler (TCD) sonography, an imaging technique that uses Doppler ultrasound to measure cerebral blood flow in basal cerebral arteries. This is also a system for automatic manipulation of the TCD probe. Because it includes a microprocessor and a mechanism for moving the probe, the



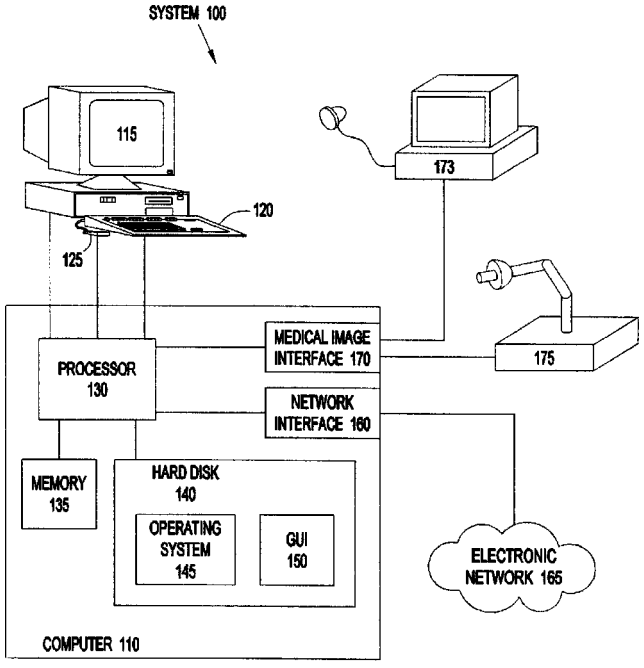
system is capable of ''learning'' probe angulations from a series of initial manual manipulations and then performing cerebral vessel insonation using electromotive forces without an operator.—DRR

6,549,802

43.80.Qf SEED LOCALIZATION SYSTEM AND METHOD IN ULTRASOUND BY FLUOROSCOPY AND ULTRASOUND FUSION

Kenneth B. Thornton, assignor to Varian Medical Systems, Incorporated
15 April 2003 (Class 600/426); filed 7 June 2001

Brachotherapy, used for treating cancer, is a radiation treatment using a solid or enclosed radioisotopic source on the surface of the body or a short distance from the area to be treated. The effectiveness of this type of treatment depends on the exact placement of the implanted radiotherapy source. The device described in this patent provides for determination of the 3-D position of implanted radiotherapy seeds with respect to an area of affected tissue, such as the prostate, so that a radiotherapy dose may be computed.



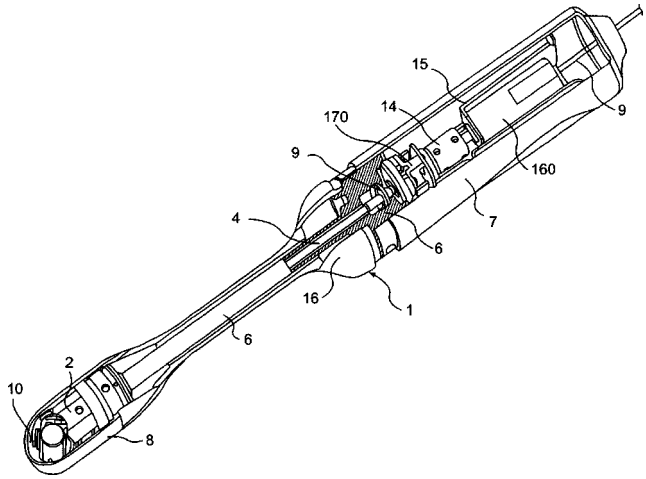
The device uses ultrasound and fluoroscopy imaging and does not require computed tomography imaging. The device also provides a way to visualize the 3-D position of implanted brachotherapy seeds by providing an interactive, computer-generated, graphical interface.—DRR

6,551,245

43.80.Qf ULTRASONIC PROBE

Kazuyoshi Irioka *et al.*, assignors to Matsushita Electric Industrial Company, Limited
22 April 2003 (Class 600/444); filed in Japan 26 November 1999

This ultrasonic probe, intended for use in medical applications for noninvasively inspecting the interior of a patient's body, contains a scan



mechanism that rotates a piezoelectric element emitting the ultrasonic wave.—DRR

6,551,247

43.80.Qf ULTRASONIC PROBE

Koetsu Saito *et al.*, assignors to Matsushita Electric Industrial Company, Limited
22 April 2003 (Class 600/459); filed in Japan 7 March 2000

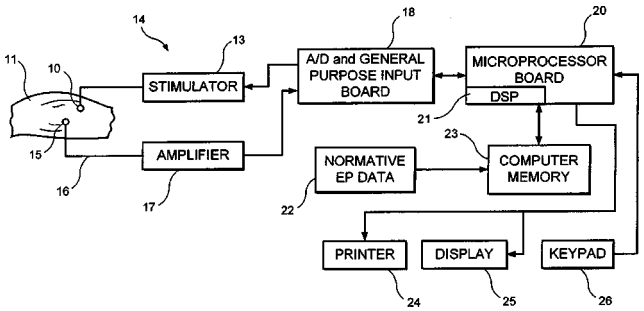
This ultrasonic probe is formed from a high-molecular-weight material having a conductive layer and is deployed between a piezoelectric element and an acoustic impedance substantially equal to that of the acoustic matching layer. The probe can be configured into a slim shape to facilitate operation. The authors argue that the probe's shape will not degrade its performance, viz. its sensitivity and frequency characteristics. It is also said that the probe is structured in such a manner that no electrical problem will occur in the event of wire breakage, even if the piezoelectric element becomes cracked by a mechanical impact or the like.—DRR

6,556,861

43.80.Qf FETAL BRAIN MONITOR

Leslie S. Prichep, assignor to New York University
29 April 2003 (Class 600/544); filed 20 November 2000

A fetal brain monitor (FBM) uses a transducer placed on a mother's abdomen, pulsed to generate auditory sounds in the form of clicks to provide brainstem-evoked responses (BAER) of a fetus. In one embodiment, the fetus's brain waves are detected by a biosensor, amplified, converted into digital format, and analyzed using a digital comb filter to increase the signal/



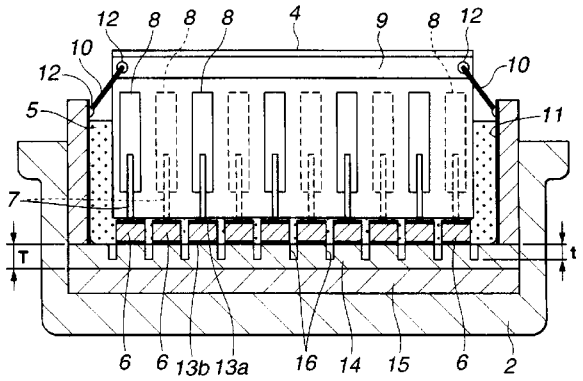
noise ratio. A complementary computer system applies a quantitative EEG analysis to compare the data from the fetus to normative data or to prior states of the fetus's own data.—DRR

6,558,323

43.80.Qf ULTRASOUND TRANSDUCER ARRAY

Katsuhiro Wakabayashi *et al.*, assignors to Olympus Optical Company, Limited
6 May 2003 (Class 600/437); filed in Japan 29 November 2000

The aim of this array design is a progressive ultrasound transducer array that minimizes the occurrence of cross talk and in which a common connection of the ground electrodes of piezoelectric elements can be reliably secured. This is to be achieved by bonding a layer of independently operated piezoelectric elements **6** to the acoustic radiation surface side and using a dicing machine to form divided grooves **16**. The divided grooves prevent the



generation of cross talk. Portions of the divided grooves not in contact with piezoelectric elements may be filled in with conductive adhesive to avoid a loss of structural strength of the array and to provide a common connection between the ground electrode **13b** on the bottom surface of each piezoelectric element and a conductive matching layer.—DRR

6,558,332

43.80.Qf ARRAY TYPE ULTRASONIC PROBE AND A METHOD OF MANUFACTURING THE SAME

Yasuo Shimizu, assignor to Nihon Dempa Kogyo Company, Limited
6 May 2003 (Class 600/459); filed in Japan 30 May 2001

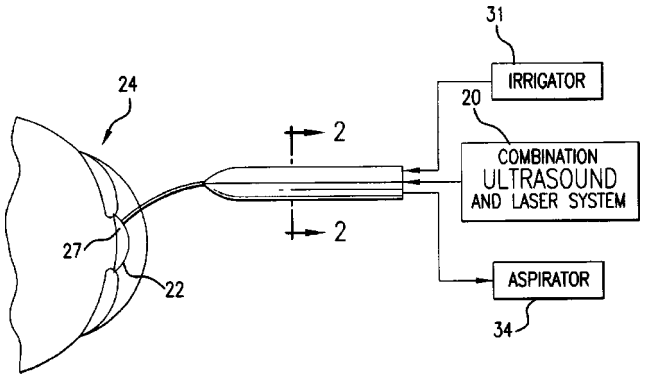
This ultrasonic probe is an array containing a number of rectangular piezoelectric elements on a backing material. Each piezoelectric element has upper and lower electrodes and an acoustic matching layer affixed to the upper face electrode. A slit at one end of a piezoelectric element electrically separates the lower face electrode. At the other end of the element, another slit extends from the piezoelectric element to the acoustic matching layer to separate the upper electrode. At the opposite ends of the piezoelectric element, a conductive adhesive electrically connects the upper and lower face electrodes.—DRR

6,544,254

43.80.Sh COMBINATION ULTRASOUND AND LASER METHOD AND APPARATUS FOR REMOVING CATARACT LENSES

Patricia Era Bath, Los Angeles, California
8 April 2003 (Class 606/6); filed 28 June 2000

This patent covers a method and apparatus for removing cataracts from the eye by the combined use of pulsed ultrasound and laser energy transmitted through an optical fiber delivery system. The parameters of the laser radiation are selected to optimize the processes of photo-phacoablation and photo-phacodisruption. The frequencies of the ultrasound are selected to



maximize cavitation to promote sono-fragmentation. Both laser radiation and ultrasound energy are delivered in effective combinations to maximize the removal of cataractous lens material.—DRR

6,551,337

43.80.Sh ULTRASONIC MEDICAL DEVICE OPERATING IN A TRANSVERSE MODE

Robert A. Rabiner *et al.*, assignors to OmniSonics Medical Technologies, Incorporated
22 April 2003 (Class 606/169); filed 19 July 2000

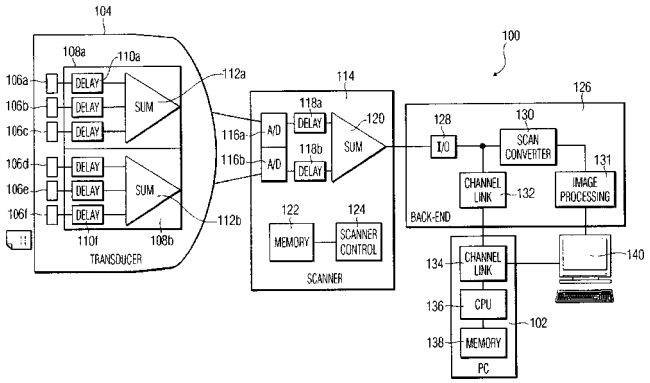
The object of this ultrasonic medical device is to destroy and emulsify tissue in a controlled fashion within a human body through cavitation by the means of a flexible probe operating in a transverse mode. The device consists of an ultrasonic vibration generator that produces vibration along its longitudinal axis, a flexible member coupled to the distal end of the transformer, and a section that supplies the longitudinal vibration at its base. The flexible member converts the longitudinal vibration into a standing wave that runs along the length of the flexible member, producing a series of nodes and anti-nodes along that member. Each of the anti-nodes generates cavitation in fluids contacting the probe. The cavitation of the fluids causes destruction of adjacent tissue and in this manner the entire length of the flexible member serves as a working surface for destroying tissue.—DRR

6,544,175

43.80.Vj ULTRASOUND APPARATUS AND METHODS FOR DISPLAY OF A VOLUME USING INTERLACED DATA

Richard M. Newman, assignor to Koninklijke Philips Electronics N.V.
8 April 2003 (Class 600/437); filed 15 September 2000

Ultrasonic signals are transmitted and echoes are received in an interleaved pattern to illuminate a number of regions at different times and the



echoes are processed to form a volumetric image that is based on the echoes from a number of frames.—RCW

6,544,178

43.80.Vj METHODS AND SYSTEMS FOR VOLUME RENDERING USING ULTRASOUND DATA

Stephen Michael Grenon *et al.*, assignors to Volumetrics Medical Imaging; Duke University
8 April 2003 (Class 600/443); filed 6 November 2000

An object is scanned to obtain an initial three-dimensional data set that represents the object and an instrument at an initial position. A volume rendering of at least a portion of the object and instrument is displayed. The position of the instrument can be adjusted relative to the object in the volume rendering. An additional three-dimensional data set representing the object and instrument in the second position is then obtained.—RCW

6,544,179

43.80.Vj ULTRASOUND IMAGING SYSTEM AND METHOD HAVING AUTOMATICALLY SELECTED TRANSMIT FOCAL POSITIONS

Daniel C. Schmiesing *et al.*, assignors to Koninklijke Philips Electronics N.V.
8 April 2003 (Class 600/447); filed 14 December 2001

A signal processor in this system analyzes image data and determines the number of transmit focal positions as well as the location for each focus based on an analysis of image data.—RCW

6,544,182

43.80.Vj ULTRASONIC NONLINEAR IMAGING AT FUNDAMENTAL FREQUENCIES

Michalakis Averkiou, assignor to Koninklijke Philips Electronics N.V.
8 April 2003 (Class 600/455); filed 20 November 2001

Nonlinear tissue or contrast agent response to an ultrasonic beam is detected below the second harmonic frequency band by combining echoes from multiple, differently modulated transmit pulses. The frequency band of received echoes may overlap the frequency band of the transmitted signals. Amplitude, phase, or polarity modulation, or both amplitude and phase modulation may be employed.—RCW

6,544,184

43.80.Vj IMAGING WITH REDUCED ARTIFACTS FOR MEDICAL DIAGNOSTIC ULTRASOUND

Ismayil M. Guracar, assignor to Acuson Corporation
8 April 2003 (Class 600/458); filed 28 August 2001

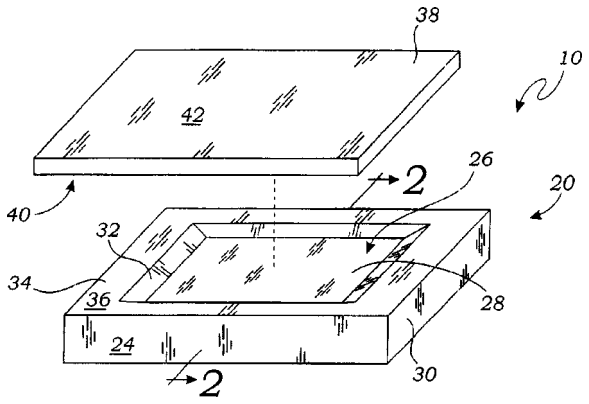
Echoes along a number of scan lines are received in response to a transmission. This process is repeated for the same scan lines to determine the loss of correlation or movement. Coherent echoes prior to detection along two or more scan lines are combined to remove differences. The combined data represents a synthesized line that is detected. Line data representing the presence or absence of contrast agents may be produced.—RCW

6,544,185

43.80.Vj ULTRASOUND IMAGING MARKER AND METHOD OF USE

Valentino Montegrande, Irvine, California
8 April 2003 (Class 600/458); filed 23 October 2001

The body of this marker is made from an ultrasound resonant material



and includes a resonant pocket.—RCW

6,547,733

43.80.Vj ULTRASOUND IMAGING APPARATUS AND METHOD USING GOLAY CODES WITH ORTHOGONAL PROPERTY

Jae Sub Hwang and Tai Kyong Song, assignors to Medison Company, Limited
15 April 2003 (Class 600/437); filed in the Republic of Korea 16 May 2001

The orthogonal property of Golay codes is used to transmit a number of ultrasound signals simultaneously to obtain an improved receive signal-to-noise ratio without a decrease in frame rate.—RCW

6,547,735

43.80.Vj PARTIAL RAYLINE VOLUMETRIC SCANNING ULTRASONIC DIAGNOSTIC IMAGING SYSTEM

Derek Henderson, assignor to Koninklijke Philips Electronics N.V.
15 April 2003 (Class 600/443); filed 5 December 2001

The acquired data set in this system is comprised of raylines extending over the full depth of field and raylines for which near-field image data are omitted. When scanning is accomplished with radially steered beams, this

results in a more uniform spatial sampling of both near and far regions. Processing and storage requirements are correspondingly reduced.—RCW

6,547,738

43.80.Vj METHODS AND APPARATUS FOR USING ULTRASOUND WITH CONTRAST AGENT

Peter Lysyansky, assignor to GE Medical Systems Global Technology Company, LLC
15 April 2003 (Class 600/458); filed 3 May 2001

Pulsed ultrasonic beams illuminate one target zone that has a higher concentration of contrast agent than another target zone. The two target zones have different ultrasonic responses. An image is formed showing the different responses in a single frame—RCW

6,551,244

43.80.Vj PARAMETRIC TRANSMIT WAVEFORM GENERATOR FOR MEDICAL ULTRASOUND IMAGING SYSTEM

Albert Gee, assignor to Acuson Corporation
22 April 2003 (Class 600/437); filed 17 October 2000

This generator uses stored parameters to define a complex transmit waveform. The waveform is represented in sections using a set of quadratic parameters for each section. Multiple transmit waveforms may be combined in a single channel and individual channels may be combined prior to excitation of transducer elements.—RCW

6,551,246

43.80.Vj METHOD AND APPARATUS FOR FORMING MEDICAL ULTRASOUND IMAGES

Kutay F. Ustuner *et al.*, assignors to Acuson Corporation
22 April 2003 (Class 600/447); filed 19 April 2001

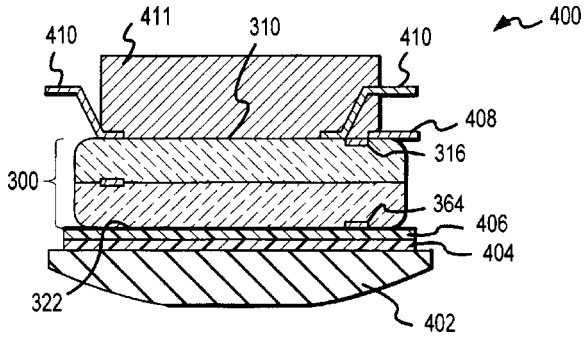
Echoes from an unfocused or weakly focused transmit beam are stored on each receive channel. The stored echoes are delayed and apodized to form individual beams to image desired areas in the region illuminated by the transmit beam. Images are synthesized using multiple receive beams.—RCW

6,552,471

43.80.Vj MULTI-PIEZOELECTRIC LAYER ULTRASONIC TRANSDUCER FOR MEDICAL IMAGING

Sanjay Chandran and David Chartrand, assignors to Parallel Design, Incorporated
22 April 2003 (Class 310/328); filed 27 January 2000

This transducer is tuned to a desired impedance by using multiple piezoelectric layers with patterns of conductive plating. The piezoelectric layers are placed so that positive and negative conducting regions on each



layer contact positive and negative regions on other layers. The layers are joined, for example, using epoxy.—RCW

6,554,770

43.80.Vj MEDICAL DIAGNOSTIC ULTRASOUND IMAGING METHODS FOR EXTENDED FIELD OF VIEW

Thilaka S. Sumanaweera *et al.*, assignors to Acuson Corporation
29 April 2003 (Class 600/443); filed 25 August 2000

Essentially coplanar b-scan images are aligned to form an extended field of view and variations in characteristics of each image are used to enhance the resulting extended field of view image.—RCW

6,554,771

43.80.Vj POSITION SENSOR IN ULTRASOUND TRANSDUCER PROBE

Vincentius Paulus Buil *et al.*, assignors to Koninklijke Philips Electronics N.V.
29 April 2003 (Class 600/459); filed 18 December 2001

The position of an ultrasonic probe on a subject is sensed relative to a region being imaged. Optical images of the probe and the surface of the subject are shown. The ultrasonic image and the probe position are combined into a single image. The probe position information is used to control the probe.—RCW

6,556,695

43.80.Vj METHOD FOR PRODUCING HIGH RESOLUTION REAL-TIME IMAGES, OF STRUCTURE AND FUNCTION DURING MEDICAL PROCEDURES

Douglas L. Packer and Richard A. Robb, assignors to Mayo Foundation for Medical Education and Research
29 April 2003 (Class 382/128); filed 16 September 1999

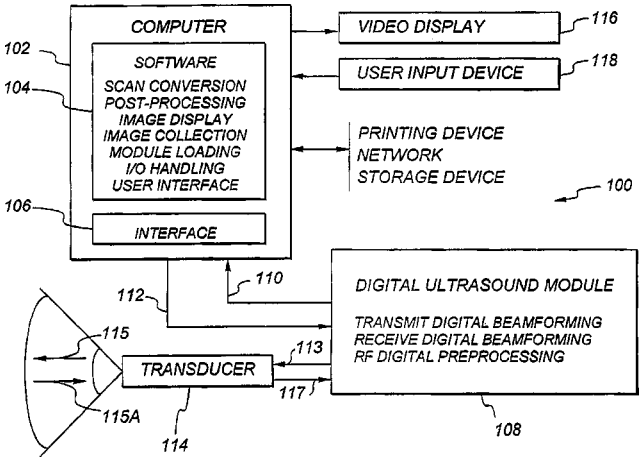
Real-time ultrasonic images are produced by an ultrasonic transducer inserted into the heart during a medical procedure such as endocardial physiology mapping and ablation. A high-resolution heart model is registered with the ultrasonic images and used to produce dynamic images for display during the medical procedure. An electrical activation map that depicts the spatial distribution of heart wall electrical activity is merged with anatomic images to facilitate cardiac ablation therapy.—RCW

6,558,326

43.80.Vj ULTRASOUND IMAGING SYSTEM

Laurent Pelissier, assignor to Ultrasonix Medical Corporation
6 May 2003 (Class 600/443); filed 5 September 2001

A module in this system contains independently programmable units that provide initial processing of signals. Real-time data collection, scan conversion, and display, as well as interfacing with a user, are accomplished



by a computer in the system. The system can be used for a wide range of imaging applications and can be integrated in a network of ultrasound devices having a client-server architecture.—RCW

6,558,328

43.80.Vj METHOD AND APPARATUS USING CODED EXCITATION IN ULTRASONIC PERFUSION IMAGING

Richard Yung Chiao *et al.*, assignors to General Electric Company
6 May 2003 (Class 600/447); filed 4 June 2001

Ultrasonic pulses that destroy contrast microbubbles in a transmit focal zone are scanned over a region of interest in one or more frames and are followed by low-amplitude encoded imaging pulses that are scanned in each frame. The low-amplitude encoded pulse train is used to image the contrast agent. The low amplitude prevents contrast bubbles in the transmit focal zone from being destroyed while they are being imaged and the coded excitation provides a suitable receive signal-to-noise ratio.—RCW

6,558,329

43.80.Vj MEDICAL ULTRASOUND RECEIVE ARCHITECTURE

Ronald Dennis Gatzke, assignor to Koninklijke Philips Electronics N.V.
6 May 2003 (Class 600/457); filed 28 April 2000

Data processing that includes continuous-wave Doppler processing is performed in this architecture that accepts an analog echo signal and produces a digital representation of the analog signal with an A/D conversion dynamic range of 154 dB.—RCW

On the assessment of shooting sounds: Loudness-level weightings versus A- and C-weighted sound exposure levels (L)

Joos Vos[a] and Frank W. M. Geurtsen
TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands

(Received 14 June 2002; revised 11 July 2003; accepted 21 July 2003)

As an alternative to the A-weighted sound exposure level (ASEL) Schomer *et al.* [J. Acoust. Soc. Am. **110**, 2390–2397 (2001)] used the equal-loudness level contours as a dynamic filter to determine the loudness-level-weighted sound exposure level (LLSEL). From their analyses they concluded that the LLSEL better orders and assesses traffic and impulsive sounds than the ASEL does. In the present study LLSELs were determined for a great variety of shooting sounds investigated by Vos [J. Acoust. Soc. Am. **109**, 244–253 (2001)]. It was concluded that for the assessment of shooting sounds, the use of loudness-level weightings is less adequate than the use of A- and C-weighted sound exposure levels as proposed in Vos's article. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1608961]

PACS numbers: 43.50.Pn, 43.50.Ba, 43.50.Qp [DKW]

I. INTRODUCTION

As a method of rating environmental sounds with respect to the expected community response, both ISO/R 1996 (1982) and ANSI S12.9 (1996) recommend the A-weighted equivalent sound pressure level (ALEQ) as the basic quantity. This does not mean, however, that aircraft, road-traffic, and railway sounds at numerically equal A-weighted average sound levels yield the same degree of annoyance also.

In the early 1980s it had already been shown that at equal A-weighted day–night levels, aircraft sounds are more annoying than road-traffic sounds, and that this difference increases with sound level (Hall *et al.*, 1981; Kryter, 1982). In contrast, again at comparable levels, railway sounds were found to be less annoying than road-traffic sounds (e.g., see Fields and Walker, 1982; Knall and Schümer, 1983). With the noise dose expressed as A-weighted levels, the various transportation sounds can only be consistently assessed by adding adjustments to the measured or calculated levels. The results from meta-analyses of field survey data reported by Finegold *et al.* (1994) and by Miedema and Vos (1998) confirmed the need for these adjustments, and suggested the appropriate size of these corrections.

As the C-weighted curve is very roughly the inverse of the 100-phon equal-loudness contour, so the A- and B-weighted curves are roughly the inverse of the 40- and 70-phon equal-loudness contours, respectively. These A-, B-, and C-weightings do not change with sound pressure level.

As an alternative to A-weighting, Schomer and colleagues (Schomer, 2000; Schomer *et al.*, 2001) recently suggested the use of the equal-loudness level contours as a dynamic filter that changes with sound level. Sound recordings from a variety of sources and conditions were analyzed to calculate loudness-level-weighted sound exposure levels (LLSELs), A-weighted sound exposure levels (ASELs; L_{AE}), and the differences between these levels.

Next, for the majority of these sources and conditions, the human response was expressed as the adjustment that had to be added to the (outdoor) ASEL to equate the ASEL of these sounds to the ASEL of an equally annoying vehicle sound. For the remaining conditions, these adjustments were estimated from relevant sets of dose-response relations (A-weighted day–night level versus percentages of ''highly annoyed'' residents) obtained in the meta-analyses from Finegold *et al.* (1994) and Miedema and Vos (1998).

From the high correlation between the adjustments and the differences between the LLSEL and ASEL relative to road traffic, Schomer *et al.* (2001) concluded that the LLSEL better orders and assesses transportation noise sources than the ASEL does. Moreover, they concluded that with the addition of an extra correction of 12 dB, the LLSEL also better orders and assesses highly impulsive sounds in relation to road-traffic sounds than the ASEL does.

In a previous study on the annoyance caused by impulse sounds produced by small, medium-large, and large firearms (Vos, 2001), it was shown that in the outdoor rating conditions, the annoyance was almost entirely determined by ASEL. The explained variance, r^2, in the mean ratings by ASEL was 0.95. In the indoor rating conditions with the windows closed, however, a satisfactory prediction of the annoyance could be obtained only if both the (outdoor) ASEL and the (outdoor) C-weighted sound exposure level (CSEL; L_{CE}) were taken into account. In the latter conditions, the explained variance in the annoyance ratings by ASEL was significantly increased from $r^2=0.87$ to $r^2=0.97$ by adding the product $(L_{CE}-L_{AE})(L_{AE}-45)$ as a second predictor. In a subsequent study (Vos, 2003), it was shown that the benefit of the second predictor was even much higher for façade attenuation types with higher degrees of sound isolation than applied in Vos (2001).

The results reported by Schomer *et al.* (2001) suggest that also for the shooting sounds investigated in Vos (2001), the predictability of the annoyance ratings from LLSELs might be superior to that from ASELs. The test provided in the present paper is especially interesting for the indoor rating conditions. If the LLSEL is much better in predicting the

[a]Electronic mail: vos@tm.tno.nl

TABLE I. Outdoor loudness-level-weighted sound exposure levels (LLSELs in dB) for the set of 56 impulses investigated in Vos (2001). The LLSELs were calculated both from the equal-loudness level contours defined in the current (Cur.) and from those defined in the proposed ISO 226 (Prop.).

Impulse type no.	Firearm/ammunition	Outdoor A-weighted sound exposure level							
		$L_{AE}=40$ dB		$L_{AE}=50$ dB		$L_{AE}=60$ dB		$L_{AE}=70$ dB	
		Cur.	Prop.	Cur.	Prop.	Cur.	Prop.	Cur.	Prop.
1	Pistol 9 mm	42	40	52	50	62	60	72	70
2	Rifle 7.62 mm	44	40	53	50	63	60	72	70
3	Rifle 0.30 in.	46	39	56	50	66	60	76	71
4	Howitzer 155 mm, charge 5M4[a]	48	40	58	51	68	61	78	72
5	Cannon 25 mm DST 127	48	38	59	50	69	61	78	72
6	Machine gun 0.5 in.	48	38	58	49	67	60	76	71
7	Howitzer 155 mm, charge 5M4[a]	48	36	60	49	69	61	79	72
8	Cannon 35 mm	50	38	61	50	71	61	80	72
9	Cannon 35 mm	49	37	60	50	71	61	80	72
10	Cannon 35 mm	49	35	61	48	71	61	80	72
11	Howitzer 155 mm, charge 5M4[a]	48	31	61	47	71	60	80	72
12	Hand grenade	49	29	62	43	74	59	85	73
13	Antitank weapon 84 mm	49	29	63	43	75	59	86	74
14	Howitzer 155 mm, charge 5M4	49	29	62	45	72	59	82	72

[a]With a spectral modification described in Table I of Vos (2001).

annoyance than the ASEL, then it might be redundant to apply Vos's second predictor.

In the analyses described in Schomer *et al.* (2001) the current equal-loudness level contours were used as the dynamic filter. Results of 12 independent experimental studies on loudness perception have led to a new Final Draft International Standard (ISO, 2003) that will replace the current ISO 226 (1987). The majority of these studies were reported between about 1990 and 2000, and the experimental conditions, such as sound field, stimuli, subject criterion, and psychometric method, satisfied the requirements set by the expert members of the pertinent Working Group.

Especially for low frequencies, the curves in the proposed ISO 226 are significantly higher than those given in the current ISO 226. For the sake of comparison with the results from Schomer *et al.* (2001), and on the assumption that the equal-loudness level contours in the proposed ISO 226 better match loudness perception than the curves in the current ISO 226 do, the suggested benefit of the LLSEL is in the present paper tested for both sets of contours.

II. METHODS

In line with the procedure followed by Schomer *et al.* (2001), the acoustic spectra of the shooting sounds investigated in Vos (2001) were determined for $\frac{1}{3}$-oct bands between 20 and 12 500 Hz with a Larson-Davis spectrum analyzer (Model 3200) in the position ''fast'' (integration time equal to 100 ms) and with a sampling time of 100 ms. The computer programs for further data processing were provided by Paul Schomer.

Outdoor LLSELs were determined for the complete set of 56 impulses (14 impulse types×4 levels). Moreover, indoor LLSELs were determined for the 35 impulses that had been presented in the indoor rating condition with the windows closed.

III. RESULTS

A. Indoor ratings as a function of outdoor levels

The mean indoor ratings for the 35 impulses [see Table II in Vos (2001)] and the corresponding outdoor LLSELs given in Table I were subjected to linear regression analyses.

The portion of the variance in the mean annoyance ratings explained by the LLSEL, as calculated from the current ISO 226 contours ($LLSEL_{cur}$), was equal to 89%. Expressed as decibel values, the root-mean-square (rms) of the differences between the obtained LLSEL and the LLSEL predicted by the regression line for the same annoyance rating was equal to 2.6 dB. With the LLSEL calculated from the proposed ISO 226 contours as the predictor ($LLSEL_{prop}$), the explained variance in the mean annoyance ratings was equal to 84% (rms=4.2 dB).

In Sec. I it was already mentioned that with the ASEL the explained variance in the annoyance ratings was equal to 87%, and with both the ASEL and CSEL as the predictors, the portion of the variance explained in the annoyance ratings was 97%. Here, the rms was as small as 1.6 dB. In the latter analysis, the rating sound level (L_r) was given by $L_r = L_{AE} + 12\ \text{dB} + 0.015(L_{CE} - L_{AE})(L_{AE} - 45)$ dB.

Consequently, the predictability of the annoyance ratings from the $LLSEL_{cur}$ was higher than that from the $LLSEL_{prop}$ and that from the ASEL, but lower than that from the ASEL and CSEL together.

B. Outdoor ratings as a function of outdoor levels

The mean outdoor ratings for the 56 impulses [see Table II in Vos (2001)] and the corresponding outdoor LLSELs given in Table I were subjected to linear regression analyses.

The portion of the variance in the mean annoyance ratings explained by the $LLSEL_{cur}$ was equal to 88% (rms=4.5 dB). With the $LLSEL_{prop}$ as the predictor, the explained variance in the mean annoyance ratings was equal to 93% (rms =3.7 dB).

TABLE II. Indoor loudness-level-weighted sound exposure levels (LLSELs in dB) for the set of 35 impulses investigated in Vos (2001). The LLSELs were calculated both from the equal-loudness level contours defined in the current (Cur.) and from those defined in the proposed ISO 226 (Prop.).

Impulse type no.	Outdoor A-weighted sound exposure level					
	$L_{AE}=50$ dB		$L_{AE}=60$ dB		$L_{AE}=70$ dB	
	Cur.	Prop.	Cur.	Prop.	Cur.	Prop.
1	⋯	⋯	32	31	42	40
2	⋯	⋯	38	33	47	43
3	⋯	⋯	42	34	53	44
4	⋯	⋯	43	36	53	46
5	⋯	⋯	46	36	55	46
6	⋯	⋯	44	35	54	45
7	⋯	⋯	46	36	55	46
8	37	27	48	37	58	48
9	37	26	48	37	57	48
10	38	26	49	36	59	47
11	39	24	49	36	58	46
12	42	23	55	35	67	49
13	43	23	56	36	68	50
14	41	23	52	35	62	47

With the outdoor ASEL, the explained variance in the annoyance ratings was equal to 95%, and with both the ASEL and CSEL as the predictors $[L_r=L_{AE}+12\,\text{dB}+0.015(L_{CE}-L_{AE})(L_{AE}-56.5)\,\text{dB}]$, the portion of the variance explained in the annoyance ratings was 96% (rms=2.8 dB).

As a result, the predictability of the annoyance ratings both from the $LLSEL_{cur}$ and from the $LLSEL_{prop}$ was lower than that from the ASEL and that from the ASEL and CSEL together.

C. Indoor ratings as a function of indoor levels

The mean indoor ratings for the 35 impulses and the corresponding indoor LLSELs given in Table II were subjected to linear regression analyses.

The portion of the variance in the mean annoyance ratings explained by the indoor $LLSEL_{cur}$, was equal to 76% (rms=4.9 dB). With the $LLSEL_{prop}$ as the predictor, the explained variance in the mean annoyance ratings was equal to 87% (rms=3.1 dB).

With the indoor ASEL, the explained variance in the annoyance ratings was equal to 88% (rms=2.5 dB), and with both the ASEL and CSEL as the predictors ($y=-6.95+0.35L_{AE}-0.049L_{CE}$), the portion of the variance explained in the annoyance ratings was 91%. With the ASEL and CSEL together, the rms was estimated to be equal to 2.2 dB. Table IV in Vos (2001) shows the indoor ASELs and CSELs.

Consequently, the predictability of the annoyance ratings from the $LLSEL_{prop}$ using the proposed contours was higher than that from the $LLSEL_{cur}$ and almost as high as that from the ASEL. With the ASEL and CSEL together, the predictability was 4 percent points higher than that with the $LLSEL_{prop}$.

IV. DISCUSSION

A. Predictive power of the acoustic measures

In the conditions in which the annoyance was predicted from the sound levels measured at the ears of the subjects (see Secs. III B and III C), the power of the $LLSEL_{prop}$ was higher than that of the $LLSEL_{cur}$.

With adoption of the proposed contours, the explained variance was 5–11 percent points higher than with adoption of the current contours, and the rms of the differences between predictions and observations was 0.8–1.8 dB smaller. However, the predictive power of the $LLSEL_{prop}$ was lower than that of the ASEL (in terms of explained variance 1–2 percent points) and that of the ASEL and CSEL together (in terms of explained variance 4–7 percent points). With the $LLSEL_{prop}$ the rms of the differences between predictions and observations was 0.9 dB higher than the rms obtained for the predictions on the basis of the ASEL and CSEL together.

In the condition in which the indoor annoyance was predicted from outdoor levels (see Sec. III A), i.e., the condition that is usually regarded as most relevant to noise zoning and land-use planning, the predictive power of the $LLSEL_{cur}$ was only 2 percent points higher than that of the ASEL, and 8 percent points lower than that of the ASEL and CSEL together. The variance in the annoyance ratings explained by the $LLSEL_{prop}$ was 13 percent points lower than the variance explained by the ASEL and CSEL together. In terms of rms-values, the accuracy of the LLSEL calculated from the contours that will be the standard in the near future was $4.2-1.6=2.6$ dB lower than that of the model that takes both the ASEL and CSEL into account.

B. Nature of the errors in the predictions

1. Sound levels measured at the ears of the subjects

In the conditions in which the outdoor ratings were related to outdoor levels (Sec. III B), $LLSELs_{cur}$ considerably underestimated the annoyance caused by the bangs produced by small firearms. With the $LLSEL_{prop}$ considerable underestimations were found for the bangs from the large firearms. Especially at outdoor ASELs of 50 and 60 dB, both LLSELs calculated from the current and from the proposed contours overestimated the annoyance for a number of bangs from medium-large firearms.

With ASEL and CSEL as the predictors, the annoyance was overestimated for a few bangs from small firearms. Moreover, the annoyance was underestimated in a portion of the impulses from medium-large firearms with bangs containing both projectile and muzzle components.

In the conditions in which the indoor ratings were related to indoor levels (Sec. III C), $LLSELs_{cur}$ strongly underestimated the annoyance caused by the bangs from the small firearms, and overestimated the annoyance caused by the bangs from the large firearms presented at outdoor ASELs of 50 and 60 dB. $LLSELs_{prop}$ underestimated the annoyance for the bangs from the large firearms, and overestimated the annoyance obtained for most of the bangs produced by the medium-large firearms.

With the ASEL and CSEL as the predictors, clear trends in the differences between predictions and observations were absent. The lack of such trends must be related to the low overall rms value obtained.

2. Indoor ratings as a function of outdoor levels

In the conditions in which the indoor annoyance was predicted from outdoor levels (Sec. III A), $LLSELs_{cur}$ underestimated the annoyance from the bangs produced by the small firearms, whereas $LLSELs_{prop}$ (strongly) underestimated the annoyance caused by the bangs from the large firearms.

Both $LLSELs_{cur}$ and $LLSELS_{prop}$ overestimated the annoyance caused by the bangs from medium-large firearms, although for $LLSELs_{cur}$ these overestimations were largely restricted to the bangs containing both projectile and muzzle components.

With ASEL and CSEL as the predictors, consistent trends in prediction errors could not be found. Again, the lack of such trends can be understood from the overall rms value, which in this condition was as low as 1.6 dB.

C. Discrepancy between current and proposed ISO 226 contours

Although the new Final Draft International Standard of ISO 226 (ISO, 2003) is not explicit about the differences among the proposed and current equal-loudness level contours, it is highly relevant to note that especially for the low frequencies between about 25 and 200 Hz, these differences are very large. For the frequency range indicated, the sound pressure levels required to yield loudness sensations of 20–70 phon are at least 10 dB higher in the proposed than in the current standard. The implications of these discrepancies are illustrated in Tables I and II of the present paper.

From the data given in these tables, it can be seen that the LLSELs calculated from the proposed ISO 226 contours were all lower than those calculated from the current contours, and that, consistent with the discrepancies described in more detail in Sec. IV B, the differences increased both with increasing firearm caliber and with decreasing overall ASEL of the bangs corresponding to greater distances between the source and the receiver. These effects can be understood from the relative contribution of the low-frequency components in the sound spectrum of the bangs, which increases both with weapon caliber and with source-receiver distance.

For a number of environmental sounds, illustrations of the differences in phon levels in the various $\frac{1}{3}$-oct bands as calculated from the proposed and current contours may be found in Schomer (2001).

V. CONCLUSIONS

(1) For the assessment of shooting sounds, the use of loudness-level weightings is less adequate than the use of A- and C-weighted sound exposure levels as proposed in Vos (2001).

(2) For the situation that is usually regarded as most relevant to noise zoning and land-use planning, i.e., the condition in which the indoor annoyance was predicted from outdoor levels, the variance in the annoyance ratings explained by LLSELs calculated from the proposed contours was 13 percent points lower than the variance explained by ASELs and CSELs together. In terms of rms values, the accuracy of LLSELs calculated from the contours that will be the standard in the near future was 2.6 dB lower than that of the model proposed in Vos (2001).

(3) Replacement of the current ISO 226 equal-loudness level contours by the proposed ISO 226 contours implies a considerable revision of loudness perception in the frequency range between about 25 and 200 Hz.

ACKNOWLEDGMENTS

The authors are grateful to Paul Schomer for providing the computer programs to determine the loudness-level-weighted sound exposure levels in line with the current and the proposed versions of ISO 226.

ANSI (**1996**). ANSI S12.9 Part 4, ''American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 4: Noise Assessment and Prediction of Long-term Community Response'' (Acoustical Society of America, New York).

Fields, J. M., and Walker, J. G. (**1982**). ''Comparing the relationships between noise level and annoyance in different surveys: a railway vs. aircraft and road traffic comparison,'' J. Sound Vib. **81**, 51–80.

Finegold, L. S., Harris, S. C., and von Gierke, H. E. (**1994**). ''Community annoyance and sleep disturbance: Updated criteria for assessing the impacts of general transportation noise on people,'' Noise Control Eng. J. **42**(1), 25–30.

Hall, F. L., Birnie, S. E., Taylor, S. M., and Palmer, J. E. (**1981**). ''Direct comparison of community response to road traffic noise and to aircraft noise,'' J. Acoust. Soc. Am. **70**, 1690–1698.

ISO (**1982**). ''Acoustics—Description and Measurement of Environmental Noise,'' ISO/R 1996 International Organization for Standardization, Geneva, Switzerland.

ISO (**1987**). ''Acoustics—Normal equal-loudness level contours,'' ISO 226, International Organization for Standardization, Geneva, Switzerland.

ISO (**2003**). ''Acoustics—Normal equal-loudness level contours,'' ISO/FDIS 226, International Organization for Standardization, Geneva, Switzerland.

Knall, V., and Schümer, R. (**1983**). ''The differing annoyance levels of rail and road traffic noise,'' J. Sound Vib. **87**, 321–326.

Kryter, K. D. (**1982**). ''Community annoyance from aircraft and ground vehicle noise,'' J. Acoust. Soc. Am. **72**, 1222–1242.

Miedema, H. M. E., and Vos, H. (**1998**). ''Exposure-response relationships for transportation noise,'' J. Acoust. Soc. Am. **104**, 3432–3445.

Schomer, P. D. (**2000**). ''Loudness-level weighting for environmental noise assessment,'' Acustica **86**, 49–61.

Schomer, P. D. (**2001**). ''Use of the new ISO 226 equal loudness contours as a filter to assess noise annoyance,'' in Proceedings Internoise 2001, The Hague, The Netherlands, pp. 1811–1816.

Schomer, P. D., Suzuki, Y., and Saito, F. (**2001**). ''Evaluation of loudness-level weightings for assessing the annoyance of environmental noise,'' J. Acoust. Soc. Am. **110**(5), 2390–2397.

Vos, J. (**2001**). ''On the annoyance caused by impulse sounds produced by small, medium-large, and large firearms,'' J. Acoust. Soc. Am. **109**, 244–253.

Vos, J. (**2003**). ''A- and C-weighted sound levels as predictors of the annoyance caused by shooting sounds, for various façade attenuation types,'' J. Acoust. Soc. Am. **113**, 336–347.

Random walk approach to wave propagation in wedges and cones

Bair V. Budaev[a] and David B. Bogy[b]
Department of Mechanical Engineering, University of California, Berkeley, California 94720

(Received 14 November 2002; revised 14 June 2003; accepted 11 July 2003)

Two- and three-dimensional Helmholtz equations in wedge-shaped and conical domains are addressed by the random walk method. The solutions of the Dirichlet problems in such domains are represented as mathematical expectations of specified functionals on trajectories of multidimensional random motions whose radial components run in a complex space while the angular components remain real valued. This technique is applied to the Sommerfeld problem of diffraction by a semi-infinite screen which is explicitly solved here in the probabilistic form. The numerical results confirm the efficiency of the random walk approach to the analysis of wave propagation. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1605413]

PACS numbers: 43.20.Bi, 43.20.Fn [JGH]

I. INTRODUCTION

It was recently shown[1–4] that problems of wave propagation can be addressed by a combination of the ray method,[5,6] which is a common tool of the theory of wave propagation, and of the random walk method which delivers exact Feynman–Kac solutions[7,8] of second-order differential equations. This combination approach starts similarly to the ray method in the sense that the solution of the Helmholtz equation

$$\nabla^2 U + k^2 U = 0, \tag{1}$$

is represented in the Liouville form

$$U = u e^{ikS}, \tag{2}$$

with the eikonal $S(x)$ satisfying the eikonal equation

$$(\nabla S)^2 = k^2, \tag{3}$$

which can be constructively solved by the Hamilton–Jacobi method.[9,6] Then, instead of following the ray method, where the amplitude $u(x)$ is determined from the approximate first-order transport equation $\vec{\nabla} S \cdot \vec{\nabla} u + (\nabla^2 S) u = 0$, this amplitude is determined from the complete transport equation

$$\frac{1}{2ik} \nabla^2 u + \vec{\nabla} S \cdot \vec{\nabla} u + \frac{1}{2} (\nabla^2 S) u = 0, \tag{4}$$

whose exact solution is delivered by the Feynman–Kac formula, and which is equivalent to the Helmholtz equation (1) considered together with (1) and (2).

The first probabilistic solutions of partial differential equations were obtained in the 1920–1930s[10–13] as a result of the analysis of the Laplace equation by the finite difference scheme. Rapid progress in the development of probabilistic methods in partial differential equations was made in the 1950s, after the publication of landmark papers of Feynman[14,15] and Kac.[16] These theories have a long record of successful applications to numerical simulation of evolutions of quantum systems, and quite recently attempts have been made to apply path integral methods to acoustics[17] and electromagnetics.[18] Probabilistic methods have also been applied for the analysis of transport of energy by waves propagating in random media.[19,20]

In our previous papers it was shown that the probabilistic solutions of the Helmholtz equation, as the exact solutions, make it possible to provide a unified description of different phenomena of wave propagation. Thus, in papers[1,3,4] the probabilistic formulas were used for derivation of well-known asymptotes of the ray method. It was also demonstrated[2] that the random walk method provides a natural explanation of such typical phenomena of wave propagation as backscattering, which is predicted neither by the ray theory nor by a more general method of parabolic equations.[21,22] Here, we continue our development of the random walk approach to wave propagation, but, unlike the previous publications dealing mostly with general aspects of the method, the focus of the present paper is shifted to specific problems arising in the analysis of two and three-dimensional waves radiating from wedge-shaped or conical domains. In the next paper[23] the technique developed here is applied to an important three-dimensional problem of diffraction by a perfectly reflecting plane sector of arbitrary angle.

II. PROBABILISTIC FORMULAS FOR DIFFERENTIAL EQUATIONS WITH COMPLEX COEFFICIENTS

It is well known that many problems involving second-order elliptic differential equations admit explicit probabilistic solutions.[7,8] Thus, the classical Dirichlet problem

$$\sum_{n=1}^{N} \left(\frac{\sigma_n^2}{2} \frac{\partial^2 u}{\partial x_n^2} + A_n \frac{\partial u}{\partial x_n} \right) + Bu = 0, \quad u|_{\partial G} = f, \tag{5}$$

with real-valued coefficients σ_n, A_n has the solution

$$u(x) = \mathbf{E}\{ f(\xi_\tau) e^{\int_0^\tau B(\xi_s) ds} \}, \tag{6}$$

where $\mathbf{E}$ denotes the average computed over trajectories of the random motions

[a]Electronic mail: budaev@cml.me.berkeley.edu
[b]Electronic mail: dbogy@cml.me.berkeley.edu

$$\xi_t=(\xi_t^1,\xi_t^2,\ldots,\xi_t^N), \tag{7}$$

running across the N-dimensional domain G in accordance with the rules described below. The motion ξ_t starts at $t=0$ from the point $\xi_0=x$, and it stops at the exit time τ defined as the first instant when ξ_t touches the boundary ∂G. In the time interval from $t=0$ to $t=\tau$, this motion is controlled by a system of N stochastic differential equations

$$\mathrm{d}\xi_t^n=\sigma_n(\xi_t)\mathrm{d}w_t^n+A_n(\xi_t)\mathrm{d}t,$$
$$\xi_0^n=x_n, \quad n=1,2,\ldots,N, \tag{8}$$

enumerated by the index n and driven by independent one-dimensional Brownian motions w_t^n. The representation (6)–(8) is widely known as the Feynman–Kac solution, and in a monograph[24] one may find an extensive bibliography providing references to numerous proofs of these formulas based on different ideas and applying different hypotheses on the regularity of the coefficients σ_n, A_n, B, and on the regularity of the boundary ∂G from (5).

The theory of wave propagation customarily leads to equations of the type (5) with complex-valued, analytic coefficients. As shown in the previous papers,[3,4] solutions of such equations can be represented by the Feynman–Kac formulas. However, since the random walks corresponding to the equations with complex coefficients may run in the complex space, the applications of the Feynman–Kac formulas to boundary value problems with complex coefficients is not straightforward and has specific features which are briefly discussed below.

Let $\mathfrak{G}\subset\mathbb{C}^N$ be a $2N$-dimensional domain whose intersection with the real-space $\mathbb{R}^N$ coincides with G. Then, as shown in Refs. 3 and 4, the analytic solution of the problem (5) can be represented by the formulas (6), (8) with the exit time τ defined as the first moment when ξ_t hits $\partial\mathfrak{G}$. These formulas, however, do not necessarily deliver the solution of the Dirichlet problem (5) with complex-valued coefficients. Indeed, if σ_n and A_n are not real, then the trajectories ξ_t defined by (6) leave the real space and, therefore, the right-hand side of (6) involves values of $f(x)$ at complex points $x=\xi_\tau\in\partial\mathfrak{G}$, where $f(x)$ may not be defined because the boundary condition from (5) assumes that $f(x)$ is defined on the real part $\partial G=\partial\mathfrak{G}\cap\mathbb{R}^N$ of the boundary $\partial\mathfrak{G}$. This difficulty is rather fundamental and it cannot be overcome by analytic continuation, which cannot extend $f(x)$ from the $(N-1)$-dimensional boundary $\partial G\subset\mathbb{R}^N$ to the $(2N-1)$-dimensional surface $\partial\mathfrak{G}\subset\mathbb{C}^N$. The analytic continuation may extend $f(x)$ to the $(2N-2)$-dimensional analytic continuation ∂G^c of the boundary ∂G, but, in general, a $(2N-2)$-dimensional object ∂G^c has zero probability of being hit by a one-dimensional trajectory ξ_t running in the $2N$-dimensional space $\mathbb{C}^N$.

The last comment identifies the main obstacle in the application of the random walk method to the analysis of boundary value problems (5) with complex coefficients. In many cases, however, this obstacle can be overcome by the following generalization of the Feynman–Kac formulas.

Let us multiply Eq. (5) by an arbitrary function $q^2(x)$ that is nonvanishing inside G. Then, treating the product as another equation of the type (5), we apply (6), (8) and derive the expression

$$u(x)=\mathbf{E}\{f(\xi_\tau)\mathrm{e}^{\int_0^\tau q^2(\xi_s)B(\xi_s)\mathrm{d}s}\}, \quad \xi_\tau\in\partial\mathfrak{G}, \tag{9}$$

where the averaging is computed over the trajectories of the random motion $\xi_t=(\xi_t^1,\xi_t^2,\ldots,\xi_t^N)$ governed by the the system of N stochastic differential equation

$$\mathrm{d}\xi_t^n=q(\xi_t)\sigma_n(\xi_t)\mathrm{d}w_t^n+q^2(\xi_t)A_n(\xi_t)\mathrm{d}t,$$
$$\xi_0^n=x_n, \quad n=1,2,\ldots,N, \tag{10}$$

enumerated by the index n. If $q(\xi)$ is real valued, then the random motions determined by the systems of stochastic equations (8) and (10) differ from each other only by the time scale, so that the representation (9) with real-valued $q(\xi)$ does not offer anything new compared to (6). On the other hand, if $q(x)$ is complex valued, then Eqs. (8) and (10) determine different sets of trajectories, which makes the representations (6) and (9) distinctly different because they involve values of $f(\xi_t)$ computed at different sets of exit points.

The difference between trajectories (10) with real or complex factors $q(x)$ is well illustrated by a degenerate one-dimensional example with $N=1$, $\sigma_1=0$, $A_1=1$, and $q=\mathrm{const}$. In this case, Eqs. (8) and (10) deterministically define uniform motions $\xi_t=x+t$ and $\xi_t=x+q^2t$ along straight rays. If q is real, then both of these rays coincide with the positive half of the real coordinate axis. However, if q is complex these rays have different complex directions and have only one common point $\xi_0=x$.

The presence of an arbitrary complex-valued factor $q(x)$ makes formulas (9)–(10) much more versatile than the original Feynman–Kac formulas (6)–(8). In particular, in many cases $q(x)$ can be selected by rules which ensure that the trajectories ξ_t controlled by (10) exit the domain $\partial\mathfrak{G}\subset\mathbb{C}^N$ through the analytic continuation ∂G^c of boundary ∂G, where the function $f(x)$ is defined as an analytic continuation of the Dirichlet data $f(x)$ from (5). The rules of such selection of $q(x)$ are based on a few simple principles that are presented in the following sections, where we consider some particular, but representative, problems of the type (5) with complex coefficients.

III. WAVES RADIATING FROM A PLANE WEDGE

Here, we focus on the computation of waves radiating from an infinite wedge

$$G=\{r,\theta:\ r>0,\ \alpha_1<\theta<\alpha_2\}, \tag{11}$$

defined in terms of standard polar coordinates (r,θ). Such waves are described by the solution of the Helmholtz equation

$$\nabla^2U+k^2U=0, \quad k=\mathrm{const}, \tag{12}$$

satisfying the Dirichlet boundary conditions

$$U(r,\alpha_1)=F_1(r), \quad U(r,\alpha_2)=F_2(r), \tag{13}$$

and the radiation condition at infinity

$$U(r,\theta)=C(\theta)\frac{e^{ikr}}{\sqrt{kr}}(1+o(1)),\quad r\to\infty. \tag{14}$$

We seek the solution of the Helmholtz equation in the Liouville form

$$U(r,\theta)=u(r,\theta)e^{ikr}, \tag{15}$$

where u is a new unknown function. Straightforward substitution of (15) into (12) leads to the requirement that $u(r,\theta)$ has to solve the Dirichlet problem

$$\begin{aligned}&\frac{i}{2k}\nabla^2 u-\nabla r\cdot\nabla u-\frac{1}{2}(\nabla^2 r)u=0,\\&u(r,\alpha_m)=F_m(r)e^{-ikr},\quad m=1,2,\end{aligned} \tag{16}$$

with the boundary values completely determined by (13). Finally, writing (16) in polar coordinates we arrive at the Dirichlet problem

$$\frac{i}{2k}\left(\frac{\partial^2 u}{\partial r^2}+\frac{1}{r^2}\frac{\partial^2 u}{\partial\theta^2}\right)+\left(\frac{i}{2kr}-1\right)\frac{\partial u}{\partial r}-\frac{u}{2r}=0,$$

$$u|_{\partial G}=f, \tag{17}$$

where the unified boundary function $f(r,\theta)$ is introduced as

$$f(r,\alpha_m)=F_m(r)e^{-ikr},\quad m=1,2. \tag{18}$$

If the functions $F_1(r)$ and $F_2(r)$ from (13) are analytic, then the solution $u(r,\theta)$ of the problem (17) is also analytic with respect to both arguments r and θ. Therefore, this solution can be represented by the Feynman–Kac formula (9), which has the structure

$$u(r,\theta)=\mathbf{E}\left\{u(\xi_\tau^1,\xi_\tau^2)\exp\left(-\int_0^\tau\frac{q^2(\xi_s^1,\xi_s^2)ds}{2\xi_s^1}\right)\right\},$$

$$(\xi_\tau^1,\xi_\tau^2)\in\partial\mathfrak{G}, \tag{19}$$

where $q(r,\theta)$ is an indefinite factor; $\mathfrak{G}\subset\mathbb{C}^2$ is a four-dimensional domain whose intersection with $\mathbb{R}^2$ coincides with G from (11); and $\xi_t=(\xi_t^1,\xi_t^2)$ is the random process governed by the stochastic equations

$$\begin{aligned}&d\xi_t^1=q(\xi_t^1,\xi_t^2)\sqrt{\frac{i}{k}}dw_t^1+q^2(\xi_t^1,\xi_t^2)\left(\frac{i}{2k\xi_t^1}-1\right)dt,\\&\xi_0^1=r,\\&d\xi_t^2=q(\xi_t^1,\xi_t^2)\sqrt{\frac{i}{k}}\frac{dw_t^2}{\xi_t^1},\qquad \xi_0^2=\theta,\end{aligned} \tag{20}$$

driven by the standard one-dimensional Brownian motions w_t^1 and w_t^2.

In general, formula (19) does not solve the problem (17), because its right-hand side involves values $u(\xi_\tau^1,\xi_\tau^2)$ of the unknown function on the three-dimensional boundary $\partial\mathfrak{G}$ which extends outside of the line ∂G, where $u(r,\theta)$ is defined by the boundary conditions from (17). However, we will see below that the appropriate selections of the factor $q(r,\theta)$ and the domain $\mathfrak{G}$ make it possible to obtain the solution of the problem (17) in the form (19).

Let $q(r,\theta)$ be defined by

$$q^2(r)=-ikr^2. \tag{21}$$

Then, Eq. (20) controlling the random motions ξ_t^1 and ξ_t^2 are simplified to the forms

$$d\xi_t^1=\xi_t^1 dw_t^1+\xi_t^1(\tfrac{1}{2}+ik\xi_t^1)dt,\quad \xi_0^1=r, \tag{22}$$

$$d\xi_t^2=dw_t^2,\quad \xi_0^2=\theta, \tag{23}$$

which implies that ξ_t^2 is the standard one-dimensional Brownian motion along the real axis and that ξ_t^2 may leave the strip $\alpha_1<\mathrm{Re}(\xi_t^2)<\alpha_2$ only through one of two points, $\theta=\alpha_1$ or $\theta=\alpha_2$. Therefore, introducing the domain $\mathfrak{G}\subset\mathbb{C}^2$ as

$$\mathfrak{G}=\{r,\theta:\quad r\in\mathbb{C},\quad \alpha_1<\mathrm{Re}(\theta)<\alpha_2\}, \tag{24}$$

we rearrange (19) to the form

$$u(r,\theta)=\mathbf{E}\left\{f(\xi_\tau^1,\xi_\tau^2)\exp\left(\frac{ik}{2}\int_0^\tau\xi_s^1 ds\right)\right\},\quad \xi_\tau^2=\alpha_{1,2}, \tag{25}$$

where τ is the "exit time" defined as the moment when ξ_τ^2 from (23) leaves the interval (α_1,α_2). If the functions $f(r,\alpha_1)$ and $f(r,\alpha_2)$ admit analytic continuations with respect to r from the real semiaxis to a complex domain accessible to the random motion ξ_t^1, then the mathematical expectation (25) represents the solution of the Dirichlet problem (17) in an infinite wedge (11).

It is remarkable that the solution (25) admits an alternative representation in the form

$$u(r,\theta)=\frac{1}{\sqrt{r}}\mathbf{E}\{f(\xi_\tau^1,\xi_\tau^2)\sqrt{\xi_\tau^1}e^{-(1/2)dw_\tau^1}\},\quad \xi_0=(r,\theta), \tag{26}$$

which emphasizes the structure of $u(r,\theta)$ as $r\to\infty$, suggested by the radiation condition (14). To show this we note that Eq. (22) yields

$$ik\xi_t^1 dt=\frac{d\xi_t^1}{\xi_t^1}-\frac{1}{2}dt-dw_t^1, \tag{27}$$

and Ito's formula of stochastic differentiation[7,25] generates the identity

$$d\ln(\xi_t^1)=\frac{d\xi_t^1}{\xi_t^1}-\frac{1}{2}\left(\frac{d\xi_t^1}{\xi_t^1}\right)^2\equiv\frac{d\xi_t^1}{\xi_t^1}-\frac{dt}{2}, \tag{28}$$

whose combination with (27) leads to the expression

$$ik\xi_t^1 dt=d\ln(\xi_t^1)-dw_t^1. \tag{29}$$

Then, integrating (29) and taking into account the initial condition $\xi_0^1=r$, we get the expression

$$\int_0^\tau ik\xi_t^1\,dt=\ln(\xi_\tau^1)-\ln(r)-\int_0^\tau dw_t^1, \tag{30}$$

whose substitution into (25) generates (26).

Before proceeding to applications of the formulas (25) or (26), it is instructive to discuss in more detail the structure of the trajectories of the random motion ξ_t^1 controlled by the

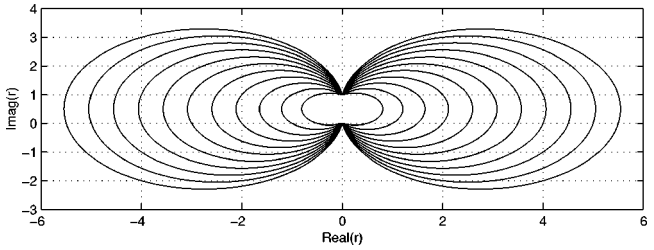


FIG. 1. Radial complex rays.

stochastic equation (22). In particular, it is important to have an indication as to where on the complex plane the point ξ_t^1 may be expected to appear.

First, we notice that the trajectories of ξ_t^1 never intersect the imaginary axis, which means that if the motion ξ_t^1 is launched from the positive real point $\xi_t^1 = r > 0$ then the inequality $\mathrm{Re}(\xi_t^1) \geqslant 0$ remains valid at any time t. Indeed, if at some moment $t = t_0$ the point ξ_t^1 appears on the imaginary axis, then ξ_t^1 can be represented as $\xi_t^1 = \mathrm{i}\eta_t$, with η_t determined by the Cauchy problem

$$\mathrm{d}\eta_t = \eta_t[\mathrm{d}w_t^1 + (\tfrac{1}{2} + k\eta_t)\mathrm{d}t], \quad \eta_{t_0} = -\mathrm{i}\xi_{t_0}^1 \in \mathbb{R}. \tag{31}$$

Since the coefficients and initial data of the problem (31) are real valued, its solution η_t is also real valued, and, therefore, points $\xi_t^1 = \mathrm{i}\eta_t$ are all on the imaginary axis.

It is well-known that the random motion ξ_t^1 controlled by the stochastic equation (22) can be regarded as a superposition of random fluctuations with a deterministic drift along the integral lines of the ordinary differential equation

$$\mathrm{d}\zeta_t = \zeta_t(\tfrac{1}{2} + \mathrm{i}k\zeta_t)\mathrm{d}t, \tag{32}$$

corresponding to the last term of Eq. (22). Comparing the derivation of Eq. (32) with the definition of rays customarily used in the ray method,[6] one may interpret the lines determined by (32) as radial components of complex rays associated with the Helmholtz equation in a wedge.

From the general solution of the equation (32)

$$\zeta_t = \frac{\zeta_0 \mathrm{e}^{t/2}}{1 + 2\mathrm{i}k\zeta_0(1 - \mathrm{e}^{t/2})}, \tag{33}$$

we easily see that all of its integral lines can be divided into two groups: three trivial straight lines and a family of generic curves. One trivial line runs from $\zeta = \mathrm{i}\infty$ down to the singular point $\zeta = \mathrm{i}/2k$, the second trivial line runs from $\zeta = 0$ to $\zeta = -\mathrm{i}\infty$, and the last trivial line goes from $\zeta = 0$ up to $\zeta = \mathrm{i}/2k$. All other integral lines of (32) start from the point $\zeta = 0$, cross the real axis at $\zeta = \zeta_0$, and end at $\zeta = \mathrm{i}/2k$, as shown in Fig. 1, corresponding to the case with $k = 1/2$. It is worth mentioning that all trajectories of (33) launched from the real axis are localized in the upper semiplane $\mathrm{Im}(\zeta_t) > 0$.

Figure 2 shows typical examples of trajectories of the random motion ξ_t^1 described by Eq. (22) with different values of the parameter k. The first diagram corresponds to $k = 2$ and the second diagram corresponds to $k = 10$. As expected, trajectories ξ_t^1 determined by Eq. (22) with the larger value of k are less random, and they follow closer to the

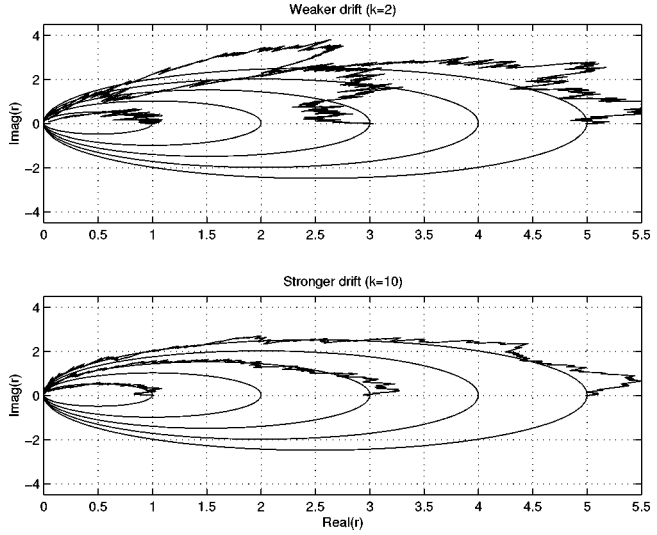


FIG. 2. Samples of radial random walks.

deterministic integral lines of (33), also shown in Fig. 2. These examples clearly show that the random trajectories of ξ_t^1 originating from the real axis are mostly located in the domain $\mathrm{Im}(\xi_t^1) > 0$, which is explained by the drift of the random motion ξ_t^1 along the integral lines of (33) running, as mentioned above, from the real line to the upper half-plane. Such a trend of the trajectories of ξ_t^1 is important because it provides convergence of the integration in (25).

To verify the solution (25) and to assess its practicability we used it for the probabilistic simulation of the Hankel function $U(r, \theta) = H_0^1(kr)$. We considered this function as the solution of the Helmholtz equation (12) satisfying the radiation condition (14) and the boundary conditions (13) with $F_1(r) = F_2(r) = H_0^{(0)}(kr)$, imposed on the faces $\theta = \alpha_1$ and $\theta = \alpha_2$ of the wedge G from (11). For numerical experiments we set the wave number to the value $k = 1$, and the boundaries of the wedge G were fixed by the conditions $\alpha_1 = 0$ and $\alpha_2 = 360°$. Continuous one-dimensional random motions ξ_t^1 and ξ_t^2 defined by the stochastic equation (22) were approximated by the discrete walks $r \equiv \xi_0^1 \to \cdots \to \xi_n^1 \to \xi_{n+1}^1 \to \cdots$ and $\theta \equiv \xi_0^2 \to \cdots \to \xi_n^2 \to \xi_{n+1}^2 \to \cdots$, consisting of instant jumps

$$\xi_n^1 \to \xi_{n+1}^1 = \xi_n^1 \pm \varepsilon\xi_n^1 + \varepsilon^2\xi_n^1(\tfrac{1}{2} + \mathrm{i}\xi_n^1),$$

and (34)

$$\xi_n^2 \to \xi_{n+1}^2 = \xi_n^2 \pm \varepsilon,$$

following one after another with the time increment $\Delta t = \varepsilon^2$.

Figure 3 presents results of the numerical simulation of $U(r, \theta) = H_0^{(1)}(r)$ along the ray $\theta = 60°$. The first two diagrams show real and imaginary parts of this function, while the third diagram shows the absolute error of the probabilistic simulation. Exact values are plotted by continuous lines and simulated values are marked by dots. The mathematical expectation in (25) was approximated by averaging over trajectories of 2000 discrete random walks (34) with $\varepsilon = 0.1$. Computations were stable over the entire trial interval from $r = 0.5$ to $r = 50$, and the error of simulation was under 0.01

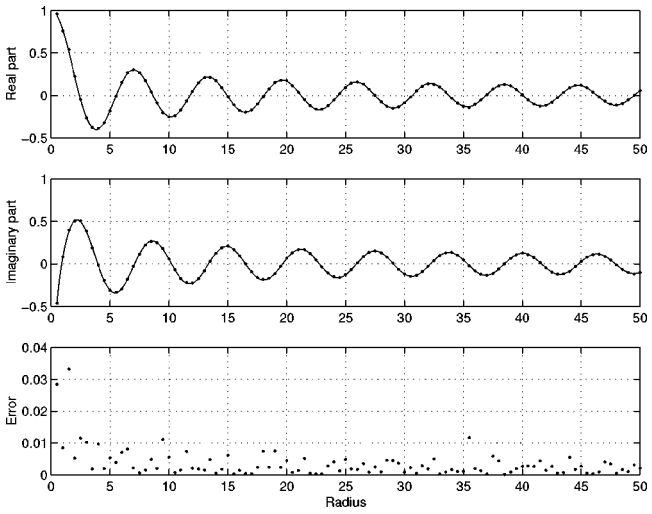


FIG. 3. Random walk simulation of $H_0^{(1)}(r)$.

virtually anywhere, except in a close range from the origin $r=0$, where $H_0^{(1)}(r)$ has a logarithmic singularity.

IV. THE SOMMERFELD PROBLEM OF DIFFRACTION

The Sommerfeld problem of diffraction is often formulated as the problem of computing the solution $U(r,\theta)$ of the Helmholtz equation

$$\nabla^2 U + k^2 U = 0, \tag{35}$$

that is defined in the domain $r>0$, $0<\theta<2\pi$, satisfies the boundary conditions

$$U(r,0) = U(r,2\pi) = 0, \tag{36}$$

and admits the decomposition

$$U = U_0 + U_1 + U_d, \tag{37}$$

comprising two predefined piecewise continuous components

$$U_0(r,\theta;\theta_0) = \begin{cases} e^{-ikr\cos(\theta-\theta_0)}, & \text{if } 0<\theta<\pi+\theta_0<2\pi, \\ 0, & \text{if } \pi+\theta_0<\theta<2\pi, \end{cases} \tag{38}$$

$$U_1(r,\theta;\theta_0) = \begin{cases} -e^{-ikr\cos(\theta+\theta_0)}, & \text{if } 0<\theta<\pi-\theta_0<2\pi, \\ 0, & \text{if } \pi-\theta_0<\theta<2\pi, \end{cases} \tag{39}$$

and an unknown component $U_d(r,\theta;\theta_0)$, satisfying the radiation condition

$$U_d(r,\theta;\theta_0) = C(\theta)\frac{e^{ikr}}{\sqrt{kr}}(1+o(1)), \quad r\to\infty. \tag{40}$$

The known components $U_0(r,\theta;\theta_0)$ and $U_1(r,\theta;\theta_0)$ from (38) and (39) are usually referred to as the incident and reflected waves, respectively, while $U_d(r,\theta;\theta_0)$ is called the diffracted wave. As for the parameter θ_0, it is defined here as the polar angle of the ray along which the incident wave arrives from infinity.

The expressions (38) and (39) show that the sum $\tilde{U} = U_0 + U_1$ of the incident and reflected waves satisfies the equation $\nabla^2\tilde{U} + k^2\tilde{U} = 0$ as well as the boundary conditions $\tilde{U}(r,0;\theta_0) = \tilde{U}(r,2\pi;\theta_0) = 0$. This observation makes it clear that the Sommerfeld problem can be reformulated as the Dirichlet problem for $U_d(r,\theta;\theta_0)$

$$\nabla^2 U_d + k^2 U_d = 0, \quad U_d(r,0;\theta_0) = U_d(r,2\pi;\theta_0) = 0, \tag{41}$$

accompanied by the radiation condition (14) and by the interface conditions

$$U_d\big|_{\theta=\theta_m+0} = U_d\big|_{\theta=\theta_m-0} + (-1)^m e^{ikr}, \quad m=1,2, \tag{42}$$

$$\left.\frac{\partial U_d}{\partial\theta}\right|_{\theta=\theta_m+0} = \left.\frac{\partial U_d}{\partial\theta}\right|_{\theta=\theta_m-0}, \qquad m=1,2, \tag{43}$$

formulated on the rays

$$\theta=\theta_m, \quad \text{where } \theta_1 = \pi+\theta_0, \quad \theta_2 = \pi-\theta_0. \tag{44}$$

It should be mentioned that although, according to (42), the diffracted field $U_d(r,\theta;\theta_0)$ has discontinuities on the interfaces $\theta=\theta_m$, the total field $U(r,\theta;\theta_0) = \tilde{U}(r,\theta;\theta_0) + U_d(r,\theta;\theta_0)$ remains continuous in the entire domain $0<\theta<2\pi$.

We seek the solution of (41)–(43) in the product form

$$U_d(r,\theta;\theta_0) = u(r,\theta;\theta_0)e^{ikr}, \tag{45}$$

similar to (15). Then, substituting (45) into (41) we find that $u(r,\theta;\theta_0)$ must satisfy the equation

$$\frac{1}{2}\left(r^2\frac{\partial^2 u}{\partial r^2} + \frac{\partial^2 u}{\partial\theta^2}\right) + r\left(\frac{1}{2}+ikr\right)\frac{\partial u}{\partial r} + \frac{ikr}{2}u = 0, \tag{46}$$

which coincides with the equations from (17) multiplied by the factor $q^2 = -ikr^2$ from (21). Additionally, $u(r,\theta)$ has to obey the radiation condition at infinity

$$u(r,\theta;\theta_0) = O(1/\sqrt{r}), \quad r\to\infty, \tag{47}$$

equivalent to (40), as well as the boundary and interface conditions

$$u(r,0;\theta_0) = u(r,2\pi;\theta_0) = 0, \tag{48}$$

$$u(r,\theta_m+0;\theta_0) = u(r,\theta_m-0;\theta_0) + (-1)^m, \quad m=1,2, \tag{49}$$

$$u'_\theta(r,\theta_m+0;\theta_0) = u'_\theta(r,\theta_m-0;\theta_0), \quad m=1,2, \tag{50}$$

equivalent to (36), (42) and (43).

It will be shown below that the solution of the problem (46)–(47) can be represented by the explicit probabilistic formula

$$u(r,\theta;\theta_0) = \frac{1}{2}\mathbf{E}\left\{\sum_{\tau_\nu<\tau}(-1)^{m_\nu}\delta(\xi^2_{\tau_\nu}) \times\exp\left(\frac{ik}{2}\int_0^{\tau_\nu}\xi^1_s\,ds\right)\right\}, \quad \xi^2_{\tau_\nu} = \theta_{m_\nu}, \tag{51}$$

where ξ_t^1 and ξ_t^2 are independent random motions governed by Eqs. (22) and (23). The exit time τ is defined in (51) as the first moment when the angular motion ξ_t^2 hits either of the boundary rays $\theta=0$ or $\theta=2\pi$. The "crossing" instants τ_ν are defined as the times when the motion ξ_t^2 touches the ray $\theta=\theta_{m_\nu}$, where

$$m_\nu=\begin{cases}1, & \text{if } \xi_{\tau_\nu}=\theta_1,\\ 2, & \text{if } \xi_{\tau_\nu}=\theta_2,\end{cases} \tag{52}$$

is a "crossing index" which has the value $m_\nu=1$ or $m_\nu=2$ depending on which of the rays $\theta=\theta_1$ or $\theta=\theta_2$ is touched. As for the factors $\delta(\xi_{\tau_\nu})$ appearing in (51) they are defined by the rule

$$\delta(\xi_{\tau_\nu})=\begin{cases}1, & \text{if } \xi_{\tau_\nu-0}>\xi_{\tau_\nu},\\ -1, & \text{if } \xi_{\tau_\nu-0}<\xi_{\tau_\nu},\end{cases} \tag{53}$$

which means that $\delta(\xi_{\tau_\nu})=1$ if at the moment $t=\tau_\nu$ the one-dimensional Brownian motion ξ_t hits either of the interfaces $\theta=\theta_m$ from the right, and that $\delta(\xi_{\tau_\nu})=-1$ if the interface $\theta=\theta_m$ is hit from the left.

To derive formula (51), we first observe that the interface conditions (49), (50) imply that

$$u(r,\theta_m\pm 0;\theta_0)=\pm\tfrac{1}{2}+\tfrac{1}{2}[u(r,\theta_m+\varepsilon)+u(r,\theta_m-\varepsilon)] +o(\varepsilon),\quad \varepsilon\to 0, \tag{54}$$

where θ_m are the angles from (44). Then, assuming for definiteness that $0<\theta<\theta_1$, we apply (25) and represent $u(r,\theta)$ in the form

$$u(r,\theta;\theta_0)=\mathbf{E}\left\{\chi(\tau-\tau_1)u(\xi^1_{\tau_1},\xi^2_{\tau_1}-0)\times\exp\left(\frac{ik}{2}\int_0^{\tau_1}\xi_s^1\,ds\right)\right\}, \tag{55}$$

where τ is the first time when the angular motion ξ_t^2 governed by (23) touches the boundary $\xi=0$, the crossing time τ_1 is defined as the first time when ξ_t^2 reaches the interface $\xi=\theta_1$, which may be approached from the side $\xi<\theta_1$, and

$$\chi(t)=\begin{cases}1, & \text{if } t\geqslant 0,\\ 0, & \text{if } t<0,\end{cases} \tag{56}$$

is the standard Heaviside function which insures the vanishing of the factor $\chi(\tau_1-\tau)$ in the case where ξ_t^2 hits the external boundary $\xi=0$ before reaching the interface $\xi=\theta_1$. After straightforward substitution of (54) into (55), we arrive at the representation

$$u(r,\theta;\theta_0)=\mathbf{E}\left\{-\frac{1}{2}\chi(\tau-\tau_1)\exp\left(\frac{ik}{2}\int_0^{\tau_1}\xi_s^1\,ds\right)\right\} +\mathbf{E}\left\{\chi(\tau-\tau_1)\frac{u(r,\xi^2_{\tau_1}+\varepsilon)+u(r,\xi^2_{\tau_1}-\varepsilon)}{2} \times\exp\left(\frac{ik}{2}\int_0^{\tau_1}\xi_s^1\,ds\right)\right\}+o(\varepsilon), \tag{57}$$

which may be rearranged to the form

$$u(r,\theta;\theta_0)=\mathbf{E}\left\{-\frac{1}{2}\chi(\tau-\tau_1)\exp\left(\frac{ik}{2}\int_0^{\tau_1}\xi_s^1\,ds\right)\right\} +\mathbf{E}\left\{\chi(\tau-\tau_1)u(r,\xi^2_{\tau_1}\pm\varepsilon) \times\exp\left(\frac{ik}{2}\int_0^{\tau_1}\xi_s^1\,ds\right)\right\}+o(\varepsilon), \tag{58}$$

where the second averaging is extended over the trajectories of the random motions ξ_t^1, ξ_t^2 and over the random selection of the sign in (58).

From the theory of Brownian motion it follows that if $\varepsilon\to 0$ then the points $\xi^2_{\tau_1}\pm\varepsilon$ can be considered as the positions of the Brownian motion ξ_t^2 at the time $t=\tau_1+\varepsilon^2$, so that the value $u(r,\xi^2_{\tau_1}\pm\varepsilon)$ in (58) can be replaced by $u(r,\xi^2_{\tau_1+\Delta t})$ with $\Delta t=\varepsilon^2$. Then, $u(r,\xi^2_{\tau_1+\Delta t};\theta_0)$ itself can be computed by a formula similar to (58), and continuing this process until the motion ξ_t^2 eventually leaves the interval $0<\xi_t^2<\theta$, we arrive at the solution (51).

The Sommerfeld problem of diffraction (35)–(40) provides an opportunity to verify the random walk approach to wave propagation by comparing numerical results generated by the probabilistic solution (51) against the results generated by the well-known[26,5,27] conventional solution, which represents the diffracted field U_d from (37) by the Sommerfeld integral

$$U_d(r,\theta;\theta_0)=\frac{1}{8\pi i}\int_{C_0}[\Phi(w+\theta;\theta_0) -\Phi(w-\theta;\theta_0)]e^{ikr\cos w}dw, \tag{59}$$

in which

$$\Phi(w;\theta_0)=\cot\left(\frac{w-\theta_0-\pi}{4}\right)-\cot\left(\frac{w+\theta_0-\pi}{4}\right), \tag{60}$$

and C_0 is the contour running from $-(\pi/2)+i\infty$ to $(\pi/2)-i\infty$.

To verify the obtained probabilistic solution of the Sommerfeld problem, we used it for the numerical simulation of the diffracted field $U_d(r,\theta;45°)$ generated by the incident plane wave arriving from infinity along the ray $\theta=45°$. Figure 4 shows the real and imaginary parts of $U_d(r,\theta)$ computed along the circle $r=2\pi$. The results obtained by the integration in (59) are plotted as a continuous line, while the results obtained by the random walk method are marked by small circles. The mathematical expectation (51) was ap-

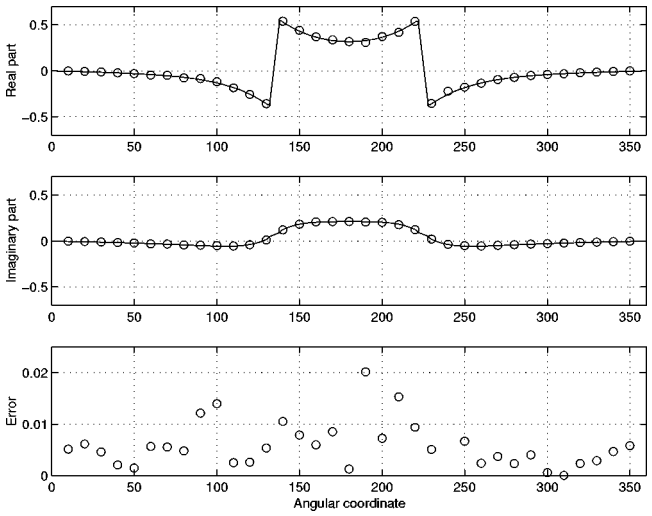


FIG. 4. Random walk simulation of the diffracted field.

proximated by averaging 1000 discrete random walks (34) with the spatial increment $\varepsilon = 0.1$. The computations were stable over the entire trial interval from $\theta = 0$ to $\theta = 360°$, and the difference between the results obtained by the probabilistic and analytic methods did not exceed 0.02, with the average difference noticeably below the 0.01 level.

V. WAVES RADIATING FROM A CONICAL DOMAIN

Here, we demonstrate that the technique employed in the previous sections for obtaining random walk solutions of boundary value problems in two-dimensional wedge-shaped domains can also be applied to the analysis of waves radiating from three-dimensional conical domains.

Let (r, θ, ϕ) be spherical coordinates defined by

$$r > 0, \quad -\frac{\pi}{2} \leqslant \theta < \frac{\pi}{2}, \quad 0 \leqslant \phi < 2\pi. \tag{61}$$

Then, a conical domain $G \subset \mathbb{R}^3$ can be specified by the condition

$$G: \quad r > 0, \quad (\theta, \phi) \in g, \tag{62}$$

where $g \subset S_2$ may be viewed as a subdomain of the two-dimensional rectangular region

$$S_2 = \left\{ \theta, \phi: \ -\frac{\pi}{2} \leqslant \theta \leqslant \frac{\pi}{2}, \ 0 \leqslant \phi < 2\pi \right\}. \tag{63}$$

We are interested in the solution of the Helmholtz equation

$$\nabla^2 U + k^2 U = 0, \quad k = \text{const}, \tag{64}$$

which is defined in G and satisfies the Dirichlet boundary conditions

$$U(r, \theta, \phi)\big|_{(\theta, \phi) \in g} = F(r, \theta, \phi), \tag{65}$$

together with the radiation condition at infinity

$$U(r, \theta, \phi) = C(\theta, \phi) \frac{e^{ikr}}{kr} (1 + o(1)), \quad r \to \infty. \tag{66}$$

We seek the solution of the Helmholtz equation in the Liouville form

$$U(r, \theta, \phi) = u(r, \theta, \phi) e^{ikr}, \tag{67}$$

where u is a new unknown function. Substitution of (67) into (64) and evaluation of the derivatives in spherical coordinates requires $u(r, \theta, \phi)$ to satisfy the equation

$$\frac{i}{2k}\left[\frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\left(\frac{\partial^2 u}{\partial \theta^2} - \tan\theta \frac{\partial u}{\partial \theta}\right) + \frac{1}{r^2 \cos^2\theta}\frac{\partial^2 u}{\partial u^2}\right] - \frac{\partial u}{\partial r} - \frac{u}{r} = 0. \tag{68}$$

Then, multiplying (68) by $q = -ikr^2$ we arrive at the final equation

$$\frac{r^2}{2}\frac{\partial^2 u}{\partial r^2} + r(1 + ikr)\frac{\partial u}{\partial r} + \frac{1}{2}\frac{\partial^2 u}{\partial \theta^2} - \frac{\tan\theta}{2}\frac{\partial u}{\partial \theta} + \frac{1}{2\cos^2\theta}\frac{\partial^2 u}{\partial \phi^2} + ikru = 0, \tag{69}$$

which has to be solved subject to the boundary conditions

$$u(r, \theta, \phi)\big|_{(\theta, \phi) \in g} = f(r, \theta, \phi) \equiv F(r, \theta, \phi) e^{-ikr}, \tag{70}$$

following from (65) and (67).

If the boundary function $F(r, \theta, \psi)$ is analytic with respect to r, then the solution of the problem (70), (69) can be represented by the Feynman–Kac formula (6) applied in the domain

$$\mathfrak{G} = \{(r, \theta, \phi): \quad r \in \mathbb{C}, \quad (\mathrm{Re}(\theta), \mathrm{Re}\,\phi) \in g\}. \tag{71}$$

As a result, we obtain the solution

$$u(r, \theta, \psi) = \mathbf{E}\left\{ f(\xi_\tau^1, \xi_\tau^2, \xi_\tau^3) \exp\left\{ ik \int_0^\tau \xi_s^1 \, ds \right\} \right\}, \quad (\xi_\tau^2, \xi_\tau^3) \in \partial g, \tag{72}$$

with the mathematical expectation computed over trajectories of the three-dimensional random motion $\xi_t = (\xi_t^1, \xi_t^2, \xi_t^3)$ governed by the stochastic differential equations

$$d\xi_t^1 = \xi_t^1 \, dw_t^1 + \xi_t^1 (1 + ik\xi_t^1) dt, \qquad \xi_0^1 = r, \tag{73}$$

$$d\xi_t^2 = dw_t^2 - \tfrac{1}{2}\tan(\xi_t^2) dt, \qquad \xi_0^2 = \theta, \tag{74}$$

$$d\xi_t^3 = dw_t^3 / \cos(\xi_t^2), \qquad \xi_0^3 = \phi, \tag{75}$$

driven by three independent one-dimensional Brownian motions w_t^1, w_t^2, and w_t^3. As for the exit time τ, it is defined as the first time when the point (ξ_t^2, ξ_t^3) touches the boundary ∂g.

It is obvious from (74) and (75) that ξ_t^2 and ξ_t^3 remain real valued at any time t, so that the pair (ξ_t^2, ξ_t^3) may be considered as a random standard Brownian motion on a two-dimensional sphere S_2 from (63). As for the radial random motion ξ_t^1 controlled by (73), it has the structure similar to that of the radial random motion studied in Sec. IV.

Finally, it is instructive to observe that the solution (72)–(75) admits transformation to the alternative form

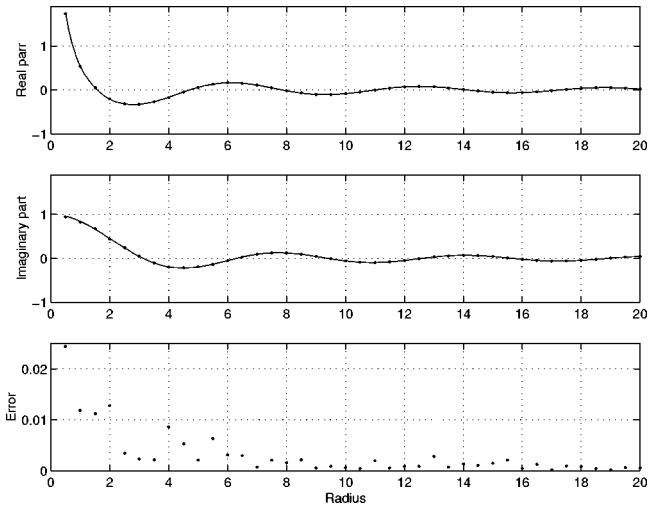


FIG. 5. Random walk simulation of the diffracted field.

$$u(r,\theta,\psi)=\frac{1}{r}\mathbf{E}\{f(\xi_\tau^1,\xi_\tau^2,\xi_\tau^3)\xi_\tau^1 e^{-w_\tau^1-\tau/2}\}, \tag{76}$$

emphasizing the structure of $u(r,\theta,\psi)$ at infinity suggested by the condition (66).

To derive (76), we note that Eq. (73) yields the relationship

$$ik\xi_t^1\,dt=\frac{d\xi_t^1}{\xi_t^1}-dw_t^1-dt, \tag{77}$$

whose combination with the Ito's formula (28) leads to the expressions

$$ik\xi_t^1\,dt=d\ln(\xi_t^1)-dw_t^1-\tfrac{1}{2}\,dt, \tag{78}$$

and

$$\int_0^\tau ik\xi_t^1\,dt=\ln(\xi_\tau^1)-\ln(r)-dw_\tau^1-\frac{1}{2}\tau, \tag{79}$$

the latter of which is obtained by the integration of (78) with the initial condition $\xi_0^1=r$ taken into account. Then, substituting (79) into (72) we arrive at (76).

Figure 5 presents results of a numerical simulation of the function $U(r,\theta,\phi)=e^{ir}/r$ considered as the solution of the equation (64) with $k=1$ in a cone

$$G:\ r>0,\quad -60°<\theta<60°,\quad 0°<\phi<120°, \tag{80}$$

with the boundary conditions (65), where

$$f(r,\theta,\psi)=\frac{1}{r}. \tag{81}$$

The first two diagrams show the values of the real and imaginary components of $U(r,\theta,\phi)=e^{ir}/r$ computed along the ray $\theta=0°$, $\phi=60°$. The computations were performed twice: by the exact formula, and by the probabilistic formulas (67), (76) with the mathematical expectation in (72) approximated by the averaging over 3000 discrete random walks driven by discrete Brownian motions with the spatial increment $\varepsilon=0.05$. Computations were stable over the entire trial interval from $r=0.5$ to $r=20$, and the error of simulation was under 0.01 virtually anywhere, except in a close range near the origin $r=0$, where $U(r,\theta,\phi)$ has a pole.

VI. CONCLUSION

The versions of the random walk methods presented above provide simple exact formulas describing two and three-dimensional waves radiating from wedge-shaped and conical domains. In general, the probabilistic formulas derived in Refs. 1–4 determine solutions of the Helmholtz equation as mathematical expectations of values of specified functionals computed over trajectories of the random motion composed of deterministic drifts along complex rays and random complex-valued fluctuations. However, here we employed specific features of the polar and spherical coordinate systems and obtained solutions which include averaging over the trajectories whose angular components remain real valued. This simplification significantly improves the efficiency of the random walk approach and makes it possible to derive the probabilistic solution of the classic Sommerfeld problem of diffraction.

As with analytic or asymptotic methods, probabilistic solutions provided by the random walk method are local in the sense that they make it possible to compute functions of interest at individual points without computing them on dense meshes. Moreover, such solutions admit simple perfectly scalable implementations with practically unlimited capability for parallel processing, and such solutions admit meaningful physical interpretation which does not contradict but compliments elementary models of wave propagation employed by the ray theory. All of these features together make the random walk method attractive both for qualitative and numerical analysis, and the examples discussed in the present paper suggest that this method might be developed into a powerful tool for the analysis of general wave propagation phenomena.

ACKNOWLEDGMENTS

This research was supported by NSF Grant CMS-0098418 and by the William S. Floyd, Jr. Distinguished Professorship in Engineering held by D. Bogy.

[1] B. V. Budaev and D. B. Bogy, "Probabilistic solutions of the Helmholtz equations," J. Acoust. Soc. Am. **109**(5), 2260–2262 (2001).
[2] B. V. Budaev and D. B. Bogy, "Analysis of one-dimensional wave scattering by the random walk method," J. Acoust. Soc. Am. **111**(6), 2555–2560 (2002).
[3] B. V. Budaev and D. B. Bogy, "Application of random walk methods to wave propagation," Q. J. Mech. Appl. Math. **55**(2), 209–226 (2002).
[4] B. V. Budaev and D. B. Bogy, "Random walk methods and wave diffraction," Int. J. Solids Struct. **39**(21–22), 5547–5570 (2002).
[5] L. B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves*, Prentice-Hall Microwaves and Fields Series (Prentice-Hall, Englewood Cliffs, NJ, 1972).
[6] V. P. Maslov and M. V. Fedoriuk, *Semi-classical Approximation in Quantum Mechanics* (Reidel, Dordrecht, 1981).
[7] E. B. Dynkin, *Markov Processes*, Number 121,122 in Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen (Springer, Berlin, 1965).
[8] M. Freidlin, *Functional Integration and Partial Differential Equations*, Number 109 in The Annals of Mathematics Studies (Princeton University Press, Princeton, NJ, 1985).
[9] I. V. Arnold, *Mathematical Methods in Classical Mechanics*, 2nd ed.,

Number 60 in Graduate Texts in Mathematics (Springer, New York, 1989).
[10] R. Courant, K. Friedrichs, and K. H. Lewy, ''Uber die partiellen differenzengleichungen der mathematischen physik,'' Math. Ann. **100**, 32–74 (1928).
[11] I. G. Petrovsky, ''Uber das irrahrtproblem,'' Math. Ann. **109**(4), 425–444 (1934).
[12] H. P. Philips and N. Wiener, ''Nets and Dirichlet problem,'' J. Math. Phys. **2**, 105–124 (1923).
[13] N. Wiener, ''Differential space,'' J. Math. Phys. **2**, 131–174 (1923).
[14] R. P. Feynman, ''The principle of least square action in quantum mechanics,'' PhD thesis, Princeton University, Princeton, NJ, 1942.
[15] R. P. Feynman, ''Space-time approach to non-relativistic quantum mechanics,'' Rev. Mod. Phys. **20**, 367–387 (1948).
[16] M. Kac, ''On the distribution of certain Wiener functionals,'' Trans. Am. Math. Soc. **65**(1), 1–13 (1949).
[17] R. B. Schlottmann, ''A path integral formulation of acoustic wave propagation,'' Geophys. J. Int. **137**(2), 353–363 (1999).
[18] R. D. Nevels, J. A. Miller, and R. E. Miller, ''A path integral time-domain method for electro-magnetic scattering,'' IEEE Trans. Antennas Propag. **48**(4), 565–573 (2000).
[19] G. Bal, J. B. Keller, G. Papanicolaou, and L. Ryzhik, ''Transport theory for acoustic waves with reflection and transmission at interfaces,'' Wave Motion **30**(4), 303–327 (1999).
[20] G. Bal, G. Papanicolaou, and L. Ryzhik, ''Probabilistic theory of transport processes with polarization,'' SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **60**(5), 1639–1666 (2000).
[21] V. A. Fock, *Electromagnetic Diffraction and Propagation Problems*, Vol. 1 of International Series of Monographs on Electromagnetic Waves (Pergamon, Oxford, 1965).
[22] V. A. Fock and M. A. Leontovich, ''Solution of the problems of propagation of electromagnetic waves along the Earth's surface by the method of parabolic equation,'' J. Phys. (Moscow) **10**, 1–13 (1946).
[23] B. V. Budaev and D. B. Bogy, ''Diffraction by a plane sector,'' Proc. R. Soc. London, Ser. A (submitted).
[24] B. Simon, *Functional Integration and Quantum Physics*, Number 86 in Pure and Applied Mathematics: A Series of Monographs and Textbooks (Academic, New York, 1979).
[25] W. Feller, *An Introduction to Probability Theory and its Applications*, Wiley Series in Probability and Mathematical Statistics (Wiley, New York, 1967).
[26] *Electromagnetic and Acoustic Scattering by Simple Shapes*, edited by J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi (Hemisphere, New York, 1969).
[27] A. Sommerfeld, ''Mathematische theorier der diffraction,'' Math. Ann. **47**, 317–374 (1898).

Experimental setup for measurement of acoustic power dissipation in lined ducts for higher order modes propagation with air mean-flow conditions

Jean-Michel Ville[a] and Felix Foucart
Laboratoire Roberval UMR UTC-CNRS no6066, Université de Technologie de Compiègne, BP 20529 F60205 Compiègne cedex, France

(Received 14 November 2002; revised 11 July 2003; accepted 28 July 2003)

A flow duct acoustic facility was developed to measure liner efficiency in attenuating higher order acoustic duct modes propagation conditions with mean air flow. The method is based on measurement, upstream and downstream of a liner, of the acoustic power produced by a periodic source. Directly measured total or modal acoustic powers are deduced from the local measurement, in both cross sections, of acoustic pressure, axial acoustic particle velocity, and axial mean flow velocity which are supplied by a probe made of a microphone and a single hot film. In this paper, the equipment, signal processing, and the data treatment process of this facility are first described. Then, information on the accuracy of the methodology is provided by a validation test performed with a rigid wall duct section. Finally, the results of an experiment carried out with a locally reacting liner and a mean flow velocity of 20 m/s will be presented. Measurements of the main attenuation frequency and of the main total acoustic power dissipated agree with the values for which the liner was designed. These results point out the limitations of the method presented to sources with high-level periodic sounds to provide a sufficient signal-to-noise ratio, the noise being produced by fluctuations of the turbulent flow.
[DOI: 10.1121/1.1610461]

PACS numbers: 43.20.Mv, 43.60.Qv, 43.58.Fm [DKW]

I. INTRODUCTION

Since fan noise will be a major contributor to the exterior noise of new turbofan engines, the aerospace industry still would benefit from new technologies[1,2] that reduce the acoustic power radiated from the inlet or control its directivity patterns.[3,4] Propagation of higher order acoustic modes occurs in presence of air flow in the duct where reduction elements are placed. To thoroughly understand and to describe accurately the effects of these elements on sound propagation, in duct acoustic total and modal power measurements upstream and downstream these elements can be very useful for optimizing inlets or validating theoretical models.

Previously, sound-power measurement methods using pairs of flush wall-mounted microphones were developed for plane-wave propagation conditions in a realistic flow found in the exhaust of an automobile.[5,6] Also, in-duct standardized procedures were studied and are applied for measuring acoustic sound power radiated by fans for heating, ventilating, and air-conditioning equipment[7] where acoustic pressure is distributed over higher order modes. Because the two microphones' sound intensity probe can measure sound intensity only when mean flow and sound propagation are colinear,[8] which is invalid for higher order mode propagation conditions and also because intensity measurement is time consuming, acoustic sound power is deduced from a local pressure measurement at a radial position and corrections are applied to take into account the effects of flow and higher order modes pressure distribution.[9,10]

In previous works a pressure–velocity probe made of a microphone and a hot film was studied[11] including an acoustic calibration[12] and modal decomposition[13] procedures. It was used to measure, in front of an axial cowled fan, the acoustic potential per mode radiated on blade passing frequencies in presence of a mean flow.[11] Computation of the radiated field was performed with input data supplied by the experiment, a good agreement was found between experimental and predicted pressure radiated fields. In the present paper, a new method of measuring the axial component of the acoustic sound power is introduced. It is based on the direct measurement of local sound axial intensity and mean flow with this pressure velocity probe. A duct acoustic flow facility and the experimental procedure developed to measure the total and modal acoustic axial intensities and powers in two cross sections located downstream and upstream of an element under test with the pressure–velocity probe are described. The results of measurements with a hard wall configuration and a locally reacting liner in mean flow velocity condition for which the velocity is actually much lower than that encountered in real aircraft engines are discussed.

II. THE DUCT ACOUSTIC FLOW FACILITY

In-duct acoustic power measurements are conducted in the duct flow facility presented in Fig. 1 and Fig. 2. Its function is to create the aero-acoustic conditions of the experiment and to collect local acoustic and aerodynamic data in two cross sections located on each side of the test section.

[a]Electronic mail: Jean-Michel Ville, jean-michel.ville@utc.fr

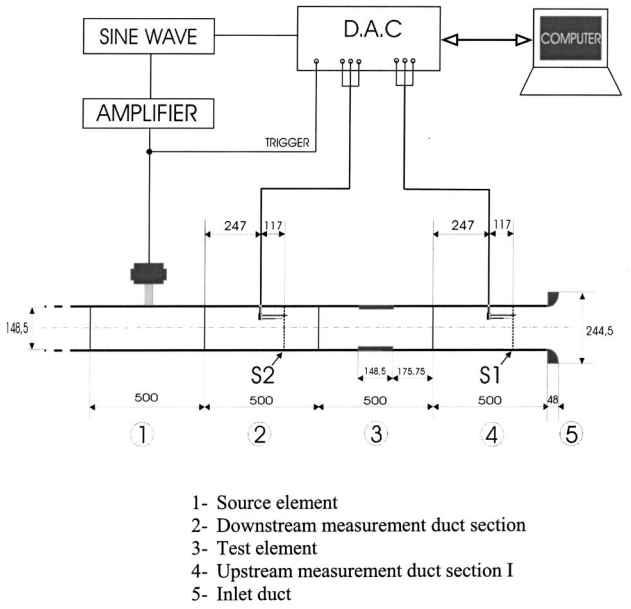


FIG. 1. Schematic of the experimental setup. Dimensions are in mm.

A. Hardware setup

Flow conditions are produced by a centrifugal fan located outside the anechoic chamber. Flow entering the duct through an inlet comes from an opening located in the upper part of the chamber. The rest of the facility, installed inside the Compiegne University of Technology anechoic chamber to conduct future measurements of radiated acoustic field, is shown in Fig. 2. It is made of six duct sections (0.50 m long and 0.15 m in diameter) placed on a support. From the right to the left in the photograph (Fig. 2) are shown

(i) A muffler used to reduce noise radiated from the fan and reflection of acoustic waves from the face of the fan;
(ii) The source section on which acoustic drivers are flush mounted;
(iii) A hard wall section;
(iv) A downstream measurement hard wall section described below;
(v) A test section where the liner is mounted; and
(vi) An upstream measurement hard wall section identical to the downstream one.

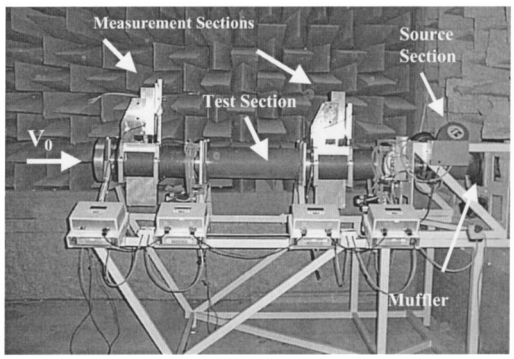


FIG. 2. Duct flow facility in the anechoic chamber.

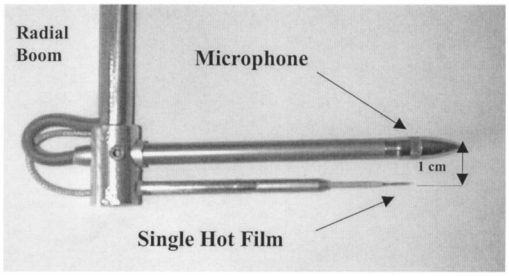


FIG. 3. Pressure–velocity probe used in the experiment.

Like the other rigid wall duct sections, the walls of the measurement duct sections are made of 0.015-m-thick steel. Local measurements are carried out in a cross-duct section at 15 radial positions 5.4 mm equally spaced over the duct radius range for 16 angular locations equally spaced from 0 to 337.5°. Displacements of the probe are supplied by stepping motors which rotate both measurement duct sections and traverse the radial probe supports. A computer station automatically operates and controls all these displacements. Both measurement cross sections are 1 m apart.

The probe (Fig. 3) consists of two transducers 0.01 m spaced on the radial boom: a 0.003 17-m diameter microphone with its nose cone and a single hot film. This probe, developed and used during previous studies,[11] provides local measurements of acoustic pressure, axial acoustic particle velocity, and axial mean flow velocity. Calibration of the hot film is performed in two steps by an original experimental procedure[12] consisting of aerodynamic calibration by a classical experiment against a pitot tube in a small jet and acoustic calibration.

B. Signal acquisition and data reduction process

A work station with its acquisition interface performs discrete frequency signal generation on an acoustic driver through a power amplifier and simultaneously acquires six signals issued from transducers located in both sections.

Measurement of amplitude and phase of acoustic pressure p and axial particle velocity v_z is performed by triggering all acoustic signals by the signal generator. Furthermore, each acoustic signal is time averaged to avoid flow noise before being Fourier transformed. Axial mean flow velocity V_0 is deduced by time averaging the dc hot film signal. As acoustic pressure is not measured at the same radial point as the acoustic axial particle velocity; a correction of the acoustic particle velocity based on a third-order interpolation process is applied.[11] Also, as the microphone cannot move as close as 0.013 m to the duct wall, an extrapolation procedure is applied to deduce measurement near the duct wall.[11] This data collection procedure is performed for the 240 positions (16 angular×15 radial) in both upstream and downstream measurement cross sections.

III. ACOUSTIC POWER MEASUREMENT

A data treatment procedure (Fig. 4) was developed to deduce the following quantities from local aero-acoustic quantities collected from the flow facility:

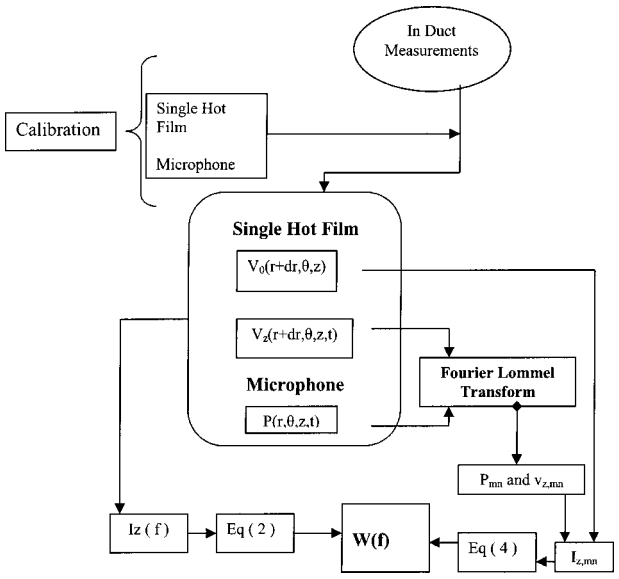


FIG. 4. Flow chart of the experimental determination of the acoustic powers.

(i) Local axial intensity;
(ii) Modal axial intensity; and
(iii) Total axial acoustic powers deduced by integration of axial sound intensity over the cross section and by summation of modal axial intensity.

The computations are based on definitions, theoretical basis, and the data treatment procedure described below.

A. Directly measured total acoustic power

In an irrotational and isentropic uniform flow, at a given frequency f, the axial component of acoustic intensity $I_z(f)$ is given by the following relationship:[14]

$$I_z(f)=\frac{1}{2}(1+M_0^2)\mathrm{Re}(p\cdot v_z^*)+\frac{\rho_{\cdot 0\cdot}V_0}{2}|v_z|^2 + \frac{V_0}{2\rho_0 c_0^2}|p|^2, \tag{1}$$

where * means complex conjugate, M_0 is the Mach number deduced by integrating the mean-flow velocity profile over the cross section, and ρ_0 and c_0 are, respectively, the air density and speed of sound. The acoustic power $W_{\mathrm{total}}(f)$ through a duct cross section is then calculated by integrating axial acoustic intensity over the duct cross section S

$$W_{\mathrm{total}}(f)=\int_S I_z(f)\,dS. \tag{2}$$

B. Modal intensity and total acoustic power in the propagating modes

Assuming that the duct walls in the measurement duct sections are rigid, modes of propagation in the cylindrical duct represented by two integers (m,n) are orthogonal.[15] Therefore, for each frequency, the Fourier–Lommel's transform[13] applied to the spatial distribution of acoustic pressure $p(r,\Theta,f)$ and axial acoustic particle velocity $v_z(r,\Theta,f)$ on duct cross sections yields the wave number spectra p_{mn} and $v_{z,mn}$. These spectra are limited to $m=-7$ to 7 because of the angular mesh and to $n=0$ to 3.[11,13] The indices (0,0) represent the plane-wave mode. The axial component of intensity for mode (m,n), $I_{z,mn}(f)$ can then be deduced from the following relationship:

$$I_{z,mn}(f)=\frac{1}{2}(1+M_0^2)\mathrm{Re}(p_{mn}\cdot v_{z,mn}^*)+\frac{\rho_{\cdot 0\cdot}V_0}{2}|v_{z,mn}|^2 + \frac{V_0}{2\rho_0 c_0^2}|p_{mn}|^2. \tag{3}$$

The total acoustic power in the propagating modes $W_{\mathrm{modal}}(f)$ through the duct measurement cross section is then calculated by adding acoustic intensity of all propagating modes

$$W_{\mathrm{modal}}(f)=\sum_n\sum_m N_{mn}I_{z,mn}(f), \tag{4}$$

where N_{mn} the normalization factor is defined by

$$N_{mn}=SJ_m^2(\chi_{mn})\left[1-\frac{m^2}{\chi_{mn}^2}\right], \tag{5}$$

with J_m Bessel's function of first kind and order m, χ_{mn} the nth zero of the derivative of J_m. If the distribution over propagating modes represents the overall acoustic power, the following equality should be valid:

$$W_{\mathrm{total}}(f)=W_{\mathrm{modal}}(f). \tag{6}$$

The flow chart of the experimental determination of the acoustic power through both measurement duct cross sections is shown in Fig. 4.

C. The liner acoustic power dissipation

The efficiency of the liner is then given by two expressions.

(i) For each mode (m,n) the modal intensity attenuation defined by

$$\Delta I_{z,mn}(f)=I_{z,mn}^{\mathrm{downstream}}(f)-I_{z,mn}^{\mathrm{upstream}}(f). \tag{7}$$

(ii) Total power attenuations defined by

$$\Delta W(f)=W^{\mathrm{downstream}}(f)-W^{\mathrm{upstream}}(f), \tag{8}$$

where $W^{\mathrm{downstream/upstream}}(f)$ can be total—Eq. (2)—or modal—Eq. (4).

IV. EXPERIMENTAL RESULTS

Two liner configurations were mounted in the test section and tested with a mean flow velocity of 20 m/s for the following purposes:

(i) First, in order to control the ability of the procedure to verify the conservation law of total power between both measurement sections. The experiment was then conducted with a hard wall test element for only five

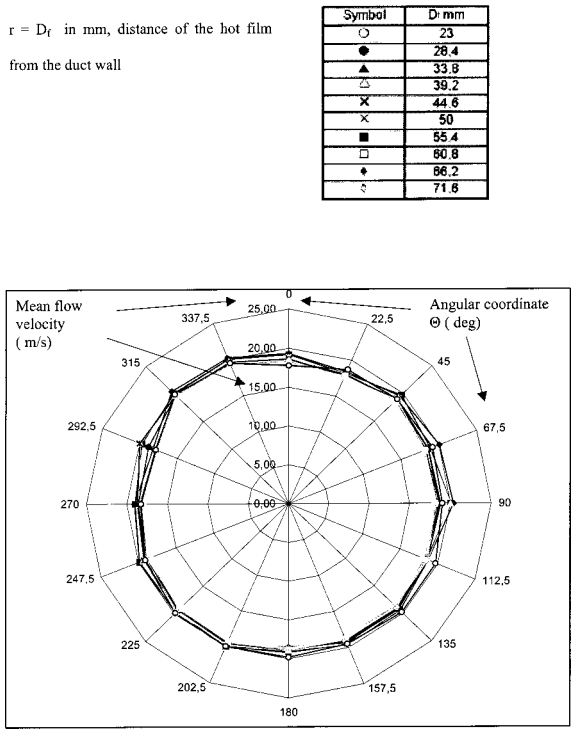


FIG. 5. $V_0(r,\theta)$, mean-flow velocity profile (m/s) in polar coordinates in upstream duct cross section (hard wall).

frequencies, starting at 1190 Hz and then each 840 Hz thereafter. Except at 2870 Hz which is close to mode (0,1) cutoff frequency, all other frequencies are far enough from mode cutoff frequencies to reduce reflection problems by the inlet (see the discussion in Secs. IV B 1 and IV B 2).

(ii) Second, in order to measure the acoustic power dissipated by a locally reacting liner. The experiment was conducted for nine frequencies surrounding 4800 Hz, the expected specific frequency of the liner, and starting also from 1190 Hz.

A. Mean-flow velocity profile measurement

Mean-flow velocity profiles are measured by the hot films in both measurement duct cross sections. Typical results (measured in the upstream section) are plotted in polar coordinates in Fig. 5 for all radial positions. These curves demonstrate that the mean-flow velocity is about 20 m/s and verify the uniformity of the air flow in the measurement cross section.

B. Acoustic power measurements

1. Modal decomposition validation

During each experiment, the modal decomposition process was controlled by comparing levels of ''total energies'' $E^p(f)$, $E^v(f)$ and ''coherent energies'' $E^{CP}(f)$, $E^{Cv}(f)$ on both acoustic pressure and acoustic axial particle velocity. They are defined by[16]

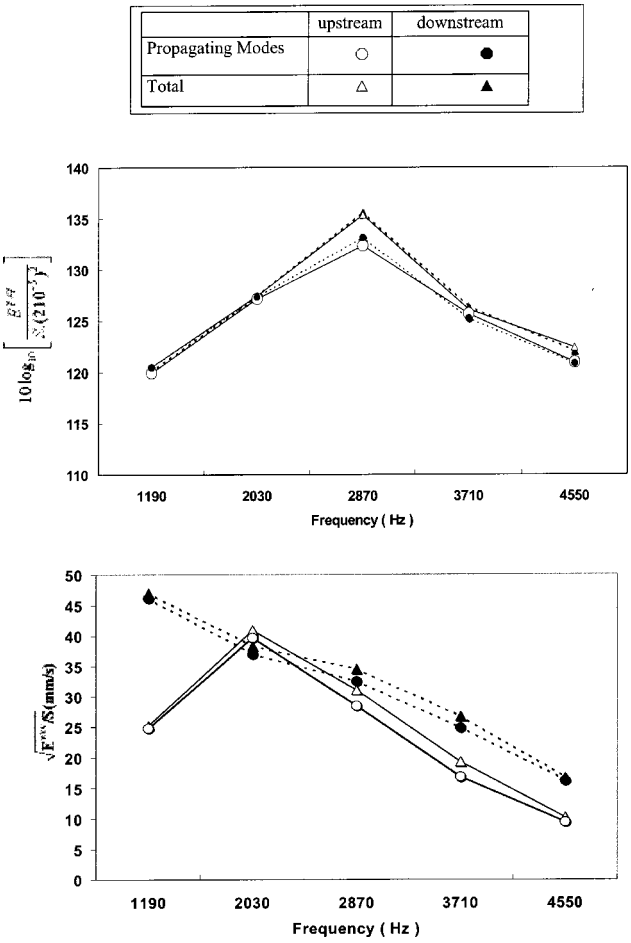


FIG. 6. Total and coherent energy levels in upstream and downstream measurement cross sections (hard wall, mean-flow velocity: 20 m/s).

$$E^P(f)=\int_S |P(f,r,\theta)|^2 dS$$

$$=E^{CP}(f)=\sum_n \sum_m N_{mn}[P(f,m,n)]^2, \tag{9}$$

$$E^v(f)=\int_S |v_z(f,r,\theta)|^2 dS$$

$$=E^{CV}(f)=\sum_n \sum_m N_{mn}[v_z(f,m,n)]^2. \tag{10}$$

Typical results of this validation procedure are shown in Fig. 6, when a hard wall duct test section configuration was installed with a mean-flow velocity of 20 m/s. In both measurement cross sections, total energy level $E^p(f)$, $E^v(f)$ are higher than coherent energy levels $E^{CP}(f)$, $E^{Cv}(f)$. Evanescent modes and ''numerical noise'' during the computation of Lommel's transform[11] cause this underestimation. On pressure, the difference is less than 2 dB and on axial acoustic particle velocity less than 2 mm/s. Therefore, the decomposition of acoustic quantities into propagating modes is assumed to be validated.

Energy levels for both acoustic pressure and axial acoustic particle velocity are higher in the downstream section,

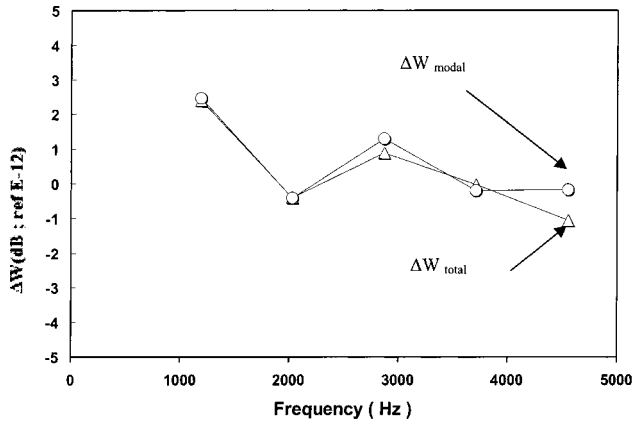


FIG. 7. Verification of the acoustic power conservation law (hard wall, mean-flow velocity: 20 m/s).

which is the closest one to the source section, except for acoustic velocity at 2030 and 1190 Hz.

Figure 6 also shows that the difference between total pressures measured in both sections is lower than 1 dB, but this difference is nearly 5 dB (a ratio close to 2) on particle velocity. This points out the dependence of the axial location of the transducer on measurement which is due to sound reflections occurring at the end of the duct. This reflection can also cause, at some frequencies, sound pressure or axial particle velocity nodes in a measurement section, leading to signal-to-noise ratio troubles. The number of these frequencies will be more important in the downstream section because it is located farther away from the inlet than in the upstream one.

The modulus of the reflection coefficients depends on the modal distribution, and it is well known that the modal reflection coefficient decreases from 1 when the difference between ka and χ_{mn} increases. In a previous study[17] modal reflection coefficients of a more reflective inlet than the one shown in Fig. 1 were measured and found to be lower than 0.3 in modulus for all ka studied in the present experiment except at 2870 Hz ($ka=3.98$ is close to $\chi_{01}=3.83$), resulting in a reflection coefficient $R_{01,01}$ close to 1.

At 1190 Hz, $ka=1.64$, only the plane wave propagates and the reflection at the end of the duct is described by only $R_{00,00}$, the plane-wave reflection coefficient, which can be assumed[17] to be around 0.3.

2. Verification of the conservation of total acoustic power

Before applying this method to study liner efficiency, an experiment was performed with a hard wall test section configuration to verify the conservation of the total acoustic power between upstream and downstream sections. As shown in Fig. 7, the conservation law deduced from directly measured total acoustic power [Eq. (2)] as well as modal acoustic power [Eq. (4)] is verified. The absolute value of $\Delta W(f)$ is generally lower than 1 dB, although it increases to 2.5 dB at 1190 Hz. This difference is due to the difference between particle velocities measured in both sections. As explained in the previous paragraph, this discrepancy comes from the reflection at the end of the duct.

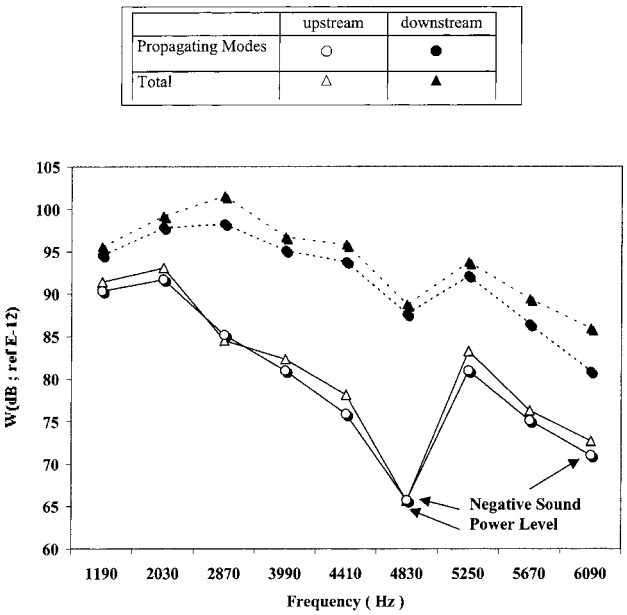


FIG. 8. The directly measured total acoustic power and the total acoustic power in the propagating modes. Measurements are shown for the upstream and downstream cross sections in the presence of a hard wall. The mean flow velocity was 20 m/s.

This result gives an estimation of the reliability of this method in determining efficiency of a liner duct section with flow.

3. Measurement with a locally reacting liner

A liner barrel, 1 diameter long, is mounted in the middle of the duct test section. It consists of a honeycomb of hexagonal, 10-mm-deep cells. The face of the liner is made of a perforated plate covered with a wire mesh. It was designed and manufactured by an industrial partner from the aeronautical industry[1] to reduce noise around 4800 Hz by 20 dB.

Overall acoustic powers deduced from Eq. (2) and Eq. (4) upstream and downstream of the liner are plotted in Fig. 8. Total and modal powers are nearly identical. Downstream acoustic power is the acoustic power coming in the lined wall duct test section, and the upstream power represents the acoustic power coming out of this section and radiating outside the duct through the inlet. Because of the liner, the acoustic power in the upstream section is lower than the acoustic power in the downstream section for all frequencies. At 4830 Hz a very low acoustic power level of 65 dB was found in upstream section but with a negative sign on both the total power in the propagating modes and the directly measured total power. Also, at 6090 Hz a value of -70 dB was found only on the total modal power. In Fig. 8, the absolute values of these negative powers were plotted to provide a better readability of the figure. This sign problem is most probably due to the uncertainty on determination of phase difference between acoustic pressure and axial acoustic particle velocity at low level which introduces low signal to noise ratio. Previous studies[18] have pointed out that the dynamic range of a hot film wire is limited, particularly when the level is low. Increasing the pressure noise level

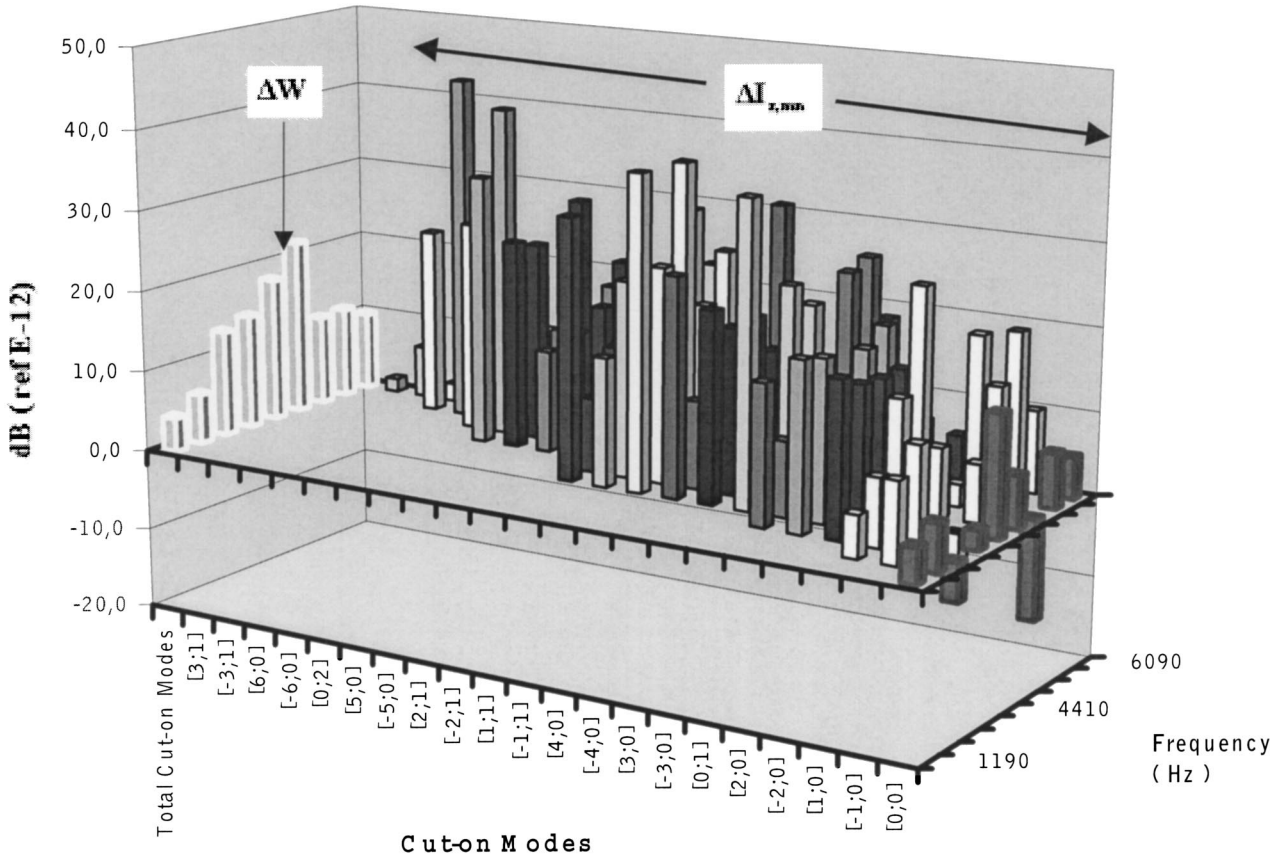


FIG. 9. Acoustic intensity dissipation per mode (*m*,*n*) and total acoustic power in propagating modes dissipation through the locally reacting liner. The mean-flow velocity was 20 m/s.

provided by the acoustic drivers would probably solve this problem.

The total acoustic power in the propagating modes and the ''modal intensity for each mode'' dissipated by the liner are deduced from Eq. (6). Results are plotted in Fig. 9 for the nine frequencies. At 6090 Hz 21 modes are propagating. A maximum of 21 dB on the acoustic power dissipation occurs at 4830 Hz. Both values of frequency and level of attenuation agree with those given by the designer of the liner.

Results on the acoustic power dissipated per propagating mode point out that there is a high level attenuation at frequencies near mode cutoff frequencies. Also, the plane-wave mode appears to be the least attenuated mode by the liner.

V. CONCLUSION

A duct flow facility, utilizing local measurements of the acoustic pressure, axial components of the acoustic particle, and mean-flow velocity, was constructed. With this original experimental equipment, direct measurement of modal and total acoustic powers in the presence of mean flow was performed and studies of the effect of the liner on the overall acoustic power and on the modal distribution are made possible.

This new facility first was tested with a simple hard wall section. The distributions of propagating modes of acoustic pressure, axial particle velocity, and acoustic power in both measurement duct cross sections were controlled by comparing results from modal summation with directly measured acoustic energies and powers. Also, the acoustic power was verified within 2-dB accuracy to be conserved when waves propagate between the two measurement sections surrounding this rigid wall section.

Results with a locally reacting liner were found to agree with the characteristics expected from the design only for mean-flow velocities up to 20 m/s.

This experimental setup requires an acoustic source (a loudspeaker or a long-length coherent periodic source as the blade passing frequency of a fan) that provides a signal to ambient flow noise ratio on measurements from the microphone and from the hot film which is limited in measuring low-level acoustic particle velocity.

Future studied will be devoted to

(i) improving the experimental duct flow facility (increasing the mean flow Mach number capability, etc.);
(ii) improving signal enhancement to validate the method in realistic flow found in aircraft engines;
(iii) conducting radiation measurement; and
(iv) studying new concepts of active and passive reduction technologies.

ACKNOWLEDGMENTS

This work was carried out under a contract awarded by European Commission, Contract Number BRPR-CT97-0479. The authors wish to express their thanks to all partners who work with them on this Ducat program.

[1] E. R. Rademaker, ''Publishable synthesis report DUCAT,'' Report DUCAT-NL-01-T5-9, NLR-TR-2001-170 (2001).
[2] A. Selamet, N. S. Dickey, and J. M. Novak, ''The Herschel-Quincke tube: A theoretical, computational, and experimental investigation,'' J. Acoust. Soc. Am. **96**, 3177–3185 (1994).
[3] J. M. Tyler and T. G. Sofrin, ''Axial flow compressor noise studies,'' SAE Trans. **7096**, 309–322 (1960).
[4] E. J. Rice, ''Optimum wall impedance for spinning modes-a correlation with mode cutoff ratio,'' J. Aircr. **16**(5), 336–343 (1979).
[5] K. R. Holland, P. O. A. L. Davies, and D. C. van der Walt, ''Sound power flux measurements in strongly exited ducts with flow,'' J. Acoust. Soc. Am. **112**, 2863–2871 (2002).
[6] P. O. A. L. Davies, ''Practical flow duct acoustics,'' J. Sound Vib. **124**, 91–115 (1988).
[7] ISO/DIS 5136, ''Acoustics-determination of sound power radiated into a duct by fans and other moving devices in duct method'' (1999).
[8] D. H. Munro and K. U. Ingard, ''On acoustic intensity measurements in the presence of mean flow,'' J. Acoust. Soc. Am. **65**, 1402–1406 (1979).
[9] W. Neise and F. Arnold, ''On sound power determination in flow ducts,'' J. Sound Vib. **244**, 481–503 (2001).
[10] U. Bolleter and M. J. Crocker, ''Theory and measurement of modal spectra in hard walled cylindrical ducts,'' J. Acoust. Soc. Am. **51**, 1439–1477 (1972).
[11] Z. Lazreq, M. Ben Tahar, J. M. Ville, and J. M. Auger, ''A mixed experimental and theoretical method for noise propagation and radiation prediction of a cowled axial fan,'' J. Acoust. Soc. Am. **100**, 166–177 (1996).
[12] Z. Lazreq and J. M. Ville, ''Acoustic calibration of a pressure-velocity probe,'' J. Acoust. Soc. Am. **100**, 364–371 (1996).
[13] J. M. Auger and J. M. Ville, ''Measurement of liner impedance based on the determination of duct eigenvalues by Fourier–Lommel's transform,'' J. Acoust. Soc. Am. **88**(1), 19–22 (1990).
[14] C. L. Morfey, ''Sound transmission and generation in ducts with flow,'' J. Sound Vib. **14**, 37–55 (1971).
[15] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[16] P. Harel and M. Perulli, ''Measurement in a duct of space-structure of the discrete frequency noise generated by an axial compressor,'' J. Sound Vib. **23**(4), 487–506 (1972).
[17] M. Akoum and J. M. Ville, ''Measurement of the reflection matrix of a discontinuity in a duct,'' J. Acoust. Soc. Am. **103**(1), 2463–2468 (1998).
[18] M. Benhammou and J. M. Ville, ''Measurement of intensity vector in duct with flow,'' Internoise, Leuven, Belgium, pp. 1243–1246 (1993).

Nonlinear Fresnel diffraction of weak shock waves

François Coulouvrat[a] and Régis Marchiano
Laboratoire de Modélisation en Mécanique, Université Pierre et Marie Curie & CNRS (UMR 7607), 4 place Jussieu, 75252 Paris Cedex 05, France

(Received 7 June 2003; revised 24 July 2003; accepted 28 July 2003)

Fresnel diffraction at a straight edge is revisited for nonlinear acoustics. Considering the penumbra region as a diffraction boundary layer governed by the KZ equation and its associated jump relations for shocks, similarity laws are established for the diffraction of a step shock, an ''N'' wave, or a periodic sawtooth wave. Compared to the linear case described by the well-known Fresnel functions, it is shown that weak shock waves penetrate more deeply into the shadow zone than linear waves. The thickness of the penumbra increases as a power of the propagation distance, power 1 for a step shock, or $\frac{3}{4}$ for an N wave, as opposed to power $\frac{1}{2}$ for a periodic sawtooth wave or a linear wave. This is explained considering the frequency spectrum of the waveform and its nonlinear evolution along the propagation, and is confirmed by direct numerical simulations of the KZ equation. New formulas for the Rayleigh/Fresnel distance in the case of nonlinear diffraction of weak shock waves by a large, finite aperture are deduced from the present study. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1610454]

PACS numbers: 43.25.Cb, 43.25.Jh [MFH]

I. INTRODUCTION

Diffraction occurs in case of localized singularities of the geometrical approximation. These singularities can be of different types (Berry, 1981): infinite amplitude (caustics), discontinuity of the amplitude (shadow boundary produced by either a smooth or a sharp edge) or of the phase (so-called phase dislocations), and, for optical waves, polarization singularities. In the case of caustics, Berry (1976) has fruitfully related it to the mathematical theory of catastrophes (Thom, 1972).

Belonging to the class of amplitude discontinuities, one of the simpliest and most studied singularities is Fresnel diffraction of a wave at a straight edge. The geometrical approximation predicts an abrupt discontinuity, the undiffracted wave on the illuminated side suddenly vanishing when entering the shadow zone. Diffraction effects take place around the shadow boundary, inside the *penumbra*, to match continuously the geometrical wave to the shadow zone. Following his 1816 memoir, awarded a prize by the Paris Science Academy in 1818, on diffraction fringes in the shadow zone, Fresnel, in subsequent works (collected in 1866), calculated the diffraction caused by straight edges, small apertures, and screens, and introduced the functions now bearing his name. For modern presentations, the reader is referred to the textbooks of Born and Wolf for optics (1975) or Pierce (1981) for acoustics.

The Fresnel function can also be recovered using the method of matched asymptotic expansions. Considering the penumbra region around the shadow boundary as a diffraction boundary layer, in this elongated region, the paraxial approximation of the wave equation describes the inner solution, which is matched to the undiffracted plane wave (the outer solution) far off the penumbra (Buchal and Keller, 1960; Zauderer, 1970). This presentation will be briefly recalled in Sec. II, as it will be further used as a reference case to which nonlinear effects will be compared. The method of matched asymptotic expansion has also been successfully applied to other types of singularities. The description of the field around a fold caustic (the simplest of any caustic) by means of matched asymptotic expansions was introduced in the same article of Buchal and Keller (1960). This approach has since been extensively used, and numerous examples can be found in the books of Bouche and Molinet (1994), or Borovikov and Kinber (1994).

A key advantage of the matched asymptotic expansion is that it can also be extended to nonlinear acoustics. The nonlinear counterpart of the paraxial approximation is the well-known KZ equation (Zabolotskaya and Khokhlov, 1969). It is recalled in Sec. III, and completed by original derivation of the associated shock jump relations, as we are especially interested in weak shock waves here and neglect any absorption phenomenon which would have here only a negligible effect. The KZ equation was introduced initially to model finite-amplitude bounded beams (for a review, refer to Hamilton, 1998). A main advantage of the paraxial approximation is the subsequent one-way assumption, allowing one to solve the KZ equation numerically by advancing plane by plane away from the source. Also, omitting diffraction, the KZ equation reduces to the inviscid Burgers' equation, a very well-known equation for which reliable numerical algorithms have been developed, as proved by comparisons with exact solutions of weak shock theory. The coupling of both advantages has been used for the design of efficient codes, either in the frequency (Frøysa *et al.*, 1993) or in the time domain (Lee and Hamilton, 1995). Finally, despite the paraxial approximation, several comparisons with experiments have shown the high precision of the KZ equation, which is able to reproduce the finest details of the field radiated by large emitters at the fundamental and harmonic fre-

[a]Electronic mail: coulouvr@ccr.jussieu.fr

quencies, either the main axis oscillations or the sidelobes (Hamilton, 1998).

However, the scope of application of the KZ equation is much more general, as it can model many diffraction effects localized along singularities, such as the Fresnel diffraction (as shown in the present study), the tip of finite fold caustics (Marchiano, 2003), or cusped caustics (Cramer and Seebass, 1978; Coulouvrat, 2000). A generalized version of the KZ equation models nonlinear diffraction in the shadow zone of an upward-refracting atmosphere (Coulouvrat, 2002). Omitting the propagation term in this generalized equation, it reduces to the nonlinear Tricomi equation modeling the focusing of weak shock waves at fold caustics (Guiraud, 1965). Another key point is that nonlinear similarity laws can easily be deduced by this approach. The so-called Guiraud's similitude shows that the maximum amplitude of a step shock focusing at a fold caustic varies as the power 4/5 of the amplitude of the incoming wave, obviously a nonlinear law. The same power law can be demonstrated at the tip of a finite fold caustic. For a cusp caustic, there is a similar law (Coulouvrat, 2000), which deviates even more from the linear law with a power 2/3. Therefore, there are numerous examples where intimate coupling between diffraction and nonlinearities occurs.

The objective of the present study is to provide a new example of such a coupling, devoted to nonlinear penumbra diffraction around a shadow boundary. The ingredients of the study will be the same as those described above: the nonlinear paraxial KZ equation and similitude rules for simple wave profiles. Compared to caustics, however, the singularity of the geometrical approximation is smoother here, the amplitude being discontinuous instead of infinite. Therefore, the coupling will be different too. For caustics, nonlinear effects play a key role in limiting the amplitude of the focused shock. For straight-edge diffraction, nonlinearities will be shown to modify the spatial extent of the diffracted zone (penumbra). Indeed, it will be demonstrated (Sec. IV) that the way a weak shock wave is diffracted strongly depends on the overall time waveform. For cases of importance (''N'' wave or step shock), the diffraction boundary layer around the shadow boundary thickens more rapidly for weak shock waves than for linear waves. In other words, and this is the key result of this study, diffracted shock waves penetrate deeper into the shadow zone. Revisiting Fresnel laws almost two centuries later, we could say that nonlinearities enhance diffraction. The degree of this enhancement can be related to the nonlinear evolution of the frequency spectrum of the undiffracted wave. This theoretical result, obtained without solving explicitly the KZ equation, will be nicely confirmed in Sec. V by direct numerical simulations. With this in view, a new time-domain algorithm solving the inviscid KZ equation has been designed, with special emphasis on nonlinear effects (linear diffraction being handled in the same way as by Lee and Hamilton, 1995) which are solved exactly using weak shock theory for the potential, as suggested by Hayes, Heafeli, and Kulsrud (1969) for sonic boom applications. Finally, new expressions for the Rayleigh/Fresnel distance of an aperture of finite size will be deduced from the results of the present study. In the future, this may elucidate some nonlinear behaviors of finite-amplitude beams, thus returning to the original application of the KZ equation.

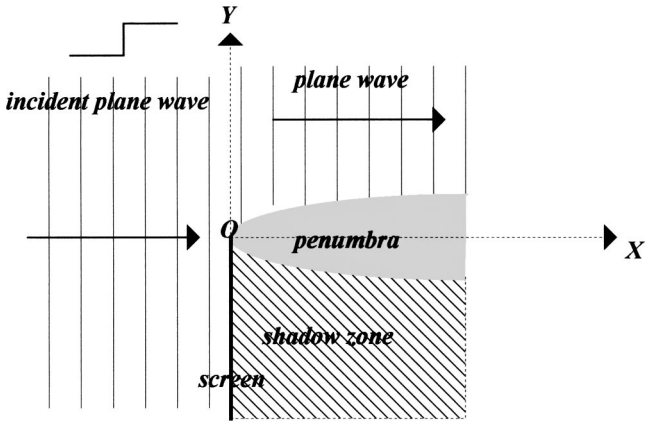


FIG. 1. Weak shock wave diffraction: Geometry of the problem.

II. LINEAR FRESNEL DIFFRACTION AT A STRAIGHT EDGE

In a two-dimensional, homogeneous, and inviscid fluid of ambient density ρ_0 and sound speed c_0, a plane wave is incident normally on a perfectly thin, rigid, and semi-infinite screen. The Ox axis is oriented towards the direction of the incident wave, while the Oy axis is parallel to the screen, the origin being at its edge (Fig. 1). According to the geometrical acoustics approximation, the plane wave would propagate unaffected by diffraction in the upper half-space $y>0$, while no acoustic field would exist in the shadow zone $y<0$. The shadow boundary $y=0$ would be a singular line, the field amplitude being discontinuous there. Indeed, this is only a crude approximation: across this boundary, diffraction effects take place to match continuously the plane wave to the shadow zone. This *penumbra* region can easily be described by matched asymptotic expansions. As the penumbra (the inner region) is close to the shadow boundary, the paraxial approximation of the wave equation can be used at first order

$$\frac{\partial^2 P}{\partial X \partial \tau} = \frac{\partial^2 P}{\partial Y^2}, \tag{1}$$

where the dimensionless acoustic pressure is $P = p_a/\rho_0 c_0 U_0$, p_a being the acoustic pressure and U_0 the velocity amplitude of the incident field. The dimensionless retarded time is denoted $\tau=\omega(t-x/c_0)$, ω being the wave pulsation. The transverse variable y is scaled by a (yet unspecified) length L: $Y=y/L$, and the longitudinal one by the associated Fresnel distance $X=x/D$, with $D=2\omega L^2/c_0$. For a pure monochromatic wave, one has $P(X,Y,\tau) = \hat{P}(X,Y)\exp(-i\tau)$, so that

$$-i\frac{\partial \hat{P}}{\partial X} = \frac{\partial^2 \hat{P}}{\partial Y^2}. \tag{2}$$

Indeed, the problem does not depend on any transverse scale, so the solution should be independent from the arbitrary quantity L. Therefore, the pressure field should be sought as a function of $\eta = Y/\sqrt{X}$, a combination independent of L. This yields the following equation:

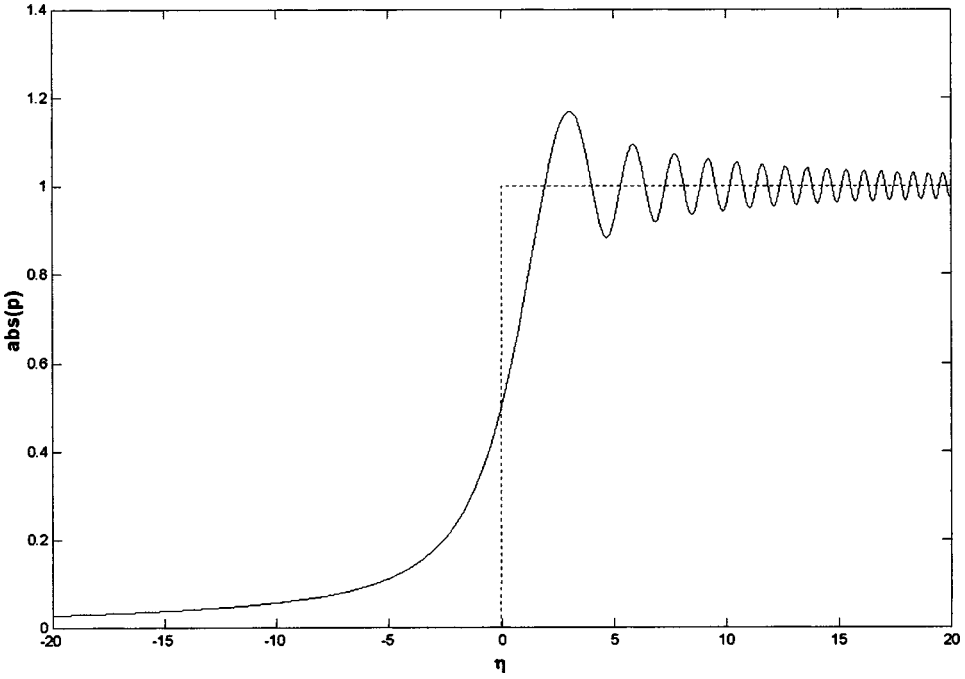


FIG. 2. Fresnel self-similar linear solution (continous line) and geometrical approximation (dotted line).

$$i\eta\frac{d\hat{P}}{d\eta}=2\frac{d^2\hat{P}}{d\eta^2}, \tag{3}$$

with the boundary conditions

$$\hat{P}(\eta\to+\infty)=1, \tag{4}$$

matching the field far from the shadow boundary to the transmitted plane wave unaffected by diffraction (the geometrical approximation, or in terms of matched asymptotic expansion, the outer expansion). The boundary condition in the shadow zone is simply

$$\hat{P}(\eta\to-\infty)=0. \tag{5}$$

The solution of Eq. (3) is expressed in terms of the Fresnel integrals

$$\hat{P}=a\left[C\left(\frac{\eta}{\sqrt{2\pi}}\right)+iS\left(\frac{\eta}{\sqrt{2\pi}}\right)\right]+b, \tag{6}$$

where $C(x)=\int_0^x\cos(\pi t^2/2)dt$ and $S(x)=\int_0^x\sin(\pi t^2/2)dt$. The two constants a and b are determined according to the boundary conditions Eqs. (4) and (5), leading to the self-similar solution

$$\hat{P}=\frac{1}{2}\left(1+(1-i)\left[C\left(\frac{\eta}{\sqrt{2\pi}}\right)+iS\left(\frac{\eta}{\sqrt{2\pi}}\right)\right]\right). \tag{7}$$

The fact that the pressure field is self-similar and depends on the variable η only, illustrates the existence of a *diffraction boundary layer* around the shadow boundary, whose thickness grows as the square root of the distance from the tip of the edge. Right on the singular line ($\eta=0$) $\hat{P}=1/2$, the pressure field is simply half the incident wave. Deep inside the shadow zone ($\eta\to-\infty$), one has

$$\hat{P}\approx\frac{-1}{\eta\sqrt{\pi}}\exp\left[i\left(\frac{\pi+\eta^2}{4}\right)\right]:$$

the shadow field emanates from the edge wave diffracted at the tip of the screen and decaying as a cylindrical wave according to Huygens principle. This well-known Fresnel solution Eq. (7) is illustrated in Fig. 2, showing the characteristic diffraction fringes near the shadow boundary observed by Fresnel himself.

III. THE KZ EQUATION AND THE ASSOCIATED SHOCK RELATIONS

In nonlinear acoustics, the paraxial approximation of the wave equation is the well-known KZ equation (Zabolotskaya and Khokhlov, 1969)

$$\frac{\partial^2 P}{\partial X\partial\tau}-\frac{\partial^2 P}{\partial Y^2}=\frac{\partial^2}{\partial\tau^2}\left(\frac{P^2}{2}\right). \tag{8}$$

Contrary to the linear case, in the nonlinear case there does exist a physical scale D for propagation along the acoustical axis, which is the shock formation distance $D=1/\beta kM$, with $\beta=1+B/2A$, B/A being the nonlinearity parameter and $M=U_0/c_0$ the acoustical Mach number. The corresponding transverse scale L is therefore given by: $L=1/k\sqrt{2\beta M}$.

In the nonlinear, inviscid case, shock waves are likely to happen. Therefore, the KZ propagation equation is insufficient to solve the problem and must be supplemented by shock jump relations. These relations could be deduced from the general Rankine–Hugoniot shock relations (see, for instance, Landau and Lifshitz, 1959), applying on them the same asymptotic process enabling one to derive the KZ equation from the Euler fluid equations. However, a more straightforward derivation is proposed below, following the method of Whitham (1974). Indeed, an equivalent weak formulation of a balance law $(\partial a/\partial t)+\mathrm{div}\,\mathbf{b}=0$ (where generally, but not necessarily, t is time and $\mathbf{x}$ is space) compatible with shock waves is obtained by integrating it over a volume V of surface S: $\partial/\partial t\int\int\int_V a(\mathbf{x},t)d\mathbf{x}=\int\int_S\mathbf{b}(\mathbf{x},t)\cdot\mathbf{n}\,dS$. Now,

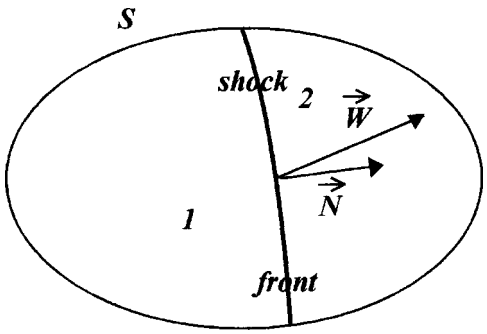


FIG. 3. Notations for shock jump relations.

assuming the existence of a shock wave located on surface Σ with unit normal vector $\mathbf{N}$ oriented from side 1 to side 2 and moving at speed $\mathbf{W}$ (Fig. 3), the usual jump relation satisfied across the shock wave can be proved

$$-W[a]=[\mathbf{b}]\cdot\mathbf{N}, \tag{9}$$

where $[f]=f_2-f_1$ is the notation for the jump of any quantity f across the shock and $W=\mathbf{W}\cdot\mathbf{N}$ is the shock normal speed.

The KZ equation can be expressed in the form of two balance equations of the form below:

$$\frac{\partial P}{\partial X}=\frac{\partial U}{\partial Y}+\frac{\partial}{\partial\tau}\left(\frac{P^2}{2}\right), \tag{10a}$$

$$0=\frac{\partial P}{\partial Y}-\frac{\partial U}{\partial\tau}, \tag{10b}$$

in which the variables (t,x,y) are identified with (X,Y,τ), $a=P$, and $\mathbf{b}=(U,P^2/2)$ for the first conservation equation Eq. (10a), and $a=0$ and $\mathbf{b}=(P,-U)$ for the second one, Eq. (10b). Physically, U is simply the (dimensionless) transverse acoustic velocity in the direction Y associated with diffraction effects, and Eq. (10b) is the usual linearized momentum equation in the transverse direction. Applying general jump relations (9) to Eqs. (10) yields

$$\begin{aligned}0&=[P]N_Y-[U]N_\tau\\ -W[P]&=[U]N_Y+[P^2/2]N_\tau.\end{aligned} \tag{11}$$

For the parabolic equation, the role of time is played by the propagation variable X. Therefore, the shock wave is considered as a line $Y_S(\tau,X)$ in the two-dimensional space (Y,τ), with normal vector $\mathbf{N}=(N_Y,N_\tau)$, evolving with the propagation variable X. From Eqs. (11), it is possible to eliminate the jump of transverse velocity, so as to get

$$-WN_\tau=N_Y^2+\langle P\rangle N_\tau^2, \tag{12}$$

where $\langle P\rangle=(P_1+P_2)/2$ is the mean value of the pressure across the shock. In the one-dimensional case, $\mathbf{N}=(0,1)$ and the jump relation Eq. (12) reduces to the well-known weak shock relation for the inviscid Burgers' equation $d\tau_S/dX=-\langle P\rangle$, or its equivalent geometrical form known as the "law of equal areas."

IV. SIMILITUDE LAWS FOR DIFFRACTED WEAK SHOCK WAVES

In a way similar to the linear self-similar Fresnel solution Eq. (7), we are looking for self-similar solutions of the KZ Eq. (8) and associated shock relation Eq. (12) in the case of an incident weak shock wave. Three cases are studied: a step shock, an "N" wave, and a periodic sawtooth wave. The geometrical approximation (the outer expansion) for these three cases would be on the illuminated side $(Y>0)$

$$P(X,Y>0,\tau)=\begin{cases}0 & \tau<-X/2\\ 1 & \tau>-X/2,\end{cases} \tag{13a}$$

$$P(X,Y>0,\tau)=\begin{cases}-\tau/(1+X) & |\tau|<\sqrt{1+X}\\ 0 & \text{otherwise},\end{cases} \tag{13b}$$

$$P(X,Y>0,\tau)=-\tau/(1+X)\quad|\tau|<\pi \tag{13c}$$

(in the last equation, the pressure is periodic of period 2π), and 0 on the shadow side $Y<0$. Sufficiently far from the edge of the screen $(X\gg1)$, the quantity $X+1$ can be identified with X in Eqs. (13b)–(13c). Undiffracting solutions (13) satisfy the inviscid Burgers equation and weak shock theory.

The inner solution in the penumbra must be solution of the KZ Eq. (8) and of the associated shock relation Eq. (12). Far from the penumbra $Y\to+\infty$, it must match the geometrical approximation Eq. (13) and vanish in the shadow zone $Y\to-\infty$. Searching for self-similar solutions, we introduce the following rescaling:

$$P\to P^*P,\quad X\to X^*X,\quad Y\to Y^*Y,\quad\text{and}\quad\tau\to\tau^*\tau, \tag{14}$$

the quantities * denoting the rescaling amplitude of each corresponding variable. The rescaling (14) will leave the KZ equation invariant provided $P^*=\tau^*/X^*$ and $Y^*=\sqrt{X^*\tau^*}$. The invariance of the matching condition to the geometrical approximation implies the conditions:

$P^*=1$ and $\tau^*=X^*$, and therefore $Y^*=X^*$ for the step shock Eq. (13a); $\tau^*=\sqrt{X^*}$ and $P^*=1/\sqrt{X^*}$, and therefore $Y^*=X^{*3/4}$ for the N wave Eq. (13b); $\tau^*=1$ and $P^*=1/X^*$, and therefore $Y^*=\sqrt{X^*}$ for the sawtooth wave Eq. (13c).

Hence, the self-similar solutions must be sought in the form

$$P(X,Y,\tau)=Q\left(\eta=\frac{Y}{X},\theta=\frac{\tau}{X}\right), \tag{15a}$$

$$P(X,Y,\tau)=\frac{1}{\sqrt{X}}Q\left(\eta=\frac{Y}{X^{3/4}},\theta=\frac{\tau}{\sqrt{X}}\right), \tag{15b}$$

$$P(X,Y,\tau)=\frac{1}{X}Q\left(\eta=\frac{Y}{\sqrt{X}},\theta=\tau\right), \tag{15c}$$

respectively for an incident step shock Eq. (15a), N wave Eq. (15b) and periodic sawtooth wave Eq. (15c). Substituting the above forms in the KZ equation, we get the following equations to be satisfied by the self-similar solutions, respectively, for the step shock Eq. (16a), the N wave Eq. (16b), and the periodic sawtooth wave Eq. (16c):

$$\frac{\partial Q}{\partial \theta}+\theta\frac{\partial^2 Q}{\partial \theta^2}+\eta\frac{\partial^2 Q}{\partial \eta\partial\theta}+\frac{\partial^2 Q}{\partial \eta^2}+\frac{1}{2}\frac{\partial^2 Q^2}{\partial \theta^2}=0, \tag{16a}$$

$$\frac{\partial Q}{\partial \theta}+\frac{1}{2}\theta\frac{\partial^2 Q}{\partial \theta^2}+\frac{3}{4}\eta\frac{\partial^2 Q}{\partial \eta\partial\theta}+\frac{\partial^2 Q}{\partial \eta^2}+\frac{1}{2}\frac{\partial^2 Q^2}{\partial \theta^2}=0, \tag{16b}$$

$$\frac{\partial Q}{\partial \theta}+\eta\frac{\partial^2 Q}{\partial \eta\partial\theta}+\frac{\partial^2 Q}{\partial \eta^2}+\frac{1}{2}\frac{\partial^2 Q^2}{\partial \theta^2}=0, \tag{16c}$$

with the following boundary conditions to match the geometrical acoustics:

$$Q(\eta\to+\infty,\theta)=\begin{cases}0 & \theta<-1/2\\ 1 & \theta>-1/2,\end{cases} \tag{17a}$$

$$Q(\eta\to+\infty,\theta)=\begin{cases}-\theta & |\theta|<1\\ 0 & |\theta|>1,\end{cases}$$

$$Q(\eta\to+\infty,\theta)=-\theta \quad |\theta|<\pi. \tag{17c}$$

In the shadow zone, the field vanishes in all cases

$$Q(\eta\to-\infty,\theta)=0. \tag{18}$$

In time, for the step shock, the field is zero before the wave arrives (causality) and tends to a constant value after it has passed

$$Q(\eta,\theta\to-\infty)=0, \quad \partial Q/\partial\theta(\eta,\theta\to+\infty)=0. \tag{19a}$$

For the N wave, the field is equal to zero before the wave arrives (causality) and returns to zero after it has passed

$$Q(\eta,\theta\to\pm\infty)=0. \tag{19b}$$

For the sawtooth wave, the signal is periodic

$$Q(\eta,\theta)=Q(\eta,\theta+2\pi). \tag{19c}$$

Self-similar solutions must satisfy not only the KZ equation, but also the jump relation Eq. (12), as all solutions exhibit shock waves. This is checked by using the self-similar form Eqs. (15) and introducing the instant of shock $\theta_S(\eta)$ as a function of the self-similar distance across the diffraction boundary layer η. This yields the following equation of the shock line, respectively, for the step shock, the N wave, and the sawtooth wave:

$$(Y_S,\tau_S)=(\eta X,\theta_S(\eta)X), \tag{20a}$$

$$(Y_S,\tau_S)=(\eta X^{3/4},\theta_S(\eta)X^{1/2}), \tag{20b}$$

$$(Y_S,\tau_S)=(\eta X^{1/2},\theta_S(\eta)), \tag{20c}$$

where the variable η parametrizes the shock line at a fixed position X. From this comes the shock-front speed $\mathbf{W}=(\partial Y_S/\partial X,\partial\tau_S/\partial X)$, the shock-front tangent vector $\mathbf{T}=(\partial Y_S/\partial\eta,\partial\tau_S/\partial\eta)$, and the shock-front unit normal vector $\mathbf{N}=(-\partial\tau_S/\partial\eta,\partial Y_S/\partial\eta)/|\mathbf{T}|$. Substituting this into Eq. (12) yields the shock equation satisfied by $\theta_S(\eta)$ for the penumbra diffraction of, respectively, the step shock Eq. (21a), the N wave Eq. (21b), and the sawtooth wave Eq. (21c)

$$\theta_S(\eta)-\eta\frac{d\theta_S}{d\eta}+\left(\frac{d\theta_S}{d\eta}\right)^2+\langle Q\rangle=0, \tag{21a}$$

$$\frac{1}{2}\theta_S(\eta)-\frac{3}{4}\eta\frac{d\theta_S}{d\eta}+\left(\frac{d\theta_S}{d\eta}\right)^2+\langle Q\rangle=0, \tag{21b}$$

$$-\frac{1}{2}\eta\frac{d\theta_S}{d\eta}+\left(\frac{d\theta_S}{d\eta}\right)^2+\langle Q\rangle=0. \tag{21c}$$

These first-order differential equations must be completed by the boundary condition matching the diffracted shock position $\theta_S(\eta)$ to the undiffracted one as $\eta\to+\infty$, $\theta_S(+\infty)=-1/2$ for the step shock, $\theta_S(+\infty)=\pm1$ for the N wave, and $\theta_S(+\infty)=\pi$ for the periodic sawtooth wave. This completes the set of Eqs. (15)–(19) satisfied by the self-similar solutions.

The main result of the present study is given by Eqs. (15). It demonstates that the linear Fresnel laws of diffraction are substantially modified in nonlinear acoustics. Especially, it is proved that the diffraction of weak shock waves is strongly dependent on the shape of the waveform. While in linear acoustics, the diffraction boundary layer thickness (the penumbra) increases as the power $\frac{1}{2}$ of the distance from the diffracting edge, this result may not be true anymore in nonlinear acoustics. It remains so for the periodic sawtooth wave [Eq. (15c)], but the penumbra thickens as the power $\frac{3}{4}$ of the distance in case of an N wave [Eq. (15b)], and even linearly with the distance in case of a step shock [Eq. (15a)]. Hence, because of nonlinear effects, a diffracted step shock or N wave penetrates much deeper into the shadow zone than a linear or periodic sawtooth wave. For a step shock, that result could have been inferred easily from a basic dimension argument. Indeed, the dimensionless variables (X,Y) of the KZ equation both depend linearly on a typical wave number k. But, a step shock has no characteristic duration, and the problem should not depend on this arbitrary wave number. Therefore, it can depend only on the ratio Y/X, which is independent from k. The same reasoning explains the self-similar time variable τ/X, also independent from any frequency ω. For an undiffracting N wave, the duration of the signal increases (and therefore the frequency and the wave number decrease) with the square root of the distance. As, according to the linear Fresnel law, the penumbra thickness is proportional to $\sqrt{x/k}$, it becomes proportional to $x^{3/4}$ since k decreases with $x^{1/2}$: the diffraction boundary layer has to thicken faster than in the linear case to compensate for the nonlinear shift of the spectrum towards low frequency. On the contrary, the frequency spectrum of the periodic sawtooth wave remains constant, which explains why the penumbra thickness is similar to the linear Fresnel law in this case.

For more complex incident waveforms, the present study based on self-similarity cannot be used as such. However, significant effects are qualitatively expected. For instance, after nonlinear evolution, a finite pulse will be an intermediate between the N wave, similar to the start and the end of the signal, and a periodic sawtooth wave, for its main part. According to the above analysis, we can expect the start and end of the signal to penetrate deeper into the shadow zone, where they will be reinforced relatively to the main part so that the envelope of the pulse will flatten there. A more detailed analysis, likely to depend on the duration of the signal and the shape of its envelope, is beyond the scope

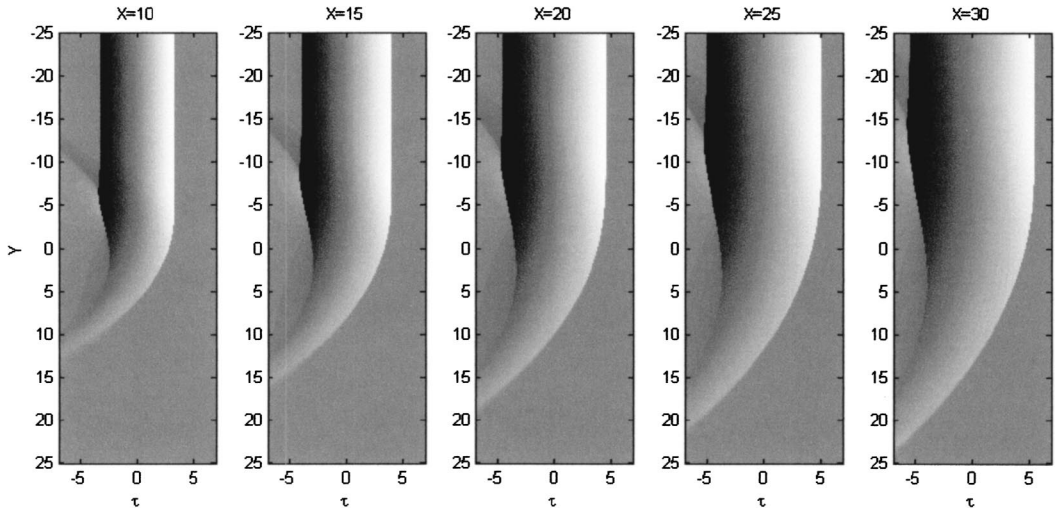


FIG. 4. Diffraction of an N wave. The penumbra solution of the nonlinear KZ equation is represented in physical dimensionless variables (horizontal: retarded time τ, vertical: transverse distance Y) at different distances X from the screen. The pressure level is represented in gray scales.

of the present analysis and would deserve further study relying on the numerical solver of the KZ equation presented below. Finally, despite its essentially theoretical objective, this study nevertheless demonstrates in practice that screens are likely to be less efficient against weak acoustical shock waves than could be expected from a purely linear analysis. Their design to protect from such type of loud sounds should necessarily take into account nonlinear effects.

V. NUMERICAL SIMULATION

To validate the theoretical results presented above, we compare them to a numerical simulation of the KZ equation Eq. (8). The incoming pressure field is specified in the plane of the screen $X=0$

$$P(X=0,Y,\tau)=\begin{cases} F(\tau) & Y>0 \\ 0 & \text{otherwise,} \end{cases} \tag{22}$$

with the incident waveform

$$F(\tau)=\begin{cases} 0 & \tau<0 \\ 1 & \tau>0, \end{cases} \tag{23a}$$

$$F(\tau)=\begin{cases} -\tau & |\tau|<\sqrt{1} \\ 0 & \text{otherwise,} \end{cases} \tag{23b}$$

$$F(\tau)=-\tau \quad |\tau|<\pi, \tag{23c}$$

respectively, for a step shock Eq. (23a), an N wave Eq. (23b), and a periodic sawtooth wave Eq. (23c). For a numerical simulation, the KZ equation is expressed in terms of the potential $\Phi(X,Y,\tau)=\int_{-\infty}^{\tau}P(X,Y,u)du$

$$\frac{\partial^2\Phi}{\partial X\partial\tau}-\frac{\partial^2\Phi}{\partial Y^2}=\frac{1}{2}\frac{\partial}{\partial\tau}\left(\left(\frac{\partial\Phi}{\partial\tau}\right)^2\right), \tag{24}$$

and is solved numerically advancing plane by plane away from the screen by small steps ΔX. For each step, the split-step method is used. At first, only diffraction effects are considered, and nonlinear ones on the right-hand side of Eq. (24) are omitted. This linear equation is discretized in the time domain by finite differences according to the scheme of Lee and Hamilton (1995). This provides a first estimation of the solution in the new plane, taking into account only diffraction effects.

This estimation is improved by then omitting diffraction effects and recovering nonlinear ones, so that the KZ equation reduces to the inviscid Burgers' equation for the potential

$$\frac{\partial\Phi}{\partial X}=\frac{1}{2}\left(\frac{\partial\Phi}{\partial\tau}\right)^2. \tag{25}$$

An analytical solution of Eq. (25) is known, based on the implicit Poisson solution. In case of shocks, this solution may, however, be multivalued. Then, the physically admissible, single-valued solution is simply the maximum value of all multiple values, a test which is much simpler to implement from a numerical point of view than the weak shock theory, as in the Pestorius (1973) algorithm, or the law of equal areas. This procedure for solving the Burgers' equation via the potential was used by Hayes, Haefeli, and Kulsrud (1969) to determine sonic boom waveform distortion along acoustical rays launched by a supersonic aircraft. A full demonstration of the properties of the potential is given by Coulouvrat (2003). It has been applied to nonlinear diffraction problems in the case of shock wave focusing at fold caustics (Marchiano, Coulouvrat, and Grenon, 2003), where the pressure field satisfies the nonlinear Tricomi equation, which is a mixed-type transonic equation (elliptic/hyperbolic). Comparisons to experimental simulations (Marchiano, Thomas, and Coulouvrat, 2003) demonstrate the validity of this numerical approach.

Figures 4 and 5 present, respectively, the pressure field for a diffracted step shock or N wave at different distances X from the screen. The horizontal axis is τ, the vertical one is Y, the pressure levels are indicated by a gray scale. The wave diffraction in the penumbra, and its penetration inside the shadow zone is clearly observable in both cases. The self-similar aspect as expected from Eq. (15) is demonstrated by Figs. 6 and 7, where the *same* numerical solution is repre-

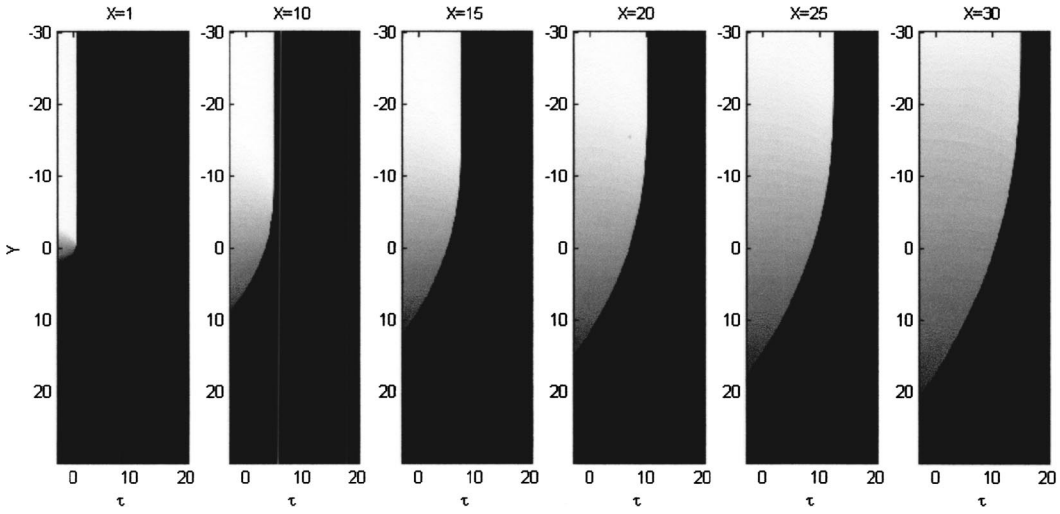


FIG. 5. Diffraction of a step shock. The penumbra solution of the nonlinear KZ equation is represented in physical dimensionless variables (horizontal: retarded time τ, vertical: transverse distance Y) at different distances X from the screen. The pressure level is represented in gray scales.

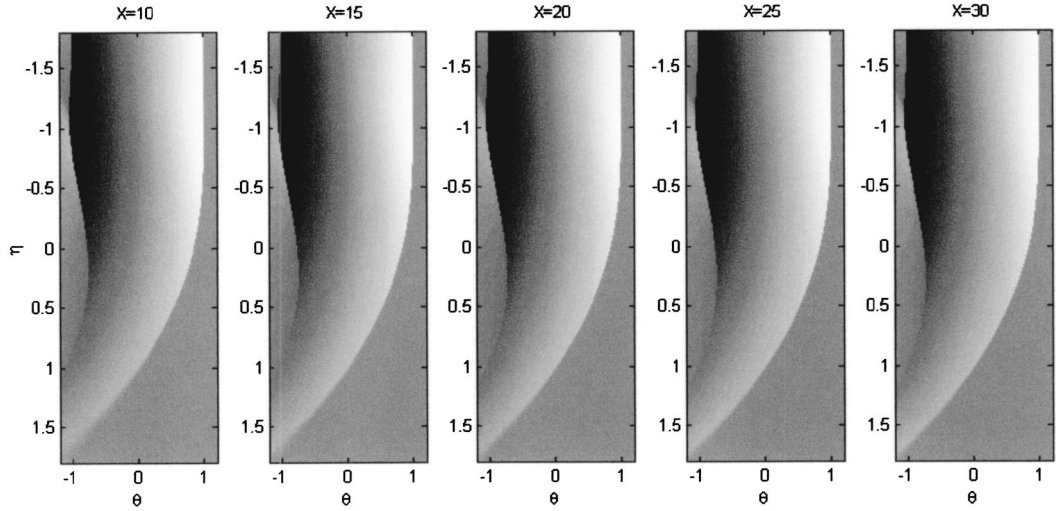


FIG. 6. Diffraction of an N wave. The penumbra solution of the nonlinear KZ equation is represented in self-similar variables (horizontal: θ, vertical: η) at different distances X from the screen. The pressure level is represented in gray scales. The invariance of the solution demonstrates the numerical solution is self-similar, in agreement with theory.

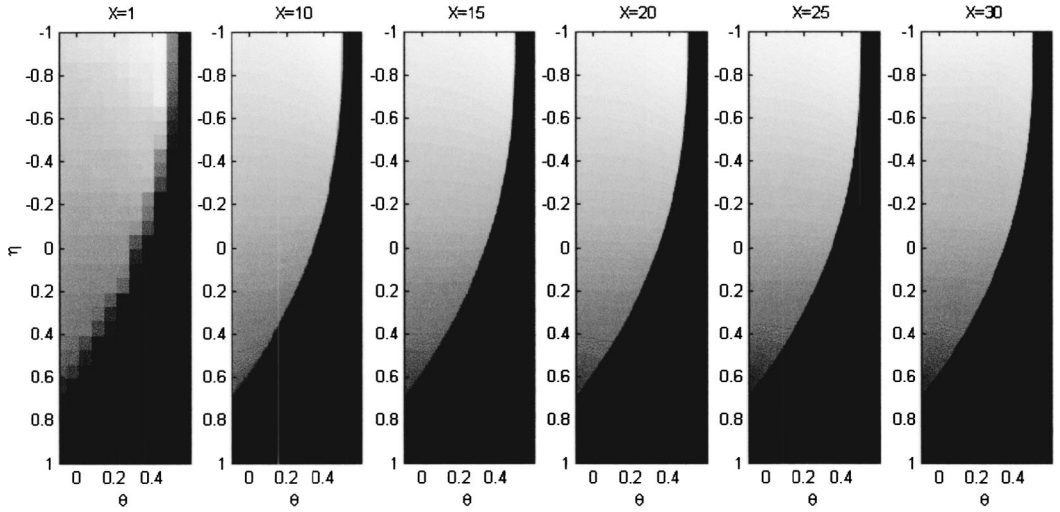


FIG. 7. Diffraction of step shock. The penumbra solution of the nonlinear KZ equation is represented in self-similar variables (horizontal: θ, vertical: η) at different distances X from the screen. The pressure level is represented in gray scales. The invariance of the solution demonstrates the numerical solution is self-similar, in agreement with theory.

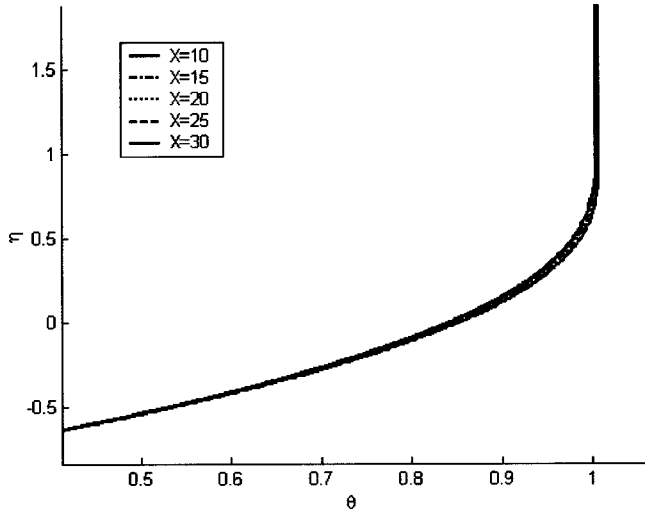


FIG. 8. Position of the first shock front of the diffracted N wave (numerical solution of the KZ equation) represented in self-similar variables (horizontal: θ, vertical: η) at different distances X from the screen. The superposition of the different curves demonstrates the numerical solution is self-similar, in agreement with theory.

sented in terms of the corresponding self-similar variables η and θ. In these variables, it is obvious that the numerical solution of the KZ equation is independent from the propagation distance. In other words, it is indeed self-similar, in agreement with the theoretical laws, Eq. (15). A more quantitative visualization of the self-similar aspect of the numerical solutions is illustrated by Figs. 8 and 9, showing the position of the initial shock front $\theta_s(\eta)$ reported as a function of the self-similar variables η and θ for different values of X. The almost-perfect superposition of the different curves proves that the numerical simulations of the KZ equation satisfy the nonlinear self-similar rules, Eq. (15).

VI. CONCLUSION

The theoretical analysis of Sec. III, confirmed by numerical simulations of Sec. IV, demonstrates that diffraction

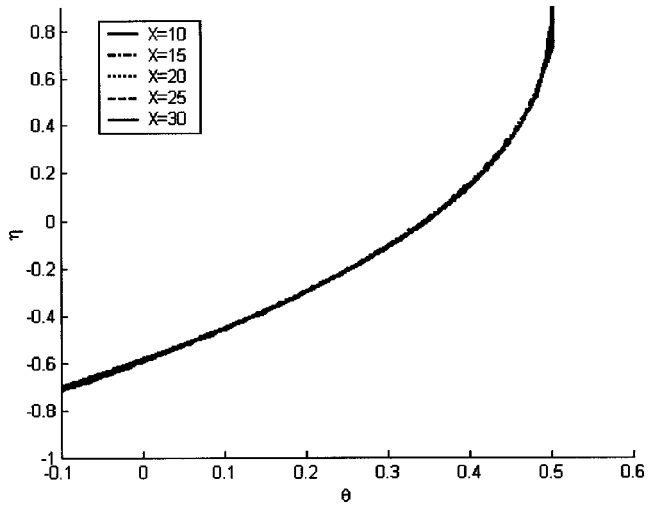


FIG. 9. Position of the shock front of the diffracted step shock (numerical solution of the KZ equation) represented in self-similar variables (horizontal: θ, vertical: η) at different distances X from the screen. The superposition of the different curves demonstrates the numerical solution is self-similar, in agreement with theory.

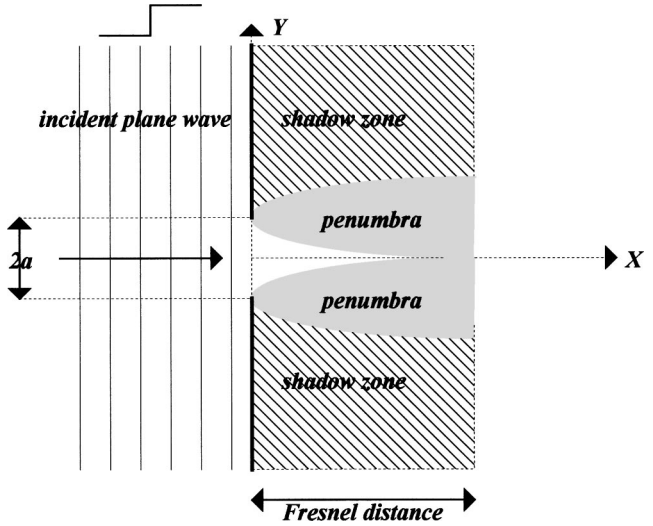


FIG. 10. Fresnel distance defined as the junction of the two penumbras emanating from the two edges of the slit.

of weak shock waves is intrinsically nonlinear and obeys similitude rules different from those of usual linear acoustics. Especially, it is proved that the thickness of the penumbra is dependent on the incident waveform, as it has to adapt to the evolution of the wave spectrum propagating nonlinearly far off the shadow boundary. Consequently, a step shock or an N wave penetrates deeper into the shadow zone than a periodic sawtooth wave: for weak shock waves, nonlinearities enhance diffraction.

An important consequence of this is a redefinition of the Fresnel/Rayleigh length for the case of diffraction by a slit of finite aperture $l=2a$ large compared to the wavelength. The Fresnel/Rayleigh length F can be estimated at the distance at which the two diffraction boundary layers emanating from each edge of the aperture do contact one another (Borovikov and Kinber, 1994) (Fig. 10). All points located beyond this distance will ''feel'' simultaneously both edge waves, which interfere in a transition between the near field and the far field. In linear acoustics, or for a periodic sawtooth wave, according to Fresnel self-similar solution, this gives $1=\eta=y\sqrt{2k/x}$ for $y=a$ and $x=F$, so that $F=kl^2/2$. However, in nonlinear acoustics, the same reasoning would give $F=l/\sqrt{2\beta M}$ for the Fresnel/Rayleigh distance of a step shock, and $F=k^{1/3}l^{4/3}/(4\beta M)^{1/3}$ for an N wave. In the case of a step shock, it is of course independent from any frequency and depends only on the slit aperture (linearly) and wave amplitude, while for an N wave it involves an unexpected combination of frequency, aperture, and amplitude. For these last two cases, as diffraction is enhanced by nonlinearities, the Fresnel/Rayleigh distance varies as a smaller power with the aperture (power 1 or 4/3 instead of 2), so that the near field will be of lesser extent than in the linear case.

ACKNOWLEDGMENTS

The authors are grateful to Jean-Louis Thomas (Lab. Milieux Désordonnés et Hétérogènes, Université Pierre et Marie Curie, UMR CNRS 7603) for useful discussions on the topic.

Berry, M. V. (**1976**). ''Waves and Thom's theorem,'' Adv. Phys. **25**, 1–26.
Berry, M. V. (**1981**). ''Singularities in waves and rays,'' in *Physics of Defects*, edited by R. Balian, M. Kleman, and J. P. Poirier (North-Holland, Amsterdam), pp. 453–543.
Born, M., and Wolf, E. (**1975**). *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of light*, 5th ed. (Pergamon, Oxford), pp. 428–435.
Borovikov, V. A., and Kinber, B. Ye. (**1994**). *Geometrical Theory of Diffraction* (The Institution of Electrical Engineers, London), Chap. V, pp. 92–154.
Bouche, D., and Molinet, F. (**1994**). *Méthodes Asympotiques en Electromagnétisme (Asymptotic Methods in Electromagnetism)* (Springer, Berlin), pp. 132–135 (in French).
Buchal, R. N., and Keller, J. B. (**1960**). ''Boundary layer problems in diffraction theory,'' Commun. Pure Appl. Math. **13**, 85–114.
Coulouvrat, F. (**2000**). ''Focusing of weak acoustic shock waves at a caustic cusp,'' Wave Motion **32**, 233–245.
Coulouvrat, F. (**2002**). ''Sonic boom in the shadow zone: A geometrical theory of diffraction,'' J. Acoust. Soc. Am. **111**, 499–508.
Coulouvrat, F. (**2003**). ''Propagation acoustique non linéaire en milieu inhomogène: De la modélisation à la simulation (Nonlinear acoustical propagation in heterogeneous media: From modeling to simulation),'' Ecole de Printemps de Mécanique des Fluides Numérique, Fréjus (France), 2–7 June 2003 (in French).
Cramer, M. S., and Seebass, A. R. (**1978**). ''Focusing of weak shock waves at an arête,'' J. Fluid Mech. **88**, 209–222.
Fresnel, A. J. (**1816**). ''Mémoire sur la diffraction de la lumière, où l'on examine particulièrement le phénomène des franges colorées que présentent les ombres des corps éclairés par un point lumineux (Memoir on the diffraction of light, where we especially examine the phenomenon of coloured fringes shown by the shadows of bodies illuminated by a bright point),'' Ann. Phys. Chimie **1**, 239–281 (in French).
Fresnel, A. J. (**1866**). *Oeuvres Complètes (Complete Works)*, Vol. 1 (Imprimerie Impériale, Paris) (fac-similé, Bergeret (Bordeaux, 1995) (in French).
Frøysa, K.-E., Tjøtta, J. N., and Berntsen, J. (**1993**). ''Finite amplitude effects in sound beams. Pure tone and pulsed excitation,'' in *Advances in Nonlinear Acoustics*, edited by H. Hobæk, 13th International Symposium on Nonlinear Acoustics, Bergen (Norway), 28 June–2 July (World Scientific, Singapore), pp. 233–238.
Guiraud, J.-P. (**1965**). ''Acoustique géométrique, bruit balistique des avions supersoniques et focalisation (Geometrical acoustics, ballistic noise of supersonic aircraft and focusing),'' J. Mec. **4**, 215–267 (in French).
Hamilton, M. F. (**1998**). ''Sound beams,'' in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, San Diego), pp. 233–261.
Hayes, W. D., Haefeli, R. C., and Kulsrud, H. E. (**1969**). ''Sonic boom propagation in a stratified atmosphere with computer program,'' NASA CR-1299.
Landau, L. D., and Lifshitz, E. M. (**1959**). *Fluid Mechanics* (Pergamon, London), pp. 317–319.
Lee, Y.-S., and Hamilton, M. F. (**1995**). ''Time-domain modeling of pulsed finite-amplitude sound beams,'' J. Acoust. Soc. Am. **97**, 906–917.
Marchiano, R. (**2003**). Private communication.
Marchiano, R., Coulouvrat, F., and Grenon, R. (**2003**). ''Numerical simulation of shock wave focusing at fold caustics, with application to sonic boom,'' J. Acoust. Soc. Am. (accepted).
Marchiano, R., Thomas, J.-L., and Coulouvrat, F. (**2003**). ''Experimental simulation of supersonic superboom in a water tank: Nonlinear focusing of weak shock waves at a fold caustic,'' Phys. Rev. Lett. (accepted).
Pestorius, F. M. (**1973**). ''Propagation of plane acoustic noise of finite amplitude,'' Technical Report ARL-TR-73-23, Applied Research Laboratories, The University of Texas at Austin.
Pierce, A. D. (**1981**). *Acoustics: An Introduction to its Physical Principles and Applications* (McGraw-Hill, New York), pp. 239–243.
Thom, R. (**1972**). *Stabilité Structurelle et Morphogenèse* (Benjamin, Reading), pp. 72–107 (in French), English translation (1975): *Structural Stability and Morphogenesis* (Benjamin, Reading).
Whitham, G. B. (**1974**). *Linear and Nonlinear Waves* (Wiley, New York).
Zabolotskaya, E. A., and Khokhlov, R. V. (**1969**). ''Quasi-plane waves in the nonlinear acoustics of confined beams,'' Sov. Phys. Acoust. **15**, 35–40.
Zauderer, E. (**1970**). ''Boundary layer and uniform asymptotic expansions for diffraction problems,'' SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **19**, 575–600.

Numerical simulation of shock wave focusing at fold caustics, with application to sonic boom

Régis Marchiano[a] and François Coulouvrat
Laboratoire de Modélisation en Mécanique, UMR CNRS 7607, Université Pierre et Marie Curie, 8 rue du Capitaine Scott, 75015 Paris, France

Richard Grenon
Département d'Aérodynamique Appliquée, DAAP/ACI, ONERA, 29 Avenue de la division Leclerc, BP2, 92322 Châtillon Cedex, France

(Received 13 February 2003; revised 11 July 2003; accepted 28 July 2003)

Weak shock wave focusing at fold caustics is described by the mixed type elliptic/hyperbolic nonlinear Tricomi equation. This paper presents a new and original numerical method for solving this equation, using a potential formulation and an ''exact'' numerical solver for handling nonlinearities. Validation tests demonstrate quantitatively the efficiency of the algorithm, which is able to handle complex waveforms as may come out from ''optimized'' aircraft designed to minimize sonic booms. It provides a real alternative to the approximate method of the hodograph transform. This motivated the application to evaluate the ground track focusing of sonic boom for an accelerating aircraft, by coupling CFD Euler simulations performed around the mock-up on an adaptated mesh grid, atmospheric propagation modeling, and the Tricomi algorithm. The chosen configuration is the European Eurosup mock-up. Convergence of the focused boom at the ground level as a function of the matching distance is investigated to demonstrate the efficiency of the numerical process. As a conclusion, it is indicated how the present work may pave the way towards a study on sonic superboom (focused boom) mitigation. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1610459]

PACS numbers: 43.25.Cb, 43.25.Jh [MFH]

I. INTRODUCTION

Sonic booms remain a community acceptance problem that may jeopardize the development of future civil supersonic aircraft (either supersonic transport or business jet). The most intense sonic boom is the focused sonic boom due to the aircraft transonic acceleration from Mach 1 to cruise speed, and it cannot be avoided by realistic maneuvers. It leads to an amplification of ground pressures up to two to three times the carpet boom shock strength. Therefore, to comply with a future international regulation on sonic boom, it is important to predict accurately the level of focused booms. The present work will demonstrate such a prediction is now within reach. It is a complex task, as it requires a precise CFD simulation of the pressure field far enough from the aircraft, a correct matching between aerodynamical evaluations and acoustical propagation modeling in the atmosphere, and a validated modeling of shock waves focusing around the caustic at ground level.

According to the classic theory (Hayes *et al.*, 1969), sonic booms are computed within the framework of geometrical acoustics. The eikonal (phase) function is determined by the ray path, and the signal amplitude by the ray-tube area. Nonlinear effects along each ray entail the pressure signal to evolve from the complicated shock flow around the body of the aircraft down to the ultimate ''N'' wave reached at ground level for long propagation distances. The input pressure signal describing the aircraft flow field was initially estimated from the linear aerodynamic theory for slender bodies (Whitham, 1952), a theory that also provides the correct matching to geometrical acoustics. However, Whitham's linear approximation lacks precision for relatively low Mach numbers such as those of acceleration focusing (around Mach 1.2 to 1.4). Nowadays numerical means enable CFD simulations to be reliable far enough from the aircraft (one or several fuselage lengths). Direct matching between CFD simulations and acoustical propagation can be achieved for cruise (steady) flight, with a good convergence at ground level provided the matching is performed far enough from the aircraft. More sophisticated ways of matching can be achieved closer to the aircraft, but have not been fully explored yet (Plotkin and Page, 2002).

Caustics are surfaces, and their ground intersection are lines, where the ray-tube area vanishes and the geometrical acoustics approximation neglecting diffraction breaks down. Around regular smooth caustics surfaces [''fold'' caustics according to the terminology of catastrophe theory (Thom, 1972; Berry, 1976)], diffraction must be taken into account inside a ''diffraction boundary layer'' around the caustic (Buchal and Keller, 1960) to be matched to geometrical acoustics. Within this layer, the pressure satisfies the linear Tricomi equation, whose generic solution in the frequency domain is the well-known Airy function, in agreement with the catastrophe theory. If, as for sonic booms, the incoming signal possesses shock waves, the amplified signal near the caustics and the outgoing signal exhibit a ''U'' shape resulting from the $\pi/2$ phase jump through the caustics. This shape is substantiated by flight tests (Wanner *et al.*, 1972), but the

[a]Electronic mail: marchi@lmm.jussieu.fr

 0001-4966/2003/114(4)/1758/14/$19.00

linear theory fails to predict finite peaks for the ''U'' wave. To recover finite amplitudes, nonlinearities must be taken into account as an additional ''limiting'' mechanism. The resulting equation satisfied by the pressure field is the so-called *nonlinear Tricomi equation* (Guiraud, 1965) which is a mixed-type (elliptic/hyperbolic) equation (Sec. II). The process of linear diffraction being the dominant mechanism around caustics, supplemented by nonlinearities, is supported by laboratory-scale experiments (Sturtevant and Kulkarny, 1976; Marchiano *et al.*, 2003) at small Mach numbers. The objective of this paper is to present a new numerical method to solve the *nonlinear Tricomi equation* using a pseudospectral method, and to apply it for predicting ground track sonic boom focusing coupled with CFD nearfield simulations of a realistic high-speed supersonic transport.

The new algorithm for solving the nonlinear Tricomi equation derives from a previous version (Auger and Coulouvrat, 2002). The reader is referred to this work for an extensive bibliography on numerical methods applied to the nonlinear Tricomi equation. Let us only cite the most recent work of Cheng and Hafez (2002). Compared to this previous code, two very substantial improvements have been brought (Sec. III). The equation is now solved for the potential field instead of the pressure field, and nonlinear effects are treated with an ''exact'' solver that removes artificial numerical dissipation and avoids any stability condition. This results into an innovative combination of numerical methods with potential applications for other equations in nonlinear acoustics (such as the KZ equation). As a consequence, the number of iterations and the computation time have been reduced dramatically by a factor of (roughly) 40, the convergence of the maximum peak amplitude with mesh refinement is now perfectly reached, and the theoretical Guiraud's similitude is satisfied numerically with a good precision. This constitutes a complete and quantitative validation of the algorithm (Sec. IV), which is able to handle complex waveforms as may come out from ''optimized'' aircraft designed to minimize sonic booms. It provides a real alternative to the approximate method of the hodograph transform developed by Seebass (1971) and Gill and Seebass (1973), and currently applied for the numerical evaluation of the superboom (Plotkin, 2002). This motivated the application to evaluate the ground track focusing of the sonic boom of an accelerating aircraft, by coupling CFD Euler simulations performed far away from the aircraft on an adaptated mesh grid, atmospheric propagation modeling, and the Tricomi algorithm (Sec. V). Convergence of the focused boom at ground level as a function of the matching distance is investigated to demonstrate the efficiency of the numerical process. As a conclusion, we indicate how the present work may pave the way towards a study on sonic superboom (focused boom) mitigation.

II. THE PHYSICAL PROBLEM: THE NONLINEAR TRICOMI EQUATION

From a physical point of view, a caustic is a region where sound is amplified. In terms of geometrical acoustics, it is the locus of points where the ray-tube area vanishes and where the geometrical approximation breaks down. It is also an envelope of rays, tangent to acoustical rays at any point.

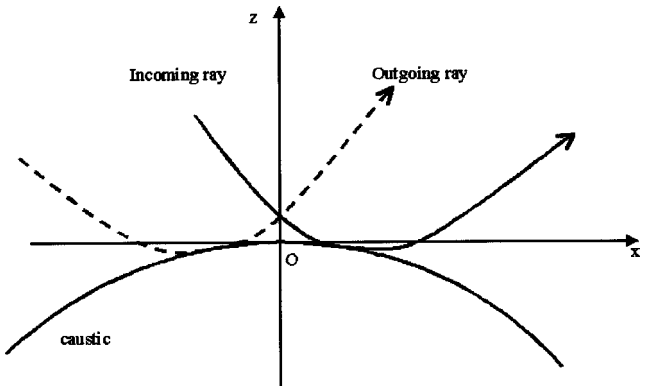


FIG. 1. Geometry of the caustic.

So, near a regular caustic [a fold caustic in the terminology of catastrophe theory (Thom, 1972)], either two or zero acoustical rays go through a given point, depending on which side (''illuminated'' or ''shadow'' side) of the caustic this point lies on (Fig. 1). For a point on the illuminated side, among the two rays going through this point, one (the so-called *incoming ray*) has not yet tangented the caustic, whereas the other one (the so-called *outgoing ray*) has already tangented the caustic. Sufficiently far from the caustics, these two rays do not interact and sound propagates according to the laws of geometrical acoustics. However, close to the caustics, the two rays become indiscernable [according to the terminology of Kravtsov and Orlov (1993)]. There, diffraction effects must be taken into account to limit the amplitude of the acoustical field. The thickness δ of the diffraction boundary layer (Buchal and Keller, 1960) around the caustic can be chosen as the distance at which the arrival times of the incoming and outgoing signals differ by less than the period of the signal (if periodic), or its duration (for a pulse). Let us choose the origin O at some point of the fold caustic, and introduce the Cartesian coordinate system $(Oxyz)$, with Ox being tangent to the caustic at point O and oriented towards the direction of the ray tangenting the caustic at this point, and Oz being oriented in the normal direction to the caustic, towards the illuminated side. At point O, we denote by ρ_0 the ambient density, c_0 the ambient sound speed, and p_0 the ambient pressure. The radius of curvature of the intersection of the caustics with the (Oxz) plane is R_{sec}, while R_{ray} is the radius of curvature of the projection of the acoustical ray on the (Oxz) plane and $R_{\mathrm{cau}}=1/(1/R_{\mathrm{sec}}+1/R_{\mathrm{ray}})$ is the radius of curvature of the acoustical ray relative to the caustics. To formulate the problem in a dimensionless form, three characteristic quantities are introduced:

(1) the characteristic duration of the acoustical signal: T_{ac},
(2) the thickness of the diffraction boundary (Buchal and Keller, 1960): $\delta=(2/c_0^2T_{ac}^2R_{\mathrm{cau}})^{-1/3}$, and
(3) the maximal pressure p_{ac} of the incoming acoustical signal at distance δ from the caustic.

According to the catastrophe theory (Berry, 1976), the pressure near the fold caustic can be shown in the high-frequency limit to be a function of two independent variables only: the distance to the caustic z, and the phase of the signal. The corresponding dimensionless variables are

(1) the dimensionless delayed time: $\bar{\tau}=[t-x(1-z/R_{\mathrm{sec}})/c_0]/T_{\mathrm{ac}}$,
(2) the dimensionless distance to the caustic: $\bar{z}=z/\delta$, and
(3) the dimensionless pressure: $\bar{p}=(p-p_0)/p_{\mathrm{ac}}$.

In writing the Euler equations in terms of these variables, the pressure field can be shown to satisfy the nonlinear Tricomi equation (Guiraud, 1965; Hayes, 1968):

$$\frac{\partial^2\bar{p}}{\partial\bar{z}^2}-\bar{z}\frac{\partial^2\bar{p}}{\partial\bar{\tau}^2}+\frac{\mu}{2}\frac{\partial^2(\bar{p}^2)}{\partial\bar{\tau}^2}=0, \tag{1}$$

where coefficient μ measures the nonlinear effects relative to the diffraction effects:

$$\mu=2\beta\mathrm{M}_{\mathrm{ac}}\left[\frac{R_{\mathrm{cau}}}{(2c_0T_{\mathrm{ac}})}\right]^{2/3}, \tag{2}$$

with $\beta=1+B/2A$ the nonlinearity parameter of the medium, and M_{ac} the acoustical Mach number ($\mathrm{M}_{\mathrm{ac}}=P_{\mathrm{ac}}/(\rho_0c_0^2)$).

The associated boundary conditions of the nonlinear Tricomi equation are written below.

(1) In time, for a transient signal, the medium is not perturbed before or after the acoustical wave has passed:

$$\bar{p}(\bar{z},\bar{\tau}\to\pm\infty)=0; \tag{3}$$

or, for a periodic signal (with a period T), one simply gets

$$\bar{p}(\bar{z},\bar{\tau}+T)=p(\bar{z},\bar{\tau}). \tag{4}$$

(2) Away from the caustic in the shadow zone, the acoustical pressure decays exponentially:

$$\bar{p}(\bar{z}\to-\infty,\bar{\tau})\to0. \tag{5}$$

(3) Away from the caustic on the illuminated side, the field matches the geometrical acoustics approximation:

$$\bar{p}(\bar{z}\to+\infty,\bar{\tau})=\bar{z}^{-1/4}[F(\bar{\tau}+\tfrac{2}{3}\bar{z}^{3/2})+G(\bar{\tau}-\tfrac{2}{3}\bar{z}^{3/2})]. \tag{6}$$

The F function is the (dimensionless) time waveform of the *incoming* signal, before it is affected by diffraction while tangenting the caustic. Therefore, the F function is supposed to be known. On the contrary, the G function is the time waveform along the outgoing ray. Unlike the incoming signal F, the outgoing signal G has undergone the diffraction effects after having tangented the caustic, and is unknown. To eliminate this unknown function, the matching boundary condition Eq. (6) can be written as a ''radiation condition,'' by a combination of its derivatives with respect to $\bar{z}$ and $\bar{\tau}$:

$$\bar{z}^{1/4}\frac{\partial\bar{p}}{\partial\bar{\tau}}+\bar{z}^{-1/4}\frac{\partial\bar{p}}{\partial\bar{z}}\xrightarrow[\bar{z}\to+\infty]{}2\frac{dF}{d\bar{\tau}}\left(\bar{\tau}+\frac{2}{3}\bar{z}^{3/2}\right). \tag{7}$$

In Eq. (7), the term $-\frac{1}{4}\bar{z}^{-3/2}[F(\bar{\tau}+\frac{2}{3}\bar{z}^{3/2})+G(\bar{\tau}-\frac{2}{3}\bar{z}^{3/2})]=O(1/z^{3/2})$ coming from the derivative of the slowly varying amplitude is omitted since it is negligibly small compared to other terms $O(1)$ of Eq. (7) emanating from the derivative of the rapidly varying phase.

However, in this new formulation now appears the time derivative of the incoming signal. In the case of an incoming shock wave, this leads to a boundary condition with a sharp singularity (delta Dirac distribution), which is not well suited to a numerical treatment. In order to avoid this difficulty, the problem will now be formulated in terms of potential, instead of pressure.

III. THE NUMERICAL ALGORITHM

A. The equation for the potential

From now on, we drop all subscripts—for the dimensionless variables. The acoustical potential is related to the acoustical pressure by the expression:

$$p=\frac{\partial\phi}{\partial\tau}\Leftrightarrow\phi=\int_{-\infty}^{\tau}p(\tau')d\tau', \tag{8}$$

the (arbitrary) value of the potential at large negative times being chosen equal to 0.

The equation satisfied by the potential field near the caustic now is

$$\frac{\partial^2\phi}{\partial z^2}-z\frac{\partial^2\phi}{\partial\tau^2}+\frac{\mu}{2}\frac{\partial}{\partial\tau}\left[\left(\frac{\partial\phi}{\partial\tau}\right)^2\right]=0. \tag{9}$$

It turns out to be exactly the nonlinear Tricomi Eq. (1), except for the nonlinear term. Equation (9) is called the Tricomi equation for potential.

The associated boundary conditions in time Eqs. (3) and (4) and in the shadow zone Eq. (5) are identical for the potential as for the pressure field. Only the matching boundary condition with geometrical acoustics [Eq. (6) or (7)] is different:

$$z^{1/4}\frac{\partial\phi}{\partial\tau}+z^{-1/4}\frac{\partial\phi}{\partial z}\xrightarrow[\bar{z}\to+\infty]{}2F\left(\tau+\frac{2}{3}z^{3/2}\right). \tag{10}$$

In this ''weak'' formulation, the original incoming signal F now appears instead of its time derivative. Therefore, this formulation is better suited to a numerical resolution in case the incoming signal displays some shocks, as for sonic booms.

B. The iterative scheme

A usual way for solving nonlinear equations involving only boundary conditions (and no initial condition) is to build an iterative scheme, starting from an arbitrary initial condition and converging after several iterations towards the solution of the problem. Following Auger and Coulouvrat (2002), we introduce the so called ''*pseudo time*'' t, through an additional term in the nonlinear Tricomi equation:

$$\frac{\partial^2\phi}{\partial\tau\partial t}=\frac{\partial^2\phi}{\partial z^2}-z\frac{\partial^2\phi}{\partial\tau^2}+\frac{\mu}{2}\frac{\partial}{\partial\tau}\left[\left(\frac{\partial\phi}{\partial\tau}\right)^2\right]. \tag{11}$$

with the unmodified boundary conditions Eqs. (3)–(5) and (10).

The additional term $\partial^2\phi/\partial\tau\partial t$ is supposed to tend toward zero when the iterative solution $\phi(t,\tau,z)$ has reached its ''steady'' limit for large values of the pseudo time t. In this case, $\phi(t\to+\infty,\tau,z)$ satisfies the right-hand side of Eq. (11), i.e., the nonlinear Tricomi equation for potential Eq. (9). The new Eq. (11) is called the ''unsteady'' nonlinear Tricomi equation. It is of hyperbolic type, generally consid-

ered as easier to solve numerically than a mixed (elliptic/hyperbolic) type equation such as the original nonlinear Tricomi equation. The artificial variable t is labeled as a pseudo time for its use to design the iterative process. However, the unsteady nonlinear Tricomi Eq. (11) can be shown to effectively model the diffraction of nonlinear acoustical waves in the shadow zone created by atmospheric refraction (Coulouvrat, 2002). In this case, the (now physical) variable t has the meaning of the penetration distance of the wave inside the shadow zone. Moreover, if the $z\partial^2\phi/\partial\tau^2$ term is omitted in Eq. (11), this one reduces to the well-known KZ equation (Zabolotskaya and Khokhlov, 1969) describing the diffraction of finite amplitude paraxial sound beams. The KZ equation is also a model equation for the diffraction of nonlinear acoustical waves at cusp caustics (Coulouvrat, 2000), the caustics immediately following the fold caustics in the hierarchy of catastrophe theory.

The numerical resolution of the unsteady nonlinear Tricomi Eq. (11) is achieved by means of the fractional steps method (Ames, 1977). Over a single integration step Δt for pseudo time, the equation is split into two simpler equations. The first one takes into account the linear diffraction effects (unsteady linear Tricomi equation):

$$\frac{\partial^2\phi}{\partial\tau\partial t}=\frac{\partial^2\phi}{\partial z^2}-z\frac{\partial^2\phi}{\partial\tau^2}, \tag{12}$$

while the second one takes into account the nonlinear effects (inviscid Burgers' equation for the potential):

$$\frac{\partial\phi}{\partial t}=\frac{\mu}{2}\left(\frac{\partial\phi}{\partial\tau}\right)^2. \tag{13}$$

Starting from the previous iteration i, first Eq. (12) is solved numerically over the pseudo time step Δt, thus providing the intermediate iteration $i+1/2$. This one is used as a starting point for solving Burgers' Eq. (13) over the same pseudo step Δt and finally obtaining iteration $i+1$. This process is repeated until convergence. As a result, both physical mechanisms (linear diffraction and nonlinearities) are taken into account over one pseudo time step. The method of fractional steps is rather common and has been applied successfully for standard algorithms in nonlinear acoustics solving the KZ equation, either in the frequency (Frøysa *et al.*, 1993) or in the time domain (Lee and Hamilton, 1995). Here, it is all the more suited as intermediate steps have no physical meaning, and are just a way to reach convergence toward the solution of the Tricomi's equation. The criterium for convergence simply consists in comparing two successive iterations. If they differ by less than a small parameter ε, then the program is stopped, otherwise the iterative process goes on. Finally, once convergence is reached, the pressure field is derived from the potential by using a standard finite difference scheme (centered derivatives). The choice of the arbitrary parameter ε is discussed in Sec. IV.

C. Resolution of the unsteady linear Tricomi equation

The unsteady linear Tricomi Eq. (12) with boundary conditions Eqs. (3)–(5) and (10) is solved in the frequency domain, using a FFT algorithm. For each frequency, derivatives with respect to $\bar{z}$ are approximated by finite differences precise at second order (centered differences, except for the boundary conditions). Estimation of iteration $i+1/2$ is obtained through a first-order, implicit finite scheme in variable t. Finally, the resulting quasi-tridiagonal linear matrix system is solved using a standard algorithm, and the potential in the time domain is recovered after an inverse Fourier transform. As the procedure is completely identical to the one used by Auger and Coulouvrat (2002), further details are not repeated here. The only difference is that, in the boundary condition Eq. (10), there now appears the incoming time waveform $F(\tau)$ instead of its time derivative in the algorithm of Auger (2001). For incoming shock waves (such as an N wave for sonic boom), the time waveform is much less singular than its derivative. Indeed, as the potential is a continuous function, solving the problem in terms of potential instead of pressure leads to a suppression of Gibbs oscillations that resulted from the Fourier transform of the discontinuous pressure field. While these oscillations could be sharply reduced through the iterative process of Auger and Coulouvrat (2002) by adding some small numerical viscosity in the discretization of Burgers' equation, this reduction was achieved only after a sufficient number of iterations, and convergence required several thousands of iterations. Now, with the new algorithm for the potential, convergence is reached after a few tens of iterations only.

The choice of an algorithm in the frequency domain is motivated by the transsonic aspect of the Tricomi equation. Let us recall that a Tricomi equation is of mixed type, either hyperbolic if $z-\mu p>0$, or elliptic otherwise, the sonic line $z-\mu p=0$ separating the illuminated (or supersonic) side of the caustics where sound "propagates," from the shadow (or subsonic) side where sound is evanescent. In the elliptic domain below the sonic line, there is no oriented flow of information in time τ, contrarily to the hyperbolic domain where information propagates along the sense of ascending times. This difference of behavior was the key point of the original algorithm of Murman and Cole (1971) for transsonic problems, which introduced a switch in the way of discretizing the transsonic equations, using upwind finite differences in the hyperbolic domain but centered finite differences in the elliptic domain. Such a way of discretizing is therefore incompatible with a time domain algorithm for the KZ equation such as the one used by Lee and Hamilton (1995), which relies heavily on the hyperbolic character of the KZ equation and the oriented flow of information along ascending times. The use of an algorithm in the frequency domain implies an artificial periodicity of the problem (the period being the size of the computation domain in time). Imposing this periodicity, it forces some reverse flow of information, hence removing the problem of specifically handling the flow of information in the elliptic domain. This can be seen as an equivalent to the hyperbolic wave equation (with its associated initial conditions in time) being transformed in the frequency domain into the elliptic Helmholtz equation (with no initial conditions). The numerical cost to pay for this is that the method requires a sufficiently large domain in time compared to the effective duration of the signal, so that the artificial periodization of the signals interferes as little as pos-

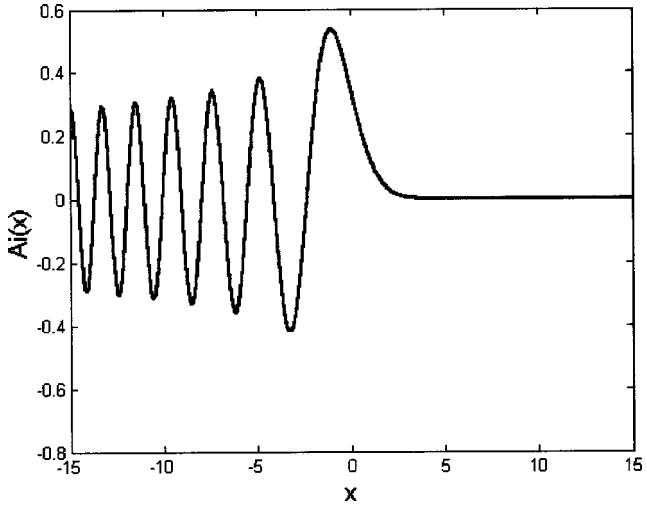


FIG. 2. The Airy function.

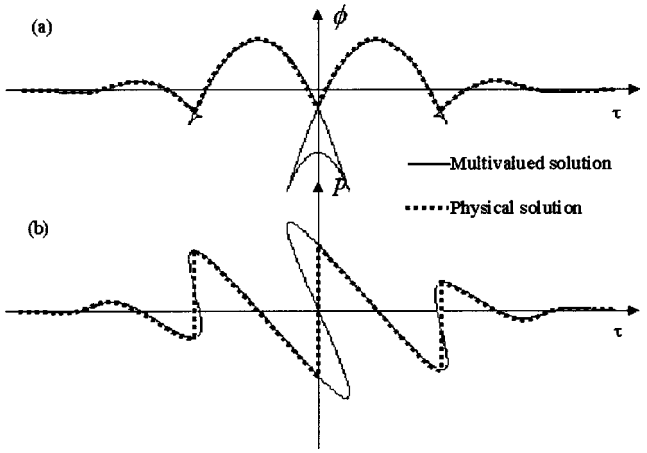


FIG. 3. Multivalued and physical solutions of the inviscid Burgers' equation (a) for potential and (b) for pressure.

sible with the signal. This appears perfectly manageable from a computational point of view by choosing for an incoming "N" wave of duration 1 a time domain equal to $[-8/3;+11/3]$. A second justification for the choice of an algorithm in the frequency domain is that, in the linear case $\mu=0$, the exact solution of the Tricomi equation can be obtained as a superposition of Airy functions (Gill and Seebass, 1973). In catastrophe theory, the Airy function (1838) is the generic solution for the pressure field near fold caustics (Berry, 1976) and reflects the mixed elliptic/hyperbolic type of the Tricomi equation through its oscillating or exponentially decaying character, depending on the sign of its argument (Fig. 2). Moreover, for sonic boom, the nonlinear parameter μ in Tricomi equation is generally small, of the order of 0.1. This means that the main effect of nonlinearities is to limit the amplitude of the "U" wave (unbounded in the linear case), according to the scheme intuited by Guiraud (1965). Therefore the solution of the nonlinear Tricomi equation deviates only weakly from the linear solution, which is "naturally" given in the frequency domain by the Airy function.

D. Analytical solution of the inviscid Burgers equation

In the algorithm of Auger and Coulouvrat (2002), the inviscid Burgers' equation for the pressure field was solved numerically using a *shock capturing* algorithm of McDonald and Ambrosiano (1984). The chosen scheme discretizes the Burgers' equation by using either first-or second-order finite differences. Generally, more precise second-order differences are used, however, near shock waves, first-order differences are preferable as they introduce numerical viscosity that stabilizes the algorithm and enable to go through (to capture) the shock. The "switch" between first- and second-order differences is made through a filter that guarantees monotonicity. As an explicit scheme, it imposes a CFL stability condition that prevents the use of relatively large Δt steps. In the present version of the algorithm, the *shock* capturing scheme has been replaced by an "exact" *shock fitting* algorithm. It is based on the graphical method used by Hayes *et al.* (1969) in the so-called ARAP sonic boom propagation code for solving Burgers' equation along the acoustical rays launched by the aircraft. The algorithm is based on the exact solution of the Burgers' Eq. (13) expressed for the pressure field:

$$\frac{\partial p}{\partial t}=\frac{\mu}{2}\frac{\partial (p)^2}{\partial \tau}. \tag{14}$$

Note that variable t (pseudo time) plays the role of a distance of propagation in the usual formulation of the inviscid Burgers' equation in nonlinear acoustics. For an initial condition $p(t=0,\tau)=p_0(\tau)$, the solution of Eq. (14) is given under the implicit form by the Poisson's solution (Blackstock *et al.*, 1998):

$$\begin{aligned} &p(t,\tau)=p_0(\theta)\\ &\tau=\theta-\mu t p_0(\theta), \end{aligned} \tag{15}$$

where θ is the nonlinear characteristic variable associated to the distorsion of the wave profile. If the distance of propagation t is higher than the shock formation distance t_c, then this solution is multivalued, and hence physically meaningless. Shocks must be introduced, their position being determined according to the weak shock theory [such as in Pestorius algorithm (1973)], or equivalently by applying the law of equal areas (Landau, 1945). However, these methods are rather complex to implement numerically, especially for dealing with the emergence of new shocks, or the merging of several shocks into a single one, phenomena that are likely to occur for the Tricomi equation, especially near the sonic line where the shock pattern may be rather complex. The method of Hayes based on the potential turns out much simpler to code. First, Poisson's solution for the potential Eq. (13) with the initial condition $\phi(t=0,\tau)=\phi_0=\int_{-\infty}^{\tau}p_0(\tau')d\tau'$ is [θ being implicitly defined by relation (15)]

$$\phi(t,\tau)=\phi_0(\theta)-\frac{\mu t}{2}(p_0(\theta))^2. \tag{16}$$

As for the pressure, if the propagation "distance" is larger than the shock formation distance $t>t_c$, this solution is multivalued. However, Hayes *et al.* (1969) noted that, among the multiple possible values of the potential, the physically

meaningful one is the maximum of potential [Fig. 3(a)]. This result is directly related to the second principle of thermodynamics and the fact that shock waves of Burgers' equation can only be compression shock waves. The positions of shock waves are therefore automatically determined as the intersections of the several branches of the multivalued Poisson's solution. In short, the exact solution, including shock waves, is simply

$$\phi(t,\tau)=\max\left\{\phi_0(\theta)-\frac{\mu t}{2}(p_0(\theta))^2\right\}. \tag{17}$$

From a numerical point of view, finding the maximum of several values is much more efficient than solving the differential equation governing the position of the shocks according to weak shock theory, or determining the position that makes the surface of the two lobes equal, as for the law of equal areas. The only numerical cost is due to the discretization. Indeed, the initial condition is given numerically on an equally spaced grid. This means that the exact Poisson's solution is known for equally spaced discrete values of the characteristic variable θ. As a counterpart, the discrete values of the physical variable τ are unequally spaced. To compare the different values of the multivalued solution Eq. (16) for the *same* value of τ, it is therefore necessary to reinterpolate the multivalued solution on a *fixed*, equally spaced grid for the τ variable. This is the only numerical approximation that is introduced into the scheme, otherwise the method is analytically exact. In particular, it can be applied for whatever value of the pseudo time step Δt. There is no CFL condition.

Note that this method allows us to find the physical pressure too, since the physical points for the Poisson's solution in pressure are the points corresponding to the physical potentiel [Fig. 3(b)]. More details on the overall algorithm illustrated by a diagram can be found in Auger and Coulouvrat (2002). In the present up-dated version, the only changes in that diagram consist in replacing the pressure by the potential, and the MacDonald and Ambrosiano algorithm by the analytical solution Eq. (17).

IV. VALIDATION OF THE ALGORITHM

A. Convergence down to machine precision

Convergence tests have shown that the convergence parameter ε can be chosen arbitrarily small, down to machine precision ($\varepsilon \approx 10^{-15}$) (Fig. 4). This is a crucial test to check the effective convergence of the whole numerical process (iterative unsteady Tricomi equation+split step +discretisation) towards the solution of the nonlinear Tricomi equation. However, in practice, such an extreme precision is not necessary, and we routinely choose $\varepsilon \approx 10^{-7}$. Using this parameter the computation time is dramatically reduced with a factor of 40 in comparison with the existing code of Auger and Coulouvrat (2002). This factor grows exponentially with the criterion of convergence. This is an important difference between the two methods, that proves the potential formulation associated with an explicit treatment of nonlinearity is more pertinent than the associated pressure formulation.

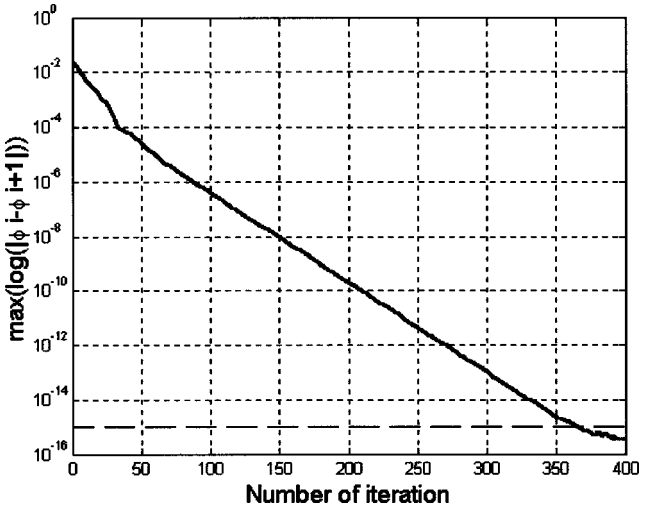


FIG. 4. Logarithm of the maximal value of the difference between two successive iterations i and $i+1$ versus the number of iterations, for an incoming N wave with $\mu=0.05$.

B. Convergence with mesh grid refinement

When nonlinearities are neglected ($\mu=0$), the solution of the linear Tricomi's equation for pressure for an incoming shock wave is singular (logarithmic singularity on the illuminated side, power $-\frac{1}{6}$ on the geometrical caustic). This singularity is impossible to reproduce numerically because of the filtering induced by discretization, but we observe a regular increase of the maximal pressure amplitude with the number of discretization points. As the exact mathematical solution is singular, the numerical one does not converge with grid refinement (Fig. 5). This behavior is of course physically meaningless and, according to Guiraud (1965), nonlinearities must be introduced to remove the singularity when the incoming wave has a shock. This expected behavior is observed numerically in Fig. 5, where we clearly get numerical convergence for a sufficiently fine grid (1024 discretization points in time or more). Also, the nonlinear solution strongly deviates from the linear one. This result is an indirect validation of the algorithm, showing that it follows Guiraud's assumption, and better converges with time discretization than the previous algorithm of Auger and Coulouvrat (2002). Let us finally notice that, contrarily to the time discretization, the number of points in distance z does not influence very much the solution.

C. Physical behavior

For incoming ''N'' waves, the numerical solution depends only on the physical nonlinearity parameter μ. Figures 6 and 8 show the spatio-temporal field of pressure $p(z,\tau)$ in a linear grayscale calculated along a line perpendicular to the caustic versus dimensionless time for $\mu=0.05$ and $\mu=0.5$. The white line is the sonic line separating the elliptic (shadow) zone from the hyperbolic (illuminated) one. The deformation of the sonic line is very sensitive to the nonlinearity parameter. In the hyperbolic zone, cusped wavefronts are obvious, corresponding to each incoming and outgoing shock waves. For both values of μ, the incoming and outgo-

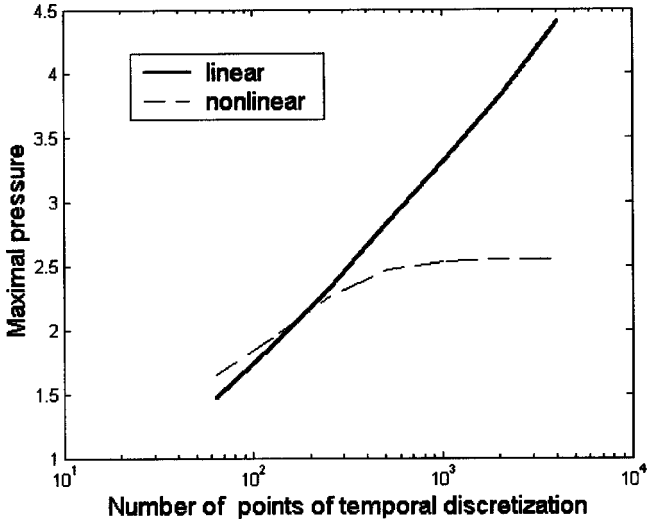


FIG. 5. Maximal value of the dimensionless pressure versus the number of points of the temporal discretization (1500 points in the spatial grid) in the linear case ($\mu=0$) and in the nonlinear case ($\mu=0.1$), for an incoming N wave.

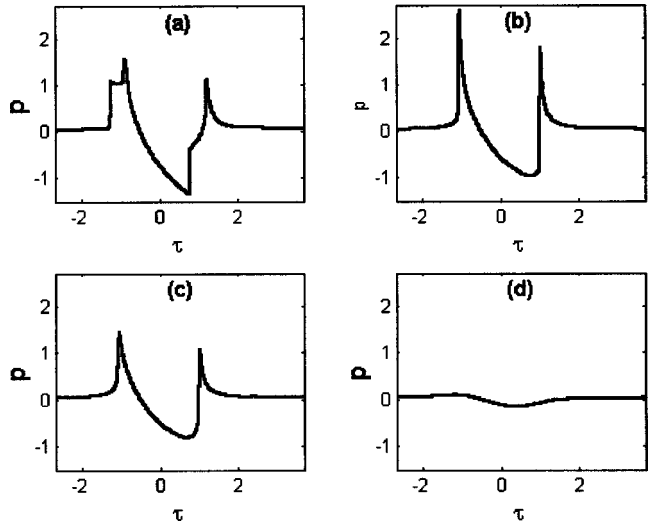


FIG. 7. Pressure versus dimensionless time for $\mu=0.05$, for four distances from the caustic: (a) $z=1$, (b) $z=z_{max}$ (distance where the maximal pressure is found), (c) $z=0$, and (d) $z=-1$.

ing shocks merge exactly on the sonic line. No triple shock is observed. In the shadow zone, no shock propagates in accordance with the elliptic nature of the equation there. For both values of μ, the signal is also plotted at several distances from the caustic (Figs. 7 and 9). At matching distance with geometrical acoustics $z=1$ [Figs. 7(a) and 9(a)] the incoming and the outgoing signals are similar to test flights measurements (Wanner *et al.*, 1972): the ''N'' wave is transformed into a ''U'' wave. The higher the nonlinearity parameter, the more the solution is distorted is. Figures 7(b) and 9(b) show the pressure versus time at the distance where pressure is maximal. The amplification factor (maximal amplitude divided by the amplitude of the ''N'' wave at the matching boundary) decreases as the nonlinearity parameter increases: it is about 2.5 for $\mu=0.05$, but only 1.5 for $\mu=0.5$. For the value $\mu=0.05$ typical for sonic boom, the amplification factor is in agreement with flight tests values. On the geometrical caustic [$z=0$, Figs. 7(c) and 9(c)], the amplitude has strongly decreased because the first peak is now in the shadow zone. This is completely different from the linear behavior, where the highest singularity is right on the geometrical caustic. In the nonlinear case, the point of maximum amplitude is generally in the elliptic domain, very close to, but not exactly on, the sonic line. Note that for strong nonlinearities [Fig. 9(c)], the signal deviates from the usual ''U'' shape, as the first peak is strongly smoothed because in the shadow zone, while the second one remains sharp because still in the hyperbolic domain. Finally, away in the elliptic zone [Figs. 7(d) and 9(d), $z=-1$], the signal is smooth, and its amplitude is very weak as expected, due to the exponential decay there. These results, showing the numerical simulations behave qualitatively as expected from the physics and in agreement with flight tests, provide an additional validation.

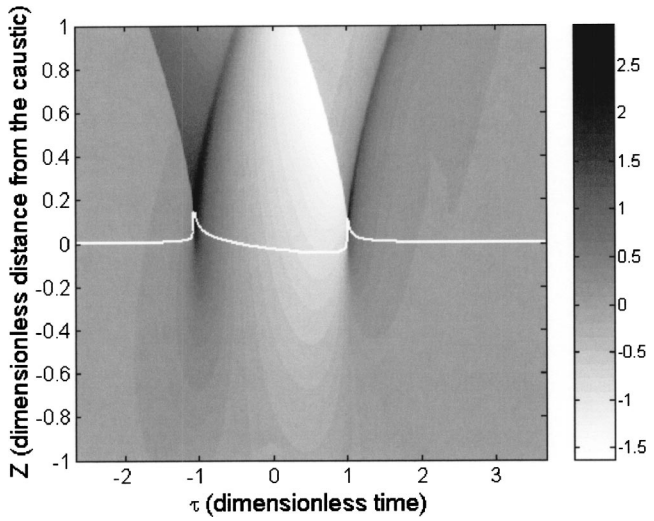


FIG. 6. Pressure field for $\mu=0.05$. The white line is the sonic line. The hyperbolic zone is above that line, the elliptic one is below. The pressure levels are indicated by the gray bar.

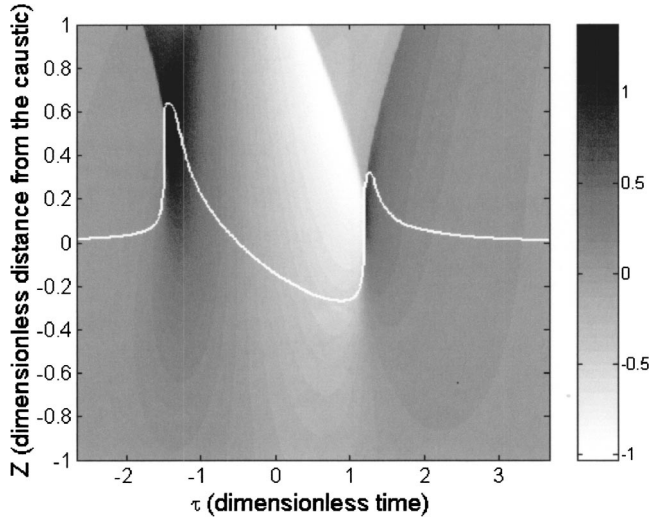


FIG. 8. Pressure field for $\mu=0.5$. The white line is the sonic line. The hyperbolic zone is above that line, the elliptic one is below. The pressure levels are indicated by the gray bar.

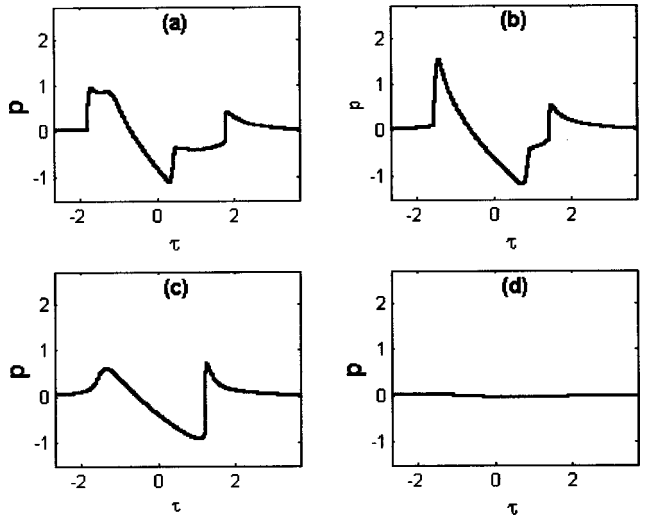


FIG. 9. Pressure versus dimensionless time for $\mu=0.5$, for four distances from the caustic: (a) $z=1$, (b) $z=z_{\max}$ (distance where the maximal pressure is found), (c) $z=0$, and (d) $z=-1$.

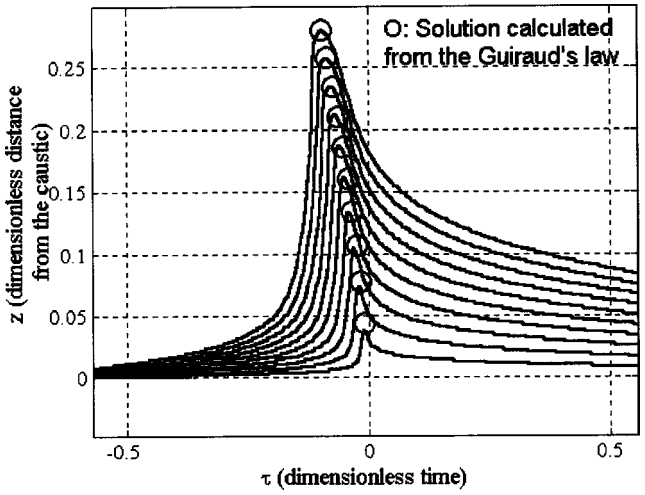


FIG. 10. Validation of Guiraud's similitude for an incoming rectangular window signal. Computed deformation of the sonic line for different values of μ between 0.01 and 0.2 (continuous lines) and comparison with Guiraud's similitude for the point of maximum deformation (circles).

D. Guiraud's similitude

It has been established previously that the numerical solution converges toward a finite value. To ensure this value is correct, and to provide a quantitative validation of the code, we check the similitude law derived by Guiraud (1965). Indeed, with the new scaling: $\phi=\mu\tilde{\phi}$, $z=\mu^{4/5}\tilde{z}$, and $\tau=\mu^{6/5}\tilde{\tau}$, the nonlinear Tricomi equation can be written

$$\frac{\partial^2\tilde{\phi}}{\partial\tilde{z}^2}-\tilde{z}\frac{\partial^2\tilde{\phi}}{\partial\tilde{\tau}^2}+\frac{1}{2}\frac{\partial}{\partial\tilde{\tau}}\left[\frac{\partial\tilde{\phi}}{\partial\tau}\right]^2=0, \tag{18}$$

and its boundary condition

$$\tilde{z}^{1/4}\frac{\partial\tilde{\phi}}{\partial\tilde{\tau}}+\tilde{z}^{-1/4}\frac{\partial\tilde{\phi}}{\partial\tilde{z}}=2F\left(\mu^{6/5}\left[\tilde{\tau}+\frac{2}{3}\tilde{z}^{3/2}\right]\right). \tag{19}$$

Note that the nonlinearity parameter is eliminated from the nonlinear Tricomi equation, but now appears in the phase of the boundary condition. However, for a ''step'' wave [$F(\tau)=0$ for $\tau<0$ and $F(\tau)=1$ for $\tau>0$], the incoming wave profile is invariant by a phase dilatation, so the whole problem gets independant of the value of μ. The objective of the validation test is to check whether the code satisfies this property. However, because of the discretization, a true step shock of infinite duration cannot be simulated. Instead, we use a rectangular window [$F(\tau)=1$ if $0<\tau<2$ and $F(\tau)=0$ everywhere else]. This function is completely identical to the step function locally around the shock, so that we can expect it to satisfy at least approximately Guiraud's scaling law. This kind of approximation is used in practice by Gill and Seebass (1973) to estimate solutions of the nonlinear Tricomi equation when the hodograph transform fails to predict the shock positions. For the pressure field, the correct similitude is $p=\mu^{-1/5}\tilde{p}$. Guiraud's similitude is tested in Fig. 10 for the amplitude and position of the maximal distorsion of the sonic line, corresponding to the cusp of the shock wavefront on the sonic line. This point is chosen as it is exactly on the shock wavefront, so that it is as little as possible influenced by the finite duration of the signal. Moreover, this point is easily spotted, and especially difficult to capture numerically, as convergence there requires a finer discretization than for other points. Therefore, it is a demanding, quantitative test. At this point, according to the scaling law,

$$p_{\max}=\mu^{-1/5}C_p, \tag{20}$$

$$z(p_{\max})=\mu^{4/5}C_z, \tag{21}$$

$$\tau(p_{\max})=\mu^{6/5}C_\tau, \tag{22}$$

where the quantities C_p, C_z, C_τ should be constant if the problem is invariant. Results are plotted in Fig. 10, where the *computed* sonic lines at different values of μ between 0.01 and 0.2 are plotted as continuous lines, while the positions of the cusp (circle) for the same values of μ are *deduced* numerically by Guiraud's scaling law Eqs. (21) and (22) from the computed position at the lowest value of μ. Figure 10 obviously demonstrates this point (circle) constantly remains on the top of the (continuous) sonic line, accordingly to the theoretical similitude. The results are also in very good agreement for the pressure amplitude, as exemplified by Fig. 11: for μ values larger than 0.02, the ratio $\mu^{1/5}p_{\max}$ deviates from less than 2% of the constant value 1.76. The larger deviation for smaller values of μ is related to unsufficient time discretization in cases where the very sharp peaks of the wave profile are difficult to capture numerically. Hence, the numerical scheme satisfies the nonlinear quantitative similitude law. The whole sonic line, and not only the positions of the wavefront cusp, could be theoretically deduced from one another through Guiraud's law. However, if applied for the *finite* signal used here, the similitude would rapidly diverge from numerical simulations. For instance, according to Guiraud's law, the ''width'' of the sonic line around his maximum deformation should be 36 times larger for the largest value of μ than for the smallest one. Clearly this is not the case, because the numerical results are constrained by the *finite* duration of the signal, which prevents too huge distorsions, especially for the phase variable. This explains why,

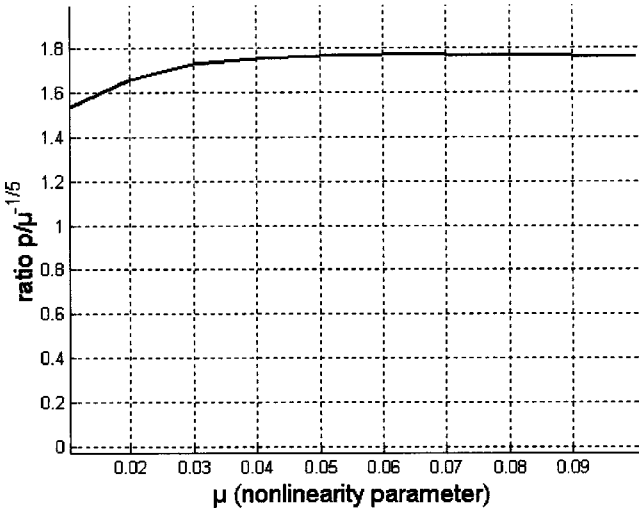


FIG. 11. Validation of Guiraud's similitude: ratio $p_{\max}/\mu^{-1/5}$ as a function of μ for an incoming rectangular window signal and for the wavefront cusp.

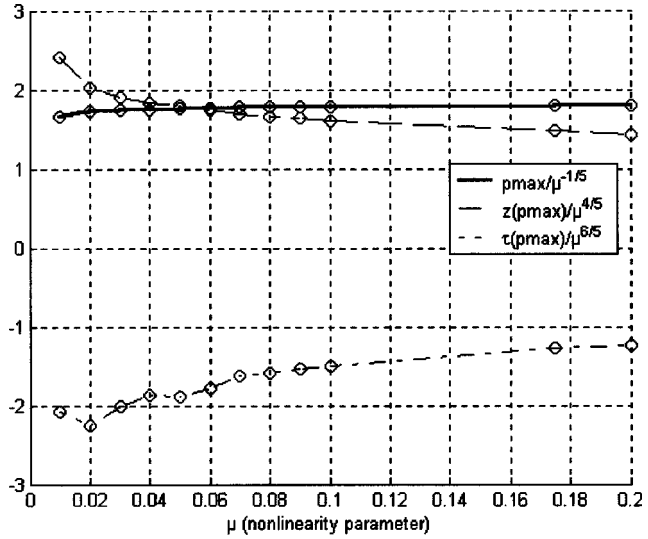


FIG. 12. Validation of Guiraud's similitude: ratios $p_{\max}/\mu^{-1/5}$, $z(p_{\max})/\mu^{4/5}$, and $\tau(p_{\max})/\mu^{6/5}$ as a function of μ for an incoming rectangular window signal and for the point of maximum amplitude.

when applied for the point of maximum amplitude which is slightly *away* from the wavefront cusp inside the elliptic domain, the similitude law remains well satisfied for the pressure amplitude, but not so well for its position (Fig. 12). Indeed, it still satisfies a similitude law, but with a power 0.63 instead of 0.8 for the distance z, and a power 0.99 instead of 1.2 for the phase τ. Similar results can be obtained when the incoming signal is an ''N'' wave. Consequently, when applying Guiraud's similitude, Gill and Seebass (1973) and Plotkin (2002) predict with a rather good accuracy the focused boom peak amplitude, but not so well the precise shape of the superboom.

V. SONIC BOOM SIMULATION OF AN ACCELERATING SUPERSONIC TRANSPORTER

A numerical simulation of focused sonic boom at the ground level has been realized in a ''realistic'' case. The chosen mock-up is the so-called Eurosup configuration (Evans and Doherty, 1997) (Fig. 13), derived from the European ESCT configuration for a future supersonic transport aircraft. It is a wing–body configuration without horizontal canard or aft-tail and without vertical fin. The double delta wing is set in a low position relative to the fuselage which has a circular section and is 89 m long. The wing twist, camber, and thickness have been numerically optimized, and a model has been built for wind-tunnel tests at supersonic and transonic speeds.

Sonic boom focusing has been numerically simulated under the assumption of a horizontal flight at altitude 35 000 ft (10 668 m), with a constant acceleration of 1 m/s^2, and in the standard atmosphere. The altitude is typical for Concorde during its transonic acceleration phase along the route Paris to New York. The acceleration is relatively large, but leads to a ground track focusing at the ground level occurring almost exactly at Mach 1.2. (1.207). Due to atmospheric conditions, the aircraft Mach number for ground track focusing is variable, and Mach 1.2 is considered as a representative ''mean'' value. Anyway, previous simulations have shown (Auger and Coulouvrat, 2002) that the acceleration has little influence on the amplitude of focused sonic boom (but it has on the caustics position). For focusing, only boom simulations in the plane of symmetry (boom emitted with a zero azimutal angle) have been realized.

The computations have been performed by coupling the numerical solver of the Tricomi equation to the numerical code **BANGV** currently being developed by Airbus France SAS and Université Pierre et Marie Curie. It is based on a standard ray model, modified to take into account the non-

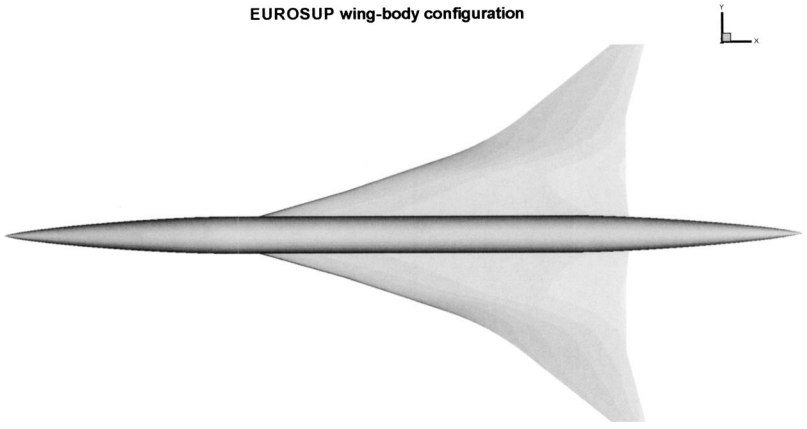


FIG. 13. The Eurosup mock-up.

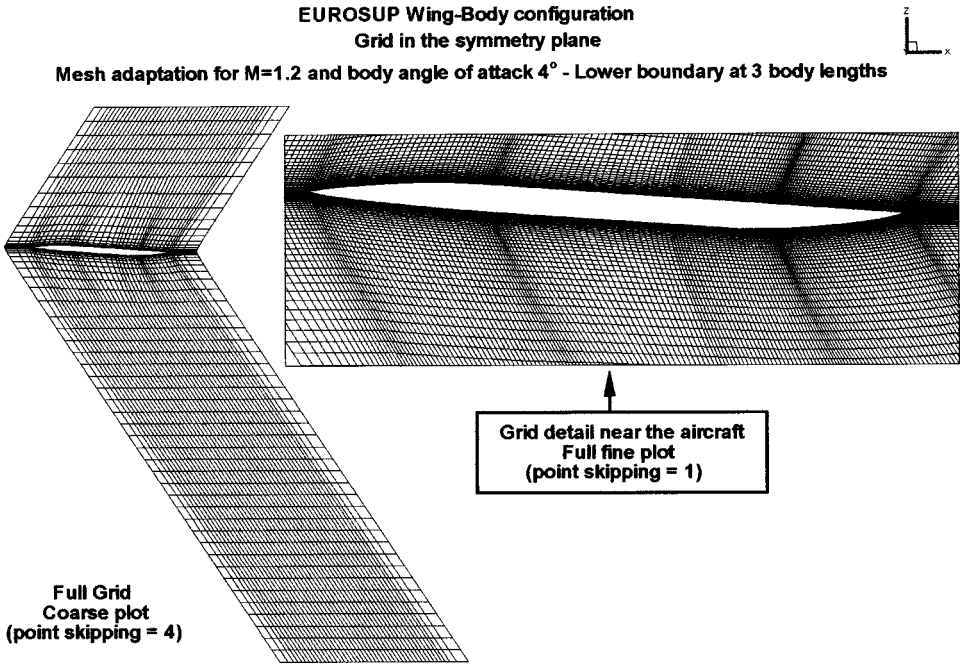


FIG. 14. Mesh grid in the mock-up symmetry plane.

linear distortion of the sonic boom waveform as it propagates from the aircraft down to the ground. Boom is computed along the ray that is tangent to the caustic at the ground until this ray reaches a point slightly above the ground. This provides the incoming waveform $F(\tau)$ and the nonlinear parameter μ that are the only input parameter for the algorithm solving the nonlinear Tricomi equation.

The source term, used as an input pressure field for the sonic boom code, has been determined using CFD simulations of the Euler equations around the mock-up. The finite volume software **elsA** (Cambier, 1999) has been used, with uncentered upwind fluxes of the Roe type. The meshing of the air volume around the mock-up was determined using an analytical method to build a structured grid with an H-type topology. The mock-up is first set at the desired angle of attack before meshing the volume around the configuration so that constant K index surfaces remain always planes parallel to the incoming flow velocity at some distance: this is intended to avoid interpolation errors when extracting the pressure fields in the planes required for the sonic boom code input. The H-type topology of the meshing is especially suited for capturing fine transverse gradients of the pressure field. The meshing was preadapted to the Mach number, by limiting the grid between the two Mach cones emanating from the fuselage at its nose and at its tail (slightly extended forward/backward to be sure to capture all the shock waves emanating from the mock-up, Fig. 14). The computational domain was extended below to 3 fuselage lengths, but to only 1 fuselage length above. For a better capture of the origin of the shock waves, grid refinement was introduced around the Mach cones emanating from the nose, the wing leading edge, the wing trailing edge and the fuselage tail. Far from the mock-up, the grid refinement has to become more regular because the exact position of the shock waves cannot be guessed in advance. This results in the type of meshing shown on Fig. 14 in the plane of symmetry. The 3D meshing was of the order of 5 million computation points.

The matching between the near-field Euler simulations and the acoustical propagation model was not sophisticated, using directly the CFD pressure waveform extracted at some distance from the aircraft as an input for the ray propagation. No ''smooth'' matching as those described by Plotkin and Page (2002) was performed.

Pressure waveforms below the mock-up in the symmetry plane are shown in Fig. 15 for five different distances H (0.5, 1, 1.5, 2, 2.5 fuselage lengths L). For comparison, the pressure fields are normalized by the ratio $\sqrt{H/L}$ to compensate for the geometrical attenuation due to the conical form of the wavefront. If the pressure field would perfectly match the classic Whitham's assumptions for sonic boom theory (slender body, far-field approximation, linear and locally axisymmetric field), these different curves should perfectly superimpose. On the figure we can clearly see three shock waves emanating from the aircraft nose, wing leading edge, and fuselage tail. At $H=0.5L$, we observe a fourth, sharp shock wave emanating from the wing trailing edge, but this one rapidly decays at larger distances. Therefore, it appears as a near-field, local effect, and the distance $H=0.5L$ seems too short for a sound matching with acoustical propagation. The end of the pressure waveform at $H=2.5L$ strongly deviates from the undisturbed value 0, probably due to a mesh boundary too close to the fuselage tail. More generally, the pressure field at the rear appears pretty complex, as confirmed by off-symmetry simulations, and the extent of the meshing is probably too close to the fuselage for capturing precisely the decaying tail of the pressure waveform. On the contrary, the two first shocks seem to superimpose almost perfectly. Small deviations (advancement of the nose shock and attenuation of the small dip ahead of the leading edge shock) can be explained by nonlinear propagation effects that are anyway taken into account in the propagation code. Especially, we do not see any strong deviation from the assumption of locally axisymmetric field, even for the leading edge shock though

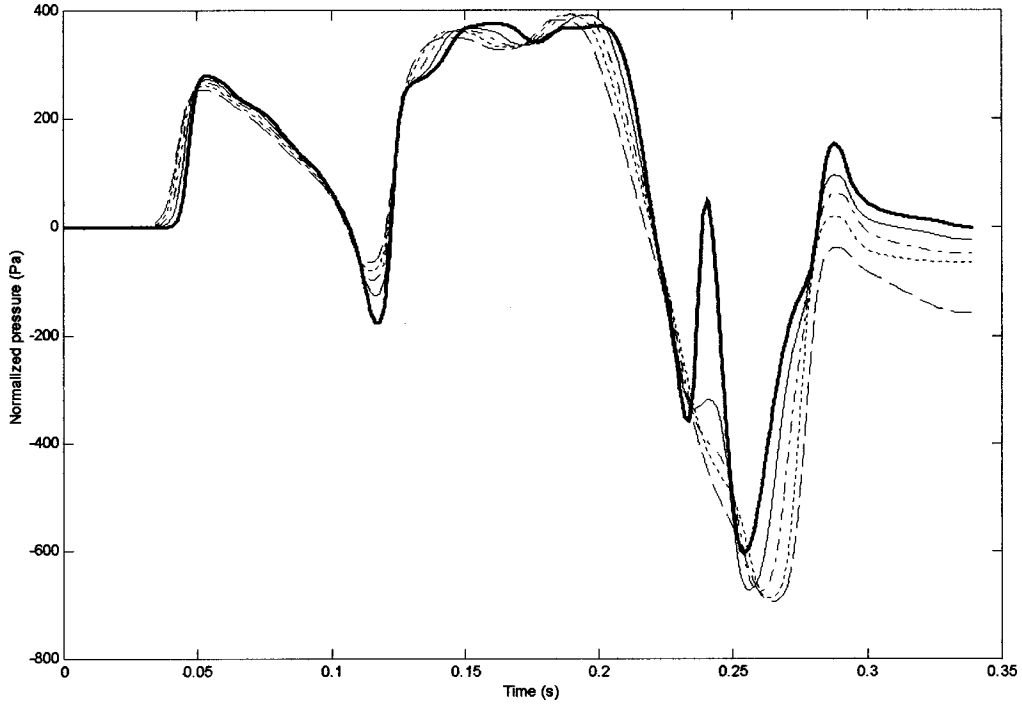


FIG. 15. Normalized pressure waveforms at five different distances below the mock-up (thick line: H/L=0.5, continuous line: H/L=1, dash dotted line: H/L=1.5, dotted line: H/L=2, and dashed line: H/L=2.5).

this one is mostly associated to lift effects that are far from axisymmetric. This point is also confirmed by off-symmetry pressure fields.

Figure 16 shows the ground sonic boom predicted in cruise condition at Mach 1.2 (no acceleration, therefore no focusing), depending on the ratio H/L at which the matching between CFD and the propagation code is realized. For the two head shocks (nose and leading edge), the agreement is almost perfect, which demonstrates the soundness of the whole numerical procedure. The results are not so good for the rear shock, especially for $H/L=0.5$ and 2.5, obviously a consequence of the nearfield simulations (local fourth shock wave at $H/L=0.5$, too short mesh at $H/L=2.5$). The agreement is acceptable for the three other intermediate values, despite some uncertainty in the rear shock position and amplitude. Also noteworthy is the triple shocks structure of the

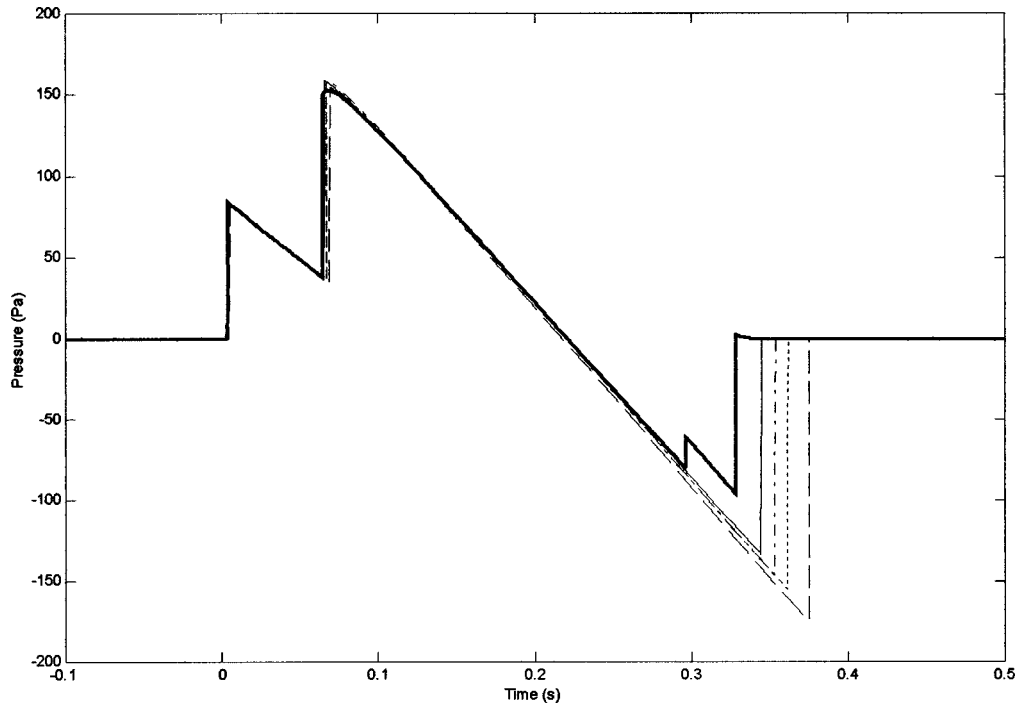


FIG. 16. Ground track sonic boom at cruise Mach number 1.2 computed by matching with CFD Euler simulations at five different distances below the aircraft (thick line: H/L=0.5, continuous line: H/L=1, dash dotted line: H/L=1.5, dotted line: H/L=2, and dashed line: H/L=2.5).

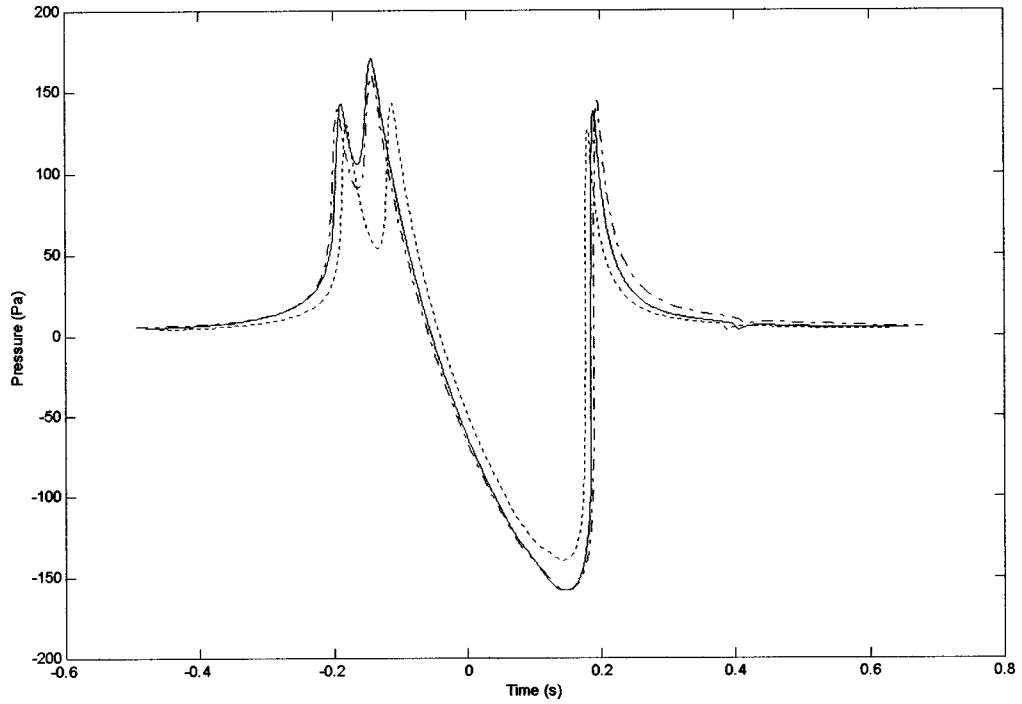


FIG. 17. Groundtrack focused pressure waveforms at the geometrical caustic ($z=0$) for three different matching distances (continuous line: H/L=1, dash dotted line: H/L=1.5, and dotted line: H/L=2).

boom waveform. This is due to the low Mach 1.2 inducing a nose shock rather strong compared to the leading edge shock (roughly 75%, Fig. 15). At Mach 1.2, fly altitude is lower than for a Mach 2 cruise, boom propagation is shorter, and nonlinear effects are not sufficient for the shocks to merge into the classic ''N'' wave boom (which would be observed at Mach 2). For computation of boom focusing, the Tricomi solver perfectly handles such ''complex'' incoming waveforms, contrarily to the aproximate method of Gill and Seebass (1973) valid for noninteracting shock waves, an assumption not satisfied here for the two close head shocks.

Figure 17 shows the computed focused boom right at the geometrical caustics ($z=0$ in the Tricomi equation), only for the three ''admissible'' distances $H/L=1$, 1.5 or 2. Due to the three-shock structure of the incoming waveform, the focused boom exhibits a more complex shape than the usual

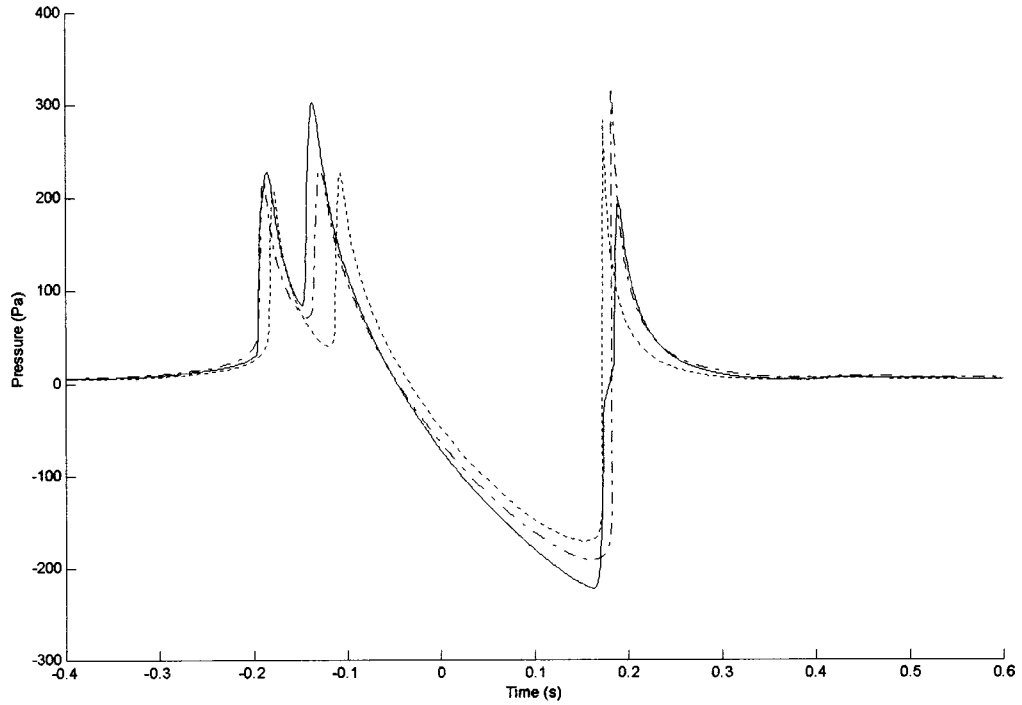


FIG. 18. Groundtrack focused pressure waveforms at the position of maximum amplitude for three different matching distances (continuous line: H/L=1, dash dotted line: H/L=1.5, and dotted line: H/L=2).

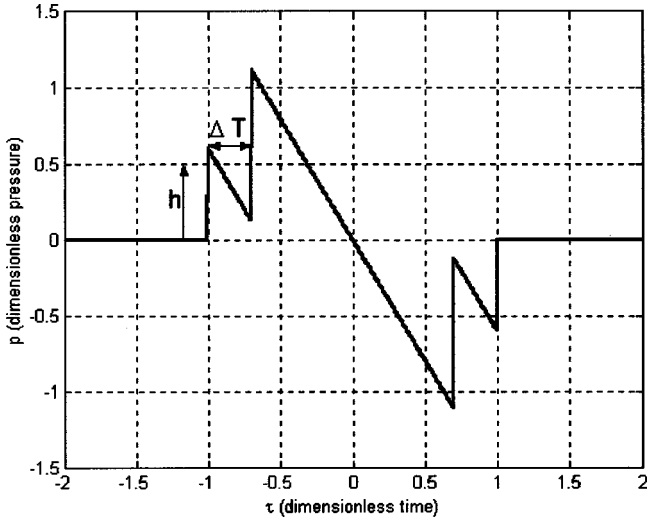


FIG. 19. ''Optimized'' signal where the nose and leading edge shocks have not merged yet.

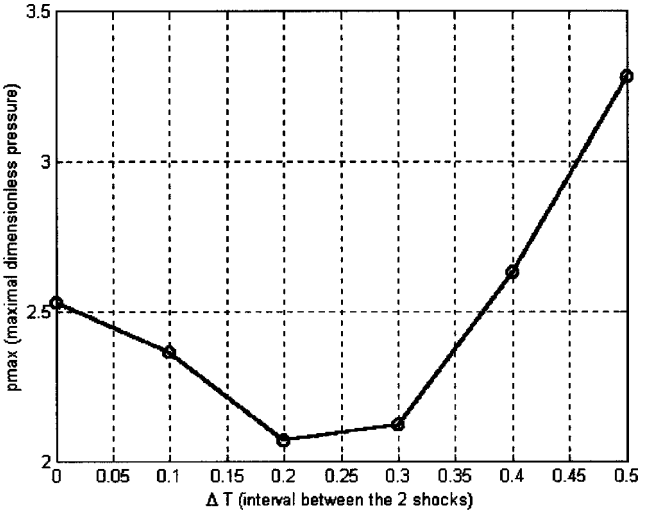


FIG. 20. Maximal overpressure as a function of the time interval between the nose and the leading edge shocks.

''U'' wave, each incoming shock wave giving rise to a peak. The agreement between the three curves turns out to be very satisfying, especially for the first and third ''peaks.'' There is some deviation, however, for the phase of the second shock in the case $H/L=2$, but it has only a marginal consequence on the peak amplitude.

Figure 18 displays the same curves, but now at the distance z_{max} from the geometrical caustics where the pressure field reaches its peak value. As z_{max} depends on the input parameter, the three curves of Fig. 18 are not calculated exactly at the same distance z_{max}, a fact that amplifies the differences between the curves compared to Fig. 17. This is especially obvious, for the case $H/L=1$, where the highest peak is associated to the second incoming shock (leading edge), while for the two other cases, it is associated to the third (tail) shock. Let us recall that for an N wave, it is always associated to the first shock! Comparing the two cases $H/L=1.5$ or 2, however, the agreement is very good, the only main difference being once again the phase of the second shock for $H/L=2$. Also to be noticed is the peak amplitude ($\sim$300 Pa for all three cases), significantly different from the one at the geometrical caustics ($\sim$150 Pa). The position z_{max} is typically of the order of 0.25 times the boundary layer thickness, which corresponds roughly to about 100–200 m. Taking into account the inclination of the caustic relative to the ground plane, this means that the point of maximum pressure may deviate from several hundred meters from the geometrical caustics, a fact that is likely to make test flights measurements of focused boom especially difficult to carry out. Precise numerical simulations such as the present one are all the more attractive.

VI. CONCLUSION: TOWARDS A MITIGATION OF SUPERBOOM?

Finally, a numerical study on the reduction of the focused sonic boom is presented. The solution of the nonlinear Tricomi equation is presented for several incoming waveforms. We choose to compare a signal where the shocks due to the nose and the leading edge have not merged yet, as for the Eurosup configuration. We also chose to keep the pressure waveform symetric. The amplitude h of the first shock is chosen equal to 0.6 and the signal energy is kept constant, so that the whole signal is entirely determined by a single parameter, the time interval Δt between the two first shocks (Fig. 19). This interval varies between 0 (the classic N wave) and 0.5 (25% of the total duration). Figure 20 shows the highest amplitude of the focused boom as a function of the interval between the two first shocks. Compared to an ''N'' wave, we first observe an amplitude decrease, the minimal value being reached for an interval between the two first shocks equal to 0.25 (12.5% of the total duration). The reduction of the amplitude is then about 20% compared to the N wave. For larger time intervals, the amplitude increases anew. This proves that potential ways for the reduction of a focused boom do exist. They require to produce incoming waves with multiple shock, both at the front *and* at the rear of the signal. The example of the Eurosup configuration shows that this objective should not be ruled out for a realistic configuration. Nevertheless, superboom reduction sonic remains a formidable challenge, especially for the rear part of the aircraft, which is especially complex from an aerodynamical point of view.

ACKNOWLEDGMENTS

Parts II, III, and IV of this investigation have been carried out under a contract awarded by the European Commission, Contract No. G4RD-CT-2000-00398. No part of this report may be used, reproduced and/or disclosed, in any form of by any means without the prior written permission of Université Pierre et Marie Curie and the SOBER project partners. 16/11/2001 All rights reserved.

Part V of this investigation was supported by Ministère de la Recherche (France), decisions No. 00 T0113 and 00 T0114.

Airy, G. B. (**1838**). ''On the intensity of light in the neighbourhood of a caustic,'' Trans. Cambridge Philos. Soc. **6**, 379–401.

Ames, W. F. (**1977**). *Numerical Methods for Partial Differential Equations* (Academic, New York), pp. 315–467.

Auger, T. (**2001**). ''Modélisation et simulation numérique de la focalisation d'ondes de choc acoustiques en milieu en mouvement. Application à la focalisation du bang sonique en accélération,'' thèse de l'Université Pierre et Marie Curie (Paris 6) (in French).
Auger, T., and Coulouvrat, F. (**2002**). ''Numerical simulation of sonic boom focusing,'' AIAA J. **40**, 1726–1734.
Berry, M. V. (**1976**). ''Waves and Thom's theorem,'' Adv. Phys. **25**, 1–26.
Blackstock, D. T., Hamilton, M. F., and Pierce, A. D. (**1998**). ''Progressive waves in lossless and lossy fluids,'' in *Nonlinear Acoustics* (Academic, San Diego), pp. 66–150.
Buchal, R. N., and Keller, J. B. (**1960**). ''Boundary layer problems in diffraction theory,'' Commun. Pure Appl. Math. **13**, 85–114.
Cambier, L. (**1999**). ''The elsA project,'' 1st ONERA/DLR Aerospace Symposium, Paris, 21–24 June 1999.
Cheng, H. K., and Hafez, M. M. (**2002**). ''The Superboom as a Tricomi problem: application,'' IUTAM Symposium Transsonicum IV, Göttingen (Germany), 2–6 September 2002 (proceedings to be published).
Coulouvrat, F. (**2000**). ''Focusing of weak acoustic shock waves at a caustic cusp,'' Wave Motion **32**, 233–245.
Coulouvrat, F. (**2002**). ''Sonic boom in the shadow zone: a geometrical theory of diffraction,'' J. Acoust. Soc. Am. **111**, 499–508.
Evans, T. P., and Doherty, J. J. (**1997**). ''The aerodynamic design of EUROSUP configuration,'' DERA/AS/ASD/CR97620/1.0.
Frøysa, K.-E., Tjøtta, J. N., and Berntsen, J. (**1993**). ''Finite amplitude effects in sound beams. Pure tone and pulsed excitation,'' in *Advances in Nonlinear Acoustics*, edited by H. Hobæk, 13th Int. Sym. Nonlinear Acoustics, Bergen, Norway, 28 June–2 July (World Scientific Singapore), pp. 233–238.
Gill, P. M., and Seebass, A. R. (**1973**). ''Nonlinear acoustic behavior at a caustic: an approximate analytical solution,'' AIAA Aeroacoustics Conference, Seattle (MIT, Cambridge), AIAA Paper 73-1037, pp. 353–386.
Guiraud, J.-P. (**1965**). ''Acoustique géométrique, bruit balistique des avions supersoniques et focalisation,'' J. Mec. **4**, 215–267 (in French).
Hayes, W. D. (**1968**). ''Similarity rules for nonlinear acoustic propagation through a caustic,'' Second Conference on Sonic Boom Research, NASA SP-180, pp. 165–171.
Hayes, W. D., Haefeli, R. C., and Kulsrud, H. E. (**1969**). ''Sonic boom propagation in a stratified atmosphere with computer program,'' NASA CR-1299.
Kravtsov, Y. A., and Orlov, Y. I. (**1993**). *Caustics, Catastrophes and Wave Fields* (Springer-Verlag, Berlin), pp. 8–33.
Landau, L. (**1945**). ''On shock waves at large distances from the place of their origin,'' J. Phys. U.S.S.R. **9**, 496–500.
Lee, Y. S., and Hamilton, M. F. (**1995**). ''Time-domain modeling of pulsed finite-amplitude sound beams,'' J. Acoust. Soc. Am. **97**, 906–917.
Marchiano, R., Thomas, J.-L., and Coulouvrat F. (**2003**). ''Experimental simulation of supersonic superboom in a water tank: nonlinear focusing of weak shock waves at a fold caustic,'' Phys. Rev. Lett. (to be published).
McDonald, B. E., and Ambrosiano, J. (**1984**). ''High order upwind flux methods for scalar hyperbolic conservation laws,'' J. Comput. Phys. **56**, 448–460.
Murman, E. M., and Cole, J. D. (**1971**). ''Calculation of plane steady transonic flows,'' AIAA J. **9**, 114–121.
Pestorius, F. M. (**1973**). ''Propagation of plane acoustic noise of finite amplitude,'' Technical Report ARL-TR-73-23 (Applied Research Laboratories, The University of Texas at Austin), AD778868.
Plotkin, K. J. (**2002**). ''State of the art of sonic boom modelling,'' J. Acoust. Soc. Am. **111**, 530–536.
Plotkin, K. J., and Page, J. A. (**2002**). ''Extrapolation of Sonic Boom Signatures from CFD Solutions,'' AIAA Pap. **2002-0922**, 1–6.
Seebass, A. R. (**1971**). ''Nonlinear acoustic behavior at a caustic,'' Third Conference on Sonic Boom Research, NASA SP-255, pp. 87–120.
Sturtevant, B., and Kulkarny, V. A. (**1976**). ''The focusing of weak shock waves,'' J. Fluid Mech. **73**, 651–671.
Thom, R. (**1972**). *Stabilité structurelle et morphogenèse* (Benjamin, Reading), pp. 72–107 (in French).
Wanner, J.-C., Vallée, J., Vivier, C., and Théry, C. (**1972**). ''Theoretical and experimental studies of the focus of sonic booms,'' J. Acoust. Soc. Am. **52**, 13–32.
Whitham, G. B. (**1952**). ''The flow pattern of a supersonic projectile,'' Commun. Pure Appl. Math. **5**, 301–348.
Zabolotskaya, E. A., and Khokhlov, R. V. (**1969**). ''Quasi-plane waves in the non-linear acoustics of confined beams,'' Sov. Phys. Acoust. **15**, 35–40.

Frequency response of nonlinear oscillations of air column in a tube with an array of Helmholtz resonators

N. Sugimoto, M. Masuda, and T. Hashiguchi
Department of Mechanical Science, Graduate School of Engineering Science, University of Osaka, Toyonaka, Osaka 560-8531, Japan

(Received 10 January 2003; revised 20 June 2003; accepted 23 June 2003)

Nonlinear cubic theory is developed to obtain a frequency response of shock-free, forced oscillations of an air column in a closed tube with an array of Helmholtz resonators connected axially. The column is assumed to be driven by a plane piston sinusoidally at a frequency close or equal to the lowest resonance frequency with its maximum displacement fixed. By applying the method of multiple scales, the equation for temporal modulation of a complex pressure amplitude of the lowest mode is derived in a case that a typical acoustic Mach number is comparable with the one-third power of the piston Mach number, while the relative detuning of a frequency is comparable with the quadratic order of the acoustic Mach number. The steady-state solution gives the asymmetric frequency response curve with bending (skew) due to nonlinear frequency upshift in addition to the linear downshift. Validity of the theory is checked against the frequency response obtained experimentally. For high amplitude of oscillations, an effect of jet loss at the throat of the resonator is taken into account, which introduces the quadratic loss to suppress the peak amplitude. It is revealed that as far as the present check is concerned, the weakly nonlinear theory can give quantitatively adequate description up to the pressure amplitude of about 3% to the equilibrium pressure. [DOI: 10.1121/1.1600719]

PACS numbers: 43.25.Gf, 43.25.Vt, 43.25.Cb [MFH]

I. INTRODUCTION

Novel methods for generation of shock-free, high-amplitude oscillations of gas in a tube (or a container) have recently attracted much attention in view of applications to thermoacoustic devices. They commonly exploit resonance in one form or another by exciting an acoustic system at a frequency close to or equal to one of its resonance frequencies (eigenfrequencies). As the amplitude of oscillations of gas becomes high, however, there emerges a shock, i.e., discontinuity in pressure, etc., so that increase in pressure amplitude of oscillations tends to be suppressed as the excitation is increased.

Whether or not the shock emerges is crucially determined by the relation between resonance frequencies and frequencies of higher harmonics of the excitation. If the resonance frequencies are ordered as multiples of the fundamental one, just as in the case of a closed tube of uniform cross section, then the tube is called being *consonant* and otherwise *dissonant*.[1,2] As the amplitude of excitation is increased in a consonant tube, each frequency of higher harmonics generated by nonlinearity hits the respective resonance frequencies so that higher modes are gradually excited and energy in the fundamental mode is pumped up into higher modes and dissipated there. This cascade process of energy flow is the mechanisms behind emergence of the shock and the resulting suppression of increase in pressure amplitude of the fundamental mode.

In order to annihilate the shock, the cascade process should be blocked by any means. At present, there are two methods confirmed experimentally at high pressure level. One is the method devised by Lawrenson *et al.*[2] and Ilinskii *et al.*[3] at MacroSonix. The essential point lies in making the axial cross section of a container nonuniform so as to render it dissonant. By vibrating the whole container on a shaker, shock-free, high-amplitude oscillations have been achieved experimentally and theoretical analysis has also been made.

The other method is to exploit wave dispersion, as devised by the present authors,[4] although it was proposed originally in a different problem.[5] Sound speed in gas is usually constant independent of frequency; therefore, no dispersion occurs in propagation. But, by connecting external agents to the tube, the phase speed can be made dependent on a frequency. In fact, this is achieved by connecting an array of Helmholtz resonators to a tube of uniform cross section axially. By sinusoidally driving the bellows mounted at one end of this tube, shock-free, high-amplitude oscillations of about 10% of the equilibrium pressure have been generated at the other closed end. Because the phase speed now depends on the frequency, resonance frequencies are no longer ordered as multiples of the fundamental one, and the tube becomes dissonant automatically without any change of cross section of the tube.

In the present context, it may be worthwhile to mention the dispersion. For propagation in a tube of nonuniform cross section, the phase speed cannot be defined strictly even in the case of an exponential horn because the amplitude decays or grows in the direction of propagation. The propagation speed is then determined by the characteristics and is still given by the sound speed. In this sense, the system still remains *hyperbolic*. In the tube with the array, on the contrary, a plane sinusoidal wave can be propagated so that the phase speed is clearly defined, which is now different from the sound speed and dependent on the frequency. The connection of the array changes the hyperbolic system originally to be *dispersive*.

 0001-4966/2003/114(4)/1772/13/$19.00

Besides the two methods, Rudenko *et al.*[6] have already proposed ideas to block the cascade process by introducing special absorbers of the second harmonics through the boundary condition. Andreev *et al.*[7] have confirmed the effects experimentally, though at even lower pressure level. For suppression or reduction of the second harmonics, Gusev *et al.*[8] have abandoned a usual monochromatic excitation to control the driver actively by adding to the fundamental sinusoidal excitation the second harmonic one with a phase difference. While these methods focus on increase in so-called *quality factor* of a resonator, it is questionable whether or not shock-free oscillations are ultimately achieved.

The purpose of this paper is to formulate nonlinear forced oscillations of an air column in the tube with the array of Helmholtz resonators and to derive theoretically a frequency response corresponding to the one obtained experimentally.[4] Supposing each resonator is small in effect, one-dimensional motion of air is assumed over the cross section of the tube except for a boundary layer on the wall and vicinity of resonator's orifices open to the tube. Supposing also that the axial spacing between neighboring resonators is small in comparison with a wavelength, the continuum approximation for the resonators is made so that the effects of the discrete distribution in array may be smeared out per axial length.

Because the pressure level observed in the experiment is still small relative to the equilibrium pressure, weakly nonlinear theory is developed by using the asymptotic method of multiple (two) scales.[9] The bellows used in the experiment are modeled as a plane piston reciprocating sinusoidally by taking account of the correspondence between the displacement of the bellows and the one of the piston. The boundary condition for the piston is usually of three types, either one of the maximum displacement, maximum speed, or maximum acceleration being held constant, the first of which is used in the present theory.

As the ratio of the maximum displacement of the piston to the tube length is much smaller than unity, so is the piston Mach number defined by the ratio of the maximum piston speed to the sound speed, which is comparable in order with the former. Denoting the piston Mach number by ε_p, and a typical acoustic Mach number in the tube by ε, respectively, the situation in the experiment corresponds to a case where ε is of order $\varepsilon_p^{1/3}$ and the relative detuning of a frequency from the resonance one is of order ε^2 $(=\varepsilon_p^{2/3})$. It is emphasized that the rigorous cubic nonlinear theory starting from the formulation based on the above assumptions is necessary to obtain a frequency response correctly up to the third order in the pressure amplitude.

The problem is formulated in Sec. II and the lossless linear theory is first described in Sec. III. Section IV is devoted to the nonlinear theory by using the method of multiple scales to derive the equation for slow modulation of a complex pressure amplitude of the fundamental mode. From the steady-state solution to the equation, the frequency response is obtained and compared with the experiment. In Sec. V, an effect of the jet loss at the throat of the resonator is considered in order to compensate the discrepancy between the theory and experiment as the amplitude becomes high.

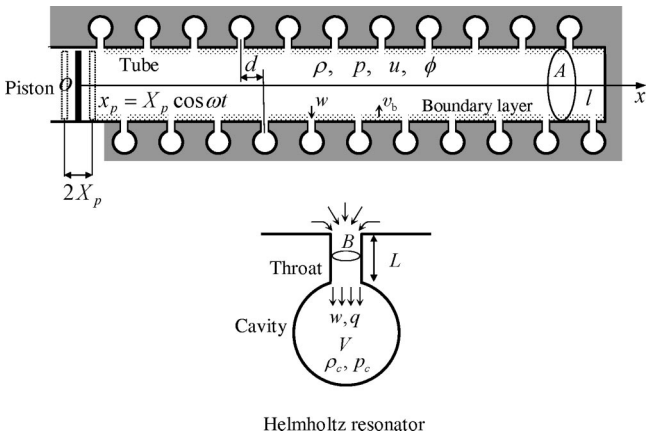


FIG. 1. Illustration of a tube with an array of Helmholtz resonators.

II. FORMULATION OF THE PROBLEM

A. Basic equations

We start by formulating the problem. Suppose a straight, rigid tube of radius R and of length l, to which an array of Helmholtz resonators is connected (see Fig. 1). One end of the tube is allowed to be displaced by a plane piston while the other end is closed by a flat plate. The tube is closed hermetically with the resonators inclusive. Each resonator is assumed small in the sense that the cavity's volume V is much smaller than the tube's volume per spacing Ad, A $(=\pi R^2)$ being the cross-sectional area of the tube and d the axial spacing. This ratio is denoted by κ $(\equiv V/Ad \ll 1)$, and called a *size parameter* of the array. Taking the axial spacing to be much smaller than a wavelength of oscillation, the continuum approximation is made for discrete distribution of the resonators to average its effect per unit axial length.

The Reynolds number is usually sufficiently high that effects of viscosity and heat conduction are limited only within the boundary layer developing on the tube wall. In the outside of the boundary layer called a region of an acoustic main flow, these lossy effects are negligible. The boundary layer is thin and the array is small $(\kappa \ll 1)$, so the acoustic main flow may be regarded as being almost one-dimensional. Under these assumptions, the basic equations for the main flow have already been presented in Ref. 10.

The equations of continuity and of motions are combined into the following equations:

$$\left[\frac{\partial}{\partial t}+(u\pm a)\frac{\partial}{\partial x}\right]\left[u\pm\frac{2}{\gamma-1}(a-a_0)\right]=\pm\frac{a}{A}\oint v_n\,\mathrm{d}s, \tag{1}$$

with the signs ordered vertically where x and t denote, respectively, the axial coordinate along the tube and the time, while u and a denote, respectively, the axial velocity of the air and the local sound speed. The latter is defined by

$$a^2=\frac{\mathrm{d}p}{\mathrm{d}\rho}=a_0^2\left(\frac{p}{p_0}\right)^{(\gamma-1)/\gamma}, \tag{2}$$

with $a_0=\sqrt{\gamma p_0/\rho_0}$ where p and ρ denote, respectively, the pressure and density of the air and the subscript 0 attached to p and ρ implies the respective values in equilibrium, γ being the ratio of the specific heats, and a_0 is the linear sound speed. Since the lossy effects are neglected in the acoustic main flow, the adiabatic relation $p/p_0=(\rho/\rho_0)^\gamma$ is assumed to hold.

Through the right-hand side of (1) are included the effects of the boundary layer and of the array of Helmholtz resonators. Here, v_n represents the velocity directed inward normal to the boundary of the axial cross section of the acoustic main-flow region, ds being the line element along the boundary. Where the tube wall exists, v_n is the velocity at the edge of the boundary layer v_b, given by

$$v_b=\left(1+\frac{\gamma-1}{\sqrt{\mathrm{Pr}}}\right)\sqrt{\nu}\,\frac{\partial^{-1/2}}{\partial t^{-1/2}}\left(\frac{\partial u}{\partial x}\right), \tag{3}$$

where ν denotes the kinematic viscosity taken constant, Pr being the Prandtl number, and the derivative of minus half-order of a function $f(x,t)$ is defined as

$$\frac{\partial^{-1/2}f}{\partial t^{-1/2}}\equiv\frac{1}{\sqrt{\pi}}\int_{-\infty}^{t}\frac{f(x,\tau)}{\sqrt{t-\tau}}\,\mathrm{d}\tau. \tag{4}$$

On the other hand, where the tube opens to the resonator, v_n is the velocity of the air flowing out of the throat into the tube, $-w$. Because the throat is much shorter than a typical wavelength, the mass flux density q averaged over the cross section of the throat may be regarded as being uniform along the throat. While no account of motions of the air is taken in the cavity, the rate of increase in the mass therein must balance with the mass flux flown into it. This requires that

$$V\frac{\partial\rho_c}{\partial t}=Bq, \tag{5}$$

where ρ_c denotes the mean density of the air in the cavity, and B denotes the cross-sectional area of the throat.

Thus, the integral in (1) consists of two contributions as follows:

$$\frac{1}{A}\oint v_n\,\mathrm{d}s=2\left(1+\frac{\gamma-1}{\sqrt{\mathrm{Pr}}}\right)\frac{\sqrt{\nu}}{R^*}\,\frac{\partial^{-1/2}}{\partial t^{-1/2}}\left(\frac{\partial u}{\partial x}\right)-\frac{\kappa}{\rho}\,\frac{\partial\rho_c}{\partial t}, \tag{6}$$

where R^* is the reduced radius of the tube defined by $R/(1-BR/2Ad)$, and q is set equal to ρw at the orifice on the tube side. The derivation of the right-hand side of (6) may be facilitated by multiplying A and ds by d, respectively, noting that $d(\mathrm{d}s)$ corresponds to the area element on the surface bounding the region of the acoustic main flow. Here, q in (5) is expressed in terms of the excess pressure p_c' $(=p_c-p_0)$ in place of the density. Assuming the adiabatic relation for the air in the cavity, ρ_c is expanded into the Taylor series with respect to p_c' to yield q in the following form:

$$q=\frac{V}{Ba_0^2}\,\frac{\partial}{\partial t}\left[p_c'-\frac{(\gamma-1)}{2\gamma p_0}p_c'^2+\frac{(\gamma-1)(2\gamma-1)}{6\gamma^2p_0^2}p_c'^3+\cdots\right]. \tag{7}$$

The behavior of the resonator is governed by (5) supplemented by the equation of motion for the air in the throat. The momentum balance of the air in the throat of length L requires that

$$L\frac{\partial q}{\partial t}=-p_c+p-\frac{2L\sqrt{\nu}}{r}\,\frac{\partial^{1/2}}{\partial t^{1/2}}(\rho_0 w), \tag{8}$$

where the last term represents the friction at the throat wall.[10] Using the lowest relation of (7) to evaluate $\rho_0 w$, substitution of (7) for q into (8) leads to the following equation:

$$\begin{aligned}\frac{\partial^2p_c'}{\partial t^2}+\frac{2\sqrt{\nu}}{r^*}\,\frac{\partial^{3/2}p_c'}{\partial t^{3/2}}+\omega_0^2p_c'&=\omega_0^2p'+\frac{\gamma-1}{2\gamma p_0}\,\frac{\partial^2p_c'^2}{\partial t^2}-\frac{(\gamma-1)(2\gamma-1)}{6\gamma^2p_0^2}\,\frac{\partial^2p_c'^3}{\partial t^2}\\&\quad+\cdots,\end{aligned} \tag{9}$$

where p' $(=p-p_0)$ is the excess pressure in the tube, r^* $(=r/c_L)$, c_L being L'/L_e with $L'=L+2r$ and $L_e=L+2\times0.82r$, is the reduced radius of the throat by taking account of the end corrections on both ends, and ω_0 $(=\sqrt{Ba_0^2/L_eV})$ is the natural angular frequency of the resonator.[4] The derivative of three-half order is defined by differentiating the derivative of minus half-order twice with respect to t.

B. Reduction by the velocity potential

Addition and subtraction of (1) with the upper and lower signs lead, respectively, to

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+\frac{2a}{\gamma-1}\,\frac{\partial a}{\partial x}=0, \tag{10}$$

and

$$\frac{2}{\gamma-1}\left(\frac{\partial a}{\partial t}+u\frac{\partial a}{\partial x}\right)+a\frac{\partial u}{\partial x}=\frac{a}{A}\oint v_n\,\mathrm{d}s. \tag{11}$$

In order to eliminate a from these equation, we introduce a velocity potential ϕ defined by

$$u\equiv\frac{\partial\phi}{\partial x}. \tag{12}$$

Substituting (12) into (10) and integrating this with respect to x, we have

$$a^2=a_0^2-(\gamma-1)\left[\frac{\partial\phi}{\partial t}+\frac{1}{2}\left(\frac{\partial\phi}{\partial x}\right)^2\right]. \tag{13}$$

This is simply the Bernoulli's theorem. Multiplying (11) by a and using (13) to eliminate a^2, we derive

$$\frac{\partial^2\phi}{\partial t^2}-a_0^2\frac{\partial^2\phi}{\partial x^2}=-\frac{\partial}{\partial t}\left(\frac{\partial\phi}{\partial x}\right)^2-(\gamma-1)\frac{\partial\phi}{\partial t}\frac{\partial^2\phi}{\partial x^2}-\frac{\gamma+1}{2}\left(\frac{\partial\phi}{\partial x}\right)^2\frac{\partial^2\phi}{\partial x^2}-\frac{a^2}{A}\oint v_n\,\mathrm{d}s. \tag{14}$$

In passing, substituting (13) into (2), p' is expressed inversely in terms of ϕ. In fact, by expanding $p/p_0=(a^2/a_0^2)^{\gamma/(\gamma-1)}$ around $a=a_0$ and truncating the series at the cubic terms, it follows that

$$p'=-\rho_0\frac{\partial\phi}{\partial t}+\frac{\rho_0}{2a_0^2}\left[\left(\frac{\partial\phi}{\partial t}\right)^2-a_0^2\left(\frac{\partial\phi}{\partial x}\right)^2\right]+\frac{\rho_0}{2a_0^2}\left[\frac{(\gamma-2)}{3a_0^2}\left(\frac{\partial\phi}{\partial t}\right)^3+\frac{\partial\phi}{\partial t}\left(\frac{\partial\phi}{\partial x}\right)^2\right]+\cdots. \tag{15}$$

The boundary conditions are imposed at the piston surface and the closed end so that the velocity of the air may be equal to the one of the respective surfaces. Taking the origin of the coordinate x at the mean position of the piston surface over one period and the closed end at $x=l$, the respective conditions are expressed as follows:

$$\frac{\partial\phi}{\partial x}=\frac{\mathrm{d}x_p}{\mathrm{d}t}=\frac{i}{2}\omega X_p\mathrm{e}^{i\omega t}+\text{c.c.}$$

$$\text{at}\quad x=x_p=\frac{1}{2}X_p\mathrm{e}^{i\omega t}+\text{c.c.}=X_p\cos(\omega t), \tag{16}$$

and

$$\frac{\partial\phi}{\partial x}=0\quad\text{at}\quad x=l, \tag{17}$$

where x_p denotes the position of the piston surface; the displacement amplitude, X_p, and the angular frequency of excitation, ω, are taken real and positive, c.c. implying the complex conjugate to *all preceding terms*, if any.

C. Normalization

We next normalize the equations and the boundary conditions by making the following replacement:

$$[x,\ t,\ \phi,\ a,\ \rho,\ \rho_c,\ p',\ p_c']=[l\overline{x},\ (l/a_0)\overline{t},\ lu_0\overline{\phi},\ a_0\overline{a},\ \rho_0\overline{\rho},\ \rho_0\overline{\rho_c},\ \rho_0a_0u_0\overline{p'},\ \rho_0a_0u_0\overline{p_c'}], \tag{18}$$

where the quantities with the overbar imply the dimensionless ones, and u_0 is a typical speed of the air, say the maximum speed in the acoustic main flow. Here, two Mach numbers are defined: one is the acoustic Mach number ε associated with the air speed and the other the piston Mach number ε_p. Both Mach numbers are assumed to be much smaller than unity, as

$$\varepsilon\equiv\frac{u_0}{a_0}\ll1\quad\text{and}\quad\varepsilon_p\equiv\frac{\omega X_p}{a_0}\ll1. \tag{19}$$

In a lossless case of a tube without the array of resonators, it is known that the air column will resonate in the lowest mode when a half-wavelength $\pi a_0/\omega$ coincides with the tube length. Using this angular frequency $\pi a_0/l$, ω is made dimensionless by introducing σ as

$$\omega=\frac{\pi a_0}{l}\sigma. \tag{20}$$

By the replacement (18), (14) with (6) is normalized as

$$\frac{\partial^2\overline{\phi}}{\partial\overline{t}^2}-\frac{\partial^2\overline{\phi}}{\partial\overline{x}^2}=\frac{\kappa\overline{a}^2}{\varepsilon\overline{\rho}}\frac{\partial\overline{\rho_c}}{\partial\overline{t}}-\varepsilon\left[\frac{\partial}{\partial\overline{t}}\left(\frac{\partial\overline{\phi}}{\partial\overline{x}}\right)^2+(\gamma-1)\frac{\partial\overline{\phi}}{\partial\overline{t}}\frac{\partial^2\overline{\phi}}{\partial\overline{x}^2}\right]-\varepsilon^2\frac{(\gamma+1)}{2}\left(\frac{\partial\overline{\phi}}{\partial\overline{x}}\right)^2\frac{\partial^2\overline{\phi}}{\partial\overline{x}^2}-\delta\overline{a}^2\frac{\partial^{-1/2}}{\partial\overline{t}^{-1/2}}\left(\frac{\partial^2\overline{\phi}}{\partial\overline{x}^2}\right), \tag{21}$$

with $\overline{\rho}=(1+\varepsilon\gamma\overline{p}')^{1/\gamma}$, $\overline{\rho_c}=(1+\varepsilon\gamma\overline{p_c}')^{1/\gamma}$, and $\overline{a}^2=1-\varepsilon(\gamma-1)[\partial\overline{\phi}/\partial\overline{t}+\varepsilon(\partial\overline{\phi}/\partial\overline{x})^2/2]$, where $\overline{a}^2/\overline{\rho}$ and $\overline{\rho_c}$ are expanded, respectively, as

$$\frac{\overline{a}^2}{\overline{\rho}}=1-\varepsilon(\gamma-2)\frac{\partial\overline{\phi}}{\partial\overline{t}}-\varepsilon^2\frac{(\gamma-2)}{2}\left[\left(\frac{\partial\overline{\phi}}{\partial\overline{t}}\right)^2+\left(\frac{\partial\overline{\phi}}{\partial\overline{x}}\right)^2\right]+\cdots, \tag{22}$$

and

$$\overline{\rho_c}=1+\varepsilon\overline{p_c}'-\varepsilon^2\frac{(\gamma-1)}{2}\overline{p_c}'^2+\varepsilon^3\frac{(\gamma-1)(2\gamma-1)}{6}\overline{p_c}'^3+\cdots. \tag{23}$$

On the other hand, (9) is normalized as

$$\frac{\partial^2\overline{p_c}'}{\partial\overline{t}^2}+\delta_r\frac{\partial^{3/2}\overline{p_c}'}{\partial\overline{t}^{3/2}}+(\pi\sigma_0)^2\overline{p_c}'=(\pi\sigma_0)^2\overline{p}'+\varepsilon\frac{(\gamma-1)}{2}\frac{\partial^2\overline{p_c}'^2}{\partial\overline{t}^2}-\varepsilon^2\frac{(\gamma-1)(2\gamma-1)}{6}\times\frac{\partial^2\overline{p_c}'^3}{\partial\overline{t}^2}+\cdots, \tag{24}$$

with $\sigma_0=l\omega_0/\pi a_0$, where δ and δ_r are the parameters representing the boundary-layer effects due to the tube wall and the throat wall, respectively, and are given by

$$\delta=2\left(1+\frac{\gamma-1}{\sqrt{\mathrm{Pr}}}\right)\frac{\sqrt{\nu l/a_0}}{R^*}\quad\text{and}\quad\delta_r=2\frac{\sqrt{\nu l/a_0}}{r^*}. \tag{25}$$

In (24), $\overline{p}'$ is given, after normalization of (18), by

$$\overline{p}'=-\frac{\partial\overline{\phi}}{\partial\overline{t}}+\frac{\varepsilon}{2}\left[\left(\frac{\partial\overline{\phi}}{\partial\overline{t}}\right)^2-\left(\frac{\partial\overline{\phi}}{\partial\overline{x}}\right)^2\right]+\frac{\varepsilon^2}{2}\left[\frac{(\gamma-2)}{3}\left(\frac{\partial\overline{\phi}}{\partial\overline{t}}\right)^3+\frac{\partial\overline{\phi}}{\partial\overline{t}}\left(\frac{\partial\overline{\phi}}{\partial\overline{x}}\right)^2\right]+\cdots. \tag{26}$$

On the other hand, the boundary conditions are normalized as

$$\frac{\partial \bar{\phi}}{\partial \bar{x}} = i\frac{\pi\sigma c}{2\varepsilon} e^{i\pi\sigma\bar{t}} + \text{c.c. at } \bar{x} = \frac{c}{2} e^{i\pi\sigma\bar{t}} + \text{c.c.}, \tag{27}$$

and

$$\frac{\partial \bar{\phi}}{\partial \bar{x}} = 0 \text{ at } \bar{x} = 1, \tag{28}$$

where c represents the dimensionless amplitude X_p/l ($\ll 1$) and $\pi\sigma c$ corresponds to ε_p.

III. LOSSLESS LINEAR THEORY

At first, it is instructive to discuss the lossless linear case. Taking the limit as c, δ, and $\delta_r \to 0$, with c/ε fixed, the linearized problem is governed by the following equations:

$$\frac{\partial^2 \phi}{\partial t^2} - \frac{\partial^2 \phi}{\partial x^2} = \kappa \frac{\partial p_c'}{\partial t}, \tag{29}$$

$$\frac{\partial^2 p_c'}{\partial t^2} + (\pi\sigma_0)^2 p_c' = (\pi\sigma_0)^2 p', \tag{30}$$

$$p' = -\frac{\partial \phi}{\partial t}, \tag{31}$$

with the boundary conditions given by

$$\frac{\partial \phi}{\partial x} = i\frac{\pi\sigma c}{2\varepsilon} e^{i\pi\sigma t} + \text{c.c. at } x = 0 \text{ and } \frac{\partial \phi}{\partial x} = 0 \text{ at } x = 1. \tag{32}$$

Here and hereafter, the overbar is omitted.

Setting

$$\begin{bmatrix} \phi \\ p' \\ p_c' \end{bmatrix} = \begin{bmatrix} \Phi(x) \\ F(x) \\ G(x) \end{bmatrix} e^{i\pi\sigma t} + \text{c.c.}, \tag{33}$$

and substituting this into (29)–(31) to eliminate F and G, we derive

$$\frac{d^2\Phi}{dx^2} + k^2\Phi = 0, \tag{34}$$

where k is given by

$$k^2 = (\pi\sigma)^2\left(1 + \frac{\kappa\sigma_0^2}{\sigma_0^2 - \sigma^2}\right). \tag{35}$$

Imposing the boundary conditions (32), the solutions are easily obtained as

$$\begin{bmatrix} \phi \\ p' \\ p_c' \end{bmatrix} = \begin{bmatrix} i/\pi\sigma \\ 1 \\ s \end{bmatrix} \frac{\pi^2\sigma^2\cos[k(x-1)]}{k\sin k}\frac{c}{2\varepsilon} e^{i\pi\sigma t} + \text{c.c.}, \tag{36}$$

with $s = \sigma_0^2/(\sigma_0^2 - \sigma^2)$.

It is noted in (36) that k becomes purely imaginary for $\sigma_0 < \sigma < \sqrt{1+\kappa}\,\sigma_0$. Then, $\cos[k(x-1)]$ and $k\sin k$ take, respectively, $\cosh[|k|(x-1)]$ and $-|k|\sinh|k|$. As σ approaches σ_0, k tends to diverge and the evanescence occurs in the limit. This is due to the side-branch resonance where each resonator reflects back the incident wave totally. As σ approaches $\sqrt{1+\kappa}\,\sigma_0$, on the other hand, k tends to vanish.

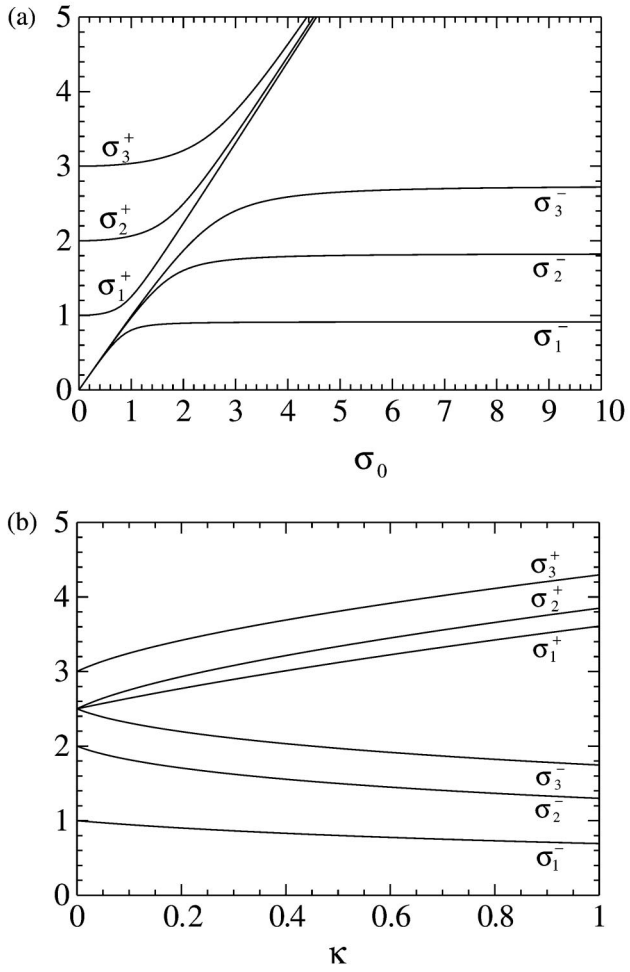


FIG. 2. Graphs of the resonance frequencies $\sigma_m^{\pm}$ ($m = 1$, 2, and 3): (a) displays the dependence on σ_0 with the value of κ fixed at 0.2, and (b) displays the dependence on κ with the value of σ_0 fixed at 2.5.

This is a resonance supported by the array of resonators where the air column oscillates in unison. It seems to be worth examining this new resonance, but we are concerned here with the resonance when $\sin k$ vanishes, i.e., $k = m\pi$ (>0) ($m = 1,2,3,\ldots$).

Unlike in the case of the tube without the array, there are two resonance frequencies $\sigma_m^{\pm}$ (>0) for a given value of m. They are determined from the quadratic equation $\sigma^4 - [m^2 + (1+\kappa)\sigma_0^2]\sigma^2 + m^2\sigma_0^2 = 0$ in σ^2 as

$$(\sigma_m^{\pm})^2 = \tfrac{1}{2}\{m^2 + (1+\kappa)\sigma_0^2 \pm \sqrt{[m^2 + (1+\kappa)\sigma_0^2]^2 - 4m^2\sigma_0^2}\}, \tag{37}$$

with the signs ordered vertically. It is easily found that $\sigma_m^- < \sigma_0 < \sqrt{1+\kappa}\,\sigma_0 < \sigma_m^+$. Figure 2 shows the dependence of $\sigma_m^{\pm}$ ($m = 1$, 2, and 3) on σ_0 or κ for either one fixed. In the tube without the array, the resonance frequencies are given by $\sigma = m$ simply. In Fig. 2(a), where κ is fixed to be 0.2, it is seen that σ_m^+ tends to m as σ_0 vanishes, whereas σ_m^- tends to vanish. On the other hand, as σ_0 becomes large, σ_m^+ and σ_m^- tend to $\sqrt{1+\kappa}\,\sigma_0$ and $m/\sqrt{1+\kappa}$, respectively, for $m \ll \sigma_0$. For a large value of σ_0, it is σ_m^-'s that correspond to the resonance frequencies in the tube without the array. But,

since the tube then becomes consonant, the merit of connection of the array is lost. For a fixed value of $\sigma_0=2.5$, on the other hand, the dependence of $\sigma_m^{\pm}$ on κ is shown in Fig. 2(b). As κ increases, the resonance frequencies σ_m^- decrease, whereas σ_m^+ increase. For a small value of κ, $\sigma_m^{\pm}$ are expressed asymptotically as

$$\sigma_m^+=\sigma_0\left[1+\frac{\kappa\sigma_0^2}{2(\sigma_0^2-m^2)}+\cdots\right], \tag{38}$$

and

$$\sigma_m^-=m\left[1-\frac{\kappa\sigma_0^2}{2(\sigma_0^2-m^2)}+\cdots\right], \tag{39}$$

for $\sigma_0>m$, while the expressions of $\sigma_m^{\pm}$ are exchanged for $\sigma_0<m$.

IV. NONLINEAR THEORY

We now develop a nonlinear theory by taking account of small but finite magnitude of ε and also lossy effects. While the magnitudes of c and of ε_p are defined clearly, the one of ε is left ambiguous and should now be related to c. When the driving frequency σ is off the resonance, the linear solution (36) suggests that ε is of the same order as c or ε_p. This is because the quantities on the left-hand side of (36) are regarded as being of order unity. But, as σ is set closer to the resonance frequency, ε becomes much larger than c, as the term ''resonance'' implies. Then, ε is determined not only by c but also by a detuning $\Delta\sigma$, indicating how far the driving frequency is set from the resonance one. Since the experiment exploits the resonance in the lowest mode, we consider a case where σ is near the lowest resonance frequency σ_1^- $(\equiv\sigma_1)$. Setting $\Delta\sigma=\sigma-\sigma_1$, $k\sin k$ in the denominator of (36) may be approximated as $-\pi(\mathrm{d}k/\mathrm{d}\sigma)|_{\sigma=\sigma_1}\Delta\sigma$, so ϕ, p', and p_c' are estimated to be of order $c/\varepsilon|\Delta\sigma|$. For this to be of order unity, ε is determined as

$$\varepsilon\approx\frac{c}{|\Delta\sigma|}. \tag{40}$$

Since we assume the lossless case here, ε diverges, of course, as $\Delta\sigma$ vanishes. The experimental result indicates that c is of order 10^{-3}, $|\Delta\sigma|$ is of 10^{-2}, and ε is of 10^{-1}. This suggests that c is of ε^3 and $\Delta\sigma$ is of ε^2. Also, δ and δ_r for the lossy effects are estimated to be of ε^2.

A. Perturbation procedures

In light of the ordering mentioned above, the exponential factor $\exp(i\pi\sigma t)$ in (27) may be written as $\exp[i\pi(\sigma_1+\varepsilon^2\sigma')t]$ by setting $\Delta\sigma=\varepsilon^2\sigma'$, σ' being a quantity of order unity. Regarding $\varepsilon^2 t$ as a slow variable compared with t, we employ the method of multiple (two) scales by introducing two variables t_0 $(\equiv t)$ and t_2 $(\equiv\varepsilon^2 t)$. Then, the boundary condition (27) may be given, without the overbar, as

$$\frac{\partial\phi}{\partial x}=i\frac{E}{2}\mathrm{e}^{i\pi\sigma_1 t_0}+\text{c.c.}+O(\varepsilon^2 E)$$

$$\text{at}\quad x=\frac{c}{2}\mathrm{e}^{i\pi\sigma_1 t_0}+\text{c.c.}+O(\varepsilon^2 c), \tag{41}$$

where we note that E $[=(\pi\sigma_1 c/\varepsilon)\exp(i\pi\sigma' t_2)]$ is dependent of t_2 and is of order ε^2.

We seek the solutions to (21) and (24) with (26) under the boundary conditions (41) and (28) in the asymptotic expansion of ε as

$$\begin{bmatrix}\phi\\ p'\\ p_c'\end{bmatrix}=\begin{bmatrix}\phi^{(0)}\\ f^{(0)}\\ g^{(0)}\end{bmatrix}+\varepsilon\begin{bmatrix}\phi^{(1)}\\ f^{(1)}\\ g^{(1)}\end{bmatrix}+\varepsilon^2\begin{bmatrix}\phi^{(2)}\\ f^{(2)}\\ g^{(2)}\end{bmatrix}+\cdots. \tag{42}$$

By introduction of t_2, the differential operator is expanded as

$$\frac{\partial}{\partial t}=\frac{\partial}{\partial t_0}+\varepsilon^2\frac{\partial}{\partial t_2}+\cdots. \tag{43}$$

The lowest-order problem of ε takes the same form as (29) to (31) with ϕ, p', and p_c' replaced by $\phi^{(0)}$, $f^{(0)}$, and $g^{(0)}$, respectively. But, they are now subjected to the homogeneous boundary conditions: $\partial\phi^{(0)}/\partial x=0$ at $x=0$ and $x=1$, since $\partial\phi/\partial x$ in (41) is expanded around $x=0$ and E is assumed to be of order ε^2. Thus, the lowest-order solutions are given by the eigensolutions of the system. Taking, in (36), the formal limit as $c/\varepsilon\to 0$, $k\to\pi$, and $\sigma\to\sigma_1$, and setting $(\pi^2\sigma^2/k\sin k)c/2\varepsilon$ to be α, the solutions are obtained as

$$\begin{bmatrix}\phi^{(0)}\\ f^{(0)}\\ g^{(0)}\end{bmatrix}=\begin{bmatrix}i/\pi\sigma_1\\ 1\\ s_1\end{bmatrix}\cos[\pi(x-1)]\alpha\mathrm{e}^{i\pi\sigma_1 t_0}+\text{c.c.}, \tag{44}$$

with $s_1=\sigma_0^2/(\sigma_0^2-\sigma_1^2)$ $(\sigma_1\neq\sigma_0)$, where $\pi^2\sigma_1^2(1+\kappa s_1)=\pi^2$ and the complex amplitude α including the detuning is assumed to depend on t_2 but unspecified at this order.

Here, we make the following remark. The constant s_1 becomes positive or negative depending on the values of σ_0 and σ_1. Therefore, the pressure in the cavity differs from the one in the tube in magnitude and phase. If σ_0 is chosen greater than σ_1, as is usually the case, the pressure amplitude in the cavity becomes higher than the one in the tube.

1. First-order problem

Let us now proceed to the first-order problem. Then, the equations take the following form:

$$\frac{\partial^2\phi^{(1)}}{\partial t_0^2}-\frac{\partial^2\phi^{(1)}}{\partial x^2}-\kappa\frac{\partial g^{(1)}}{\partial t_0}$$

$$=-\kappa\left[(\gamma-1)g^{(0)}+(\gamma-2)\frac{\partial\phi^{(0)}}{\partial t_0}\right]\frac{\partial g^{(0)}}{\partial t_0}$$

$$-\frac{\partial}{\partial t_0}\left(\frac{\partial\phi^{(0)}}{\partial x}\right)^2-(\gamma-1)\frac{\partial\phi^{(0)}}{\partial t_0}\frac{\partial^2\phi^{(0)}}{\partial x^2}, \tag{45}$$

and

$$\frac{\partial^2 g^{(1)}}{\partial t_0^2}+(\pi\sigma_0)^2 g^{(1)}-(\pi\sigma_0)^2 f^{(1)}=\frac{(\gamma-1)}{2}\frac{\partial^2 g^{(0)2}}{\partial t_0^2}, \tag{46}$$

with

$$f^{(1)}=-\frac{\partial\phi^{(1)}}{\partial t_0}+\frac{1}{2}\left(\frac{\partial\phi^{(0)}}{\partial t_0}\right)^2-\frac{1}{2}\left(\frac{\partial\phi^{(0)}}{\partial x}\right)^2, \tag{47}$$

while the boundary conditions at this order take simply

$$\frac{\partial\phi^{(1)}}{\partial x}=0 \quad \text{at} \quad x=0 \quad \text{and} \quad x=1. \tag{48}$$

Introducing (44) into the right-hand sides of (45)–(47), we find the first-order solutions should be in the following form:

$$\begin{bmatrix}\phi^{(1)}\\ f^{(1)}\\ g^{(1)}\end{bmatrix}=\begin{bmatrix}\Phi_2^{(1)}\\ F_2^{(1)}\\ G_2^{(1)}\end{bmatrix}e^{2i\pi\sigma_1 t_0}+\text{c.c.}+\begin{bmatrix}\Phi_0^{(1)}\\ F_0^{(1)}\\ G_0^{(1)}\end{bmatrix}, \tag{49}$$

where $\Phi_j^{(1)}$, $F_j^{(1)}$, and $G_j^{(1)}$ $(j=0,2)$ are functions of x and α to be determined, and the subscript j implies the coefficient of the jth harmonics, $e^{ij\pi\sigma_1 t_0}$. Substituting (49) into (45)–(47), we consider the respective harmonics separately. For the second harmonics, we eliminate $F_2^{(1)}$ and $G_2^{(1)}$ to derive the equation for $\Phi_2^{(1)}$ as follows:

$$\frac{d^2\Phi_2^{(1)}}{dx^2}+k_2^2\Phi_2^{(1)}=\frac{i}{2\pi\sigma_1}\{k_2^2\mathcal{A}_0+(k_2^2-4\pi^2)\mathcal{A}_2\times\cos[2\pi(x-1)]\}\alpha^2, \tag{50}$$

with $k_2^2=(2\pi\sigma_1)^2(1+\kappa s_2)$ and $s_2=\sigma_0^2/(\sigma_0^2-4\sigma_1^2)$ $(\sigma_1\neq\sigma_0/2)$, where $\mathcal{A}_0$ and $\mathcal{A}_2$ are given by

$$\mathcal{A}_0=\frac{\pi^2}{k_2^2}\left\{\gamma-3+\kappa\left[(s_1-s_2)\sigma_1^2-s_2+(\gamma-1)\times(s_1-1)s_1\sigma_1^2+4(\gamma-1)s_1^2s_2\frac{\sigma_1^2}{\Omega}\right]\right\}, \tag{51}$$

and

$$\mathcal{A}_2=\frac{\pi^2}{(k_2^2-4\pi^2)}\left\{\gamma+1+\kappa\left[(s_1-s_2)\sigma_1^2+s_2+(\gamma-1)\times(s_1-1)s_1\sigma_1^2+4(\gamma-1)s_1^2s_2\frac{\sigma_1^2}{\Omega}\right]\right\}, \tag{52}$$

with $\Omega=\sigma_0^2/\sigma_1^2$. Note that $k_2^2\neq4\pi^2$ but k_2 is assumed not to vanish, i.e., $\sigma_1\neq\sqrt{1+\kappa}\sigma_0/2$. The second-harmonic solutions are easily obtainable as

$$\Phi_2^{(1)}=\frac{i}{2\pi\sigma_1}\{\mathcal{A}_0+\mathcal{A}_2\cos[2\pi(x-1)]\}\alpha^2, \tag{53}$$

$$F_2^{(1)}=\left\{\mathcal{A}_0+\frac{1}{4}\left(1+\frac{1}{\sigma_1^2}\right)+\left[\mathcal{A}_2+\frac{1}{4}\left(1-\frac{1}{\sigma_1^2}\right)\right]\times\cos[2\pi(x-1)]\right\}\alpha^2, \tag{54}$$

$$G_2^{(1)}=\left\{s_2\left[\mathcal{A}_0+\frac{1}{4}\left(1+\frac{1}{\sigma_1^2}\right)-(\gamma-1)\frac{s_1^2}{\Omega}\right]+s_2\left[\mathcal{A}_2+\frac{1}{4}\left(1-\frac{1}{\sigma_1^2}\right)-(\gamma-1)\frac{s_1^2}{\Omega}\right]\cos[2\pi(x-1)]\right\}\alpha^2. \tag{55}$$

For the zeroth harmonics, on the other hand, the right-hand side of (45) contains no terms independent of t_0. Therefore, we take $\Phi_0^{(1)}=0$. From (46) and (47), we have

$$F_0^{(1)}=G_0^{(1)}=\left\{\frac{1}{2}\left(1-\frac{1}{\sigma_1^2}\right)+\frac{1}{2}\left(1+\frac{1}{\sigma_1^2}\right)\times\cos[2\pi(x-1)]\right\}|\alpha|^2. \tag{56}$$

These solutions show that while no steady streaming occurs at this order, the steady but nonuniform pressure distribution appears in the tube as well as in the cavities. The maximum $|\alpha|^2$ occurs at both ends, while the minimum $-|\alpha|^2/\sigma_1^2$ occurs in the middle. Note that the effect of the array, i.e., κ, appears only in the minimum through σ_1.

2. Second-order problem

Upon completion of the first-order problem, we proceed to the second-order problem. It is given by

$$\frac{\partial^2\phi^{(2)}}{\partial t_0^2}-\frac{\partial^2\phi^{(2)}}{\partial x^2}-\kappa\frac{\partial g^{(2)}}{\partial t_0}=K^{(2)}+L^{(2)}+M^{(2)}, \tag{57}$$

with

$$\begin{aligned}K^{(2)}=\kappa\Bigg\{&\frac{\partial g^{(0)}}{\partial t_2}-(\gamma-1)\frac{\partial}{\partial t_0}(g^{(0)}g^{(1)})\\&-(\gamma-2)\left(\frac{\partial\phi^{(0)}}{\partial t_0}\frac{\partial g^{(1)}}{\partial t_0}+\frac{\partial\phi^{(1)}}{\partial t_0}\frac{\partial g^{(0)}}{\partial t_0}\right)\\&-\frac{(\gamma-2)}{2}\left[\left(\frac{\partial\phi^{(0)}}{\partial t_0}\right)^2+\left(\frac{\partial\phi^{(0)}}{\partial x}\right)^2\right]\frac{\partial g^{(0)}}{\partial t_0}\\&+\frac{(\gamma-1)(\gamma-2)}{2}\frac{\partial\phi^{(0)}}{\partial t_0}\frac{\partial g^{(0)2}}{\partial t_0}\\&+\frac{(\gamma-1)(2\gamma-1)}{6}\frac{\partial g^{(0)3}}{\partial t_0}\Bigg\},\end{aligned} \tag{58}$$

$$L^{(2)}=-2\frac{\partial^2\phi^{(0)}}{\partial t_2\partial t_0}-\frac{\delta}{\varepsilon^2}\frac{\partial^{-1/2}}{\partial t_0^{-1/2}}\left(\frac{\partial^2\phi^{(0)}}{\partial x^2}\right), \tag{59}$$

$$\begin{aligned}M^{(2)}=&-\frac{\partial}{\partial t_0}\left(2\frac{\partial\phi^{(0)}}{\partial x}\frac{\partial\phi^{(1)}}{\partial x}\right)-(\gamma-1)\left(\frac{\partial\phi^{(0)}}{\partial t_0}\frac{\partial^2\phi^{(1)}}{\partial x^2}\right.\\&\left.+\frac{\partial\phi^{(1)}}{\partial t_0}\frac{\partial^2\phi^{(0)}}{\partial x^2}\right)-\frac{(\gamma+1)}{2}\left(\frac{\partial\phi^{(0)}}{\partial x}\right)^2\frac{\partial^2\phi^{(0)}}{\partial x^2},\end{aligned} \tag{60}$$

and

$$\begin{aligned}&\frac{\partial^2 g^{(2)}}{\partial t_0^2}+(\pi\sigma_0)^2g^{(2)}-(\pi\sigma_0)^2f^{(2)}\\&=-2\frac{\partial^2 g^{(0)}}{\partial t_2\partial t_0}-\frac{\delta_r}{\varepsilon^2}\frac{\partial^{3/2}g^{(0)}}{\partial t_0^{3/2}}+(\gamma-1)\frac{\partial^2}{\partial t_0^2}(g^{(0)}g^{(1)})\\&\quad-\frac{(\gamma-1)(2\gamma-1)}{6}\frac{\partial^2 g^{(0)3}}{\partial t_0^2},\end{aligned} \tag{61}$$

with

$$f^{(2)}=-\frac{\partial\phi^{(2)}}{\partial t_0}-\frac{\partial\phi^{(0)}}{\partial t_2}+\frac{\partial\phi^{(0)}}{\partial t_0}\frac{\partial\phi^{(1)}}{\partial t_0}-\frac{\partial\phi^{(0)}}{\partial x}\frac{\partial\phi^{(1)}}{\partial x}+\frac{(\gamma-2)}{6}\left(\frac{\partial\phi^{(0)}}{\partial t_0}\right)^3+\frac{1}{2}\frac{\partial\phi^{(0)}}{\partial t_0}\left(\frac{\partial\phi^{(0)}}{\partial x}\right)^2. \tag{62}$$

The boundary conditions now take

$$\frac{\partial\phi^{(2)}}{\partial x}=i\frac{E}{2\varepsilon^2}e^{i\pi\sigma_1 t_0}+\text{c.c. at } x=0, \tag{63}$$

and

$$\frac{\partial\phi^{(2)}}{\partial x}=0 \text{ at } x=1. \tag{64}$$

At this order, we do not necessarily have to seek the full solutions. We have only to derive the condition for the complex amplitude α from the boundary conditions. Setting the solutions in the following form:

$$\begin{bmatrix}\phi^{(2)}\\ f^{(2)}\\ g^{(2)}\end{bmatrix}=\begin{bmatrix}\Phi_1^{(2)}\\ F_1^{(2)}\\ G_1^{(2)}\end{bmatrix}e^{i\pi\sigma_1 t_0}+\begin{bmatrix}\Phi_3^{(2)}\\ F_3^{(2)}\\ G_3^{(2)}\end{bmatrix}e^{3i\pi\sigma_1 t_0}+\text{c.c.}, \tag{65}$$

and substituting this into (57)–(62), we end up with

$$\frac{d^2\Phi_1^{(2)}}{dx^2}+k_1^2\Phi_1^{(2)}=\mathcal{B}\cos[\pi(x-1)]+\text{terms in } \cos[3\pi(x-1)], \tag{66}$$

with $k_1^2=(\pi\sigma_1)^2(1+\kappa s_1)=\pi^2$ and

$$\mathcal{B}=-\mu\frac{\partial\alpha}{\partial t_2}+i\frac{S}{\varepsilon^2}\alpha+iQ_0|\alpha|^2\alpha, \tag{67}$$

where μ, S $(=S_{re}+iS_{im})$, and Q_0 $(=Q_1+Q_2+Q_3)$ are defined as

$$\mu=2+2\kappa s_1\left(1+\frac{s_1}{\Omega}\right)=\frac{2}{\pi\sigma_1}\frac{dk}{d\sigma}\bigg|_{\sigma=\sigma_1}, \tag{68}$$

$$S_{re}=-S_{im}=-\sqrt{\frac{\pi\sigma_1}{2}}\left(\frac{\delta}{\sigma_1^2}+\frac{\kappa\delta_r s_1^2}{\Omega}\right), \tag{69}$$

$$Q_1=-\frac{\pi}{\sigma_1}\left\{\gamma-1+\kappa\sigma_1^2\left[-(\gamma-3)s_1+2(\gamma-2)s_2-(\gamma-1)s_1s_2-(\gamma-1)\frac{s_1^2s_2}{\Omega}\right]\right\}\mathcal{A}_0, \tag{70}$$

$$Q_2=\frac{\pi}{2\sigma_1}\left\{\gamma+1-\kappa\sigma_1^2\left[-\left(\gamma-3+\frac{1}{\sigma_1^2}\right)s_1+2(\gamma-2)s_2-(\gamma-1)s_1s_2-(\gamma-1)\frac{s_1^2s_2}{\Omega}\right]\right\}\mathcal{A}_2, \tag{71}$$

and

$$Q_3=-\frac{3\pi(\gamma+1)}{8\sigma_1^3}+\frac{\pi\kappa\sigma_1}{8}\left\{\left[6(2\gamma-3)+\frac{(\gamma-3)}{\sigma_1^2}\right]s_1-2(\gamma-2)\left(3+\frac{1}{\sigma_1^2}\right)s_2\right\}+\frac{\pi\kappa\sigma_1(\gamma-1)}{8}\left\{6(\gamma-2)s_1^2+\left(3+\frac{1}{\sigma_1^2}\right)s_1s_2-3(2\gamma-1)s_1^3+\left[2\left(3-\frac{1}{\sigma_1^2}\right)+24(\gamma-2)s_2+\left(3+\frac{1}{\sigma_1^2}\right)s_2-3(2\gamma-1)s_1^2-12(\gamma-1)\left(1+\frac{s_1}{\Omega}\right)s_1s_2\right]\frac{s_1^2}{\Omega}\right\}. \tag{72}$$

Here, note that the fractional derivative of minus half-order of the exponential function is reduced to the Fresnel integral[11] and given simply as

$$\frac{\partial^{-1/2}}{\partial t_0^{-1/2}}e^{i\pi\sigma_1 t_0}=\frac{(1-i)}{\sqrt{2\pi\sigma_1}}e^{i\pi\sigma_1 t_0}=(i\pi\sigma_1)^{-1/2}e^{i\pi\sigma_1 t_0}. \tag{73}$$

The derivative of three-half order is given by $-(1-i)\times(\pi\sigma_1)^2/\sqrt{2\pi\sigma_1}\exp(i\pi\sigma_1 t_0)\,[=(i\pi\sigma_1)^{3/2}\exp(i\pi\sigma_1 t_0)]$. It is found from (69) that since κ is small, the friction loss due to the throat wall becomes comparably small with the one due to the tube wall. In addition, when Ω is chosen large, it is made even smaller.

The solution $\Phi_1^{(2)}$ is obtained as

$$\Phi_1^{(2)}=\frac{\mathcal{B}}{2\pi}(x-1)\sin[\pi(x-1)]+\mathcal{C}_1\cos[\pi(x-1)]+\mathcal{C}_2\sin[\pi(x-1)]+\text{terms in}\cos[3\pi(x-1)], \tag{74}$$

where $\mathcal{C}_1$ and $\mathcal{C}_2$ are integration constants to be determined. By imposing the boundary conditions (63) and (64), it follows that while $\mathcal{C}_1$, i.e., the coefficient of the homogeneous solution, is left undetermined within the present framework of the theory,[12] $\mathcal{C}_2$ must vanish, and $\mathcal{B}=iE/\varepsilon^2$. The latter gives the equation governing the behavior of α as

$$i\mu\frac{\partial\alpha}{\partial t_2}+\frac{S}{\varepsilon^2}\alpha+Q_0|\alpha|^2\alpha=\frac{E}{\varepsilon^2}. \tag{75}$$

Here, we examine the value of Q_0 for a large value of Ω. Noting that $s_1=1/(1-1/\Omega)\approx1+1/\Omega+\cdots$ and $s_2=1/(1-4/\Omega)\approx1+4/\Omega+\cdots$ so that $s_2-s_1\ll1$, and $k_2^2=4\pi^2[1+\kappa(s_2-s_1)\sigma_1^2]$, it is found that $\mathcal{A}_2$ is much larger than the other terms and therefore Q_2 is dominant in Q_0. Approximating $\mathcal{A}_2$ to be

$$\mathcal{A}_2\approx\frac{\pi^2(\gamma+1+\kappa)}{(k_2^2-4\pi^2)}, \tag{76}$$

Q_0 may be evaluated as

$$Q_0\approx Q_2\approx\frac{\pi^3(\gamma+1+\kappa)^2}{2\sigma_1(k_2^2-4\pi^2)}>0. \tag{77}$$

Thus, it is found that Q_0 is positive for $\Omega\gg1$.

We proceed to complete the second-order problem by taking account of all terms including those unspecified so far. Although the calculations are straightforward, the full expressions for $\phi^{(2)}$, $f^{(2)}$, $g^{(2)}$ are too complicated to be reproduced here. By making use of the same approximation leading to (76), we present here only the leading expressions for the respective modes

$$\Phi_1^{(2)} \approx -\frac{i(\gamma-3-\kappa)}{16\pi\sigma_1}\mathcal{A}_2 \cos[3\pi(x-1)]|\alpha|^2\alpha, \tag{78}$$

$$\Phi_3^{(2)} \approx \frac{3i\pi(\gamma-3-\kappa)}{2\sigma_1(k_3^2-\pi^2)}\mathcal{A}_2 \cos[\pi(x-1)]\alpha^3 + \frac{3i\pi(\gamma+1+\kappa)}{2\sigma_1(k_3^2-9\pi^2)}\mathcal{A}_2 \cos[3\pi(x-1)]\alpha^3, \tag{79}$$

$$F_1^{(2)} \approx -\frac{i}{\pi\sigma_1}\frac{\partial\alpha}{\partial t_2}\cos[\pi(x-1)] + \frac{1}{2}\left(1-\frac{1}{\sigma_1^2}\right)\mathcal{A}_2 \times \cos[\pi(x-1)]|\alpha|^2\alpha - \left[\frac{(\gamma-3-\kappa)}{16} - \frac{1}{2}\left(1+\frac{1}{\sigma_1^2}\right)\right]\mathcal{A}_2 \cos[3\pi(x-1)]|\alpha|^2\alpha, \tag{80}$$

$$F_3^{(2)} \approx \left[\frac{9\pi^2(\gamma-3-\kappa)}{2(k_3^2-\pi^2)} + \frac{1}{2}\left(1+\frac{1}{\sigma_1^2}\right)\right]\mathcal{A}_2 \cos[\pi(x-1)]\alpha^3 + \frac{9\pi^2(\gamma+1+\kappa)}{2(k_3^2-9\pi^2)}\mathcal{A}_2 \cos[3\pi(x-1)]\alpha^3, \tag{81}$$

$$G_1^{(2)} \approx s_1 F_1^{(2)}, \tag{82}$$

$$G_3^{(2)} \approx s_3 F_3^{(2)}, \tag{83}$$

with $k_3^2=(3\pi\sigma_1)^2(1+\kappa s_3)\neq 9\pi^2$ and $s_3=\sigma_0^2/(\sigma_0^2-9\sigma_1^2)$, where the approximation (76) is used, and s_1 and s_3 in $G_1^{(2)}$ and $G_3^{(2)}$ are left unapproximated. Here, we assume that $k_3^2 \neq \pi^2$ and $\sigma_1 \neq \sigma_0/3$. Details will be discussed later.

B. Steady-state solution

Let us examine the steady-state solution to (75). Before doing this, we transform the equation into a form suitable for comparison with the results of measurements. So far ε has been used conveniently to make a note of ordering in developing the asymptotic analysis. From the viewpoint of the experiment, however, it is not a quantity to be measured easily, while use of α is not also suitable for comparison. Because an available quantity is the excess pressure in the tube, we rewrite (75) in terms of the dimensionless excess pressure relative to p_0. In view of (18), (42), and (44), we set

$$\frac{\rho_0 a_0 u_0 \overline{p}'}{p_0} = \varepsilon\gamma\overline{p}' = \frac{1}{2}\cos[\pi(x-1)]P e^{i\pi\sigma_1 t} + \text{c.c.} + \cdots, \tag{84}$$

where $P=2\varepsilon\gamma\alpha$ ($\ll 1$) to the lowest order. Noting that $\partial\alpha/\partial t=\varepsilon^2\partial\alpha/\partial t_2+\cdots$, and using P instead of α, (75) is rewritten as

$$i\mu\frac{\partial P}{\partial t} + SP + Q|P|^2P = \Gamma e^{i\pi\Delta\sigma t}, \tag{85}$$

with $Q=Q_0/4\gamma^2$, $\Gamma=2\pi\gamma\sigma_1 c$, and $\Delta\sigma=\varepsilon^2\sigma'$.

Setting $P=|P|\exp(i\theta)$ and separating the real and imaginary parts, it follows that

$$\mu\frac{\partial|P|}{\partial t} + S_{im}|P| = \Gamma\sin\psi, \tag{86}$$

and

$$\mu|P|\frac{\partial\psi}{\partial t} - \pi\mu\Delta\sigma|P| + S_{re}|P| + Q|P|^3 = \Gamma\cos\psi, \tag{87}$$

with $\psi=\pi\Delta\sigma t-\theta$. For the steady-state solution, we drop $\partial/\partial t$ to have $S_{im}|P|=\Gamma\sin\psi$, $\theta=\pi\Delta\sigma t+\text{constant}$, and

$$\Delta\sigma = \frac{1}{\pi\mu}\left(S_{re} + Q|P|^2 \pm \sqrt{\frac{\Gamma^2}{|P|^2} - S_{im}^2}\right). \tag{88}$$

It is found that $|P|$ is bounded at Γ/S_{im}. This suggests that the peak amplitude is proportional to the magnitude of excitation Γ, i.e., c. It is also found that the peak frequency is shifted downward by the dispersion due to the wall friction while shifted upward by nonlinearity, because $S_{re}<0$ and $Q>0$ for $\Omega\gg 1$. If $\Gamma\ll 1$ so that $|P|\ll 1$, we can neglect $Q|P|^2P$ in (85) to have immediately the steady-state solution as

$$P = \frac{\Gamma e^{i\pi\Delta\sigma t}}{S-\pi\mu\Delta\sigma}. \tag{89}$$

Thus, we obtain the linear resonance curve given by

$$|P| = \frac{\Gamma}{\sqrt{(\pi\mu\Delta\sigma - S_{re})^2 + S_{im}^2}}, \tag{90}$$

which is consistent with (88) in the limit as $P\to 0$.

Incidentally, (90) can also be derived from the linear theory shown previously.[4] When the lossy effects are included, the form of p' in (36) is unaltered but k^2 is modified as

$$k^2 = \pi^2\sigma^2\left\{\frac{1+\kappa\sigma_0^2/[\sigma_0^2-\sigma^2-(1-i)\delta_r\sigma^2/\sqrt{2\pi\sigma}]}{1-(1-i)\delta/\sqrt{2\pi\sigma}}\right\}. \tag{91}$$

See the expression for k below (4) in Ref. 4. But, note the difference in the sign of argument of the exponential function for the piston displacement. This difference yields $-i$ instead of $+i$. Expanding the right-hand side of (91) around $\sigma=\sigma_0$ by setting $\sigma=\sigma_1+\Delta\sigma$, we retain the lossless terms up to the order of $\Delta\sigma$ but truncate the lossy terms to the lowest order. Thus, k is obtained as $\pi+(\pi\mu\Delta\sigma-S)\sigma_1/2$. Next, noting that α in (44) corresponds to $(\pi^2\sigma^2/k\sin k)c/2\varepsilon$, we approximate $k\sin k$ as $-\pi(k-\pi)$ and take the lowest terms for α to recover (90) after multiplying it by $2\varepsilon\gamma$ in the relation between α and P.

V. DISCUSSION OF THE RESULTS

A. Nonlinear frequency response

The nonlinear frequency response in the steady state is described by (88) theoretically. The peak amplitude P_{peak} attained in the response curve is determined to be Γ/S_{im} by the linear loss, while the peak frequency at $P=P_{\text{peak}}$ is lowered by $\Delta\sigma=S_{re}/\pi\mu$ due to the dispersive effects of the boundary layer. This result is merely an extension of the linear case. Although the linear downshift is constant, the shift dependent on the amplitude occurs through $Q|P|^2P$, which is responsible for bending (skew) of the response curve toward the higher or lower frequency side. We examine variations of Q as σ_0 is changed with κ fixed. This corresponds physically to changing cavity volume and axial spacing, with the tube length, its cross-sectional area, and the throat length held constant.

From (70)–(72) with (51) and (52), it is expected that Q may diverge at (i) $|s_1|=\infty$; (ii) $|s_2|=\infty$; (iii) $k_2^2=0$; or (iv) $k_2^2=4\pi^2$. The first case (i) where $\sigma_1=\sigma_0$ does not occur unless $\kappa=0$ or $\sigma_0=0$, as is seen from Fig. 2(a) or (37). The second case (ii) with $2\sigma_1=\sigma_0$ may occur if the frequency of the second harmonics hits σ_0, whereby σ_0 and κ must satisfy the relation $\sigma_0=\sqrt{12/(3+4\kappa)}$. But, because of evanescence at σ_0, no divergence of Q occurs there. In fact, it is confirmed that all terms proportional to s_2 in Q cancel out altogether. The third case (iii) corresponds to the one with $2\sigma_1=\sqrt{1+\kappa}\sigma_0$ for resonance, and indeed occurs when σ_0 and κ satisfy the following relation:

$$\sigma_0=\sqrt{\frac{4(3-\kappa)}{3(1+\kappa)^2}}\equiv\sigma_{cr}<2. \tag{92}$$

The fourth case (iv) corresponds to the second harmonic resonance when $2\sigma_1$ hits $2\sigma^{\pm}$. But this case does not occur because $\sigma_1\neq 0$ nor $\sigma_1\neq 1$. Besides the third case, Q diverges in the limit as $\sigma_0\to 0$. This divergence is very rapid as $-\pi(\gamma-1)(2\gamma+1)/32\gamma^2\kappa^3\sigma_0^7$. In addition, we have assumed for the second-order solutions that $k_3^2\neq\pi^2$ or $\sigma_1\neq\sigma_0/3$. The equality $k_3^2=\pi^2$ holds for $\sigma_1=\sqrt{\sigma_0/3}$, where $\sigma_0=(5\pm\sqrt{16-9\kappa})/3(1+\kappa)$. The equality $\sigma_1=\sigma_0/3$ holds for $\sigma_0=\sqrt{72/(8+9\kappa)}$. Although these cases have no direct bearing on (85), they should be avoided in the sense of seeking uniformly valid, asymptotic solutions.

Taking account of such exceptional cases, Q is drawn versus σ_0 in Fig. 3(a) for three fixed values of κ. The broken, solid, and dotted lines represent the values of Q for $\kappa=0.1$, 0.2, and 0.3, respectively. As is seen, Q takes both positive and negative values. This means that the response curve may be bent rightward or leftward, depending on the choice of σ_0. While Q diverges at $\sigma_0=\sigma_{cr}$ and at $\sigma_0=0$, it grows as σ_0 increases. In this limit, the asymptotic expression (77) for Q_0 is already available so that Q is approximated as

$$Q=\frac{Q_0}{4\gamma^2}\approx Q_a=\frac{\pi^3(\gamma+1+\kappa)^2}{8\gamma^2\sigma_1(k_2^2-4\pi^2)}. \tag{93}$$

Figure 3(b) shows the relative error $|(Q-Q_a)/Q|$ for three values of $\kappa=0.1$, 0.2, and 0.3 by using the same types of lines as in Fig. 3(a). It is found that Q_a agrees with Q very well for $\sigma>2$ and the relative errors are almost less than 1%. Even below σ_{cr}, there is a region in which Q_a still gives a good approximation to Q.

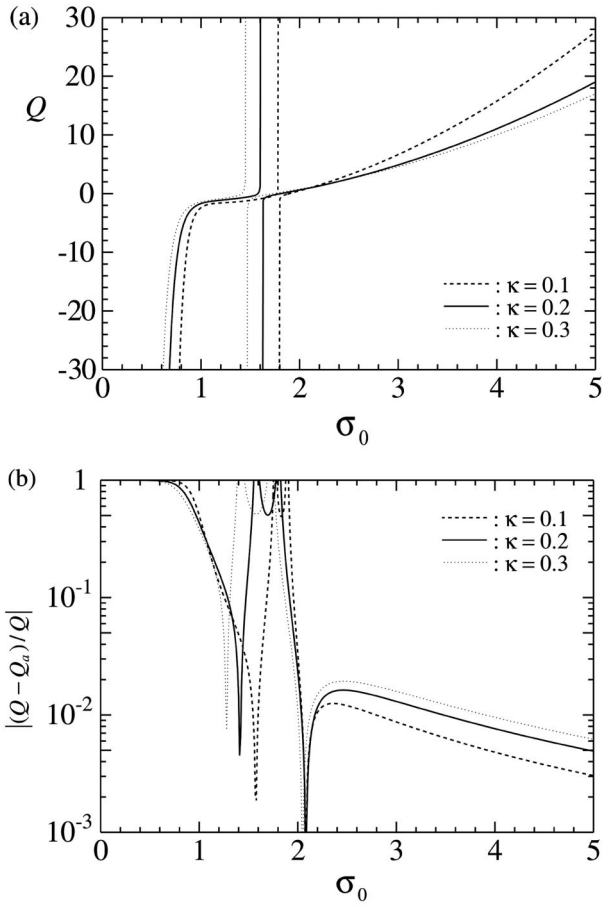


FIG. 3. Graphs of the coefficient Q versus σ_0 with κ fixed at 0.1, 0.2 and 0.3: (a) displays the values of Q where the respective values diverge at $\sigma_0=\sigma_{cr}\approx 1.79$, 1.61, and 1.46 and at $\sigma_0=0$, and (b) displays the relative errors $|(Q-Q_a)/Q|$ of the asymptotic values Q_a to Q for the three values of κ.

The other parameters μ and S are relatively simple. Figure 4 displays the graphs of μ and S_{im} versus σ_0 for $\kappa=0.1$, 0.2, and 0.3, where the meaning of D will be given later. Three families of curves labeled μ, S_{im}, and D are drawn. In each family, the broken, solid, and dotted lines represent the values of labeled quantity for $\kappa=0.1$, 0.2, and 0.3, respectively. In order to draw the graphs of S_{im}, the values of δ and δ_r are set to be 0.0282 and 0.223 in view of the experiment to be decribed in the next subsection. The value of δ_r is a little larger than the estimation. It is found that all quantities decrease as σ_0 increases, but they are almost constant for $\sigma>2$, and they decrease as κ becomes smaller.

B. Comparison with the experimental result

We now compare the frequency response with the one measured by using the tube of length l ($=3256$ mm) and of diameter $2R$ ($=80$ mm) with the Helmholtz resonator having a cavity of volume V ($=49.7\times10^{-6}$ m^3), and a throat of length L ($=35.6$ mm) and of diameter $2r$ ($=7.11$ mm) con-

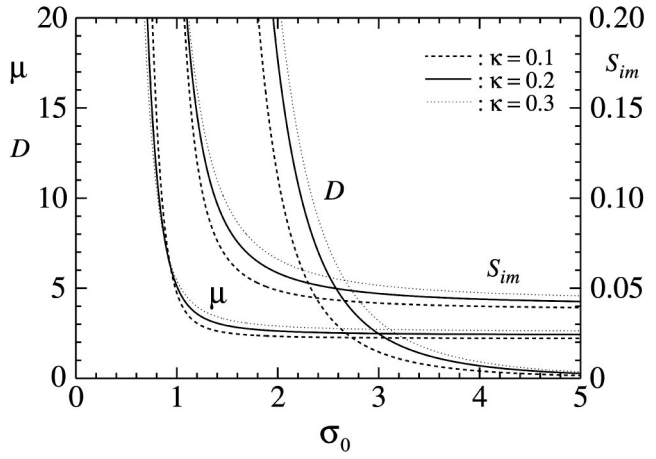


FIG. 4. Graphs of the coefficients μ, S_{im}, and D vs σ_0 for $\kappa=0.1$, 0.2, and 0.3.

nected with the axial spacing d ($=50$ mm). The size parameter κ takes the value 0.198. Figure 5 is the reproduction of the frequency response reported in Ref. 4 (see Fig. 4), where the half of the peak-to-peak value of the excess pressure measured on the flat plate at the closed end, δp, relative to the atmospheric pressure p_0 is drawn versus the frequency of excitation $\omega/2\pi$. The blank triangle, solid triangle, blank circle, and solid circle indicate the data measured at the displacement amplitude of the bellows $X_b=0.5$, 1.5, 2.5, and 3.5 mm, respectively. The equilibrium pressure and the mean temperature in the tube near the closed end are 1.007×10^5 Pa and 25.6 °C, respectively, so $a_0=347.1$ m/s, $\nu=1.567\times10^{-5}$ m^2/s, Pr$=0.7089$, and $\gamma=1.402$. At this temperature, the natural frequency of the resonator $\omega_0/2\pi$ is calculated to be 242.6 Hz with the end corrections. Thus, we find that $\sigma_0=4.55$ and $\sigma_1=0.911$.

Since $\kappa=0.198$ and $\sigma_0=4.55$ for the experiment, we have $\mu=2.43$, $Q=15.2$, $-S_{re}=S_{im}=0.0430$. In Fig. 5, the

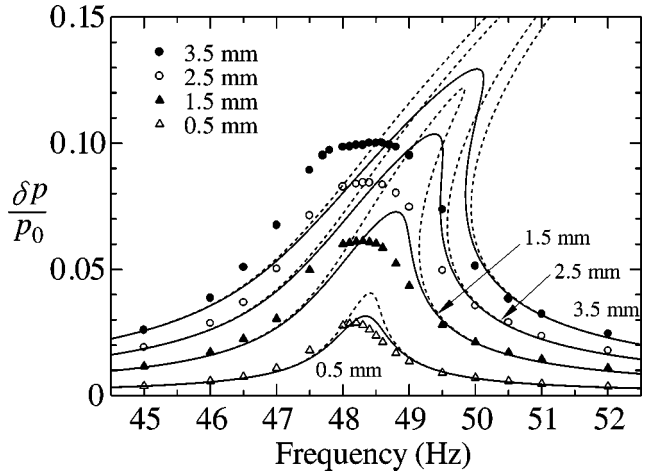


FIG. 5. Comparison of the nonlinear frequency response obtained by the theory and experiment for the displacement amplitude of the bellows $X_b \geqslant 0.5$ mm, where the broken and solid lines represent, respectively, the curves without taking account of the jet loss and with it, while the measured data are indicated by the blank triangle, solid triangle, blank circle, and solid circle for $X_b=0.5$, 1.5, 2.5, and 3.5 mm, respectively, X_b being related to the one of the piston by $X_p=1.422X_b$.

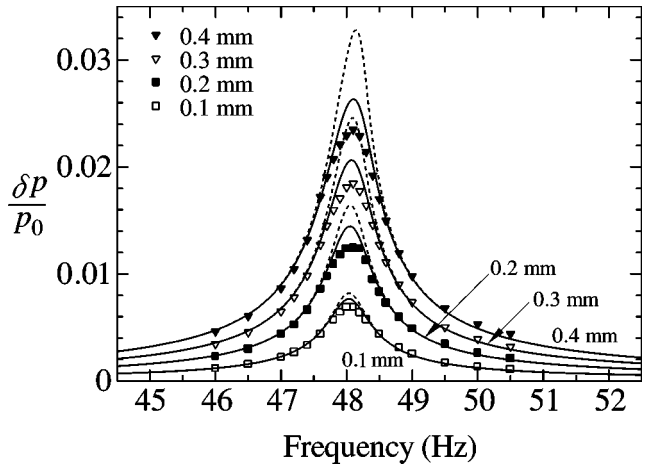


FIG. 6. Comparison of the nonlinear frequency response obtained by the theory and experiment for the displacement amplitude of the bellows $X_b<0.5$ mm, where the broken and solid lines represent, respectively, the curves without taking account of the jet loss and with it, while the measured data are indicated by the blank square, solid square, blank triangle, and solid triangle for $X_b=0.1$, 0.2, 0.3, and 0.4 mm, respectively, X_b being related to the one of the piston by $X_p=1.422X_b$.

broken lines represent $|P|$ versus the dimensional frequency $(\sigma_1+\Delta\sigma)a_0/2l$ by using (88). The displacement of the bellows is converted into the one of the piston by multiplying it with a factor 1.422,[4] so that Γ is given by $2\pi\gamma\sigma_1 X_p/l$ with $X_p=1.422X_b$. The theory overestimates not only the peak amplitudes but also the peak frequencies considerably, even for $X_b=0.5$ mm. When higher-order corrections to p' in $|P|$ on the right-hand side of (84) are taken into account, it is found that the deviation becomes worse. This is understood from the nature of asymptotic expansions that inclusion of higher-order terms does not necessarily improve the accuracy. Rather, the smallness of $|P|$ is required. Hence, it turns out that the theory based on (88) fails to describe the experimental results adequately.

So, the experimental data for X_b smaller than 0.5 mm are presented, though unpublished so far. Figure 6 shows the frequency response where the blank square, solid square, blank triangle, and solid triangle indicate the data measured at $X_b=0.1$, 0.2, 0.3, and 0.4 mm, respectively, and the equilibrium pressure and the mean temperature are 1.010×10^5 Pa and 23.1 °C, respectively. As in Fig. 5, the broken lines represent $|P|$ versus the dimensional frequency calculated by (88) where $a_0=345.6$ m/s, $\nu=1.539\times10^{-5}$ m^2/s, Pr$=0.7095$, and $\gamma=1.402$. While $\omega_0/2\pi$ is now shifted to be 241.3 Hz, note that both σ_0 and σ_1 are independent of a_0 and therefore of the temperature. It is seen in Fig. 6 that as X_b is decreased, the broken lines tend to approach the data measured. But, the theory overestimates the peak amplitude slightly even for the smallest excitation with $X_b=0.1$ mm.

C. Effect of jet loss

The discrepancy may be attributed to some effects not taken into account in the analysis. One is an effect of jet loss which occurs at the throat of the resonator. As the amplitude of oscillatory flows in the throat becomes large, the flow pattern (except for direction of flow) just entering the throat

differs from the one leaving the throat (see Fig. 1). When the air is sucked into the throat, it flows toward the orifice omnidirectionally, whereas when it is ejected, it flows in the form of jet unidirectionally. This asymmetry of flow pattern on the suction and ejection sides gives rise to the jet loss. The detailed analysis is given in the Appendix of Ref. 10 and is not reproduced here. The jet loss introduces the additional term

$$-\frac{V}{\rho_0 a_0^2 B L_e}\left|\frac{\partial p_c'}{\partial t}\right|\frac{\partial p_c'}{\partial t}, \tag{94}$$

to the right-hand side of (9) for the behavior of the resonator. Then, the normalized equation (24) is modified into

$$\frac{\partial^2 \overline{p_c'}}{\partial \overline{t}^2}+\delta_r\frac{\partial^{3/2}\overline{p_c'}}{\partial \overline{t}^{3/2}}+(\pi\sigma_0)^2\overline{p_c'}$$
$$=(\pi\sigma_0)^2\overline{p'}+\varepsilon\frac{(\gamma-1)}{2}\frac{\partial^2 \overline{p_c'}^2}{\partial \overline{t}^2}$$
$$-\varepsilon^2\frac{(\gamma-1)(2\gamma-1)}{6}\frac{\partial^2 \overline{p_c'}^3}{\partial \overline{t}^2}-\varepsilon\delta_J\left|\frac{\partial \overline{p_c'}}{\partial \overline{t}}\right|\frac{\partial \overline{p_c'}}{\partial \overline{t}}, \tag{95}$$

where $\delta_J=V/BL_e=(l/\pi\sigma_0 L_e)^2$. Since the new term is accompanied by ε, it yields additional terms in the first-order solution. In fact, it follows that

$$\left|\frac{\partial g^{(0)}}{\partial t_0}\right|\frac{\partial g^{(0)}}{\partial t_0}=-4(\pi\sigma_1 s_1)^2(|\sin\xi|\sin\xi)\{|\cos[\pi(x-1)]|$$
$$\times\cos[\pi(x-1)]\}|\alpha|^2, \tag{96}$$

with $\alpha=|\alpha|e^{i\theta}$ and $\xi=\pi\sigma_1 t_0+\theta$. Using the following Fourier expansions:

$$|\sin\xi|\sin\xi=-\frac{2i}{\pi}\sum_{n=-\infty}^{\infty}\frac{[(-1)^n-1]}{n(n^2-4)}e^{in\xi}$$
$$=-\frac{4i}{3\pi}e^{i\xi}+\frac{4i}{15\pi}e^{3i\xi}+\cdots+\text{c.c.}, \tag{97}$$

and

$$|\cos[\pi(x-1)]|\cos[\pi(x-1)]$$
$$=-\frac{4}{\pi}\sum_{n=-\infty}^{\infty}\frac{\sin(n\pi/2)}{n(n^2-4)}e^{in\eta}$$
$$=\frac{8}{3\pi}\cos[\pi(x-1)]+\frac{8}{15\pi}\cos 3[\pi(x-1)]+\cdots, \tag{98}$$

with $\pi x=\eta$ and the sums taken except for $n=0$ and $n=\pm 2$, there arises in (46) the term proportional to $\cos[\pi(x-1)]\exp(i\pi\sigma_1 t_0)|\alpha|\alpha$. Note that although $\cos[\pi(x-1)]$ is defined only in the interval $0\leqslant x\leqslant 1$ originally, it is extended periodically to the outside of it with period 2 and so is the case with $|\cos[\pi(x-1)]|$. Thus, the solution (49) must include the component $\Phi_1^{(1)}\exp(i\pi\sigma_1 t_0)$. As is inferred from the analysis for the second-order problem, this term gives rise to the equation for $\Phi_1^{(1)}$ corresponding to (66) with $\mathcal{B}$ proportional to $|\alpha|\alpha$. Applying the boundary conditions (48), it follows that α must vanish. To remedy this, it may be considered to introduce another time variable $t_1(\equiv\varepsilon t)$. But, as $\Delta\sigma$ and E are assumed to be of order ε^2, the introduction of t_1 is inconsistent.

To resolve the dilemma, we notice the following point. The parameter κ takes a small value such as 0.1 or 0.2, but it has been regarded as a quantity of order unity and independent of ε. As the magnitude of ε becomes comparable with the value of κ, however, the smallness of the latter may be taken into account. If κ is taken as small as ε, the contribution from the jet loss had rather be incorporated in the second-order problem. Then, $\mathcal{B}$ in (67) includes the additional term given by

$$-\frac{128\kappa\delta_J\sigma_1^3 s_1^3}{9\pi\varepsilon\sigma_0^2}|\alpha|\alpha, \tag{99}$$

and (75) is replaced by

$$i\mu\frac{\partial\alpha}{\partial t_2}+\frac{S}{\varepsilon^2}\alpha+i\frac{D_0}{\varepsilon}|\alpha|\alpha+Q_0|\alpha|^2\alpha=\sigma_1\frac{E}{\varepsilon^2}, \tag{100}$$

with $D_0=128\kappa\delta_J\sigma_1^3 s_1^3/9\pi\sigma_0^2$. In terms of P, finally, (85) is modified as

$$i\mu\frac{\partial P}{\partial t}+SP+iD|P|P+Q|P|^2P=\Gamma e^{i\pi\Delta\sigma t}, \tag{101}$$

with

$$D=\frac{64\kappa\delta_J\sigma_1 s_1^3}{9\pi\gamma\Omega}. \tag{102}$$

In Fig. 4, the graph of D vs σ_0 is displayed for $\kappa=0.1$, 0.2, and 0.3.

Separating the real and imaginary parts of (101) by setting $P=|P|\exp(i\theta)$, the only correction to (86) and (87) is to replace S_{im} with $S_{im}+D|P|$. Then, the frequency response is given by

$$\Delta\sigma=\frac{1}{\pi\mu}\left[S_{re}+Q|P|^2\pm\sqrt{\frac{\Gamma^2}{|P|^2}-(S_{im}+D|P|)^2}\right]. \tag{103}$$

It is found that the jet loss adds $D|P|$ to the linear one S_{im}, and then the peak amplitude is determined as

$$P_{\text{peak}}=\frac{1}{2D}(-S_{im}+\sqrt{S_{im}^2+4D\Gamma}). \tag{104}$$

For $\Gamma\ll S_{im}^2/4D$, P_{peak} is given by Γ/S_{im}, whereas for $\Gamma\gg S_{im}^2/4D$, P_{peak} is given by $\sqrt{\Gamma/D}$. The latter suggests that the peak amplitude does not increase in proportion to Γ as in the linear regime but to $\sqrt{\Gamma}$. The changeover may be defined to occur at $\Gamma=S_{im}^2/4D\equiv\Gamma_{ch}$, which gives $\Gamma_{ch}=0.0012$ for $D=0.398$. This corresponds to 0.34 mm for X_b.

The response curves taking account of the jet loss are drawn in the solid lines in Fig. 5. For the amplitude 0.5 mm, good agreement is observed for the peak amplitude but the peak frequency is slightly overestimated. The latter is determined by the accuracy of measurement of the temperature. If the temperature is higher by 1 degree, then the peak frequency is shifted by 0.09 Hz to the right. For the amplitude smaller than 0.5 mm, the solid lines in Fig. 6 shows the

response curves with the jet loss. In this case, the peak frequencies are well predicted. Although the peak amplitudes are overestimated slightly, it is found that account of the jet loss iproves the results considerably.

As the amplitude increases beyond 0.5 mm, however, there arises significant discrepancy. It may be attributed to the following two points. Although the frequency of the second harmonics never hits σ_2^-, the ratio σ_2^-/σ_1 is 1.98 and very close to 2 numerically. Even if no shock then occurs, this situation is not preferable to the theory. Secondly, the frequency of the fifth harmonics, $5\sigma_1$, hits σ_0 for evanescence in the present case. Then the energy absorption at the fifth harmonics is enhanced so that the theory might be limited. Furthermore if causes of the discrepancy are sought beyond the present framework, it is open whether or not the weakly nonlinear theory is still applicable. In addition, another effect of transverse motions of air in the cross-section might come into play, associated with emergence of acoustic streaming. To make a firm conclusion, however, it is necessary to check validity of the theory against new experiments by using tubes in geometrically different configuration taking account of some restrictions.

VI. CONCLUSIONS

The weakly nonlinear theory has been developed to obtain the frequency response of forced oscillations of the air column in the closed tube with the array of Helmholtz resonators connected axially. Since the array introduces the wave dispersion, shock-free, high-amplitude oscillations are obtainable, and for their analysis, the usual asymptotic method of multiple is successfully applicable. The analysis is based on the assumptions of one-dimensional motions in the acoustic main flow outside of the boundary layer on the tube wall, and of the continuum approximation of the discrete distribution of the resonators. Thus, no acoustic streaming involving transverse motions of air in the cross section is covered, but the steady but axially nonuniform pressure distribution of quadratic order is revealed. By completing the second-order problem, the cubic nonlinear equation describing temporal slow modulation of the complex pressure amplitude of the lowest mode has been derived.

In the linear regime, as is well known, the small dissipation and dispersion caused by the boundary layer on the tube and throat wall limit the peak amplitude and lower the peak frequency. As the amplitude of excitation is increased, the nonlinear frequency shift occurs, depending on the amplitude of oscillations and on the natural frequency of the resonator. The shift is crucially determined by the coefficient Q in the equation. It takes positive or negative values so that the response curve may be bent toward the high- or low-frequency side. But, it is usual that Q is positive and the response curve is bent toward the high-frequency side, because the natural frequency of resonator is chosen higher than the critical frequency σ_{cr}. When the theoretical curve is compared with the experiment, it is revealed that both results do not agree with each other quantitatively, although qualitatively the tendency of the bent may be explained. In particular, the theory overestimates the peak amplitude considerably.

The discrepancy has been resolved to some extent by taking account of the jet loss at the throat. It introduces the additional nonlinear dissipation to the linear one so that the peak amplitude is suppressed as the amplitude becomes large. It is found that the peak amplitude then becomes proportional to the square root of the amplitude of excitation. As far as the present experiment is concerned, the linear and weakly nonlinear theory with the jet loss can give a quantitative description up to the pressure amplitude 3% to the equilibrium pressure. In order to design a tube for generation of even higher peak amplitude, the results in the present paper provide useful guidelines.

ACKNOWLEDGMENTS

The authors acknowledge the support by the Grants-in-Aid from the Japan Society of Promotion of Science and also from The Mitsubishi Foundations, Tokyo, Japan.

[1] A. B. Coppens and A. A. Atchley, ''Nonlinear standing waves in cavities,'' in *Encyclopedia of Acoustics*, edited M. J. Crocker (Wiley, New York, 1997), Vol. 1, pp. 237–247.

[2] C. C. Lawrenson, B. Lipkens, T. S. Lucas, D. K. Perkins, and T. W. Van Doren, ''Measurements of macrosonic standing waves in oscillating closed cavities,'' J. Acoust. Soc. Am. **104**, 623–636 (1998).

[3] Y. A. Ilinskii, B. Lipkens, T. S. Lucas, T. W. Van Doren, and E. A. Zabolotskaya, ''Nonlinear standing waves in an acoustical resonator,'' J. Acoust. Soc. Am. **104**, 2664–2674 (1998).

[4] N. Sugimoto, M. Masuda, T. Hashiguchi, and T. Doi, ''Annihilation of shocks in forced oscillations of an air column in a closed tube,'' J. Acoust. Soc. Am. **110**, 2263–2266 (2001).

[5] N. Sugimoto, ''Emergence of an acoustic shock wave in a tunnel and a concept of shock-free propagation,'' in *Noise and Vibration from High-Speed Trains*, edited V. V. Krylov (Thomas Telford, London, 2001), pp. 213–247.

[6] O. V. Rudenko, ''Artificial nonlinear media with a resonant absorber,'' Sov. Phys. Acoust. **29**, 234–237 (1983).

[7] V. G. Andreev, V. E. Gusev, A. A. Karabutov, O. V. Rudenko, and O. A. Sapozhnikov, ''Enhancement of the Q of a nonlinear acoustic resonator by means of a selectively absorbing mirror,'' Sov. Phys. Acoust. **31**, 162–163 (1985).

[8] V. E. Gusev, H. Bailliet, P. Lotton, S. Job, and M. Bruneau, ''Enhancement of the Q of a nonlinear acoustic resonator by active suppression of harmonics,'' J. Acoust. Soc. Am. **103**, 3717–3720 (1998).

[9] J. Kevorkian and J. D. Cole, *Multiple Scale and Singular Perturbation Methods* (Springer, New York, 1996).

[10] N. Sugimoto, ''Propagation of nonlinear acoustic waves in a tunnel with an array of Helmholtz resonators,'' J. Fluid Mech. **244**, 55–78 (1992).

[11] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).

[12] Although C_1 should be determined, in nature, by solving an initial-value problem from start-up, it may be set equal to zero if the combination $\alpha - \varepsilon^2 i\pi\sigma_1 C_1$ on substitution of the solutions obtained into (42) is replaced by α newly. Then, (75) [and (100)] are rewritten in terms of this new α but their form is subjected to no alterations within the order concerned.

Two-dimensional streaming flows induced by resonating, thin beams

Tolga Açıkalın, Arvind Raman,[a] and Suresh V. Garimella
School of Mechanical Engineering, Purdue University, West Lafayette, Indiana 47907-2088

(Received 3 April 2003; revised 14 July 2003; accepted 28 July 2003)

Miniaturized resonating slender beams are finding increased applications as fluidic actuators for portable electronics cooling. Piezoelectric and ultrasonic ''fans'' drive a flexural mode of the beam into resonance thus inducing a streaming flow, which can be used to cool microelectronic components. This paper presents analytical, computational, and experimental investigations of the incompressible two-dimensional streaming flows induced by resonating thin beams. Closed-form analytical streaming solutions are presented first for an infinite beam. These are used to motivate a computational scheme to predict the streaming flows from a baffled piezoelectric fan. Experiments are conducted to visualize the asymmetric streaming flows from a baffled piezoelectric fan and the experimental results are found to be in close agreement with the predicted results. The findings are expected to be of relevance in the optimal design and positioning of these solid-state devices in cooling applications. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1610453]

PACS numbers: 43.25.Nm [MFH]

I. INTRODUCTION

The demand for smaller, more functional and portable electronics devices has resulted in a large number of powerful electronic components being packed into smaller spaces. Conventional methods of heat removal such as heat sinks and rotational fans are being increasingly challenged by the stringent space, power, and noise requirements of the next-generation portable electronics systems. Piezoelectric fans have recently emerged as low power, noiseless alternatives to conventional cooling technologies. A piezoelectric fan consists of a thin, cantilevered beam with a bonded patch of piezoelectric material.[1] Two prototype piezoelectric fans are shown in Fig. 1. Under an alternating voltage that matches the resonance frequency of the structure, the fan flaps back and forth like a Japanese fan. The flexural motions then drive fluid flows in the domain, thus effectively cooling the electronics with minimal power consumption. To ensure noiseless operation, these fans can be designed for infrasonic ($<$100 Hz) or ultrasonic ($>$25 kHz) frequencies, outside the audible range for humans.

The basic characteristics of heat removal and fluid mechanics of these devices have been presented by several authors. Toda[2,3] conducted experiments with piezoelectric fans that were fabricated from PVF_2 (poly vinyl difluoride) sheets. In these experiments airflow velocity measurements and thermal characterization of the piezoelectric fans were performed. Yoo *et al.*[4] developed several types of piezoelectric fans made from a PZT (Lead Zirconium Titanate) and investigated the variation with fan geometry and material choice, of the tip displacement and the induced air velocity. Schmidt[5] determined the local and average transfer coefficients on a vertical surface cooled by two piezoelectric (PVF_2) fans vibrating simultaneously out of phase. Mass transfer rates in terms of the nondimensional local Sherwood numbers were obtained using the naphthalene sublimation technique. In a recent study by Acıkalın *et al.*,[1] piezoelectric fans were found to offer enhancements in convective heat transfer coefficients of more than 100% relative to natural convection in compact spaces, with low power consumption and minimal noise. However the orientation and location of the piezoelectric fan were also found to play a crucial role on its cooling capability. None of the above works have addressed in detail the flows generated by the piezoelectric fan. Understanding the flow induced by these fans is an essential and important step in designing piezoelectric fans as effective thermal management devices.

The prediction of streaming flow over vibrating beams and cantilevers is of substantial importance in a number of other areas. For example, the knowledge of the streaming flows induced by vibrating beams and cantilevers is crucial in microcantilevers used in atomic force microscopy[6] and for the design of flexural plate wave MEMS pumps.[7] Resonant beams have also been used as propulsion devices for the development of flying insect robots[8] and as ''meso-flaps'' for the modification of boundary layers in aircraft wings.

The term ''acoustic streaming'' is used to describe the nonzero time-averaged velocity encountered in oscillating fluids. Specifically, this phenomenon is observed with two sources of oscillations: acoustic (quartz wind) and boundary (steady) streaming. In quartz wind generation, a high frequency acoustic beam penetrates a body of fluid and streaming flow is observed due to the spatial attenuation of this wave in the fluid medium. In the latter case, boundary streaming is generated due to the friction between the fluid and a solid boundary. Examples of boundary streaming include a standing wave interacting with a solid boundary, or a solid boundary oscillating in an otherwise quiescent fluid. These two cases of boundary streaming can be treated in a similar fashion as stated by Lighthill.[9] The streaming flows produced over resonating beams and piezoelectric fans fall

[a]Electronic mail: raman@ecn.purdue.edu

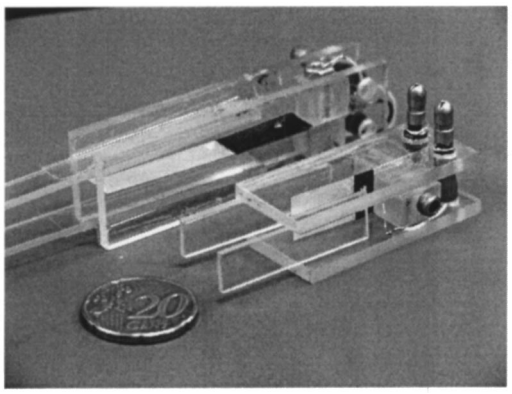


FIG. 1. Sample prototype piezoelectric fans (Ref. 1).

into this category. Recent reviews of acoustic streaming are available in Riley[10] and Nyborg.[11]

Fluid flows induced by oscillating rigid bodies such as cylinders,[12–15] concentric cylinders,[16,17] square cylinders,[18] and spheres[19] have been investigated extensively. The assumption of a rigid body enables an easy coordinate transformation to solve the boundary streaming problem. Such an approach is not generally valid for the vibrations of deformable bodies such as resonant beams and piezoelectric fans. In particular deformable bodies resonate in specific modes, which generate nonuniform velocity boundary conditions, rendering the streaming problem considerably more complicated.

Only a few authors have investigated the streaming flows induced by oscillating flexible bodies. Nyborg[20] developed an approximate solution for sonically induced steady flow. He simplified Schlichting's model[15] with an approach valid for the flow in the vicinity of the surface when the acoustic boundary layer parameter $\sqrt{\nu/\omega}$ (where ν is the kinematic viscosity and ω the angular frequency) is small relative to the oscillation velocity. Nyborg studied a number of applications, including a vibrating flexible cylindrical shell. Kildal[21] considered time-dependent as well as time-independent (streaming) flow over a vibrating flexible plate in an infinite fluid. The vibrating plate was assumed to be infinite, making possible the separation of the spatial coordinates in the momentum equation.

Loh *et al.*[22] investigated the cooling effects of acoustic streaming from an ultrasonically vibrating beam. In this study, Nyborg's theory[20] was used to generate predictions for the streaming flow, against which incompressible CFD simulations of this beam were compared. Numerical results were found to be 10 to 100 times greater than those predicted from theory. Wan and Kuznetsov[23] investigated the acoustic streaming induced by a baffled vibrating beam in the vicinity of a fixed standing plate in a compressible, viscous fluid. First and second order perturbation equations are solved numerically using finite differences. Ihara and Watanabe[24] investigated the two-dimensional flows around the ends of oscillating flexible cantilevers. The discrete vortex method, in conjunction with the singularity method, was used to numerically simulate the flow field. Flow visualization experiments were conducted to validate the numerical simulation. In each cycle of oscillation, two vortices were found to be shed from

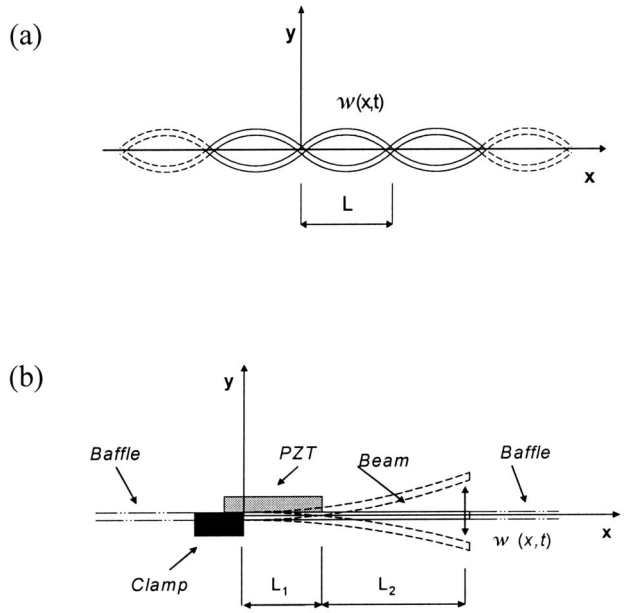


FIG. 2. Schematic representation of (a) an infinite beam, and (b) a baffled piezoelectric cantilever, vibrating in a two-dimensional incompressible, viscous fluid.

the tip of the cantilever, and these vortices traveled downstream of the tip. However their computations focused on the unsteady flows generated by large amplitude flapping of the fans. The steady streaming flows induced by such fans were not discussed explicitly.

Due to the stringent space constraints in portable electronic devices, piezoelectric fans are driven at small vibration amplitudes, and steady acoustic streaming is an important component of the flow in these applications. The present work focuses on the two-dimensional flows induced by infinite and finite, baffled resonating thin beams. The method of successive approximations is first used to derive analytical closed-form solutions to the streaming flow induced by an infinite resonating beam. To the authors' knowledge, this is the first presentation of such results. The flows induced by a finite length, baffled piezoelectric fan are then investigated. No closed-form solutions are possible in this case and a computational scheme is developed to solve the resulting perturbation equations. The computational method is shown to be able to resolve the asymmetric nature of the streaming field generated in the near and far fields of the fan. Flow visualizations obtained using a baffled commercial piezoelectric fan are then compared to the predicted solution and found to be in close agreement.

II. THEORY

A. System description

The first problem is that of an infinite, slender beam vibrating in an infinite, two-dimensional, incompressible viscous fluid [Fig. 2(a)]. The second problem deals with the streaming flow induced by a vibrating baffled, slender piezoelectric fan [Fig. 2(b)]. In both cases, the structure undergoes primarily transverse (flexural) vibrations. If $w(x,t)$ is the

TABLE I. Piezoelectric fan parameters.

Piezoelectric patch material	PZT
Piezoelectric patch thickness	0.05 cm
Piezoelectric patch length	2.9 cm
Shim material	Mylar
Shim thickness	0.025 cm
Length of the piezoelectric fan	6.5 cm
Width of the piezoelectric fan	1.3 cm
Resonance frequency	60 Hz
Vibration amplitude at the tip	0.08 cm

transverse vibration of the neutral axis of the slender structure in each case, the infinite beam supports harmonic standing wave motions, wherein the small transverse vibration velocity is given by

$$\dot{w}(x,t)=\mathrm{Re}\left[A\omega\cdot\sin\left(\frac{\pi\cdot x}{L}\right)\cdot e^{i\cdot\omega\cdot t}\right] \tag{1}$$

in which A is the vibration amplitude where $A=O(\epsilon)$ and $|\epsilon|\ll 1$. Furthermore ω is the vibration frequency and L is the half wavelength of the harmonic standing wave. For the subsequent computations for the infinite beam problem, $L=6.5$ cm, A$=0.08$ cm, and $\omega=120\pi$ rad/s, are chosen as the parameter values.

As opposed to the infinite beam, the piezoelectric fan has specific modal frequencies. A slender beam model of the piezoelectric fan was presented in Basak *et al.*[25] From Ref. 25 the first mode shape (resonance frequency of 60 Hz) of a commercial piezoelectric fan with dimensions/material properties shown in Table I is given by

$$\begin{aligned}G_1(x)&=a_1e^{-b_1x}+a_2e^{b_1x}+a_3\cos(b_1x)\\&\quad+a_4\sin(b_1x),\quad 0\leqslant x\leqslant L_1,\\G_2(x)&=a_5\cos(b_2x)+a_6\cosh(b_2x)+a_7\sin(b_2x)\\&\quad+a_8\sinh(b_2x),\quad L_1\leqslant x\leqslant L_2,\end{aligned} \tag{2}$$

where L_1 and L_2 are shown in Fig. 2(b), and the constants in Eq. (2) are

$$\begin{aligned}&&a_1&=0.072\,861,&a_5&=-0.4742,\\L_1&=0.029,&b_1&=32.8682,&a_2&=0.015\,744,&a_6&=1.7902,\\L_2&=0.065,&b_2&=49.4542,&a_3&=-0.0886,&a_7&=-0.3694,\\&&&&a_4&=0.057\,11,&a_8&=-1.755\,77.\end{aligned} \tag{3}$$

Accordingly, the transverse vibration velocity of the piezoelectric fan is given by

$$\dot{w}(x,t)=\begin{cases}\mathrm{Re}\{A\omega e^{i\omega t}G_1(x)\}, & 0\leqslant x\leqslant L_1\\ \mathrm{Re}\{A\omega e^{i\omega t}G_2(x)\}, & L_1\leqslant x\leqslant L_2\end{cases}. \tag{4}$$

B. Method of successive approximations

The governing momentum equation for an incompressible, two-dimensional flow is

$$\begin{aligned}&(\nabla^2\Psi)_{,t}+\Psi_{,y}(\nabla^2\Psi)_{,x}-\Psi_{,x}(\nabla^2\Psi)_{,y}=\nu\nabla^4\Psi,\\&\nabla^4\Psi\equiv\Psi_{,xxxx}+2\Psi_{,xxyy}+\Psi_{,yyyy},\end{aligned} \tag{5}$$

where $\Psi(x,y,t)$ is the stream function and ν is the kinematic viscosity. The compressibility of the fluid can be neglected[26] because for the problems under consideration, $\omega L/c\cong 0.07\ll 1$, where c is the speed of sound in air and ω and L are defined in Sec. II A. To solve for the streaming flow induced by the vibrating infinite beam and baffled piezoelectric fan, the method of successive approximations is used,[12]

$$\Psi(x,y,t)=\Psi_1(x,y,t)+\Psi_2(x,y,t)+\Psi_3(x,y,t)+\cdots. \tag{6}$$

Noting that the vibration velocities of the thin beams are of $O(\epsilon)$, the flow velocities induced in the surrounding fluid are also assumed to be of $O(\epsilon)$ to the leading order. Accordingly, $\Psi_i=O(\epsilon^i)$, $i=1,2,...$, and substitution of Eq. (6) into Eq. (5), and collection of terms of like order leads to the following linearized equations:

$$\begin{aligned}&O(\epsilon^1):\ (\nabla^2\Psi_1)_{,t}=\nu\nabla^4\Psi_1,\\&O(\epsilon^2):\ F(x,y,t)+(\nabla^2\Psi_2)_{,t}=\nu\nabla^4\Psi_2,\\&F(x,y,t)=\Psi_{1,y}(\nabla^2\Psi_1)_{,x}-\Psi_{1,x}(\nabla^2\Psi_1)_{,y}.\end{aligned} \tag{7}$$

While the $O(\epsilon^1)$ equation is homogeneous, the $O(\epsilon^2)$ equation contains a nonhomogeneous forcing term that depends on the $O(\epsilon^1)$ solution. Furthermore, the second order solution is decomposed into steady and unsteady components

$$\Psi_2(x,y,t)=\Psi_2^u(x,y,t)+\Psi_2^s(x,y). \tag{8}$$

In Eq. (7) Ψ_1 is purely harmonic in time, with the same frequency as the vibrating structure. Ψ_2^u is also purely harmonic in time, but with twice the frequency as the vibrating structure. Last, Ψ_2^s is the steady streaming flow. Time averaging the $O(\epsilon^2)$ equation, Eq. (8), yields

$$\begin{aligned}\nu\nabla^4\Psi_2^s&=\langle F(x,y,t)\rangle\\&=\langle\Psi_{1,y}(\nabla^2\Psi_1)_{,x}-\Psi_{1,x}(\nabla^2\Psi_1)_{,y}\rangle,\end{aligned} \tag{9}$$

where $\langle\cdot\rangle=(\omega/2\pi)\int_0^{(2\pi/\omega)}dt$ is the time averaging operator evaluated over the fundamental time period of the structural vibration.

C. Boundary conditions at the vibrating surface

Assuming a no-slip condition on the vibrating surface, the velocity of the beam matches the fluid velocity on the surface,

$$\Psi_{,y}(x,w,t)=0,\quad \Psi_{,x}(x,w,t)=-\dot{w}(x,t). \tag{10}$$

These nonlinear boundary conditions are evaluated on the deformed boundary at $y=w$. This fact is often neglected in the literature on boundary streaming which generally assumes rigid body vibrations. Under the assumption of small-amplitude deflections, $w=O(\epsilon)$, these boundary conditions can be simplified to the boundary conditions at the undeflected surface location $y=0$. Taylor series expansion of Eq. (10) about $y=0$ yields

$$\begin{aligned}\Psi_{,y}(x,w,t)&=\Psi_{,y}(x,0,t)+w\Psi_{,yy}(x,0,t)\\&\quad+\frac{w^2}{2!}\Psi_{,yyy}(x,0,t)+O(\epsilon^4)=0,\end{aligned} \tag{11}$$

$$\Psi_{,x}(x,w,t)=\Psi_{,x}(x,0,t)+w\Psi_{,xy}(x,0,t)+\frac{w^2}{2!}\Psi_{,xyy}(x,0,t)+O(\epsilon^4)=-\dot{w}(x,t). \tag{12}$$

Substitution of Eq. (6) into Eqs. (11) and (12), and collection of terms of like order leads to the following linearized boundary conditions:

$$O(\epsilon^1)\begin{cases}\Psi_{1,y}(x,0,t)=0,\\ \Psi_{1,x}(x,0,t)=-\dot{w}(x,t),\end{cases} \tag{13}$$

$$O(\epsilon^2)\begin{cases}\langle\Psi_{2,y}(x,0,t)\rangle=\langle -w(x,t)\Psi_{1,yy}(x,0,t)\rangle,\\ \langle\Psi_{2,x}(x,0,t)\rangle=\langle -w(x,t)\Psi_{1,xy}(x,0,t)\rangle=0.\end{cases} \tag{14}$$

These linearized boundary conditions are then appended to the first and second order equations, Eq. (7) for the stream function. In addition to the boundary conditions at $y=0$, the stream function at each order of perturbation needs to satisfy the farfield conditions which require that Ψ_1, Ψ_2 and their gradients vanish as $y\to+\infty$ and $x\to\pm\infty$.

III. INFINITE BEAM

For some special cases, the streaming flow solution is separable in the x and y coordinates and an exact analytical solution for the streaming flow is possible. One such case is the vibrating infinite beam. In what follows, the exact analytical solution for this case is presented. Substitution of a separable form of the first order stream function,

$$\Psi_1(x,y,t)=\mathrm{Re}\left\{Y_1(y)\cos\left(\frac{\pi x}{L}\right)e^{i\omega t}\right\} \tag{15}$$

into Eq. (7) leads to one fourth order ordinary differential equation for the unknown $Y_1(y)$. The form of Ψ_1 is chosen such that it satisfies the boundary conditions at $y=0$. The general solution for $Y_1(y)$ can be written as $c_1e^{\lambda_1^1 y}+c_2e^{\lambda_2^1 y}+c_3e^{\lambda_3^1 y}+c_4e^{\lambda_4^1 y}$ where the constants c_i and λ_i^1 are determined by the linearized boundary conditions at $y=0$, Eq. (13), and the farfield conditions at $y=\infty$,

$$\lambda_1^1=-\frac{\pi}{L},\quad \lambda_2^1=-\sqrt{\frac{\pi^2}{L^2}+\frac{i\omega}{\nu}},$$

$$c_1=\frac{AL\omega}{\pi}-\frac{A\omega}{\frac{\pi}{L}-\sqrt{\frac{\pi^2}{L^2}+\frac{i\omega}{\nu}}}, \tag{16}$$

$$c_2=\frac{A\omega}{\frac{\pi}{L}-\sqrt{\frac{\pi^2}{L^2}+\frac{i\omega}{\nu}}},$$

and $c_3=c_4=0$.

Figure 3 shows the streamlines from the analytical first order solution at $t=0$ s for $\nu=1.45\times10^{-5}$ m^2/s, $L=6.5$ cm, $A=0.08$ cm, and $\omega=120\pi$ rad/s. At this instant, the beam is at its undeflected position moving in the positive y direction, and the fluid is expelled at the antinode $x=L/2$. The fluid is either expelled or drawn in at the antinode, depending on the direction of the beam motion. Moreover the first order flow structure is symmetric about the antinode. Further in Figs. 4(a) and (b), the horizontal and vertical components, (u_1,v_1), respectively, of the first order

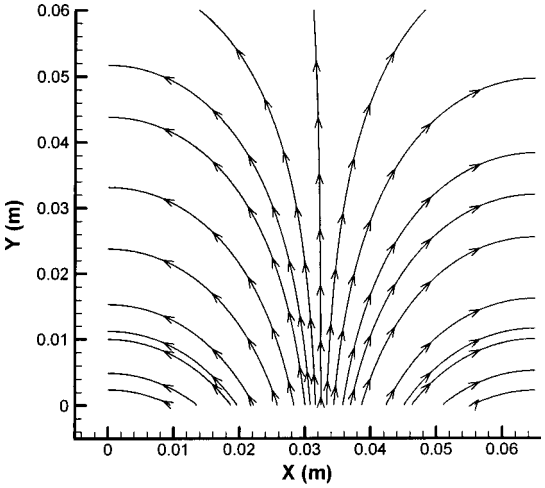


FIG. 3. Streamlines for the first order solution [Eq. (15)] of the infinite beam vibrating at 60 Hz with an amplitude of 0.08 cm at $t=0$ s.

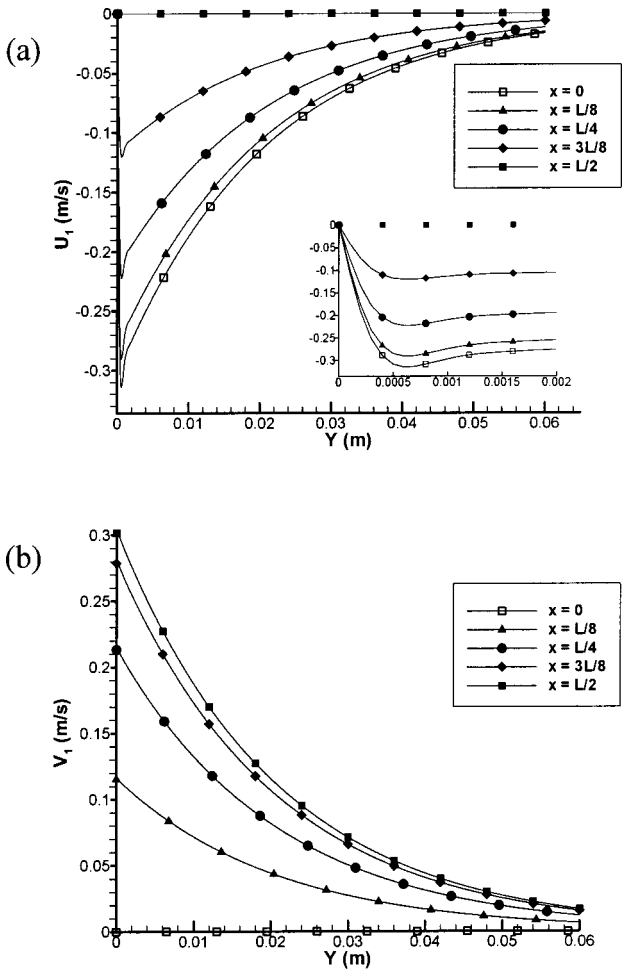


FIG. 4. First order (a) u- and (b) v-velocity profiles at various x locations induced by an infinite beam vibrating at 60 Hz with an amplitude of 0.08 cm at $t=0$ s.

velocity are plotted as a function of y at different x locations for $t=0$ s. The velocity component u_1 is zero at $y=0$ (no slip condition), reaches a maximum at around $y=0.0006$ m and vanishes as $y\to\infty$. u_1 is maximized at the nodes while v_1 reaches a maximum over the antinode. At $y=0$ v_1 equals the transverse velocity of the vibrating beam and eventually vanishes in the farfield as $y\to\infty$.

The solution for Ψ_1 is substituted into the term $F(x,y,t)$ in Eq. (9). The consequent nonhomogeneous equation for Ψ_2 is solved using separation of variables. The solution consists of a homogeneous and a particular solution and is given by

$$[c_1 e^{\lambda_1^2 y}+c_2 e^{\lambda_2^2 y}+c_3 y e^{\lambda_1^2 y}+c_4 y^2 e^{\lambda_1^2 y} + e^{\lambda_3^2 y}(c_5\cos(\beta y)+c_6\sin(\beta y))]\sin\left(\frac{2\pi x}{L}\right), \tag{17}$$

where

$$\lambda_1^2=-\frac{2\pi}{L},\quad \lambda_3^2=-\sqrt{\frac{1}{2}\left(\frac{\pi^2}{L^2}+\sqrt{\frac{\pi^4}{L^4}+\frac{\omega^2}{\nu^2}}\right)}-\frac{\pi}{L},$$

$$\lambda_2^2=-\sqrt{2\left(\frac{\pi^2}{L^2}+\sqrt{\frac{\pi^4}{L^4}+\frac{\omega^2}{\nu^2}}\right)},$$

$$\beta=-\sqrt{\frac{1}{2}\left(-\frac{\pi^2}{L^2}+\sqrt{\frac{\pi^4}{L^4}+\frac{\omega^2}{\nu^2}}\right)}.$$

Expressions for $c_1\cdots c_6$ are very involved and complex. Four of these constants are computed in Mathematica[27] from the boundary conditions at $y=0$ and the farfield conditions. The remaining two constants are calculated from the nonhomogeneous solution.

Streamlines of the second order time-averaged flow are illustrated in Fig. 5(a) and a close-up view near the boundary is shown in Fig. 5(b). Concurring with the previous literature, fluid is drawn in at the nodal points and expelled at the antinodes. Moreover, four distinct circular flow patterns form within $0\leqslant x\leqslant L$. On the left-hand side of the antinode, $0\leqslant x\leqslant L/2$, the inner streaming circulation is clockwise and the outer streaming circulation is counterclockwise. Conversely, on the other side, $L/2\leqslant x\leqslant L$, similar but opposite circulations are observed. Figure 6(a) shows the variation of u_2 as a function of y at different x locations. The y values corresponding to zero u_2 are the centers of the circulations. The centers of the inner circulations are at approximately $y=0.0002$ m and the centers of the outer circulations are at approximately $y=0.011$ m. The centers of circulation (inner and outer) are located either at $x=L/4$ or at $x=3L/4$. In Fig. 6(b) the y variation of v_2 is illustrated. At $y=0.0008$ m v_2 is zero, and it reaches a maximum at approximately $y=0.0002$ m. One important feature that may be noted is that the inner streaming is confined to the region swept by the vibrating beam, $y=0$ to 0.0008 m.

The streaming results also yield important information on the unsteady and streaming flow boundary layers in the problem.[26] The first order problem involves only normal excitation at the boundary at $y=0$, while the second order problem involves only shearing motions at $y=0$. Because of the absence in the first order problem of shearing excitation at $y=0$, the unsteady boundary layer thickness (Stokes layer) is controlled by two decay constants λ_1^1 and λ_2^1 [Eq. (16)]. The streaming boundary layer however is governed by three decay constants λ_1^2, λ_2^2, and λ_3^2. Because the real parts of all three second order constants are larger in magnitude than that of first order, the streaming boundary layer in air is smaller than the unsteady second order boundary layer (Stokes layer).

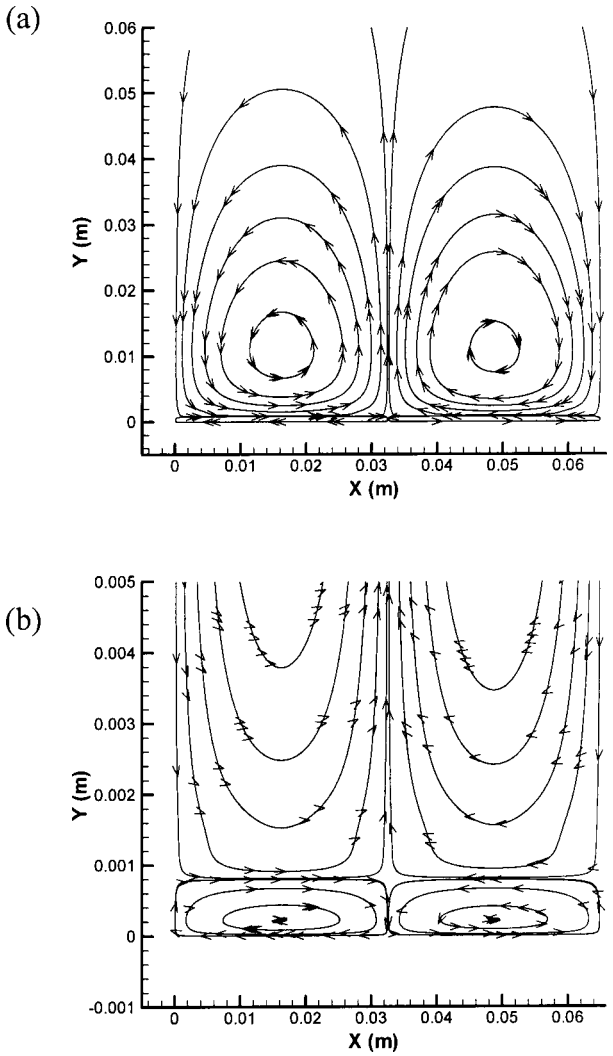


FIG. 5. Streamlines for the time-averaged second order solution [Eq. (17)] of the infinite beam vibrating at 60 Hz with an amplitude of 0.08 cm, (a) entire domain, and (b) region close to the vibrating beam. Inner circulation is confined to the region of vibration.

IV. BAFFLED PIEZOELECTRIC FAN

The streaming flow around a vibrating piezoelectric fan involves complex mass transport between the top and bottom half spaces. In order to simplify the problem a baffled piezoelectric fan resonating in its first mode is modeled. All computational results are presented for a piezoelectric fan with properties listed in Table I. Further the tip vibration amplitude is chosen to be 0.08 cm to match the amplitude encountered in the experiment. Solving for a baffled piezoelectric

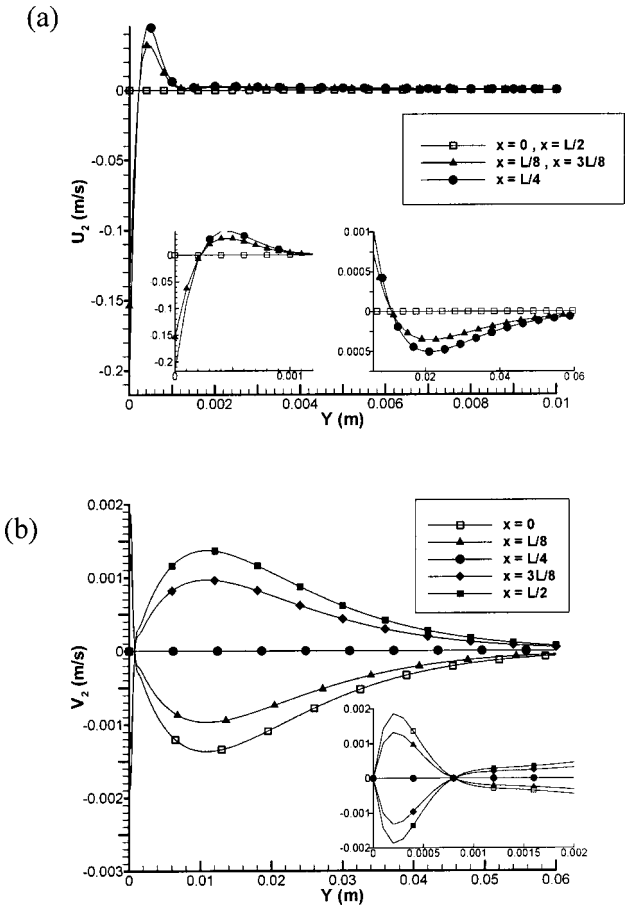


FIG. 6. Time-averaged second order (a) u- and (b) v-velocity profiles at various x locations, induced by an infinite beam vibrating at 60 Hz with an amplitude of 0.08 cm.

fan not only simplifies the boundary condition at $y=0$ but also allows restriction of the solution to a half-space.

From Eq. (4) the boundary conditions on the fluid are given by

$$\mathbf{v}(x,0,t)=\dot{w}(x,t)=\begin{cases}0, & x\leqslant 0,\\ A\omega e^{i\omega t}G_1(x), & 0\leqslant x\leqslant L_1,\\ A\omega e^{i\omega t}G_2(x), & L_1\leqslant x\leqslant L_2,\\ 0, & x>L_2\end{cases}$$
$$=A\omega G(x)e^{i\omega t}. \tag{18}$$

Because the piezoelectric excitations are assumed to be purely harmonic,

$$\Psi_1(x,y,t)=\overline{\Psi}_1(x,y)e^{i\omega t}. \tag{19}$$

Substitution of Eq. (19) into Eqs. (7) and (13), and Fourier transformation with respect to x of the resulting equations and the boundary conditions, yields

$$iw(-\alpha^2\tilde{\Psi}_1(\alpha,y)+\tilde{\Psi}_{1,yy}(\alpha,y))$$
$$=\nu(\alpha^4\tilde{\Psi}_1(\alpha,y)-2\alpha^2\tilde{\Psi}_{1,yy}(\alpha,y)+\tilde{\Psi}_{1,yyyy}(\alpha,y), \tag{20}$$

$$\tilde{\Psi}_{1,y}(\alpha,0)=0, \quad \tilde{\Psi}_1(\alpha,0)=\frac{i\omega\tilde{G}(\alpha)}{\alpha}, \tag{21}$$

where $\tilde{G}(\alpha)$ is the Fourier transform of $G(x)$. The corresponding farfield conditions are $\tilde{\Psi}_1(\alpha,\infty)=0$, $\tilde{\Psi}_{1,y}(\alpha,\infty)=0$. Solution of Eq. (20) is facilitated by the introduction of $\tilde{\Psi}_1(\alpha,y)=c(\alpha)e^{\lambda y}$. Upon solving for the characteristic multipliers,

$$\lambda_1=\alpha, \quad \lambda_2=\frac{i\sqrt{-\alpha^2\nu-i\omega}}{\sqrt{\nu}},$$

$$\lambda_3=-\alpha, \quad \lambda_4=-\frac{i\sqrt{-\alpha^2\nu-i\omega}}{\sqrt{\nu}}.$$

Further

$$\tilde{\Psi}_1(\alpha,y)=c_1(\alpha)e^{\lambda_1 y}+c_2(\alpha)e^{\lambda_2 y}+c_3(\alpha)e^{\lambda_3 y}+c_4(\alpha)e^{\lambda_4 y}, \tag{22}$$

where the constants $c_1\cdots c_4$ are computed using the boundary conditions. It can be shown that for $\alpha>0$, $c_1=c_2=0$, and for $\alpha<0$, $c_2=c_3=0$. After solving for the constants in Eq. (22) velocity components (u_1,v_1) are computed directly in the Fourier domain.

However, the complexity of the Fourier integrals necessitates the use of numerical techniques in the evaluation. As a consequence an appropriate integration range of α needs to be defined in which the improper integrals

$$u_1=\mathrm{Re}\left\{\Psi_{1,y}(x,y)e^{i\omega t}=\frac{e^{i\omega t}}{2\pi}\int_{-\infty}^{\infty}\tilde{\Psi}_{1,y}(\alpha,y)e^{i\alpha x}\partial\alpha\right.$$
$$\left.\approx\frac{e^{i\omega t}}{2\pi}\int_{-\alpha_1}^{\alpha_1}\tilde{\Psi}_{1,y}(\alpha,y)e^{i\alpha x}\partial\alpha\right\},$$
$$v_1=\mathrm{Re}\left\{-\Psi_{1,x}(x,y)e^{i\omega t}=-\frac{e^{i\omega t}}{2\pi}\int_{-\infty}^{\infty}i\alpha\tilde{\Psi}_1(\alpha,y)e^{i\alpha x}\partial\alpha\right. \tag{23}$$
$$\left.\approx-\frac{e^{i\omega t}}{2\pi}\int_{-\alpha_1}^{\alpha_1}i\alpha\tilde{\Psi}_1(\alpha,y)e^{i\alpha x}\partial\alpha\right\}$$

may be approximated suitably. The value of α_1 in the numerical computation is based on the comparison of the numerical first order solution at $y=0$ with the imposed boundary condition. Indeed the exact first order solution at $y=0$ is known from the imposed boundary condition Eq. (18). However the imposed boundary condition for the baffled piezofan features a discontinuity at $x=0.065$ m and representation of such a velocity field using a finite Fourier transform will lead inevitably to errors in the finite integral representation near this discontinuity (Gibb's phenomenon). Nonetheless, since the velocities are often at a maximum at $y=0$, it is assumed that the choice of α_1 based on matching exact velocity profiles at $y=0$ is valid throughout the fluid domain. In order to determine $u_1(x,y)$ and $v_1(x,y)$, the numerical inverse Fourier transform needs to be evaluated pointwise at each desired (x,y) location for the range of $-\alpha_1\leqslant\alpha\leqslant\alpha_1$. In this case, a domain size of $-0.15\,\mathrm{m}<x<0.2\,\mathrm{m}$ and $0<y<0.15\,\mathrm{m}$ and $\alpha_1=10\,000$ is selected to compute $u_1(x,y)$ and $v_1(x,y)$. This domain size is estimated iteratively and is also based on flow visualization experiments. The domain is divided into several zones with uniform meshes of different mesh densities, the densest being closest to the tip, in a man-

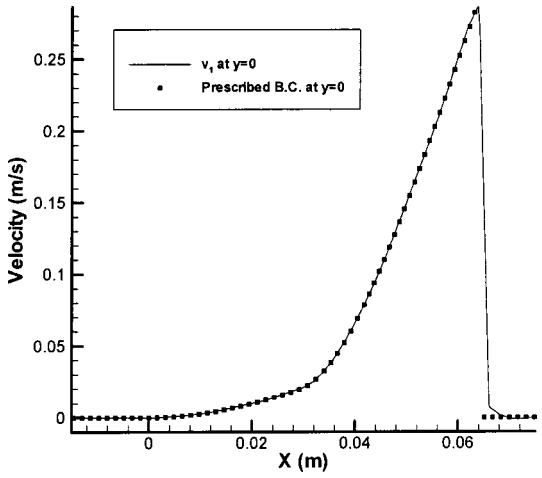


FIG. 7. Comparison of first order v-velocity component at $y=0$ with given boundary condition, Eq. (18), for the piezoelectric fan at $t=0$ s.

ner that is sufficient to resolve the key features of the flow. One advantage of this method is its parallel processing capability: $u_1(x,y)$ and $v_1(x,y)$ can be computed at every point in the domain independently of other points. Figure 7 shows a comparison of v_1 at $y=0$ with the given velocity boundary condition, Eq. (18). This demonstrates that the wave number range $\alpha_1=10\,000$ provides a least square error of less than 1% between the approximate solution and exact solution at $y=0$.

Figure 8 illustrates the first order velocity field at $t=0$ s when the fan is at its undeformed position and traveling in the positive y direction. It is observed that the first order velocity field oscillates with the same frequency as the piezoelectric fan, albeit with a slightly different phase. This phase depends on the location of the point of interest. It vanishes near the vibrating beam and is small but remains finite far from the beam. The first order velocity is dominated by v rather than u as can be seen from Figs. 9(a) and (b). This is a feature that is also common to the infinite beam problem discussed earlier. Moreover after approximately $y=0.08$ m, the x dependence of v becomes very small. From Fig. 9(a), it is seen that u changes sign slightly before the tip

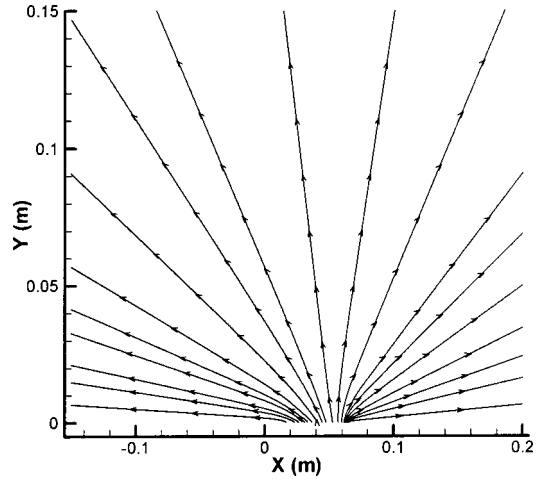


FIG. 8. Streamlines for the first order solution of a 6.5 cm long piezoelectric fan vibrating at 60 Hz with an amplitude of 0.08 cm at $t=0$ s.

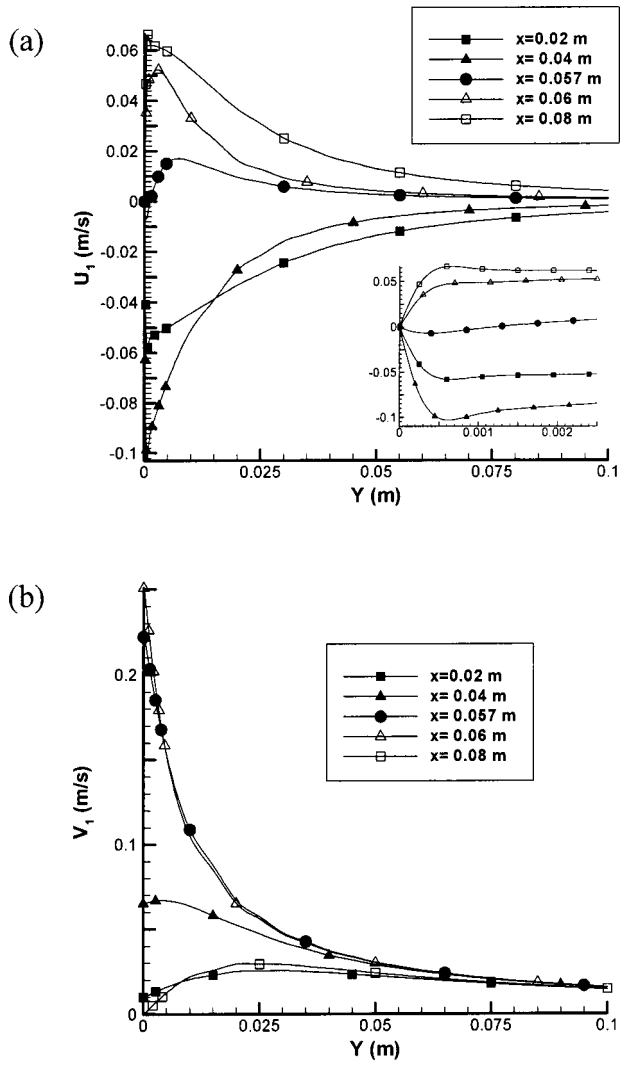


FIG. 9. First order (a) u- and (b) v-velocity profiles at various x locations induced by a 6.5 cm long piezoelectric fan, vibrating at 60 Hz with an amplitude of 0.08 cm at $t=0$ s [Eq. (23)].

at approximately $x=0.055$ m. For each wave number in the x direction, two distinct eigenvalues exist in y. Hence the induced velocity decreases away from the fan with two decay constants resulting in double boundary layers (inner and outer circulation regions).

The second order solution is now computed using the first order results computed above. Accordingly, boundary conditions for the fourth order equation at $y=0$ are

$$\tilde{\bar{\Psi}}_{2,y}(\alpha,0)=-\int_0^L \langle w(x,t)\Psi_{1,yy}(x,0,t)\rangle e^{-I\alpha x}\partial x,$$
$$\tilde{\bar{\Psi}}_2(\alpha,0)=0, \tag{24}$$

where an overbar denotes time averaging. The other farfield conditions are $\bar{\Psi}_{2,x}(x,\infty,t)=0$ and $\bar{\Psi}_{2,y}(x,\infty,t)=0$. Taking the Fourier transform of Eq. (9) on x to eliminate the partial derivates yields

$$\nu(\tilde{\bar{\Psi}}_{2,yyyy}(\alpha,y)-2\alpha^2\tilde{\bar{\Psi}}_{2,yy}(\alpha,y)+\alpha^4\tilde{\bar{\Psi}}_2(\alpha,y))=\tilde{\bar{F}}(\alpha,y). \tag{25}$$

The homogenous solution for Eq. (25) is given by

$$\tilde{\tilde{\Psi}}_2^h(\alpha,y)=c_1e^{-\alpha y}+c_2ye^{-\alpha y}+c_3e^{\alpha y}+c_4ye^{\alpha y}. \quad (26)$$

The particular solution $\tilde{\tilde{\Psi}}_2^p(\alpha,y)$ of Eq. (25) is computed numerically for a given α value in Mathematica using the following initial conditions:

$$\tilde{\tilde{\Psi}}_2(\alpha,y_1)=\tilde{\tilde{\Psi}}_{2,y}(\alpha,y_1)=\tilde{\tilde{\Psi}}_{2,yy}(\alpha,y_1)=\tilde{\tilde{\Psi}}_{2,yyy}(\alpha,y_1)\sim 0, \quad (27)$$

where $y_1=0.15\,\mathrm{m}$. For this choice of initial conditions $\tilde{\tilde{\Psi}}_2^p \to 0$ as $y\to\infty$. The total solution $\tilde{\tilde{\Psi}}_2^h(\alpha,y)+\tilde{\tilde{\Psi}}_2^p(\alpha,y)$ is then subjected to the boundary conditions, Eq. (24), and the farfield conditions $\tilde{\tilde{\Psi}}_2(\alpha,\infty)=0$ and $\tilde{\tilde{\Psi}}_{2,y}(\alpha,\infty)=0$ to yield the constants c_i. Depending on the sign of α, two of the four constants of the total solution vanish on the application of the farfield conditions. For $\alpha>0$ the total solution becomes

$$\begin{aligned}\tilde{\tilde{\Psi}}_2^+(\alpha,y)=&-\tilde{\tilde{\Psi}}_2^p(\alpha,0)e^{-\alpha y}-\alpha\tilde{\tilde{\Psi}}_{2,y}^p(\alpha,0)\\&-\tilde{\tilde{\Psi}}_2^p(\alpha,0)ye^{-\alpha y}+\tilde{\tilde{\Psi}}_2^p(\alpha,y)\\&-\int_0^L\langle w(x,t)\Psi_{1,yy}(x,0,t)\rangle e^{-I\alpha x}\partial x. \quad (28)\end{aligned}$$

Similarly for $\alpha<0$ the total solution becomes

$$\begin{aligned}\tilde{\tilde{\Psi}}_2^-(\alpha,y)=&-\tilde{\tilde{\Psi}}_2^p(\alpha,0)e^{\alpha y}+\alpha\tilde{\tilde{\Psi}}_{2,y}^p(\alpha,0)-\tilde{\tilde{\Psi}}_2^p(\alpha,0)ye^{\alpha y}\\&+\tilde{\tilde{\Psi}}_2^p(\alpha,y)\\&-\int_0^L\langle w(x,t)\Psi_{1,yy}(x,0,t)\rangle e^{-I\alpha x}\partial x. \quad (29)\end{aligned}$$

Finally, for the special case of $\alpha=0$ the total solution becomes

$$\begin{aligned}\tilde{\tilde{\Psi}}_2(\alpha,y)=&-\tilde{\tilde{\Psi}}_2^p(\alpha,0)-\tilde{\tilde{\Psi}}_2^p(\alpha,0)y+\tilde{\tilde{\Psi}}_2^p(\alpha,y)\\&-\int_0^L\langle w(x,t)\Psi_{1,yy}(x,0,t)\rangle e^{-I\alpha x}\partial x. \quad (30)\end{aligned}$$

The second order velocity components (u_2,v_2) are then computed using the relation

$$\begin{aligned}u_2(x,y)=&\frac{1}{2\pi}\left(\int_{-\alpha_2}^0\tilde{\tilde{\Psi}}_2^-(\alpha,y)_{,y}e^{i\alpha x}\,d\alpha\right.\\&\left.+\int_0^{\alpha_2}\tilde{\tilde{\Psi}}_2^+(\alpha,y)_{,y}e^{i\alpha x}\,d\alpha\right),\\v_2(x,y)=&-\frac{1}{2\pi}\left(\int_{-\alpha_2}^0 i\alpha\tilde{\tilde{\Psi}}_2^-(\alpha,y)e^{i\alpha x}\,d\alpha\right.\\&\left.+\int_0^{\alpha_2}i\alpha\tilde{\tilde{\Psi}}_2^+(\alpha,y)e^{i\alpha x}\,d\alpha\right), \quad (31)\end{aligned}$$

where the improper integral has been approximated again by a finite wave number range integral. The choice of a proper α_2 is made using the same criterion as in the first order solution. The choice of $\alpha_2=3000$ was found to match accurately the boundary conditions at $y=0$ and is used in the subsequent computations.

Following the selection of an appropriate α_2, the complete second order velocity field is computed in the chosen domain and streamlines are plotted in Fig. 10. A counter-

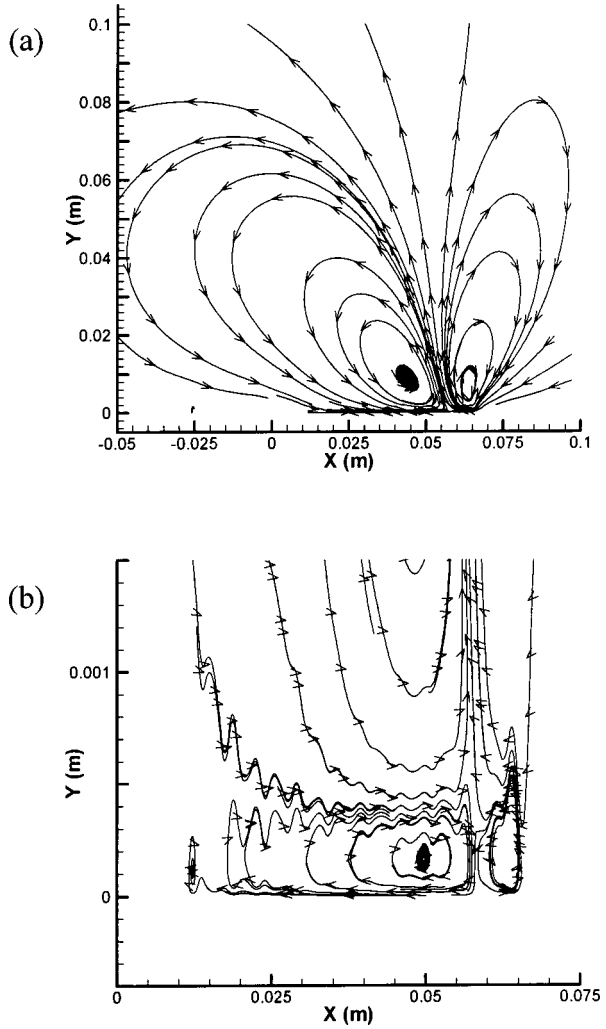


FIG. 10. Streamlines for the time-averaged second order solution of the infinite beam vibrating at 60 Hz with an amplitude of 0.08 cm, (a) entire domain, and (b) region close to the vibrating beam. In the case of a baffled piezoelectric fan, inner streaming is not uniformly confined to the region of vibration unlike the case of the infinite beam.

clockwise outer circulation is seen in the region before the tip, and a clockwise outer circulation is seen in the region after the tip. Two inner circulations with opposite directions are observed near $y=0$. The inner circulation nearer the clamp of the fan is confined to a smaller y domain compared to the circulation region on the right. Recall that in the infinite beam case, the thickness of the inner circulation region is uniform, equaling the vibration amplitude. The streaming u_2 and v_2 velocity profiles for several x locations are illustrated in Fig. 11.

In general, the features of the streaming flow field are similar to the infinite beam problem. Nevertheless, the following features distinguish the piezoelectric fan streaming flow from that generated by the infinite beam. The streaming flow field is asymmetric and the circulation centers are not located symmetrically with respect to the midpoint of the fan. Moreover the centers of both the inner and outer circulation on the left have different x-coordinates. The center of the counterclockwise outer streaming zone is at (0.044 m, 0.085 m). The other circulation zone, associated with inner streaming, on the left-hand side of the fan is clockwise and

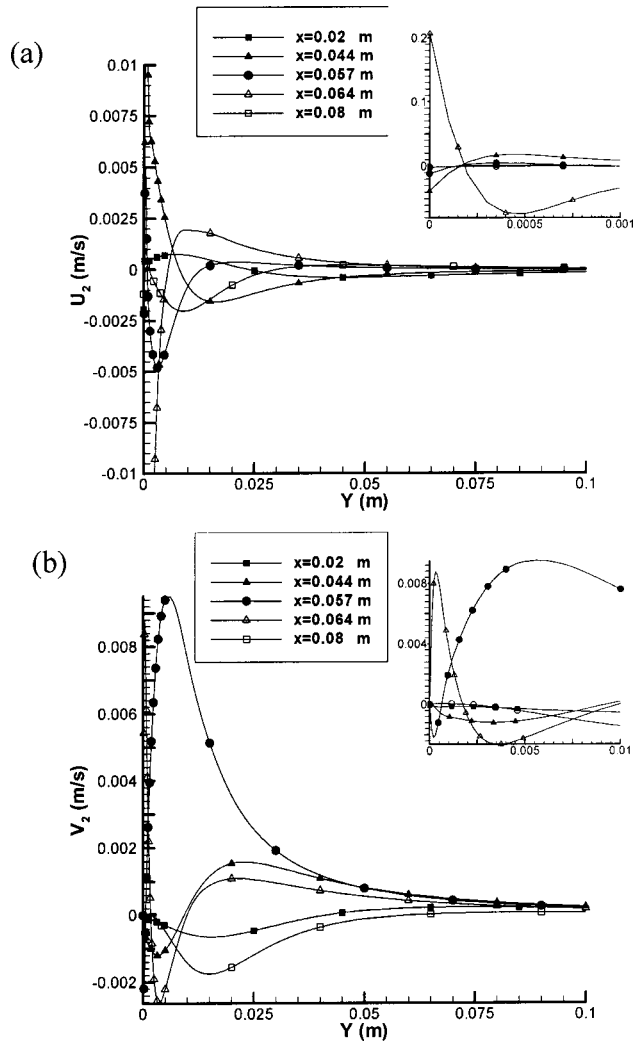


FIG. 11. Time-averaged second order (a) u- and (b) v-velocity profiles at various x locations, induced by a 6.5 cm long piezoelectric fan, vibrating at 60 Hz with an amplitude of 0.08 cm [Eq. (31)].

its center is at approximately (0.049 m, 0.00013 m). The centers of the clockwise outer circulation and counterclockwise inner circulations are approximately at (0.064 m, 0.06 m) and (0.064 m, 0.0002 m), respectively. For the baffled piezoelectric fan, the thicknesses of the inner circulation regions are nonuniform and vary with x. Unlike the infinite beam, the maximum streaming velocity is not observed at an antinode. The maximum v_2 in the outer circulation region occurs around $x=0.057$ m and not at the fan tip, $x=0.065$. Indeed, both the inner and outer circulations on the right-hand side have centers located very close to $x=0.065$. While this result may be due in part to the inaccuracies in Fourier representation near the vibrating tip, experimental results presented below also verify this phenomenon. Finally, the streamline corresponding to the maximum v is inclined to the left. This is consistent with the general asymmetric nature of the streaming induced by the baffled piezoelectric fan.

Experimental setup and comparison against numerical solution: Experiments were conducted to visualize and compare the two-dimensional flow patterns resulting from the resonant vibration of a baffled piezoelectric fan. The piezoelectric fan is sandwiched in between a black matte background and a transparent plexiglass sheet as shown in Fig. 12. The domain size of 90 cm by 80 cm adequately simulates the infinite domain problem considered in the theoretical analysis, with the walls in the experiment being far enough from the fan to have minimal effects on the flow field. The fan is located at the center as shown in Fig. 12 and a baffle is placed along the axis of the fan spanning the full length of the setup. An Ar-ion laser and a fiberoptically coupled cylindrical lens are used for the flow visualization, with smoke particles employed as tracers. The laser sheet illuminates the set up from the side and is parallel to the top plexiglass surface. All four sides of the setup are blocked to prevent smoke leakage and to minimize disturbances from the outside. A digital video camera is placed above the setup to capture the scattered light from the smoke particles inside the setup. The camera frame rate (30 fps) is less than the vibration frequency (60 Hz), resulting in an automatic experimental time averaging of the flow field. Still images extracted from the video therefore provide a good basis for comparison of the real and computed streaming flow patterns.

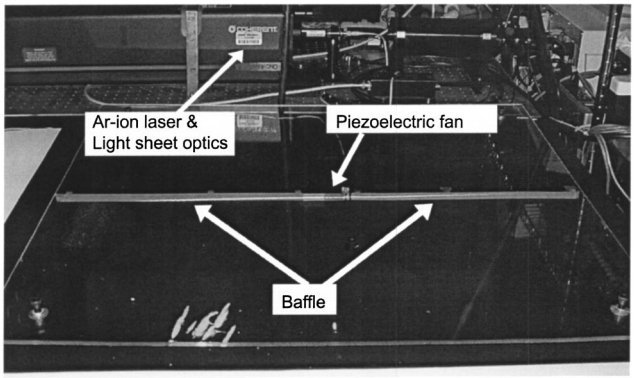


FIG. 12. Experimental setup used in the flow visualization of a baffled piezoelectric fan.

Another inherent assumption in this comparison is that the streamlines and streaklines are identical in the experimentally visualized streaming flow. This is a valid assumption as long as the flow is steady. The experimental and computed results are shown in Fig. 13. The comparison is seen to be very good as far as the geometry of the streamlines is concerned. A quantitative comparison of the magnitudes of the flow velocities was beyond the scope of this work.

Although the agreement between the experiments and theory was close, several differences can be identified. The centers of the outer streaming in the experimental results appear to be slightly higher than the predicted values from the model. Moreover the predicted outer streaming on the left is wider in x than in the outer streaming observed from the visualization. The main sources of these errors can be attributed to several reasons including:

(1) The two-dimensional flow assumption (three-dimensional effects do occur in the domain).
(2) The baffle used in the experiments is not ideal and hence allows some air passage between the two halves of the domain.
(3) Buoyancy-driven flows generated by the smoke.

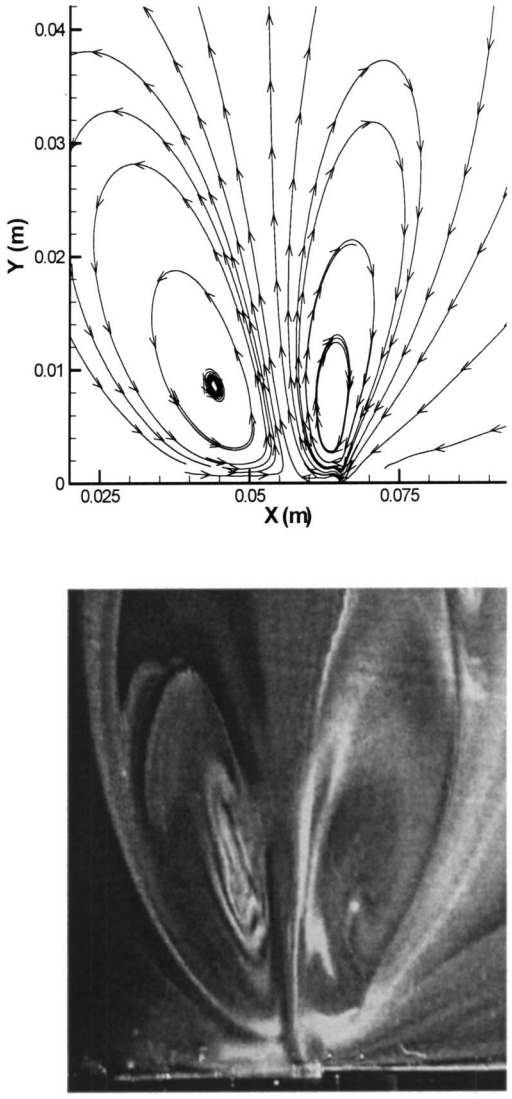


FIG. 13. Comparison between the predicted time-averaged second order flow and the flow visualization experiment.

(4) Errors in the model related to the perturbation approach and the incompressibility assumptions.
(5) Errors resulting from the numerical methods utilized in the finite Fourier transform.

V. CONCLUSIONS

This work investigates analytically, computationally, and experimentally the two-dimensional streaming flows induced by resonating thin beams such as those used in piezoelectric fans. The method of successive approximations together with other computational techniques is used to predict the streaming flows from an infinite beam and from a baffled piezoelectric fan resonating in an infinite medium. A closed form solution presented for the infinite beam problem identifies clearly the presence of an inner boundary layer in the nearfield and also describes the fluid ejection and suction at the vibration antinodes and nodes, respectively, of the beam. The extension to the finite length piezoelectric fan is more complicated and a computational method based on inverse Fourier transforms is used to approximate the stream function at each order of perturbation. The method is expected to predict accurately the fluid flow in the entire fluid domain except near the fan tip. The numerical streaming flow solution indicates an asymmetric flow field featuring two counter-rotating vortices. Further, an inner boundary layer of nonconstant thickness also arises. The predictions are compared against careful flow visualization experiments using a commercial piezoelectric fan and the results are found to be in close agreement. The results are expected to provide guidelines for the design and optimal placement in enclosures of these fans. Finally it is noted that the analysis neglects the three-dimensional effects in the flow. These will be significant in real enclosures and are the subject of continued investigation by the authors.

ACKNOWLEDGMENTS

The authors would like to thank the Cooling Technologies Research Center (CTRC) and its constituent industry members for financial support of this research.

[1] T. Açıkalın, S. M. Wait, S. V. Garimella, and A. Raman, ''Experimental investigation of the thermal performance of piezoelectric fans,'' Heat Transfer Eng. (in press).
[2] M. Toda, ''Voltage-induced large amplitude bending device-PVF_2 bimorph—Its properties and applications,'' Ferroelectrics **32**, 127–133 (1981).
[3] M. Toda, ''Theory of air flow generation by a resonant type PVF_2 bimorph cantilever vibrator,'' Ferroelectrics **22**, 911–918 (1979).
[4] J. H. Yoo, J. I. Hong, and W. Cao, ''Piezoelectric ceramic bimorph coupled to thin metal plate as cooling fan for electronic devices,'' Sens. Actuators A **79**, 8–12 (2000).
[5] R. R. Schmidt, ''Local and average transfer coefficients on a vertical surface due to convection from a piezoelectric fan,'' InterSociety Conference on Thermal Phenomena (INTHERM, 1994), Washington, D.C., 1994, pp. 41–49.
[6] J. E. Sader, ''Frequency response of cantilever beams immersed in viscous fluids with applications to the atomic force microscope,'' J. Appl. Phys. **84**, 64–76 (1998).
[7] N. M. Nguyen and R. M. White, ''Acoustic streaming in micromachined flexural plate wave devices: Numerical simulation and experimental verification,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1463–1471 (2000).
[8] A. Cox, D. Monopoli, M. Goldfarb, and E. Garcia, ''Development of piezoelectrically actuated micro-aerial vehicles,'' Proc. SPIE **3834**, 101–108 (1999).
[9] M. J. Lighthill, ''Acoustic streaming,'' J. Sound Vib. **61**, 391–418 (1978).
[10] N. Riley, ''Steady streaming,'' Annu. Rev. Fluid Mech. **33**, 43–65 (2001).
[11] W. L. Nyborg, ''Acoustic streaming,'' in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, San Diego, 1998), Chap. 7.
[12] H. Schlichting, *Boundary Layer Theory* (McGraw-Hill, New York, 1955), Chap. 15.
[13] W. P. Raney, J. C. Corelli, and P. J. Westervelt, ''Acoustical streaming in the vicinity of a cylinder,'' J. Acoust. Soc. Am. **26**, 1006–1014 (1954).
[14] J. M. Andres and U. Ingard, ''Acoustic streaming at high Reynolds numbers,'' J. Acoust. Soc. Am. **25**, 928–932 (1953).
[15] B. J. Davidson and N. Riley, ''Jets induced by oscillatory motion,'' J. Fluid Mech. **53**, 287–303 (1972).
[16] A. Bertelsen, A. Svardal, and S. Tjotta, ''Nonlinear streaming effects associated with oscillating cylinders,'' J. Fluid Mech. **59**, 493–511 (1973).
[17] E. W. Haddon and N. Riley, ''The steady state streaming induced between oscillating circular coaxial cylinders,'' Q. J. Mech. Appl. Math. **32**, 265–282 (1979).
[18] S. K. Kim and A. W. Troesch, ''Streaming flows generated by high-frequency small-amplitude oscillations of arbitrarily shaped cylinders,'' Phys. Fluids A **6**, 975–985 (1989).
[19] C. A. Lane, ''Acoustical streaming in the vicinity of a sphere,'' J. Acoust. Soc. Am. **27**, 1082–1086 (1955).

[20] W. L. Nyborg, ''Acoustic streaming near a boundary,'' J. Acoust. Soc. Am. **30**, 329–339 (1958).
[21] A. Kildal, ''Linear and nonlinear fluid motion generated by an oscillating obstacle,'' Wave Motion **19**, 171–187 (1994).
[22] B. G. Loh, S. Hyun, P. I. Ro, and C. Kleinstreuer, ''Acoustic streaming induced by ultrasonic flexural vibrations and associated enhancement of convective heat transfer,'' J. Acoust. Soc. Am. **111**, 875–883 (2002).
[23] Q. Wan and A. V. Kuznetsov, ''Numeric modeling ultrasonic acoustic streaming cooling effect on IC chips,'' Proceedings of 2001 ASME International Mechanical Engineering Congress and Exposition, New York, NY, November 11–16, 2001.
[24] A. Ihara and H. Watanabe, ''On the flow around flexible plates, oscillating with large amplitude,'' J. Fluids Struct. **8**, 601–619 (1994).
[25] S. Basak, A. Raman, and S. V. Garimella, ''Dynamic response optimization of asymmetrically configured piezoelectric fans,'' submitted to Symposium on Piezoelectric and Magnetostrictive Actuators Including MEMs Devices at the ASME 19th Biennial Conference on Mechanical Vibration and Noise, Chicago, Illinois, September 2–6, 2003.
[26] D. P. Telionis, *Unsteady Viscous Flows* (Springer-Verlag, New York, 1981).
[27] Mathematica 4, Wolfram Research, Inc.

Calculating the micrometeorological influences on the speed of sound through the atmosphere in forests

Arnold Tunick[a)]
U.S. Army Research Laboratory, Computational and Information Sciences Directorate, 2800 Powder Mill Road, Adelphi, Maryland 20783-1197

(Received 14 January 2003; revised 10 July 2003; accepted 22 July 2003)

A key element in determining point-to-point acoustic transmission within and above forests is modeling the variation (with height above ground) of the effective speed of sound. Effective speed of sound is readily derived from estimates of air temperature, relative humidity, and wind velocity. However, meteorological models for the forest canopy vary from comparatively simple to academically complex, requiring different amounts and numbers of inputs and computer capabilities. In addition, not all canopy profile models are suitable for acoustic applications. In this paper, a meteorological computer model for the forest canopy is developed to derive continuous profiles of effective sound speed from the ground to $3h$, where h is the height of the canopy. In turn, these profiles are used to make some initial approximations of short-range acoustic transmission loss through a uniform forest stand for typical clear sky, midday atmospheric conditions. Also, a radiative transfer and energy budget algorithm is incorporated into the model to obtain the appropriate heat source profile for any time of day. Thus, physics-based micrometeorology is coupled to acoustics for future applications of acoustic information in forest environments. [DOI: 10.1121/1.1608960]

PACS numbers: 43.28.Bj, 43.28.Fp [AJZ]

I. INTRODUCTION

The U.S. Army has a growing interest in the use of low-cost, non-line-of-sight acoustic sensor systems for the surveillance, detection, identification, classification, and tracking of sound-emitting targets on the battlefield (Fong and Srour, 1994; Srour, 1999). Most new Army acoustic systems use unattended microphone sensors to construct small ground-based beamforming arrays to determine line-of-bearing angles. Consequently, the U.S. Army is looking to implement the best possible computer models for determining point-to-point acoustic transmission (West *et al.*, 1991, 1992; Wilson, 1993, 2000; Salomons, 2001). At the same time, the retrieval and interpretation of acoustic signals in diverse microclimate areas, e.g., in and around forests, hilly terrain, or in cities, is greatly influenced by turbulence and refraction effects caused by finer-scale atmospheric motions over varying topography and surface energy budgets (e.g., Auvermann *et al.*, 1995; Wilson, 1998). Therefore, improved physics-based theory and computer models for micrometeorology coupled to acoustics may contribute important information on the performance of advanced combat systems.

A key element in determining point-to-point acoustic transmission is modeling the variation (with height above ground) of the effective speed of sound. Effective speed of sound is readily derived from measured or modeled estimates of air temperature, relative humidity,[1] and wind velocity. However, for the forest canopy, a recent survey of the literature (Tunick, 2002) shows that meteorological models vary from comparatively simple to academically complex, requiring different amounts and numbers of inputs and computer capabilities. In addition, not all canopy profile models are suitable for acoustic applications. For example, extinction-type profile models (Inoue, 1963; Shinn, 1971; Albini, 1981; Meyers *et al.*, 1998) may provide reasonable estimates of wind velocity through the mid-to-upper canopy layer, although in practice, these solutions do not combine very smoothly to modeled winds above the treetops or below the layer of leaves and branches. Moreover, extinction-type profile models do not address heat transfer within forest canopies.

Alternately, there are first-order closure models (Li *et al.*, 1990; Wilson *et al.*, 1998) and higher-order closure models (Wilson and Shaw, 1977; Meyers and Paw U, 1987; Wilson, 1988; Katul and Albertson, 1998) to estimate wind speeds and turbulence within and above forest canopies. However, our chief interest is in developing a meteorological computer model that will represent both the mechanical and thermodynamic influences on the speed of sound in the forest environment. For example, large-eddy simulation (LES) models would provide a fine-scale physics-based calculation of the turbulence, wind flow, and scalar fields within and above a uniform forest canopy (e.g., Shaw and Schumann, 1992; Albertson *et al.*, 2001). LES models, however, are quite computationally intensive and offer much more information than is necessary for the purposes of the present study.

Therefore, in this paper, a one-dimensional (1D), second-order turbulence closure model is presented for canopy wind flow and temperatures, to include a relatively complete radiative transfer and energy budget algorithm to compute the heat source. Steady-state, mean profiles for wind velocity and air temperature are calculated for typical midday, clear sky atmospheric conditions, from which values

a)Electronic mail: atunick@arl.army.mil

for the effective speed of sound inside and above a uniform forest canopy are derived from the ground to $3h$, where h is the height of the canopy. Then, based on the 1D forest model results for mean wind velocity and air temperature, some initial approximations for short-range acoustic transmission loss through a continuous forest stand are presented, wherein one might expect that relative sound-pressure levels will depend, in part, on the strength and locations of sound-speed profile inversions between the ground and the model top. Note, however, that the acoustic model (to be introduced in Sec. III C) does not yet account for effects like backscattering from a forest edge, transmission loss due to foliage, or scattering and transmission loss due to atmospheric turbulence. It is anticipated that mechanisms to account for these effects will be incorporated in future modeling efforts. Finally, in the Appendix, several additional (model) profiles of wind velocity and air temperature are shown in comparison to selected micrometeorological tower data collected as part of the Boreal Ecosystem–Atmosphere Study (BOREAS) (Sellers *et al.*, 1997).

II. FOREST CANOPY MODEL

A. Second-order turbulence closure model for canopy wind flow

The structure of the initial FORTRAN program code used for this research follows that described by Katul and Albertson (1998), which is based on the earlier works of Donaldson (1973), Mellor (1973), and Wilson and Shaw (1977). The model contains the one-dimensional (1D), steady-state, and horizontally homogeneous conservation (simplified Navier–Stokes) equations for the turbulent fluxes and mean variances of the winds inside and above the forest canopy. The parametrized model equations for the mean flow $\langle\overline{u}\rangle$, stress $\langle\overline{u'w'}\rangle$, longitudinal $\langle\overline{u'^2}\rangle$, lateral $\langle\overline{v'^2}\rangle$, and vertical velocity $\langle\overline{w'^2}\rangle$ variances, respectively, are as follows:

$$\frac{d\langle\overline{u}\rangle}{dt}=0=-\frac{d\langle\overline{u'w'}\rangle}{dz}-C_dA\langle\overline{u}\rangle^2, \tag{1}$$

$$\frac{d\langle\overline{u'w'}\rangle}{dt}=0=-\langle\overline{w'^2}\rangle\frac{d\langle\overline{u}\rangle}{dz}+\frac{d}{dz}\left(2q\lambda_1\frac{d\langle\overline{u'w'}\rangle}{dz}\right)+\frac{g}{\theta}\langle\overline{u'\theta'}\rangle-\frac{q}{3\lambda_2}\langle\overline{u'w'}\rangle+Cq^2\frac{d\langle\overline{u}\rangle}{dz}, \tag{2}$$

$$\frac{d\langle\overline{u'^2}\rangle}{dt}=0=-\langle\overline{u'w'}\rangle\frac{d\langle\overline{u}\rangle}{dz}+\frac{d}{dz}\left(q\lambda_1\frac{d\langle\overline{u'^2}\rangle}{dz}\right)+2C_dA\langle\overline{u}\rangle^3-\frac{q}{3\lambda_2}\left(\langle\overline{u'^2}\rangle-\frac{q^2}{3}\right)-\frac{2}{3}\frac{q^3}{\lambda_3}, \tag{3}$$

$$\frac{d\langle\overline{v'^2}\rangle}{dt}=0=\frac{d}{dz}\left(q\lambda_1\frac{d\langle\overline{v'^2}\rangle}{dz}\right)-\frac{q}{3\lambda_2}\left(\langle\overline{v'^2}\rangle-\frac{q^2}{3}\right)-\frac{2}{3}\frac{q^3}{\lambda_3}, \tag{4}$$

and

$$\frac{d\langle\overline{w'^2}\rangle}{dt}=0=\frac{d}{dz}\left(3q\lambda_1\frac{d\langle\overline{w'^2}\rangle}{dz}\right)+2\frac{g}{\theta}\langle\overline{w'\theta'}\rangle-\frac{q}{3\lambda_2}\left(\langle\overline{w'^2}\rangle-\frac{q^2}{3}\right)-\frac{2}{3}\frac{q^3}{\lambda_3}. \tag{5}$$

Here, z is the vertical coordinate, $C_d(=0.15)$ is the forest canopy drag coefficient, A (in units $\mathrm{m^2\,m^{-3}}$) is the leaf area density, $q=\sqrt{\langle\overline{u'^2}\rangle+\langle\overline{v'^2}\rangle+\langle\overline{w'^2}\rangle}$ (in units $\mathrm{ms^{-1}}$) is the turbulent kinetic energy of the mean flow, C is a constant whose value is about 0.077, g is acceleration due to gravity, θ is the ambient air potential temperature (in units K), $\langle\overline{u'\theta'}\rangle$ is the horizontal heat flux, $\langle\overline{w'\theta'}\rangle$ is the vertical heat flux, and λ_1, λ_2, and λ_3 are length scales, i.e., $\lambda_k=a_kl$, where k is an arbitrary index, and l is the mixing length. Values for these closure constants are as follows: $a_1=0.39$, $a_2=0.85$, and $a_3=16.57$. The overbars indicate a time-averaged mean, whereas the brackets, $\langle\ \rangle$, indicate horizontal averaging. Note that the lateral velocity variance $\langle\overline{v'^2}\rangle$ is calculated in the current 1D model to support the turbulent kinetic energy closure approximations for the triple product and dissipation terms. The basic equation set and modeling assumptions have been explained at length by Mellor (1973), Shaw (1977), Wilson and Shaw (1977), and Katul and Albertson (1998) and are thus not repeated here.

B. Steady-state turbulent transfer model for heat inside and above the forest canopy

Four additional conservation equations are added to the model code. They are the equations for the mean temperature $\langle\overline{\theta}\rangle$, vertical heat flux $\langle\overline{w'\theta'}\rangle$, horizontal heat flux $\langle\overline{u'\theta'}\rangle$, and temperature variance $\langle\overline{\theta'^2}\rangle$, which can be written as follows:

$$\frac{d\langle\overline{\theta}\rangle}{dt}=0=-\frac{d\langle\overline{w'\theta'}\rangle}{dz}+S_\theta, \tag{6}$$

$$\frac{d\langle\overline{w'\theta'}\rangle}{dt}=0=-\langle\overline{w'^2}\rangle\frac{d\langle\overline{\theta}\rangle}{dz}+\frac{d}{dz}\left(2q\lambda_4\frac{d\langle\overline{w'\theta'}\rangle}{dz}\right)-\frac{q}{3\lambda_5}\langle\overline{w'\theta'}\rangle+\frac{2}{3}\frac{g}{\theta}\langle\overline{\theta'^2}\rangle, \tag{7}$$

$$\frac{d\langle\overline{u'\theta'}\rangle}{dt}=0=-\langle\overline{u'w'}\rangle\frac{d\langle\overline{\theta}\rangle}{dz}-\langle\overline{w'\theta'}\rangle\frac{d\langle\overline{u}\rangle}{dz}+\frac{d}{dz}\left(q\lambda_4\frac{d\langle\overline{u'\theta'}\rangle}{dz}\right)-\frac{q}{3\lambda_5}\langle\overline{u'\theta'}\rangle, \tag{8}$$

and

$$\frac{d\langle\overline{\theta'^2}\rangle}{dt}=0=-2\langle\overline{w'\theta'}\rangle\frac{d\langle\overline{\theta}\rangle}{dz}+\frac{d}{dz}\left(q\lambda_6\frac{d\langle\overline{\theta'^2}\rangle}{dz}\right)-\frac{2q}{\lambda_7}\langle\overline{\theta'^2}\rangle, \tag{9}$$

where S_θ is the heat source,[2] and λ_4 through λ_7 are additional length scales, where $a_4=a_6=0.23$, $a_5=0.74$, and $a_7=10.10$. Values for these constants are taken from the earlier

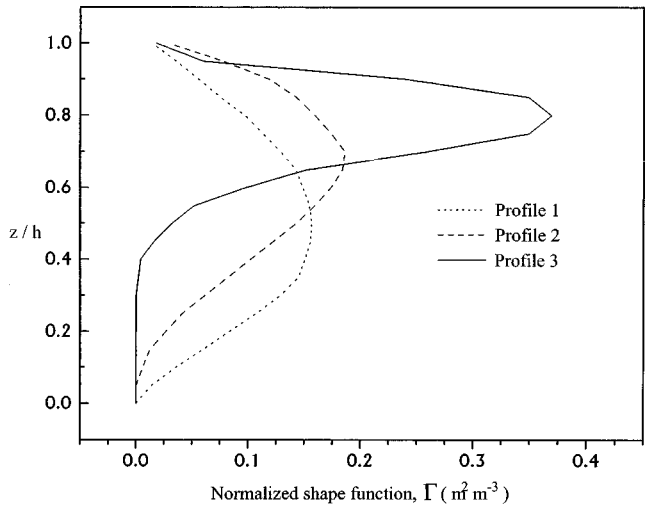


FIG. 1. Normalized vertical profiles of leaf-area distribution for forest canopies where $A(z)=\Gamma(z)\cdot\mathrm{LAI}$ (based on Meyers *et al.*, 1998 and Albertson *et al.*, 2001).

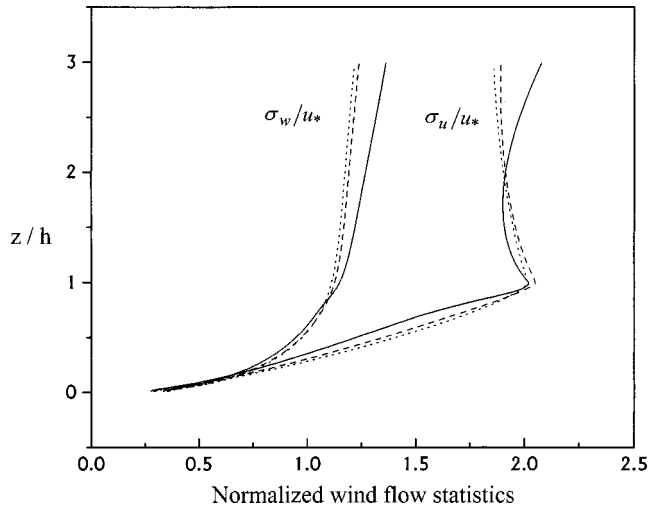


FIG. 2. Normalized wind-flow statistic profiles inside and above the forest canopy for typical clear sky, midday atmospheric conditions, where the normalized square root of the longitudinal velocity variance is σ_u/u_* and the normalized square root of the vertical velocity variance is σ_w/u_*. The three curves for each variable correspond to the three different leaf-area distributions used as input to define the canopy architecture, i.e., *profile-1* (dotted line); *profile-2* (dashed line); and *profile-3* (solid line).

works of Mellor (1973) and Mellor and Yamada (1974). Additional insights on the equation set, parametrizations, closure constants, initial conditions, boundary conditions, and numerical methods are discussed in the papers given by Meyers and Paw U (1987) and Albertson *et al.* (2001).

C. Forest canopy architecture

Canopy architecture plays an important role in defining the momentum and heat flux divergence through the forest layer. The papers given by Massman (1982) and Meyers *et al.* (1998) suggest that forest canopies may conform to one of three general leaf-area distribution profiles. Figure 1 shows that leaf-area distributions are not always symmetric about the layer of maximum foliage density (like *profile-1*) but may be more often skewed upward toward the top of the forest canopy. By definition, leaf area index is

$$\mathrm{LAI}=\int_0^h A(z)dz,$$

where $A(z)$ is the leaf-area density through the small vertical layer between z and $z+dz$ per unit surface area of ground below. Values for leaf-area index for forests vary but are reported most often in the range LAI=1 to 5 (Kaimal and Finnigan, 1994; Finnigan and Brunet, 1995). In Sec. III A, modeled profiles for wind velocity and air temperature are presented for the cases corresponding to these three different leaf-area distributions. Note, for this study, we set LAI=3.0.

D. One-dimensional radiative transfer and energy budget algorithm for the forest canopy

The heat source term (S_θ), as described by Meyers and Paw U (1987), is modeled as

$$S_\theta=2A(\overline{\theta_l}-\overline{\theta})/r_h. \tag{10}$$

Here, A (in units $\mathrm{m^2\,m^{-3}}$) is the leaf-area density, $(\overline{\theta_l}-\overline{\theta})$ is the mean leaf surface-to-ambient-air temperature difference, and r_h is the aerodynamic resistance to heat transfer. A 1D radiative transfer and energy budget algorithm is incorporated into the model calculation to make it possible to determine the heat source (S_θ) for any time of day. To start, Rachele and Tunick (1994) provide the formulations to calculate the incoming total radiation at the canopy top as a function of latitude, longitude, day of year, and time of day, i.e., these inputs are needed to determine the solar declination, hour, and zenith angles. Thus, the total downward short-wave radiation at canopy top is

$$R_{S\downarrow}=I_0T_RT_GT_WT_A\cos\theta_z, \tag{11}$$

where $I_0=1367(1+0.034\cos[[2\pi(n'-1)]/365])$ is the solar constant, n' is the day of year, θ_z is the solar zenith angle, and the short-wave transmission functions are as follows: for Rayleigh scattering (T_R), absorption by permanent gases (T_G), absorption by water vapor (T_W), and absorption and scattering by aerosols (T_A).

Then, the equations provided by Weiss and Norman (1985) are used to calculate the four spectral components for short-wave radiation (direct beam and diffuse, visible and near-infrared radiation as a function of the total downward short-wave flux at canopy top) because extinction and reflection through the forest canopy are different for each. The remainder of the radiative transfer subroutine for the forest canopy (i.e., the transmission, reflection, absorption, and emission of the solar flux necessary to compute leaf surface-to-ambient-air temperature differences) is outlined from the formulations given in the texts by Campbell (1977) and Campbell and Norman (1998).

III. MODEL RESULTS

A. Mean profiles

Based on the equation set and modeling assumptions described above, Figs. 2 and 3 show profiles inside and above the forest canopy of the modeled wind flow and temperature statistics $\langle\overline{u'^2}\rangle^{1/2}$, $\langle\overline{w'^2}\rangle^{1/2}$, and $\langle\overline{\theta'^2}\rangle^{1/2}$ for typical

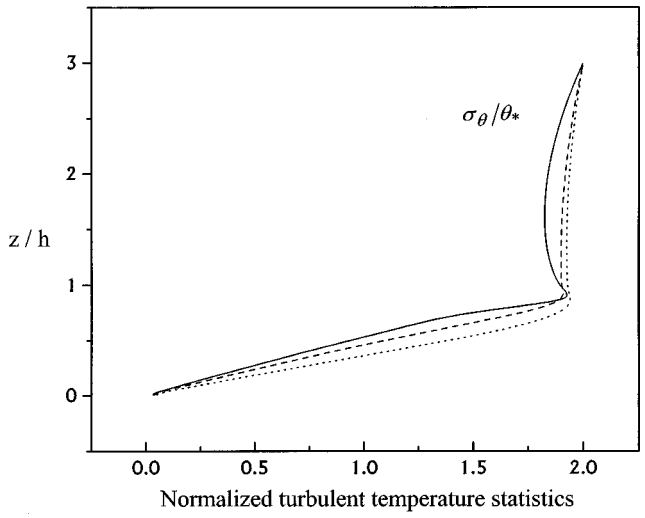


FIG. 3. Normalized square root of the temperature variance profile (σ_θ/θ_*) inside and above the forest canopy for typical clear sky, midday atmospheric conditions. The three curves are defined the same as in Fig. 2.

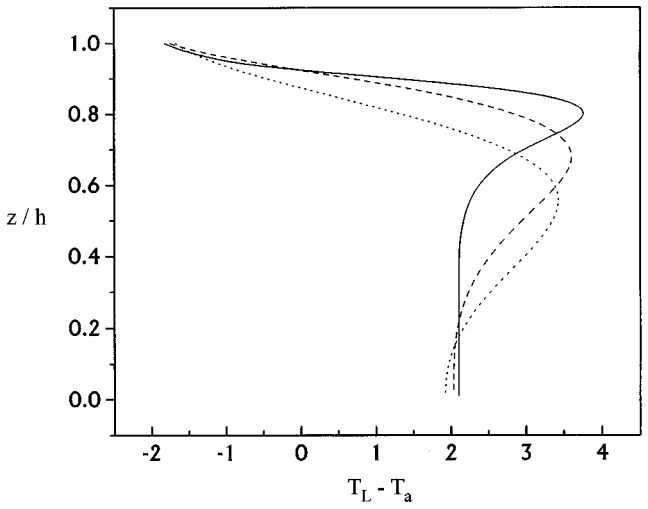


FIG. 4. Profiles of the leaf-to-ambient air temperature differences (°C). The three curves are defined the same as in Fig. 2.

clear sky, midday (i.e., 1400 LT) atmospheric conditions. The three curves for each variable correspond to the three different leaf-area distributions used as input to define the canopy architecture (see Fig. 1). For this example, upper level (i.e., model top) wind velocity is $u_{\max}=6.5\,\mathrm{ms}^{-1}$, friction velocity is $u_*\approx 1.0\,\mathrm{ms}^{-1}$, temperature scaling is $\theta_*\approx -0.5\,\mathrm{K}$, and relative humidity is $R_h=40\%$. In micrometeorology, friction velocity squared is a measure of the horizontal shearing stress and $Q=-u_*\theta_*$ is the kinematic heat flux (Busch, 1973; Wyngaard, 1973). In addition, leaf-area index is LAI=3.0, forest canopy height is $h=10\,\mathrm{m}$, and the model domain is $0.01h\leqslant z\leqslant 3h$. Not surprisingly, these results are in good agreement, in comparison with the results shown by Wilson and Shaw (1977) and Katul and Albertson (1998). The aim here is simply to demonstrate the model capability to compute reasonable profiles of the turbulent statistics. In addition, Figs. 2 and 3 show the sensitivity in these results due to variations in leaf-area density distribution. (Note that the second-order quantities shown here will not be used in the application of forest model data to the acoustic propagation code described in Sec. III C. However, these types of data may be quite important in future acoustic models to estimate scattering from atmospheric turbulence).

Figure 4 shows the modeled profiles for leaf surface-to-ambient air temperature differences, i.e., T_L-T_a, for the three cases corresponding to the three different leaf-area distributions shown in Fig. 1. In this example, leaf surface temperatures larger than ambient do not occur in the uppermost part of the canopy. Maximum values for T_L-T_a in the tree crowns are shown to be approximately 3.4 to 3.8 °C, depending on the leaf-area profile. In the trunk spaces, below the layer of leaves and branches, about a 2.0 °C temperature difference is computed for all three cases.

These estimates reflect the balance between the incoming solar energy absorbed by the canopy and the outgoing long-wave energy emitted from the surrounding air. The absorbed radiative flux depends, in part, on the amount of short-wave (visible and near-infrared) energy transmitted downward through the forest from the canopy top. [The other part, according to Campbell and Norman (1998), is the downward long-wave flux as a function of ambient air temperature and cloud cover.] As an example, Fig. 5 shows the transmission coefficient profiles calculated for the case described that corresponds to the leaf-area distribution *profile-3*. The results in Fig. 5 imply that transmission is much greater than absorption through the uppermost portion of the canopy. In this case, the total amount of short-wave (visible and near-infrared combined) energy intercepted toward the top of the canopy is not sufficient for local (leaf) surface heating to occur. At the same time, Fig. 5 suggests that the short-wave flux remaining in the lower portion of the canopy is sufficient to affect the outcome, $T_L-T_a\approx 2.0$. However, increasing the canopy leaf-area index from LAI=3 to LAI=4 or 5 reduces or eliminates the temperature differences in the trunk space, i.e., a decrease of 1.0 and 2.0 °C,

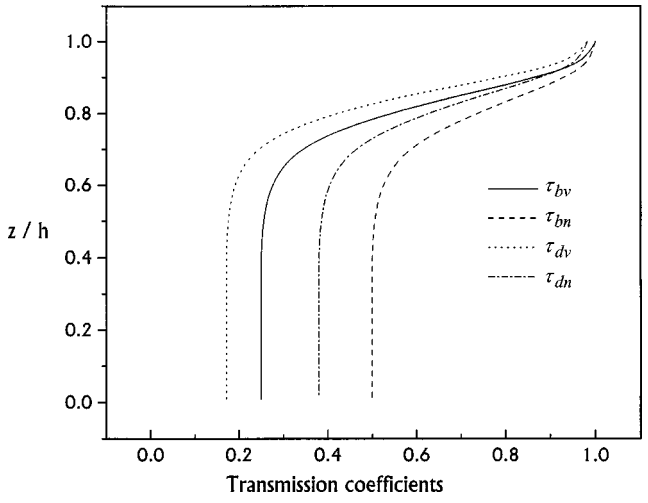


FIG. 5. Profiles of the transmission coefficients through the forest canopy layer for the following short-wave radiation components: direct beam visible (τ_{bv}, solid line), direct beam near-infrared (τ_{bn}, dashed line), diffuse visible (τ_{dv}, dotted line), and diffuse near-infrared (τ_{dn}, dash-dotted line). The transmission coefficient profiles are calculated for the case described in the text that corresponds to the assumed leaf-area distribution *profile-3*.

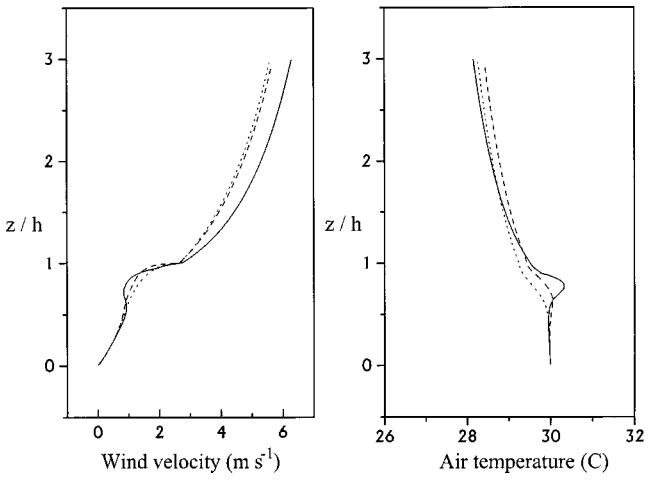


FIG. 6. Profiles of mean wind velocity and air temperature inside and above the forest canopy for the conditions described in the text. The three curves for each variable are defined the same as in Fig. 2.

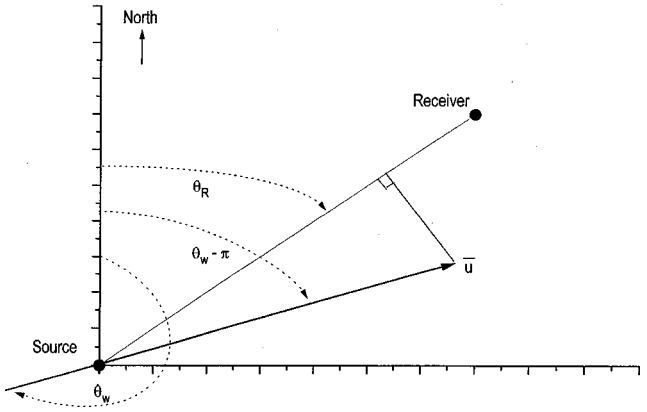


FIG. 7. Geometry schematic for the expression $c_{\text{eff}}=c_0+\bar{u}\cos(\theta_w-\pi-\theta_R)$, where c_{eff} is the effective sound speed that includes the sound-speed component due to the mean wind along the path of propagation.

respectively. This takes place because the amount of absorbed radiation is reduced through the lower canopy.

Figure 6 shows the continuous profiles of mean wind velocity and air temperature inside and above the forest canopy for the three cases previously outlined. In general, wind-speed profiles at some height above tall roughness elements are similar in form to those described for lower roughness elements, for steady-state and horizontally homogeneous conditions (Busch, 1973). As a result, the modeled wind-velocity profiles above the canopy appear quite similar to wind-velocity profiles in open fields, i.e., logarithmically increasing with height above the roughness plane. However, the model results are sensitive to variations in the assumed leaf-area distribution in the canopy below. Increasing values of leaf-area index towards the top of the canopy, i.e., as in *profile-3*, result in an increase in wind velocity above the canopy (and vice versa). This is due to a "skimming effect" of the wind over a more compact surface, resulting in an effective decrease in roughness for the canopy (Albertson *et al.*, 2001). In contrast, inside the forest, the winds are shown to decrease rapidly as momentum becomes depleted through the layers of leaves and branches. In addition, for the case that corresponds to *profile-3*, the model produces a secondary wind-speed maximum at a height of about 0.5 or $0.6h$. Shinn (1971) and Shaw (1977) discuss such low-level wind maxima in detail. They suggest that among the possible mechanisms responsible for the prominence of such wind-flow maxima is the turbulent (vertical) transport of momentum from the upper portions of the canopy, i.e., the transport term in the stress equation, $[d\langle\overline{w'u'w'}\rangle]/dz = d/dz(2q\lambda_1[(d\langle\overline{u'w'}\rangle)/dz])$. Also, they suggest that low-level wind-speed maxima occur in a transition layer between the upper canopy, extinction-type profile, and the log-law type wind-speed profile that one might expect extends from this height to the ground. Interestingly, however, the other leaf-area distribution profiles in this example do not bring about a similar feature.

Figure 6 also shows that the model produces a local maximum in air temperature within the canopy around the height of maximum leaf density (mainly for *profile-3* but also slightly for *profile-2*). This coincides nicely with the height of the local minimum in wind velocity. Generating such temperature inversions appears to be quite sensitive to variations in the leaf-area distribution profile.

Temperature lapse rates above the canopy are also influenced by variations in the assumed leaf-area distribution profile. In Fig. 6, both wind velocity and temperature gradients increase as LAI increases and roughness decreases. While these results are in good agreement in comparison to modeled profiles reported by others (e.g., Meyers and Paw U, 1987; Albertson *et al.*, 2001), an appendix to this paper is included to present additional model calculations in comparison to observed micrometeorological tower data.

B. The speed of sound through the atmosphere inside and above a forest canopy

The paper now turns its attention to deriving estimates for sound speed, from which it is possible to determine transmitted sound intensity. It is useful to define the effective sound speed (c_{eff}) from the following expression given by Noble and Marlin (1995) and Ostashev (1997):

$$c_{\text{eff}}=c_0+\bar{u}\cos(\theta_w-\pi-\theta_R), \tag{12}$$

where $c_0=\sqrt{\gamma_s RT/M}$, $R=8314.32\,\text{J mol}^{-1}\,\text{K}^{-1}$ is the universal gas constant, M is molecular mass, $\gamma_s=c_p/c_v$ is the ratio of specific heats, $\bar{u}$ is the mean of the horizontal wind, θ_w is the bearing of the wind from north, θ_R is the bearing of the receiver from the source, and $\bar{u}\cos(\theta_w-\pi-\theta_R)$ is the component of the sound speed along the direction of propagation from the source to the receiver (Fig. 7). As discussed in Ostashev (1997), the effective sound speed in Eq. (12) is valid only for nearly horizontal propagation angles. In addition, Wong and Embleton (1984) have deduced the ratio of specific heats and molar mass (γ_s/M) as a function of temperature and humidity in the following form:

$$\frac{\gamma_s}{M}=0.048\,33+(R_h-0.023)A_T, \tag{13}$$

where $A_T=9.2\times10^{-5}+5.5\times10^{-6}T+4.25\times10^{-7}T^2$, R_h is relative humidity (in this study $R_h=40\%$ and is assumed constant with height above ground), and T is air temperature

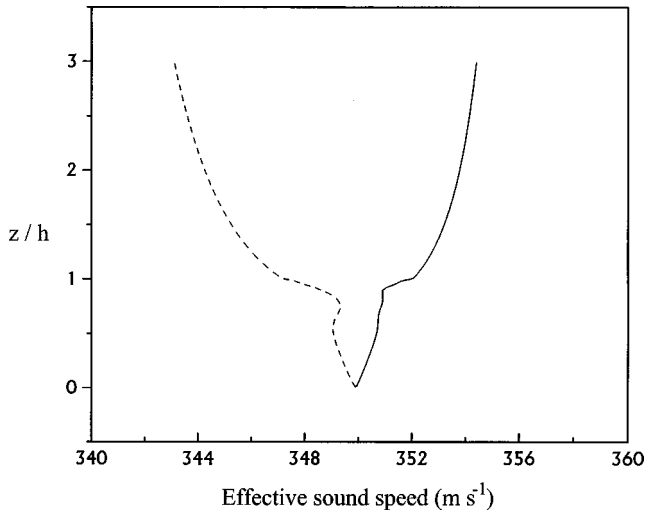


FIG. 8. Profiles of effective sound speed through the atmosphere for downwind propagation (solid) and upwind propagation (dashed) inside and above the forest canopy.

(°C). For applications in outdoor acoustics, sound-speed differences across vertical layers will cause acoustic waves to be refracted upward if the effective sound speed decreases with height and refracted downward if sound speed increases with height. Alternatively, the effective acoustic index of refraction is defined as $n_{\text{eff}}=c_0/c_{\text{eff}}$.

Figure 8 shows calculations of the effective speed of sound inside and above the forest canopy as determined from the model profiles of wind velocity and temperature corresponding to *profile-3*. *Profile-3* is selected to focus on inversions and their effects on short-range sound propagation. The effective sound-speed profiles are determined for upwind propagation (left), i.e., the bearing of the receiver from the source and the direction of the wind from north are the same, and for downwind propagation (right), i.e., the bearing angles are 180° opposite. An asymmetry exists in the graph such that the profile for downwind propagation (on the right-hand side) is slightly compressed, while the curve for upwind propagation (on the left side) is somewhat accentuated. The influence that such profile variations in sound speed have on short-range acoustic transmission loss is discussed briefly in the next section.

C. Approximation of short-range acoustic transmission loss

To produce some initial approximations for short-range acoustic transmission loss through a continuous forest stand, modeled profiles of wind speed and temperature are used as input to a flat-earth, nonturbulent acoustic propagation model called WSCAFFIP, i.e., the Windows (version) scanning fast field program. WSCAFFIP is a numerical code developed for assessing environmental effects on short-range, point-to-point acoustic transmission (Noble and Marlin, 1995). WSCAFFIP determines acoustic transmission loss as relative sound-pressure level (SPL) with range and azimuth for a given frequency, source–receiver geometry, and c_{eff} profile. WSCAFFIP contains propagation algorithms to represent the effects of atmospheric refraction, diffraction, absorption, and reflection (ground impedance) on acoustic transmission.

TABLE I. WSCAFFIP model parameters.

Parameter	Symbol	Value
Distance of receiver from source	x	500 m
Range resolution	Δx	10 m
Bearing of receiver from source	θ_R	90°
Bearing of the wind from the north	θ_w	90° (Upwind); 270° (Downwind)
Source height above ground	h_S	2 m
Receiver height above ground	h_R	1 m
Frequency of interest	f	100 Hz; 200 Hz, 300 Hz; 400 Hz
Ground porosity	Ω_p	0.72 (forest); 0.56 (open grassland)
Flow resistivity	σ_r	33 kPa s m^{-2} (forest) 110 kPa s m^{-2} (open grassland)

Note, however, that the WSCAFFIP model applied to the forest canopy does not yet contain mechanisms to account for scattering due to turbulence or scattering and absorption due to tree leaves and branches. Table I lists the model parameters for an initial approximation of short-range acoustic transmission loss in forests. The paper by Martens *et al.* (1985) provides useful values for porosity and flow resistivity for forest soils and ground cover. Figure 9 shows the WSCAFFIP results corresponding to the modeled profiles of effective sound speed described above.

Point-to-point transmission of sound waves in forests is said to involve three main phenomena: (1) interference between direct and ground reflected acoustic waves; (2) scattering by tree trunks and branches, the ground, and turbulence; and (3) absorption by the trees, leaves, branches, the ground, and the air (e.g., Burns, 1979; Fricke, 1984; Price *et al.*, 1988; Huisman and Attenborough, 1991). Also, as stated earlier, acoustic waves within and above the forest canopy will tend to be refracted downward as the effective sound speed increases with height and refracted upward as the effective sound speed decreases with height. Therefore, one might expect estimates of relative sound-pressure level

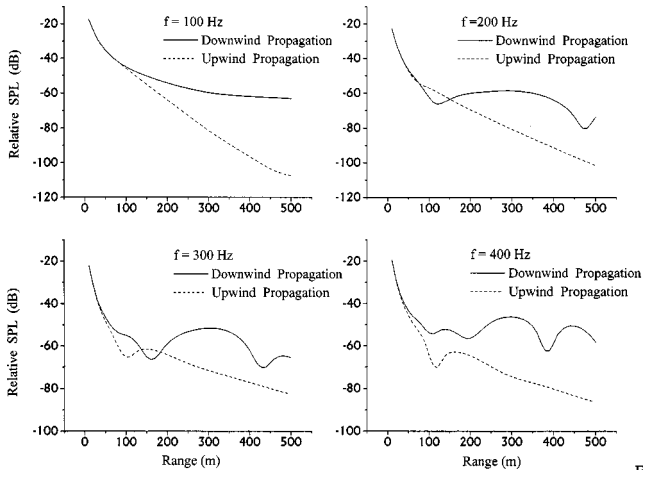


FIG. 9. Numerical approximations of short-range acoustic transmission loss within a continuous forest stand. Calculations of relative sound-pressure level (SPL) are shown at 100, 200, 300, and 400 Hz for downwind propagation (solid line) and upwind propagation (dashed line).

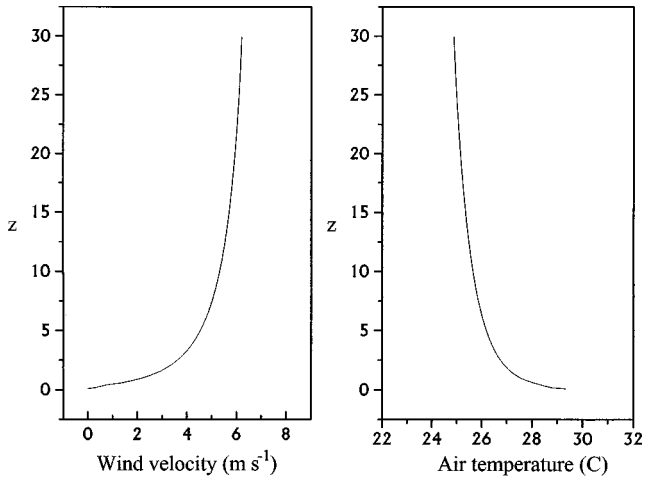


FIG. 10. Profiles of mean wind velocity and air temperature over open grassland ($h=0.5$ m).

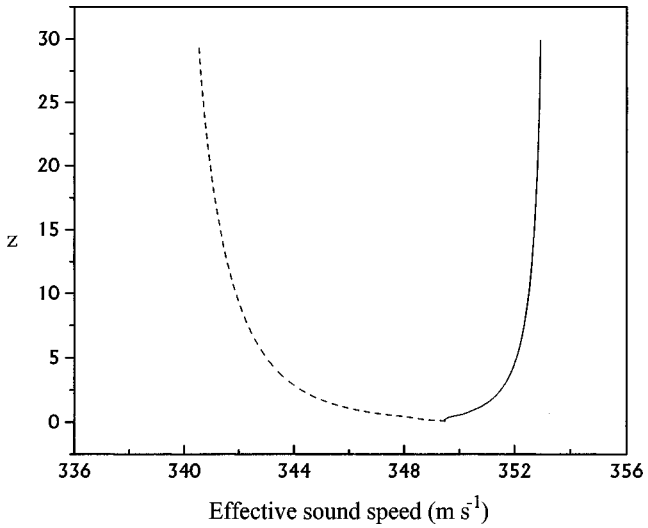


FIG. 11. Profiles of effective sound speed through the atmosphere for downwind propagation (solid) and upwind propagation (dashed) for open grassland.

(in part) to depend on the strength and locations of the sound-speed profile inversions between the ground and the model top (i.e., $z=3h$, where $h=10$ m is the height of the canopy). Naturally, acoustic transmission loss with range will also depend on ground impedance effects. Nevertheless, because generally, the effective sound-speed profile for upwind propagation has values decreasing with height, one would expect to see greater transmission loss with range at each frequency for upwind propagation as a result. Figure 9 confirms this expectation.

In contrast, the behavior of the downwind propagation transmission loss curves can be roughly understood in terms of ''ducting'' of acoustic modes. For this example, an acoustic duct exists between the surface and about 6 m. At lower frequencies (100 Hz), the transmission loss curve for downwind propagation is relatively smooth, i.e., a single mode trapped. In contrast, at 200 Hz and higher, multiple modes appear to be trapped, creating the interference pattern. Thus, ducting of acoustic energy inside and above the forest canopy is greatly affected by local variations in wind speed and temperature profile structure.

Alternately, one might expect that the impact of the sound-speed profile (and ground impedance) on acoustic transmission loss with range may be significantly different over an open field than within a forest stand. For comparison, the current 1D model was reinitialized to generate wind velocity, air temperature, and effective sound-speed profiles (Figs. 10 and 11) for an open grass-covered field ($h=0.5$ m, LAI *profile-1*, LAI=3.0). For the downwind propagation case, the sound-speed profile in Fig. 11 increases (slightly) with height above ground level. Following the same procedure as described above, Fig. 12 shows the WSCAFFIP acoustic model results for the open field in comparison to the results shown earlier for the forest stand (for downwind propagation only). At 100 Hz, relative sound-pressure levels are generally greater within the forest than over the open field. Since, as mentioned above, there may be only a single mode trapped (at 100 Hz) due to the effective sound-speed profile, it is likely that the differences in attenuation here are instead due to differences in ground impedance. This is also evidenced in Fig. 12 at $f=200$ Hz for $x<200$ m. Second, all the transmission loss curves are relatively smooth over the open field for each of the frequencies shown, despite a downward-refracting sound-speed profile. However, it was found that by extending the model range to 1500 m, one could show that refraction effects do indeed occur, but at larger distances and for higher frequencies than in the forest (Fig. 13). Finally, in order to make a distinction between refraction (meteorological profile) effects and ground impedance effects, the WSCAFFIP acoustic model was implemented using (as input) zero wind velocity and constant air temperature profiles both for the forest and open field. Figure 14 presents these additional WSCAFFIP results, wherein it is shown that acoustical properties of the ground cover and soil affect transmission loss with range primarily at lower frequencies (in this case at $f\leqslant 300$ Hz).

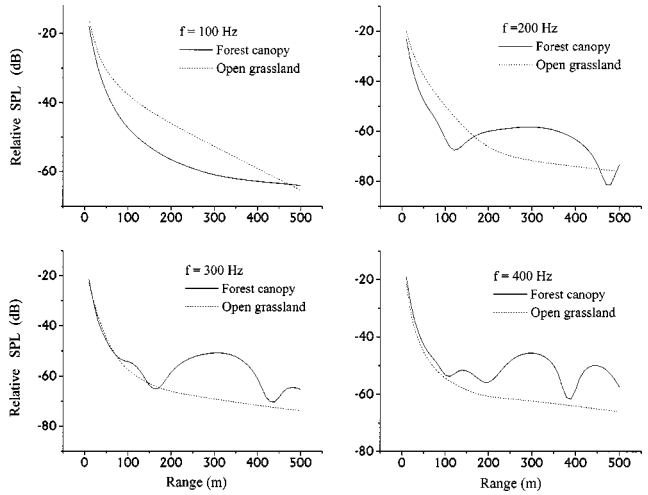


FIG. 12. Numerical approximations of downwind acoustic transmission loss within a continuous forest stand (solid) in comparison to that for open grassland (dashed). Calculations of relative sound-pressure level (SPL) are shown at 100, 200, 300, and 400 Hz.

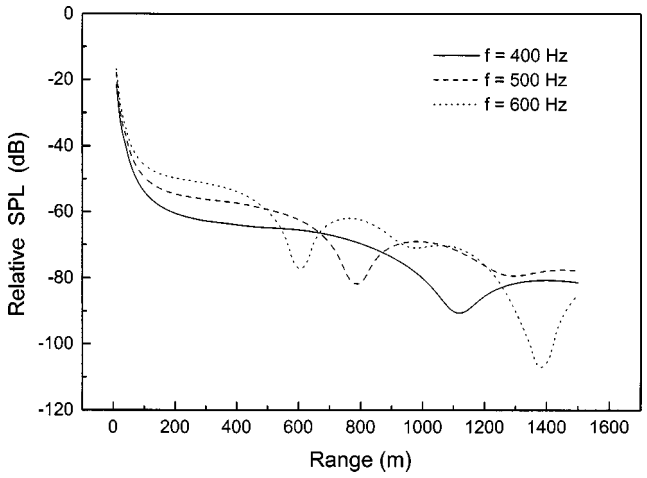


FIG. 13. Numerical approximations of downwind acoustic transmission loss with range over a tall grass field. Calculations of relative sound-pressure level (SPL) are shown at 400 Hz (solid), 500 Hz (dashed), and 600 Hz (dotted).

IV. CONCLUSIONS

A useful mathematical representation of the wind flow and temperatures inside and above a continuous forest stand can be obtained by means of a 1D steady-state, second-order turbulence closure model (with an embedded radiative transfer and energy budget algorithm to predict the heat source). This allows us to produce continuous profiles for effective sound speed that, when applied as input to an acoustic propagation code, provide realistic estimates of transmission loss and ducting of acoustic waves in the forest environment.

However, the forest canopy model does not (as yet) consider horizontal wind and temperature gradient effects on short-range acoustic transmission loss. This may be important in uneven terrain, over varying energy budgets, nonuniform forests, and forest edges. Most turbulence models, including the model described herein, are for flat surfaces. The presence of hills significantly alters the flow field inside canopies (e.g., Finnigan and Brunet, 1995). Also, it will be important for future modeling efforts to include mechanisms to account for scattering by turbulence and scattering and absorption by tree leaves and branches. In estimating line-of-bearing using acoustic sensor arrays, for example, these latter effects may have greater impact than those due to the sound-speed profile, particularly in the presence of dense understory vegetation in and around forest edges. Another problem (possibly) is how to account for the refraction of acoustic waves due to the "waving motion" of the canopy, i.e., the canopy is not a rigid element. Research in micrometeorology on the interaction between turbulence and flexible canopies has been very limited.

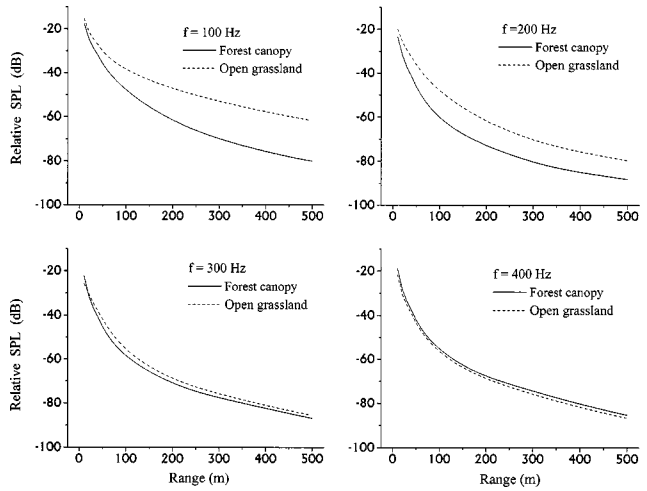


FIG. 14. Numerical approximations of downwind acoustic transmission loss for the forest (solid) and open field (dashed) derived for zero wind velocity and constant air temperature conditions. Calculations of relative sound-pressure level (SPL) are shown at 100, 200, 300, and 400 Hz.

As a final note, it would have been desirable if the model transferred heat more effectively from the leaf surface to the air to further influence the temperature lapse rate through the canopy. At midday, the model generated 3–4 °C leaf-to-air temperature differences for each case, yet the stable (inversion) layer through the upper canopy of the forest remained shallow and relatively weak ($\ll$1 °C) or nonexistent. Increasing LAI (as discussed earlier) does not improve the overall result. In any case, we would encourage further discussion on this point to obtain a better understanding of the problem.

ACKNOWLEDGMENTS

The author would like to thank Gabriel Katul and John Albertson at Duke University for providing a second-order closure, wind flow model code to perform this research. They also provided many helpful comments through the course of this study. The author also gratefully acknowledges Ronald Meyers, Keith Wilson, John Noble, and Michael Mungiole of the U.S. Army Research Laboratory for offering helpful comments on this work. Finally, K. Fred Huemmrich of the NASA Goddard Space Flight Center is acknowledged for providing helpful comments on the field data. The Department of the Army funded this research through Project B53A/61102A.

APPENDIX A: THE DATA SET

In this Appendix, modeled profiles of wind velocity and air temperature (daytime and nighttime) are presented in comparison to selected observed micrometeorological data collected as part of the Boreal Ecosystem-Atmosphere Study (BOREAS) (Sellers *et al.*, 1997). BOREAS was a very large-scale international and interdisciplinary field program conducted over a 1000×1000-km^2 region of the northern boreal forests of Canada from 1993 to 1997. Participants from several universities, the U.S. government, and Canadian government laboratories and organizations collaborated to collect data to characterize exchanges of radiative energy, sensible heat, water, CO_2, and other trace gases between the boreal forest and the lower atmosphere. The bulk of the surface-to-atmosphere flux measurements were made within two 50×50-km^2 subregions of the experimental site located near Thompson, Manitoba (the Northern Study Area), and over areas of the Prince Albert National Park, Saskatchewan (the Southern Study Area [SSA]). Data taken at each subsite included tower profile and flux measurements, hydrological

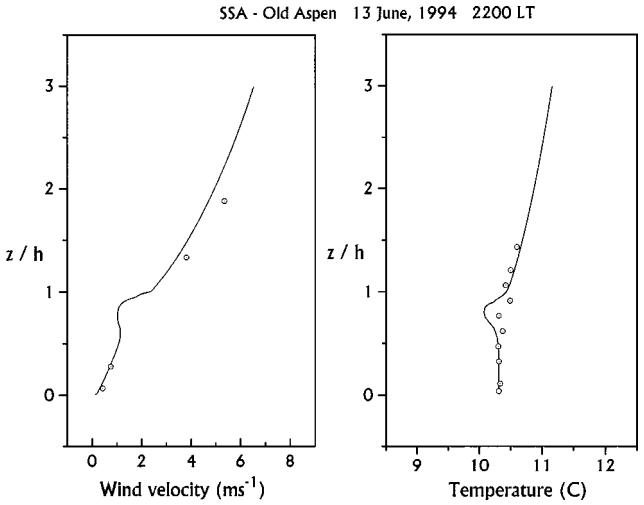


FIG. 15. Second-order closure model results (solid line) for the SSA old aspen (OA) forest canopy in comparison to micrometeorological tower data (dotted circles) taken on 13 June 1994 at 2200 local time.

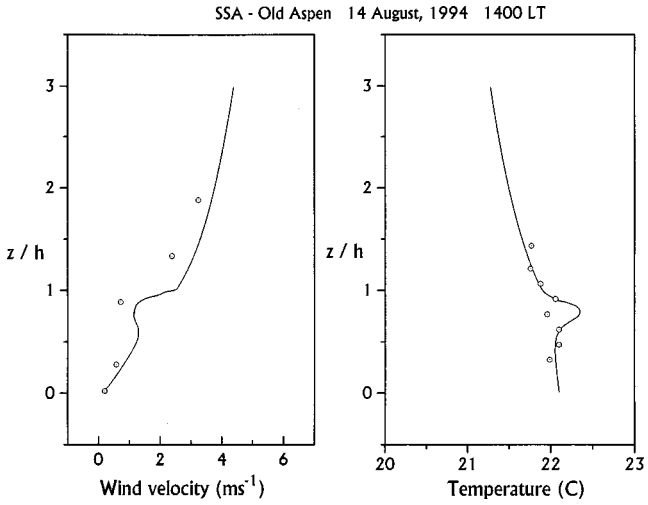


FIG. 16. Same as Fig. 15, but for 14 August 1994 at 1400 local time.

data, to include rainfall rate and soil moisture measurements, and biomass data, such as leaf area and biomass density.

The subset of the BOREAS data used here includes the measurements of wind velocity and air temperature (1/2-hour averaged) taken on the 37-m walk-up scaffold main tower located in the SSA (latitude 53.629 N; longitude, 106.198 W; elevation, 600.63 m above sea level) in an area of old growth aspen (OA) trees approximately 21 m in height (Hartog and Neumann, 2000). The trees within the boreal aspen forest had a maximum leaf area index in July 1994 of 2.3 m^2m^{-3}. The crown space for individual trees was limited to the upper 5 to 6 m of the canopy (e.g., *profile-3* as shown in Fig. 1).

APPENDIX B: COMPARISON TO OBSERVATIONS

Figure 15 shows the second-order closure model results for the SSA OA forest canopy in comparison to micrometeorological tower data taken on 13 June 1994 at 2200 local time. The computed wind-velocity profile inside and above the canopy for this nighttime case is in good agreement with the observed data (albeit there were only four data points available). The calculated temperature profile is also shown to be in good agreement with the observed data for this case. Both the modeled and observed temperature profiles show stable (lapse rate) conditions above the canopy, while isothermal or slightly unstable conditions are present below.

In contrast, Fig. 16 shows the second-order closure model results for the aspen forest site in comparison to (midday) micrometeorological tower data taken on 14 August 1994 at 1400 local time. The magnitude of the calculated wind-velocity profile above the canopy is, on average, 0.5 ms^{-1} greater than the observed data. However, the remaining model results fit the data well. Similarly, the modeled temperature profile agrees quite well to the observed data (both inside and above the canopy) for this case, with the exception of the data point at about $0.8h$. Figure 16 shows that the radiative transfer portion of the canopy model produces slight heating of the air (even under cloudy or partly cloudy conditions) through the tree crown, i.e., through the area of maximum leaf-area density. However, the data revealed that measurable rainfall was recorded on 13 August (i.e., 5–6 mm) and a trace amount of rain was recorded on 14 August (0.1–0.3 mm). The prevailing moisture then provided for large observed latent heat fluxes (evaporative cooling) both inside and above the canopy top, which suggests that the data point in question may not be erroneous. In any case, varying the model's humidity and/or cloud-cover-related parameters does not improve the result for that point.

Figure 17 shows the canopy model results for the boreal aspen forest site in comparison to (nighttime) micrometeorological profile data taken on 14 August 1994 at 2200 local time. The calculated wind-velocity profile inside and above the canopy fit the observed data quite well, i.e., this is an improvement over the daytime case. Similarly, the modeled temperature profile is in good agreement in comparison to the observed data (both inside and above the canopy), with the exception of the data point at about $0.2h$. Although an unstable inversion close to the ground would be inconsistent with the observed profile points above and below, it was suggested that on some nights, down-slope flow of cooler air (note: the aspen forest site was not completely flat) over the top of a relatively thick understory of vegetation (i.e., hazel-

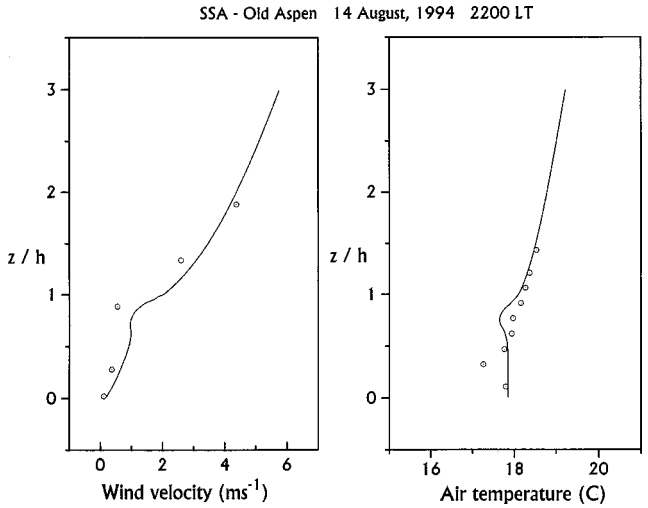


FIG. 17. Same as Fig. 15, but for 14 August 1994 at 2200 local time.

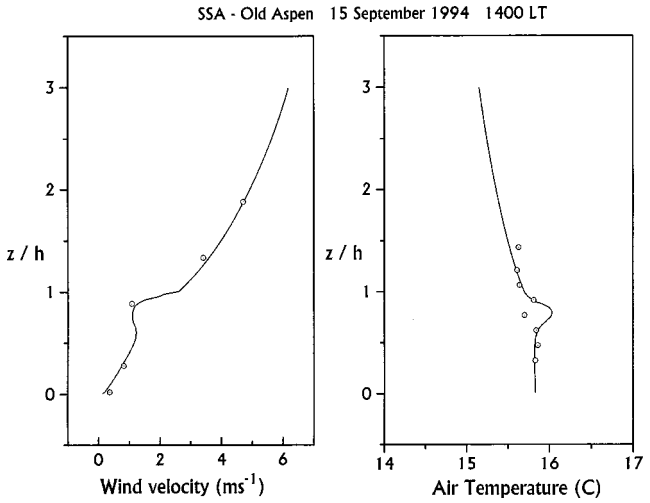


FIG. 18. Same as Fig. 15, but for 15 September 1994 at 1400 local time.

nut, about 2.0 m in height) may have provided the necessary conditions to support the observed data.

Finally, Fig. 18 shows the second-order closure model results for the SSA OA forest in comparison to (midday) micrometeorological tower data taken on 15 September 1994 at 1400 local time. The agreement is excellent between the calculated wind-velocity profile and observed data. This is an improvement over the previous (Fig. 16) daytime case. Yet again, there is an observed (temperature) data point at about $0.8h$ that is approximately 0.30 °C cooler than the modeled profile. However, in contrast to the 14 August case, values for sensible heat inside and above the canopy were found to be almost twice those recorded for latent heat. But, a trace amount of precipitation (~0.6 mm) was measured on 15 September at 1500 LT (while no measurable rain was logged on preceding days). Therefore, a large amount of evaporative cooling is not considered. Nevertheless, these evaluations compared favorably (overall) to the measured data.

[1]Variations in sound speed caused by even extreme changes in relative humidity are minimal and therefore often ignored (Noble and Marlin, 1995).
[2]The heat source (S_θ) can be expressed as a function of the leaf surface-to-ambient-air temperature differences (Meyers and Pau U, 1987), i.e., the sensible heating of the canopy sublayer.

Albertson, J. D., Katul, G. G., and Wiberg, P. (**2001**). "Relative importance of local and regional controls on coupled water, carbon, and energy fluxes," Adv. Water Resour. **24**, 1103–1118.

Albini, F. A. (**1981**). "A phenomenological model for wind speed and shear stress profiles in vegetation cover layer," J. Appl. Meteorol. **20**, 1325–1335.

Auvermann, H. J., Reynolds, R. L., and Brown, D. M. (**1995**). "Development of a multistream acoustic propagation model including scattering by turbulence," Technical Report, ARL-TR-528, U.S. Army Research Laboratory, Adelphi, MD 20783-1197. [Available from the Defense Technical Information Center, http://stinet.dtic.mil/str/index.html]

Burns, S. H. (**1979**). "The absorption of sound by pine trees," J. Acoust. Soc. Am. **65**, 658–661.

Busch, N. E. (**1973**). "On the mechanics of atmospheric turbulence," in *Workshop on Micrometeorology*, edited by D. A. Haugen (American Meteorological Society, Boston), pp. 1–65.

Campbell, G. S. (**1977**). *An Introduction To Environmental Biophysics* (Springer, New York).

Campbell, G. S., and Norman, J. M. (**1998**). *An Introduction To Environmental Biophysics*, 2nd ed. (Springer, New York).

den Hartog, G., and Neumann, H. (**2000**). "BOREAS TF-02 SSA-OA Tower Flux, Meteorological, and Precipitation Data," Available on-line [http://www.daac.ornl.gov/] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, TN.

Donaldson, C. Du P. (**1973**). "Construction of a dynamic model for the production of atmospheric turbulence and the dispersion of atmospheric pollutants," in *Workshop on Micrometeorology* (American Meteorological Society, Boston), pp. 313–392.

Finnigan, J. J., and Brunet, Y. (**1995**). "Turbulent air flow in forests on flat and hilly terrain," in *Wind and Trees*, edited by M. P. Coutts and J. Grace (Cambridge University Press, Cambridge), pp. 3–39.

Fong, M., and Srour, N. (**1994**). "Cueing of the surrogate remote sentry using an acoustic detection system," Technical Report, ARL-TR-795, U.S. Army Research Laboratory, Adelphi, MD 20783-1197. [Available from the Defense Technical Information Center, http://stinet.dtic.mil/str/index.html]

Fricke, F. (**1984**). "Sound attenuation in forests," J. Sound Vib. **92**, 149–158.

Huisman, W. H. T., and Attenborough, K. (**1991**). "Reverberation and attenuation in a pine forest," J. Acoust. Soc. Am. **90**, 2664–2677.

Inoue, E. (**1963**). "On the turbulent structure of airflow within crop canopies," J. Meteorol. Soc. Jpn. **41**, 317–326.

Kaimal, J. C., and Finnigan, J. J. (**1994**). *Atmospheric Boundary Layer Flows: Their Structure and Measurement* (Oxford University Press, New York).

Katul, G. G., and Albertson, J. D. (**1998**). "An investigation of higher-order closure models for a forested canopy," Boundary-Layer Meteorol. **89**, 47–74.

Li, Z., Lin, J. D., and Miller, D. R. (**1990**). "Air flow over and through a forest edge: A steady-state numerical simulation," Boundary-Layer Meteorol. **51**, 179–197.

Martens, M. J. M., van der Heijden, L. A. M., Walthaus, H. H. J., and van Rens, W. J. J. M. (**1985**). "Classification of soils based on acoustic impedance, air flow resistivity, and other physical soil parameters," J. Acoust. Soc. Am. **78**, 970–980.

Massman, W. J. (**1982**). "Foliage distribution in old-growth coniferous tree canopies," Can. J. For. Res. **12**, 10–17.

Mellor, G. L. (**1973**). "Analytic prediction of the properties of stratified planetary surface layers," J. Atmos. Sci. **30**, 1061–1069.

Mellor, G. L., and Yamada, T. (**1974**). "A hierarchy of turbulence closure models for planetary boundary layer," J. Atmos. Sci. **31**, 1791–1806.

Meyers, T. P., and Paw U, K. T. (**1987**). "Modeling the plant canopy micrometeorology with higher order closure principles," Agric. Forest Meteorol. **41**, 143–163.

Meyers, T. P., Finkelstein, P., Clarke, J., Ellestad, T. G., and Sims, P. F. (**1998**). "A multilayer model for inferring dry deposition using standard meteorological measurements," J. Geophys. Res. **103**, 22645–22661.

Noble, J. M., and Marlin, D. (**1995**). "User's manual for the scanning fast field program (SCAFFIP) general version 1.0," Technical Report, ARL-TR-545, U.S. Army Research Laboratory, Adelphi, MD 20783-1197. [Available from the Defense Technical Information Center, http://stinet.dtic.mil/str/index.html]

Ostashev, V. E. (**1997**). *Acoustics in Moving Inhomogeneous Media* (E&FN Spon, London), pp. 16, 67, 193–218.

Price, M. A., Attenborough, K., and Heap, N. W. (**1988**). "Sound attenuation through trees: Measurements and models," J. Acoust. Soc. Am. **84**, 1836–1844.

Rachele, H., and Tunick, A. (**1994**). "Energy balance model for imagery and electromagnetic propagation," J. Appl. Meteorol. **33**, 964–976.

Salomons, E. M. (**2001**). *Computational Atmospheric Acoustics* (Kluwer Academic, Dordrecht).

Sellers, P. J., Hall, F. G., Kelly, R. D., Black, A., Baldocchi, D., Berry, J., Ryan, M., Ranson, K. J., Crill, P. M., Lettenmaier, D. P., Margolis, H., Cihlar, J., Newcomer, J., Fitzjarrald, D., Jarvis, P. G., Gower, S. T., Halliwell, D., Williams, D., Goodison, B., Wickland, D. E., and Guertin, F. E. (**1997**). "BOREAS in 1997: Experiment overview, scientific results, and future directions," J. Geophys. Res., [Atmos.] **102**, 28731–28769.

Shaw, R. H. (**1977**). "Secondary wind speed maxima inside plant canopies," J. Appl. Meteorol. **16**, 514–521.

Shaw, R. H., and Schumann, U. (**1992**). "Large-eddy simulation of turbulent flow above and within a forest," Boundary-Layer Meteorol. **61**, 47–64.

Shinn, J. H. (**1971**). "Steady-state two-dimensional air flow in forests and the disturbance of surface layer flow by a forest wall," Technical Report, ECOM-5383, U.S. Army Electronics Command, Fort Monmouth, NJ.

[Available from the Defense Technical Information Center, http://stinet.dtic.mil/str/index.html]
Srour, N. (**1999**). ''Army acoustics needs. Sensor arrays, distributed sensors and applications,'' DARPA Air-Coupled Acoustic Micro Sensors Workshop, Crystal City, VA, 24–25 August 1999. [Available online from the Defense Advanced Research Projects Agency at http://www.darpa.mil/mto/sono/sensorarrays.html]
Tunick, A. (**2002**). ''Coupling meteorology to acoustics in forests,'' Technical Report, ARL-MR-538, U.S. Army Research Laboratory, Adelphi, MD 20783-1197. [Available online from the Defense Technical Information Center, http://stinet.dtic.mil/str/index.html]
Weiss, A., and Norman, J. M. (**1985**). ''Partitioning solar-radiation into direct and diffuse, visible and near-infrared components,'' Agric. Forest Meteorol. **34**, 205–213.
West, M., Gilbert, K., and Sack, R. A. (**1992**). ''A tutorial on the parabolic equation (PE) model used for long range sound propagation in the atmosphere,'' Appl. Acoust. **37**, 31–49.
West, M., Sack, R. A., and Walkden, F. (**1991**). ''The fast field program (FFP). A second tutorial: Application to long range sound propagation in the atmosphere,'' Appl. Acoust. **33**, 199–228.
Wilson, D. K. (**1993**). ''Sound field computations in a stratified, moving medium,'' J. Acoust. Soc. Am. **94**, 400–407.
Wilson, D. K. (**1998**). ''A prototype acoustic battlefield decision aid incorporating atmospheric effects and arbitrary sensor layouts,'' Technical Report, ARL-TR-1708, U.S. Army Research Laboratory, Adelphi, MD 20783-1197. [Available online from the Defense Technical Information Center, http://stinet.dtic.mil/str/index.html]
Wilson, D. K. (**2000**). ''A turbulence spectral model for sound propagation in the atmosphere that incorporates shear and buoyancy forcings,'' J. Acoust. Soc. Am. **108**, 2021–2038.
Wilson, J. D. (**1988**). ''A second-order closure model for flow through vegetation,'' Boundary-Layer Meteorol. **42**, 371–392.
Wilson, J. D., Finnigan, J. J., and Raupach, M. R. (**1998**). ''A first-order closure for disturbed plant-canopy flows, and its application to winds in a canopy on a ridge,'' Q. J. R. Meteorol. Soc. **124**, 705–732.
Wilson, N. R., and Shaw, R. H. (**1977**). ''A higher order model for canopy flow,'' J. Appl. Meteorol. **16**, 1197–1205.
Wong, G. S. K., and Embleton, T. F. W. (**1984**). ''Variation of specific heats and of specific heat ratio in air with humidity,'' J. Acoust. Soc. Am. **76**, 555–559.
Wyngaard, J. C. (**1973**). ''On surface-layer turbulence,'' in *Workshop on Micrometeorology* (American Meteorological Society, Boston), pp. 101–149.

Evaluation of rosette infrasonic noise-reducing spatial filters

Michael A. H. Hedlin[a)]
Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0225

Benoit Alcoverro
Commissariat a l'Energie Atomique, Departement Analyse et Surveillance de l'Environnement, BP 12, 91680 Bruyeres le Chatel, France

Gerald D'Spain
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0225

(Received 13 June 2002; revised 10 June 2003; accepted 12 June 2003)

This paper presents results from recent tests of rosette infrasonic noise-reducing spatial filters at the Pinon Flat Observatory in southern California. Data from 18- and 70-m aperture rosette filters and a reference port are used to gauge the reduction in atmospheric wind-generated noise levels provided by the filters and to examine the effect of these spatial filters on spatially coherent acoustic signals in the 0.02- to 10-Hz band. At wind speeds up to 5.5 m/s, the 18-m rosette filter reduces wind noise levels above 0.2 Hz by 15 to 20 dB. Under the same conditions, the 70-m rosette filter provides noise reduction of up to 15 to 20 dB between 0.02 and 0.7 Hz. Standing wave resonance inside the 70-m filter degrades the reception of acoustic signals above 0.7 Hz. The fundamental mode of the resonance, 15 dB above background, is centered at 2.65-Hz and the first odd harmonic is observed at 7.95 Hz in data from the large filter. Analytical simulations accurately reproduce the noise reduction and resonance observed in the 70-m filter at all wind speeds above 1.25 m/s. Resonance theory indicates that internal reflections that give rise to the resonance observed in the passband are occurring at the summing manifolds, and not at the inlets. Rosette filters are designed for acoustic arrivals with infinite phase velocity. The plane-wave response of the 70-m rosette filter has a strong dependence on frequency above 3.5 Hz at grazing angles of less than 15° from the horizontal. At grazing angles, complete cancellation of the signal occurs at 5 Hz. Theoretical predictions of the phase and amplitude response of 18- and 70-m rosette filters, that take into account internal resonance and time delays between the inlets, compare favorably with observations derived from a cross-spectral analysis of signals from the explosion of a large bolide. [DOI: 10.1121/1.1603763]

PACS numbers: 43.28.Dm, 43.28.Ra, 43.50.Cb [LCS]

I. INTRODUCTION

A. Infrasound and the new global monitoring system

A natural consequence of any global nuclear-test ban treaty is the need for global monitoring for clandestine nuclear testing activity. In the case of the recent Comprehensive Nuclear-Test-Ban Treaty, which bans nuclear explosions of any yield, four monitoring networks will be used—seismic, hydroacoustic, air acoustic (or infrasound), and radionuclide as part of an International Monitoring System (IMS). This study relates to the infrasound component of the IMS. Each of the 60 arrays planned for the infrasound network will comprise up to 8 sensors.

Because of wind-generated turbulence, the atmosphere is inherently noisy at frequencies of interest to the nuclear monitoring community (between 0.01 and 10 Hz). As a result, it is necessary to suppress infrasonic wind noise at each station in the infrasound network. IMS infrasound stations are currently being equipped with spatial noise-reducing systems (Daniels, 1959; Burridge, 1971; Grover, 1971; Alcoverro, 1998) similar to those depicted in Fig. 1. The infrasound network is to provide enough sensitivity to allow discrimination of nuclear tests from myriad other atmospheric events worldwide. Reduction of atmospheric noise is therefore of paramount interest to the monitoring community.

B. Infrasonic signals

The physics of propagation of sound through the atmosphere defines, to a large degree, the noise-reduction problem. Since the era of atmospheric testing of nuclear weapons, it has been known that the shock front from a nuclear explosion will evolve into infrasonic energy that propagates efficiently through the lower atmosphere (Landau and Lifshitz, 1959). Buried nuclear tests cause a piston-like vertical ground motion that also produces infrasonic pressure waves (Blanc, 1985; Calais *et al.*, 1998). Much of the acoustic energy from an infrasound source in the lower atmosphere (0–50-km altitude) is refracted back to the Earth's surface in the stratosphere (between ~20- to 50-km altitude) or at lower altitudes due to wind shear (Simons, 1995). A Lamb surface wave exists principally below ~30-km altitude (Lamb, 1932;

a)Electronic mail: hedlin@ucsd.edu

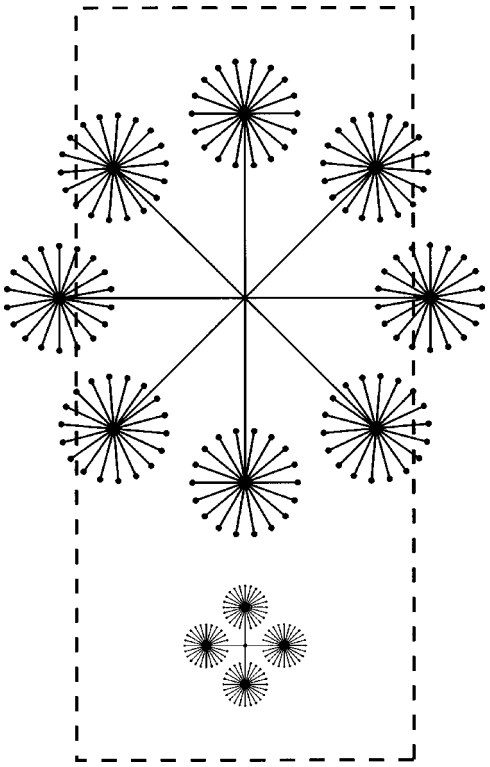


FIG. 1. Two rosette filters considered in this paper are shown to scale with a National Football League playing surface. The 18-m filter comprises 92 low-impedance inlets in 4 rosettes. The 70-m filter comprises 144 inlets arranged in 8 rosettes. In both filter designs, the inlets are connected by solid pipe to secondary summing manifolds at the center of each rosette. The secondary manifolds are connected to a primary manifold which is connected by a short pipe to the sensor. Each inlet is located the same distance along solid pipe from the sensor. Signals arriving from directly above the filter are summed in phase. Adjacent ports in the 70-m filter are separated by 2.79 m. Adjacent ports in the 18-m filter are 0.85 m apart.

Francis, 1973) and can be observed globally (Blanc, 1985). Most signals propagate across Earth's surface with phase velocities below 360 m/s (McKisic, 1997). These phase velocities are consistent with an arrival direction within 15° of the horizontal; i.e., near-grazing arrival angles are most common, particularly for long-range sources. The frequency content of infrasonic energy is strongly range dependent due to atmospheric absorption (Blanc, 1985). Most signals of interest to the monitoring community lie between 0.01 and 10.0 Hz (Zumberge *et al.*, 2003). At these frequencies, signals from remote sources are highly coherent at spacings of tens to hundreds of meters. Noise that results from atmospheric turbulence near the recording site is incoherent at spacings of tens of meters (Priestley, 1966).

C. Noise suppression

Although it is generally accepted that the study of acoustic signals in the atmosphere requires suppression of noise due to atmospheric turbulence, how this is best accomplished is a matter of vigorous debate. Spectral filtering is generally ineffective as turbulent noise spans the entire frequency band of interest to the monitoring community (Kaimal and Finnigan, 1994). It has been recognized since the 1950's that the disparate correlation lengths of most signals and noise can be exploited to improve the ratio of signal to noise. The prototypical spatial filter, proposed by Daniels (1959), consists of a single 603-m-long tapered pipe. The pipe was tapered to eliminate internal acoustic reflections back to the sensor. Acoustic energy entered the Daniels pipe via N high-impedance inlets distributed uniformly along its length. Assuming spatially incoherent noise, and perfectly spatially coherent signal, the amplitude ratio of coherent signal to incoherent noise is predicted to improve by $N^{1/2}$. The filter with 100 inlets was found to reduce atmospheric noise at the theoretical limit of 20 dB at wind speeds up to 13 m/s (Daniels, 1959), suggesting perfectly uncorrelated noise between the inlets. Although proven to be effective at improving the reception of signals, the Daniels line microphone is inherently directional if the wavelength of the signal is less than the sensor length (i.e., near 1 Hz or above).

With the possibility of clandestine nuclear testing at any azimuth from a recording station, subsequent designs use several relatively short pipes distributed in azimuth (Davidson and Whitaker, 1992). Alcoverro (1998) proposed a filter which consists of several clusters, or rosettes, of low-impedance inlets. Each inlet in a cluster is connected to a ''secondary'' summing manifold by a solid pipe. A solid pipe connects each secondary summing manifold to one primary summing manifold or directly to the sensor (Fig. 1).

D. Rationale for the present study

As development of the IMS global infrasound network surges ahead, there is a pressing need for intercomparisons of filter designs currently in use at IMS arrays, and between existing and new designs that have been proposed for use at future array sites. It is necessary to assess the utility of these filters for reducing noise due to atmospheric turbulence while preserving the coherent signals of interest to the nuclear monitoring and scientific communities. In this paper, we present our findings regarding the utility of 18- and 70-m rosette filters at attenuating noise while preserving signal. The large, 70-m, rosette filters have been built at IMS infrasound array sites near Warramunga, Australia and at Pinon Flat, California. Empirical observations of noise reduction are compared with theoretical predictions. We point out some limitations of the rosette filter design and suggest possible avenues for improving the performance of these filters.

II. THE PINON INFRASOUND TEST BED

All of our experiments have been conducted at the infrasound test bed at the Pinon Flat Observatory in southern California (Figs. 2 and 3). The observatory is located in the Anza Borrego desert in California's coastal ranges 125 km to the northeast of San Diego and 50 km to the southwest of Palm Springs. The observatory is useful for studies of infrasonic noise-reducing systems as there are strong diurnal and seasonal changes in meteorological conditions at this site. Wind speeds are often near zero at night and sometimes exceed 15 m/s during the day. Winds are most commonly from the northeast or the southwest. The desert is sparsely vegetated with Pinon pine trees and there is essentially no low-lying ground cover. The observatory is accessible at all times of the year. The International Monitoring System (IMS) in-

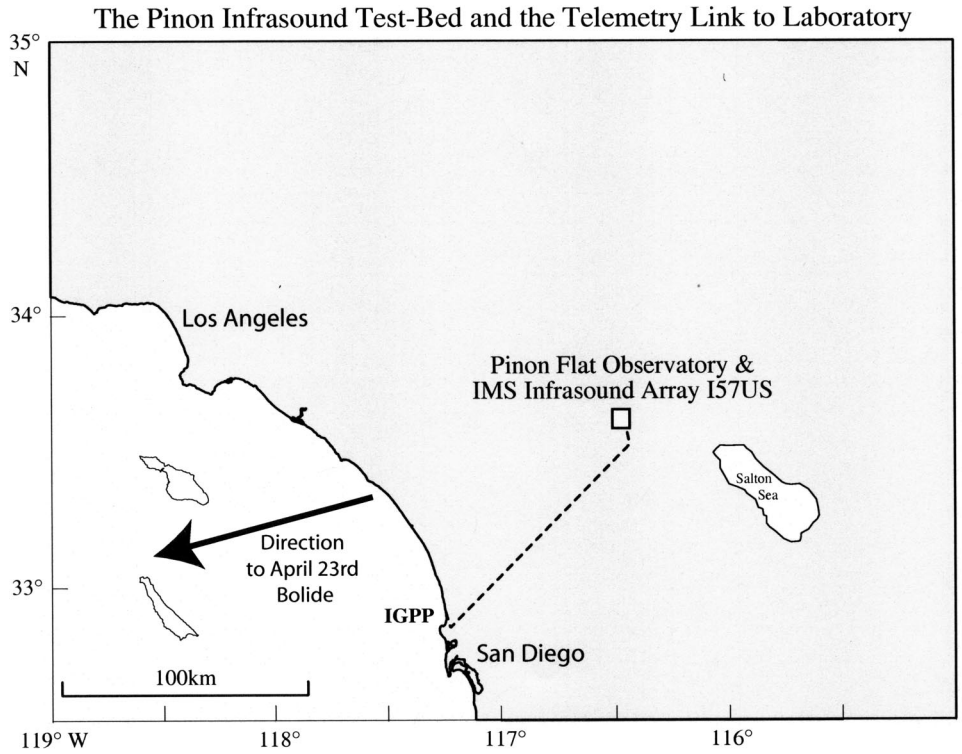


FIG. 2. The IMS infrasound array, 157US, is located in the Anza Borrego desert at the Cecil H. and Ida M. Green Pinon Flat Observatory (PFO). The infrasound test bed is located at PFO. The real-time radio-telemetry link to the laboratory (IGPP) is also shown. Signals from a large bolide that exploded to the SW of the observatory are used in this paper to calibrate the rosette filters.

frasound array I57US is located at PFO (Fig. 3). The 2-km aperture array comprises eight instruments and includes both 18-m noise-reduction rosette filters (designed to detect signals between 0.2 and 5 Hz) and 70-m-aperture filters (sensitive to low-frequency signals between 0.02 and 3 Hz). An IMS auxiliary three-component seismic station is located at the northwest corner of the observatory. There is essentially no cultural noise in the area.

III. EXPERIMENTAL LAYOUT

A typical experimental layout is pictured in Fig. 4. At each site we use the MB2000 aneroid microbarometer fabricated by the French Departement Analyze et Surveillance de l'Environnement (DASE) and now by Tekelec. These sensors provide an electrically filtered signal between 0.01 and 27 Hz (low-passed below 9 Hz by our antialiasing filter) with an adjustable sensitivity. The version used in our experiments had a sensitivity of 20 mV/Pa. The electronic noise of the sensor is 2 mPa rms, between 0.02 and 4 Hz, which is well below measured noise levels at PFO. Each sensor was placed in an insulating case and was attached to one of three systems. In total, this experiment used five sensors. Three sensors were each attached to a single low-impedance inlet located 5 cm above the ground and provided the spatially unfiltered reference data. One sensor was located near the auxiliary seismic station (Fig. 3). The two other reference sites were colocated with the 18-m rosette filter at site H1 and the 70-m rosette filter at L2 (Fig. 3). The two other sensors were attached to the rosette filters at H1 and L2. The ''reference'' sensor near the seismic station, and the sensors at H1 and L2 were colocated with an ultrasonic anemometer for wind velocity, air temperature, and humidity sensors. The temperature and humidity sensors were located 1 m above the ground. The wind sensor was 2 m above the ground at all sites. In this experiment, the infrasound and meteorological signals were digitized at 20 and 1 sps, respectively, using a 24-bit Reftek 72A-08 data acquisition system. Each system was powered by solar energy or by buried power lines. Data from three sites were transmitted in real time via a 2.2-GHz spread-spectrum radio link to our laboratory in La Jolla (Fig. 2). Data from two of the sites were transmitted along telephone lines. The experiments were preceded by calibration tests in the field at PFO, and in the laboratory at the Institute of Geophysics and Planetary Physics in La Jolla, California. The tests were conducted to ensure all field systems were robust and yielded equal digitized signals for equal input.

IV. OBSERVATIONS OF INFRASONIC NOISE REDUCTION

The first goal of this paper is to compare the capability of the different rosette filters at reducing infrasonic noise levels. To make this comparison, the filtered pressure records at each site are divided into 15-m intervals (e.g., Fig. 5). A Welch time-averaged pressure spectrum (Welch, 1967) is calculated from each 15-min segment of data. The Welch estimator reduces the variance of the power spectral estimate relative to a periodogram by averaging several spectral estimates taken from overlapping segments of data. The test of the rosette filters spanned more than 768 h and thus yielded 3074 15-min nonoverlapping segments of data. The pressure spectral estimates are binned by wind speed and stacked to reduce the statistical uncertainty inherent in the spectral estimation process. The stacked spectra clearly reveal the dependence of average noise levels on wind speed. Although wind-speed data are collected at all sites, for the purpose of comparing data collected at the same time at the different

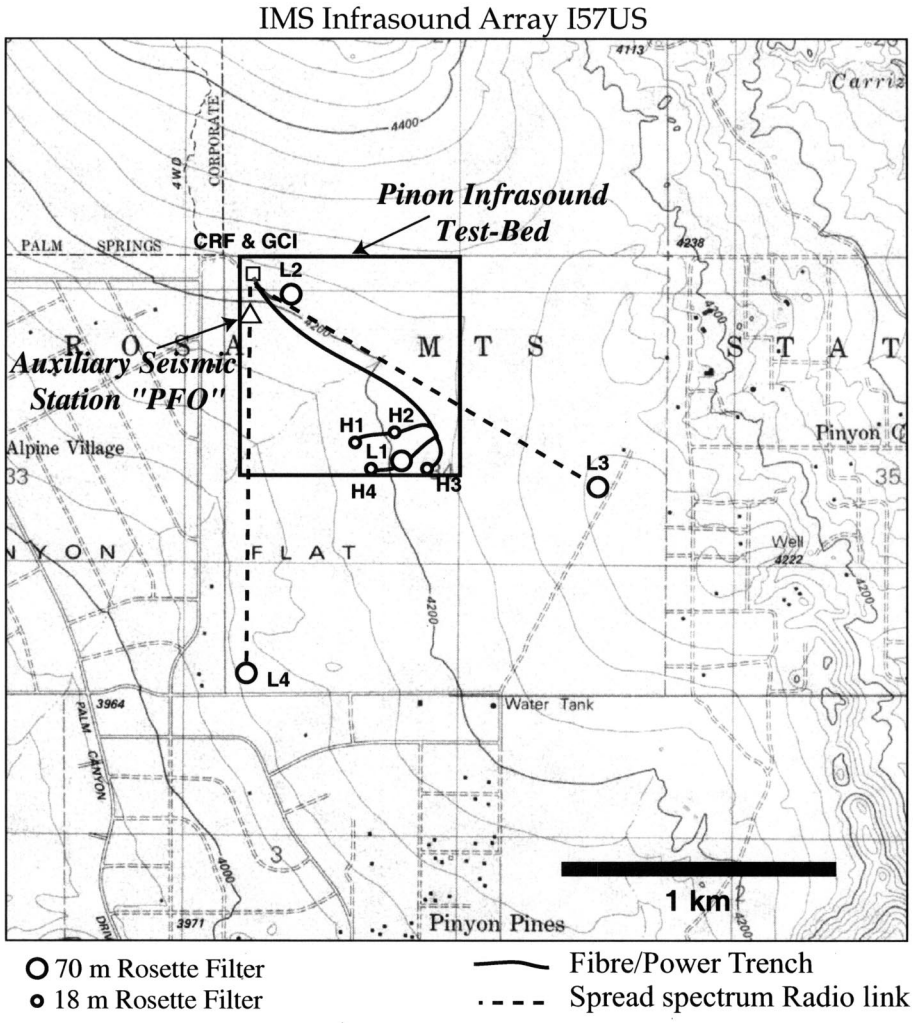


FIG. 3. The eight-element infrasound array at Pinon Flat, California comprises four long-period elements (L1 through L4) in a centered triangle spanning less than 2 km and four short-period elements (H1 through H4) in an irregular quadrilateral near the center. The terrain slopes down gently to the southwest. The contour interval is 40′.

sites, only the wind-speed data collected at 2 m at the reference port is used in the binning process. The wind speeds at the 18-m filter were reduced somewhat by the forest (Fig. 6). The 15-min average wind speeds plotted in this figure show clearly the diurnal variations in the wind speed at all sites. The experiment began during adverse weather conditions. Data were lost near day 10 due to a fault in the telemetry system.

The spectral estimates are binned into 0.5-m/s wind-speed intervals starting at 0.0 m/s (Figs. 7 and 8). The lowest curve in each panel in Fig. 7 corresponds to estimates taken when the wind at the reference site was less than 0.5 m/s. At the three sites, we see infrasonic noise increases at all frequencies with increasing wind speed. The microbarom peak (the spectral peak near 0.2 Hz) is most clearly resolved by the 70-m rosette filter. The peak is seen clearly at wind speeds up to 1.5 to 2.0 m/s and in some individual spectra at wind speeds up to 2.5 m/s. The effectiveness of this filter at reducing infrasonic noise below 1 Hz is evident in Fig. 8. Direct comparison of the stacked spectra taken from the reference port near the seismic station and 70-m rosette data streams indicates reduction of noise by 15 to 20 dB at frequencies below 1.0 Hz. The 18-m rosette filter averages noise over a smaller area and is less effective at low frequencies. The noise-reduction performance of the 18-m filter, however, surpasses the 70-m filter at frequencies above 0.5 Hz.

The stacked spectra in Fig. 7 indicate that the 18-m rosette filter is ineffective at average wind speeds (3.0 to 3.5 m/s) at frequencies below 0.1 Hz. Above 0.5 Hz, the noise reduction is near the theoretical limit for uncorrelated noise of 19.8 dB ($10 \log 10[1/N]$, where N is 96). At wind speeds above 2 m/s, and at frequencies above 1 Hz, the apparent noise reduction of the 18-m filter exceeds this limit. The reference inlet for this experiment was not colocated with the 18-m filter but was located in an area with no vegetation. The 18-m filter was surrounded by Pinon pine trees. As a result, high-frequency noise levels at the reference site are slightly higher, giving the appearance of greater noise reduction than actually occurred. The binned spectra reveal a clear increase in the frequency at which the 18-m filter reduces noise levels with increasing wind speed. At 1.0 to 1.5 m/s, noise reduction is observed at 0.05 Hz. By 5.0 to 5.5 m/s, noise reduction is not observed until 0.2 Hz.

Chi-squared statistics were calculated for the stacked spectra but are not shown here. The bins contain enough stacked spectra that the uncertainty in the spectral levels is insignificant relative to the observed differences in the noise levels between the noise-reduction systems.

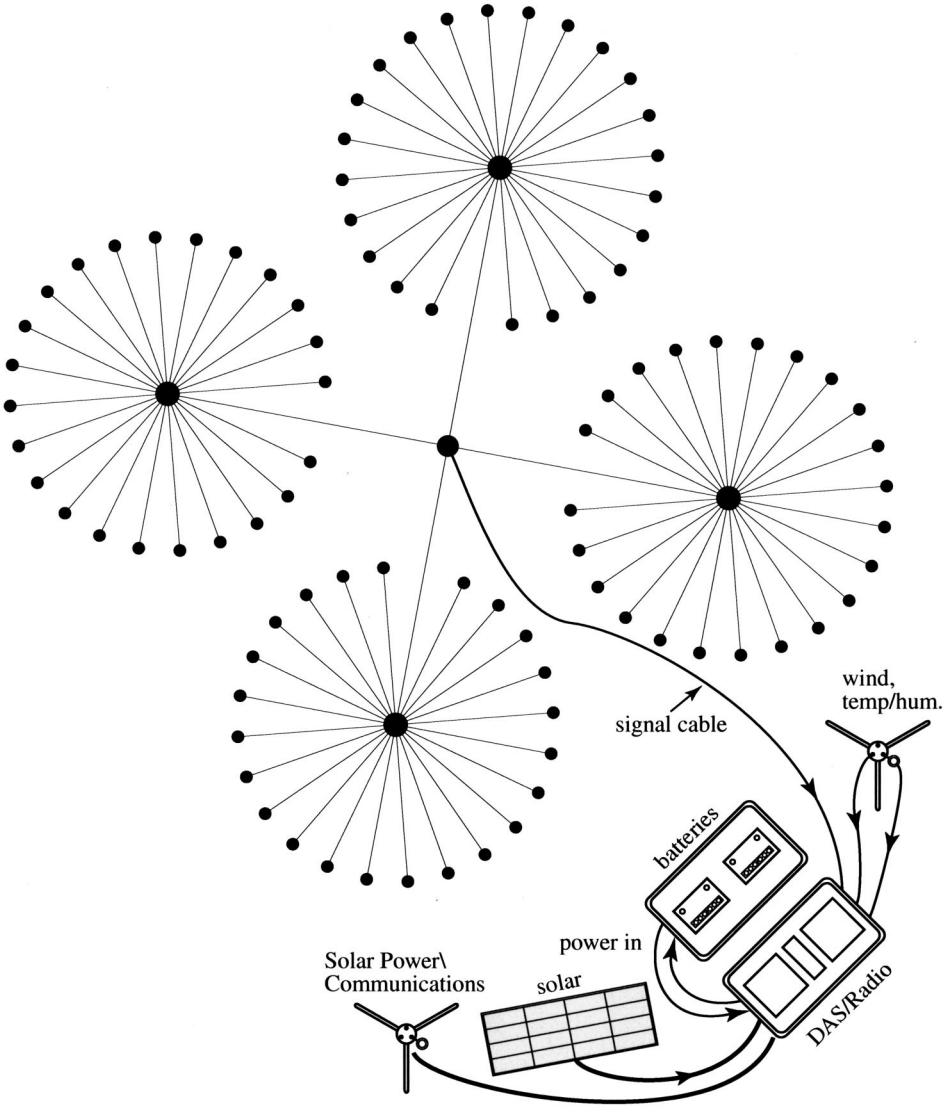


FIG. 4. A typical experimental recording site consists of a filter and a sensor. In this example, the sensor is represented by the black dot at the center of the filter. The digitizer and spread-spectrum radio are located in one enclosure, batteries are in a second. Wind speed and direction are sensed at 2-m elevation; humidity and temperature data are collected at 1-m elevation on the same mast. Data are transmitted in real-time via a 2.2-GHz spread-spectrum link back to the laboratory in La Jolla (Fig. 2). Solar energy powers all systems.

V. RESONANCE IN ROSETTE FILTERS

A. Observations of resonance

The stacked spectra reveal the effectiveness of the filters at reducing low-frequency infrasonic noise. The spectra also plainly show evidence of resonance within the rosette filters. Resonance in the 18-m filter is relatively minor and is only evident above 3.0 Hz near the Nyquist frequency. The resonance peaks in data from the 70-m filter are significant, however, and lie well within the frequency band of interest to the nuclear monitoring community. Resonance peaks are observed at 2.65 and 7.95 Hz. The second resonance peak is observed at 3.00 times the frequency of the first peak. The resonance peaks are nearly independent of wind speed (Fig. 9) and temperature. From 0 to 35 °C, the frequency of the first peak rises 1.5%, the second rises 0.8%. The very weak dependence on air temperature is expected as the pipes are buried and are therefore not exposed to the larger changes in air temperature recorded by the temperature sensor.

Resonance in the 70-m filter is seen at all wind speeds. As shown in Fig. 10, at a time when wind speeds were <0.25 m/s, the constant high-frequency resonance obscures the infrasonic pressure variations recorded via a nonresonant spatial filter. This nonresonant microporous hose filter spanning an area 70-m across is described more fully in Herrin *et al.* (2001).

B. Modeling resonance peaks in the 70-m filter

Reflections will occur inside the rosette pipe system at points where the acoustic impedance changes. These points are located at the inlets, all summing manifolds, and at the sensor. In the 70-m rosette filter tested in this experiment, the segment between the sensor and the primary summing manifold is 2.85 m long. At 27 °C the speed of sound in free space is 347 m/s. The speed of sound in narrow conduits 8–10 mm in radius at frequencies between 0.1 and 10 Hz is difficult to estimate as in this frequency band neither the large-tube or small-tube approximations of Benade (1968) are considered to be highly accurate. The dimensionless parameter r_v (which gives the ratio of the conduit radius and the viscous boundary layer thickness) ranges from 1.9 (at 0.1 Hz) to 19 (at 10 Hz). The two approximations place the propagation speed of sound at frequencies from 1 to 10 Hz between 288

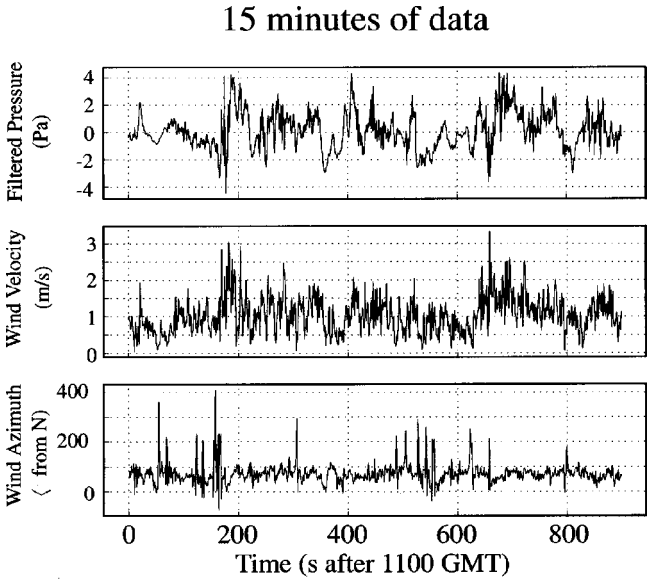


FIG. 5. Atmospheric pressure and wind data from a 15-min interval starting at 11:00 GMT at the reference site. The pressure data are filtered between 0.01 and 9.0 Hz and are sampled at 20 sps. The meteorological data are collected once per second.

and 324 m/s. The average of the two estimates is 306 m/s. Assuming this velocity for the propagation of sound waves in the pipe, the natural frequency of resonance along this segment lies far beyond the Nyquist frequency and could not produce the spectral peaks in the data. The resonance that gives rise to the spectral peaks observed in Fig. 8 occurs in the rosettes between the inlets and the secondary summing manifolds or between the primary and secondary manifolds. The rosette filter resonates at the fundamental and the first odd harmonic (at 2.65 and 7.95 Hz, respectively). Presumably, higher overtones would be observed if the data were sampled more rapidly than 20 sps. The resonance in the 70-m rosette filter is akin to organ-pipe resonance. Assuming two-way length of 54 m (double the distance between the primary and secondary summing manifolds), and a propagation velocity of 306 m/s, we would expect the fundamental mode at 2.83 Hz and the first odd overtone centered at 8.49 Hz. The closeness of these estimates to the observed values strongly suggests that the resonance that gives rise to the observed spectral peaks occurs between the primary and secondary summing manifolds. We can rule out resonance within the rosettes as the cause of the observed spectral modulations as the fundamental resonance frequency for resonance between the secondary manifolds and the inlets is predicted to lie at 9.5 Hz.

More exact modeling of the response of the 18- and 70-m rosette filters across the entire frequency band of interest to the monitoring community requires consideration of the boundary conditions at the reflection points, attenuation of propagation of sound through all elements of the rosette pipe system, and a statistical characterization of the noise field across the area spanned by the filter. This has been done by Alcoverro (1998) and will be considered in the next section.

VI. WIND-NOISE REDUCTION SIMULATIONS

As described in Alcoverro (1998), the use of nonporous hoses in the rosette noise-reduction systems allows a precise calculation of their frequency response. The calculation uses an electro-acoustic analog model for all elements constituting the system including the inlet ports, the pipes, the manifolds, and the sensor. The use of a dissipative transmission line model for pipes (Keefe, 1984) permits the introduction of viscothermal losses and time propagation in the calculation. An electro-acoustic analog schematic is generated for each noise-reduction system and the use of the matrix impedance of the schematic gives an easier computation of the pressure at each node of the circuit. The transfer function between each inlet and the output voltage of the sensor is calculated in the frequency domain. In the case of the rosette noise filters, the original principle presented in Alcoverro

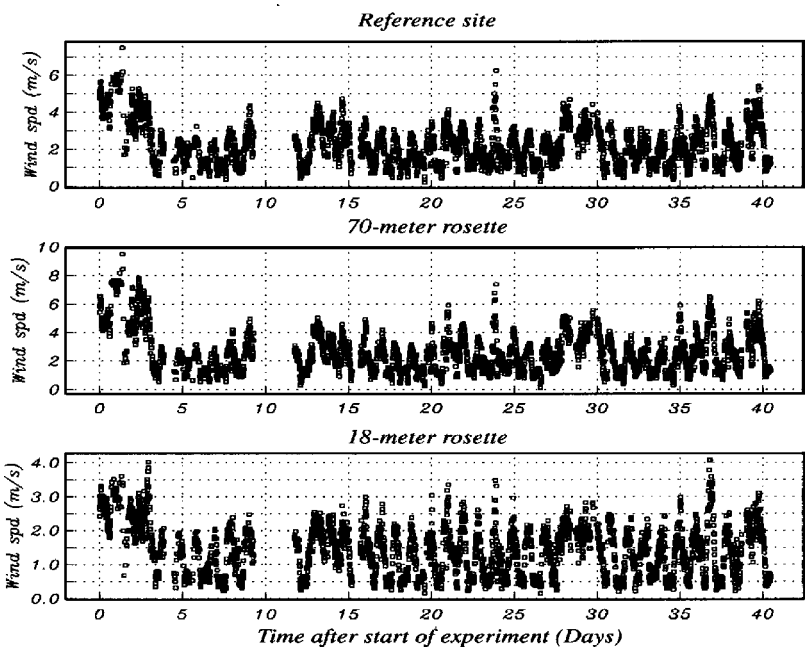


FIG. 6. Wind-speed data collected at the three sites. The panels from top to bottom represent data collected at the reference port, 70-m rosette filter and the 18-m rosette filter. Each symbol represents an average wind speed over a 15-min time interval. The experiment was interrupted briefly on day 10 by a telemetry fault. The experiment began during adverse weather conditions.

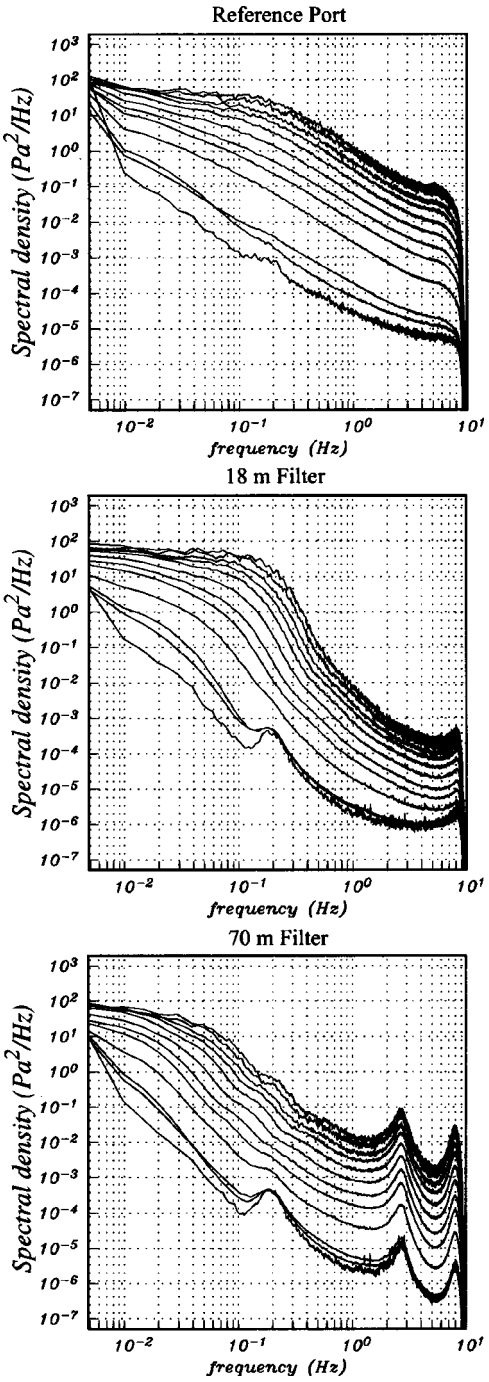


FIG. 7. The three panels in this figure show binned and stacked spectra taken at the reference port (top), 18-m rosette filter (middle) and 70-m rosette at wind speeds ranging from 0.0 to 0.5 m/s (lowest curve in each panel) up to 5.5 m/s in intervals of 0.5 m/s. The lowest curve in each panel represents spectral power at wind speeds below 0.5 m/s. The panels reveal a gradual increase in noise levels at all frequencies with wind speed at all sites. The microbarom peak is seen clearly near 0.2 Hz at all sites at wind speeds below 0.5 m/s, but is only seen in the filtered data at times when the wind speeds are above 0.5 m/s. Significant resonance peaks are seen above 1.0 Hz in the data from the 70-m filter.

(1998) gives an identical response for all inlets to ensure a perfect summation. In this case, the calculations are simplified and only one single-inlet transfer function needs to be calculated. The amplitude and phase response for 18-m/96-port and 70-m/144-port rosette filters is presented in Fig. 11. This simulation shows the frequency and amplitude for resonant peaks. In the case of a 70-m/144-port rosette filter, the first resonant peak is predicted at 2.69 Hz and the second at 8.08 Hz. It is essential that the sensor forevolume is considered in this calculation. A large forevolume acts like an active capacitance which would change the response of the overall system. The calculations presented in this paper used the known forevolume for the MB2000 sensor. The amplitude of the first resonance peak is estimated by comparing the measured noise spectra amplitudes between the this peak (at 2.65 Hz) and at 0.3 Hz. The amplitude difference, as seen in Fig. 12, is 17.9 dB. The same estimate, made from simulated noise spectra, gives an amplitude difference of 17.6 dB. The comparison between the measurements and the accurate electro acoustic calculation appears to validate the theory.

The bandpass amplitude of one inlet transfer function is equal to the inverse of the number of inlets. After summing a coherent wave passing over the noise reducer, the amplitude is equal to unity. These systems do not reduce the noise at frequencies lower than the first resonant peak. The frequency and amplitude of resonance, mainly induced by the length of the pipes, could be verified practically on the noise spectra measured with a real system.

The simulation of the noise reduction requires a wind model. The measurement of the coherence function between the noise generated by the wind at two points, separated by d meters, shows a linear increase of the coherence cutoff frequency as the wind speed increases, or the separation distance decreases. At low frequencies, the noise measured at the two points is coherent but not at high frequencies. The relation between the coherence cutoff frequency f_c, the mean wind speed v_m, and the separation distance d, is $f_c = 0.1 v_m / d$ (Alcoverro, 1998). A separate realization of random noise high-pass filtered above f_c is created for each inlet. The cutoff frequency is f_c, to respect the coherence criteria between these inlet and the others, at a given mean wind speed (Alcoverro, 1998). The synthetic noise generated for each inlet is convolved with the delayed impulse response of each transfer function (calculated in the frequency domain). The delays are determined by the geometric position of the inlet and the main wind direction. The global response of the system to a wind-speed condition is calculated by summing all convolutions. As the synthetic noise is not representative of the amplitude of noise, the results must be presented as the difference between the spectra of the noise applied at one inlet (noise generated by the wind at one point) and the spectra of the global response. The simulated noise reduction of the 70-m/144-rosette filter is presented in Fig. 12 for various mean wind speeds between 0.75 and 5.25 m/s.

Predicted noise reduction as a function of wind speed compares favorably with observed levels. The observed values are shown in the lower panel of Fig. 12. At all wind speeds, except those at or below 1.25 m/s, the theory accurately predicts the amount of noise reduction. The corner frequency of the noise reduction increases with increasing wind speed as predicted by theory. The location and width of the resonance peak is predicted accurately by theory. The location of the peak is determined by phase shifts and by the delay time between successive reflections inside the pipes.

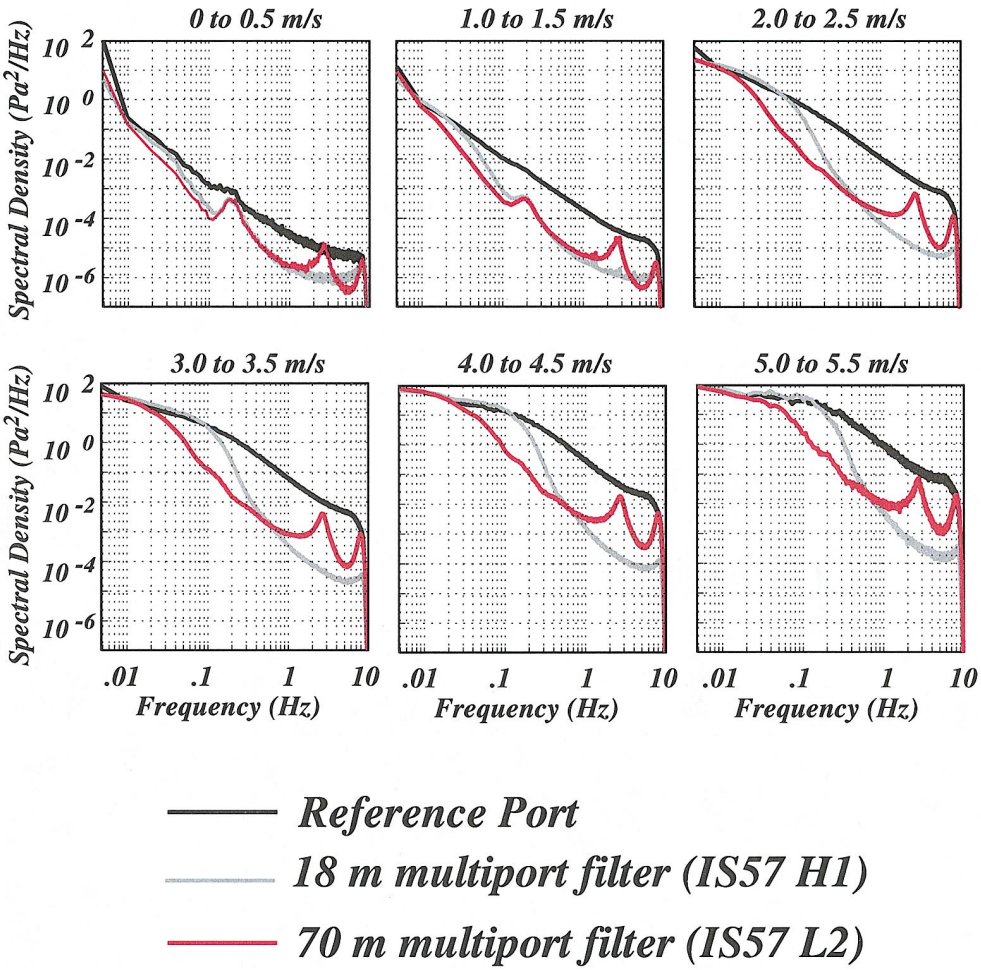


FIG. 8. A direct comparison of stacked spectra from the three different systems shows the growth of infrasonic noise levels at all sites with increasing wind speed. The increase in noise power with increasing wind speed is most significant at frequencies above 0.1 Hz. The 70-m rosette filter provides the most significant noise reduction at frequencies below 0.5 Hz. Large peaks above 1.0 Hz are seen only in the data from the 70-m filter. These peaks are not from the natural environment but are an artifact of the large filter.

The significant difference between observed and predicted noise levels near the microbarom peak at low wind speeds (1.25 m/s or less) stems from a higher level of coherence in the noise than is allowed for in the noise model.

VII. CALIBRATION OF ROSETTE FILTERS USING ACOUSTIC SIGNALS

A. Plane-wave response

Ambient noise, that has little or no coherence between inlets, can be used to study the effect of internal resonance on recorded data. Incoherent noise cannot, however, be used to calibrate the response of the filter to incoming plane waves. This response depends on time delays between the inlets in addition to resonance of the energy once it enters the pipe system. The rosette array is designed for vertically incident energy. Neglecting the resonance and dispersion inside the narrow pipes (Benade, 1968) that connect the inlets to the microbarometer, the filter should cause minimal distortion of signals arriving from directly above. The time delays between inlets become significant as the phase velocity of the signal decreases and approaches the ambient sound speed (i.e., when the signal arrives at grazing angles above the horizontal). For example, a signal that arrives with a phase velocity of 347 m/s (the sound speed at an air temperature of 27 °C) crosses an 18-m rosette filter in 0.05 s and a 70-m rosette filter in 0.2 s. Although the maximum phase delay for such energy recorded by an 18-m rosette is a quarter of the period of 5-Hz energy, this time delay is significant if the energy is recorded by a 70-m rosette.

Some atmospheric phenomena are known to produce signals with very high phase velocities. One notable example is aurora, which are known to produce acoustic waves at

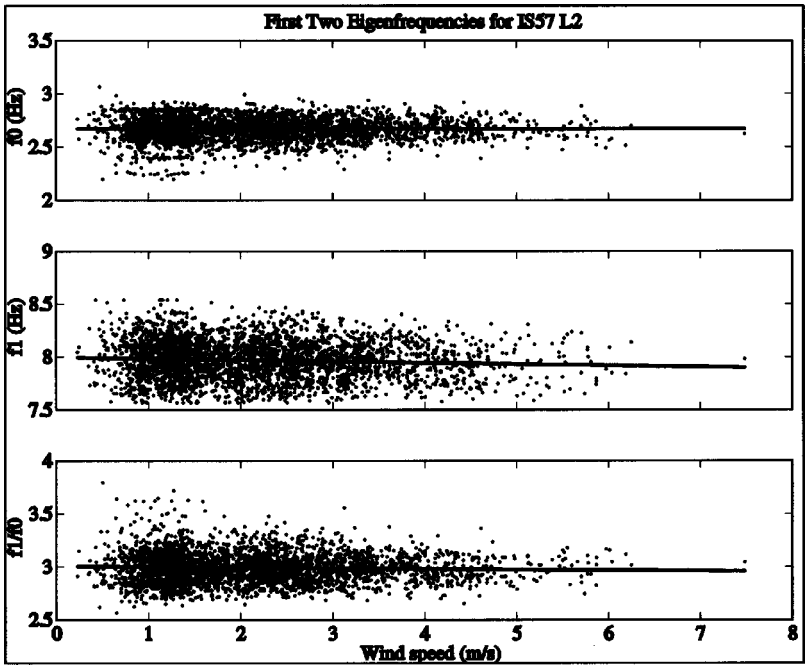


FIG. 9. Figures 7 and 8 reveal two significant peaks above 1.0 Hz in the spectra taken from the 70-m rosette filter. The center frequency of the two peaks in each of the 3074 spectral estimates is plotted in the upper two panels as a function of wind speed. The ratio of the two frequencies is given in the lower panel. The scatter observed in all panels is due to the inherent instability of the individual, unstacked, spectral estimates. We observe essentially no dependence of the peaks on wind speed. The ratio of the two frequencies is almost exactly 3.0.

phase velocities between 400 and 1200 m/s (McKisic, 1997). From ray theory, the phase velocity of an arrival equals the phase velocity at its turning point. As a result, high-phase velocity arrivals turn in the thermosphere; however, such arrivals are highly attenuated. Signals from most manmade sources and natural atmospheric phenomena in the lower atmosphere turn in the stratosphere and propagate across Earth's surface at phase velocities below 360 m/s (McKisic, 1997). At an air temperature of 27 °C, a phase velocity of 360 m/s is consistent with an arrival angle of 15° from the horizontal.

The plane-wave response of 18- and 70-m rosette filters is given in Fig. 13. As the figure shows, the 70-m filter response has a strong dependence on frequency for near-grazing angles of incidence. This frequency dependence is caused by destructive interference due to the relative time delays associated with propagation across the horizontal aperture of the space filters. At 5 Hz, the relative phase delays associated with the time delays of horizontal propagation

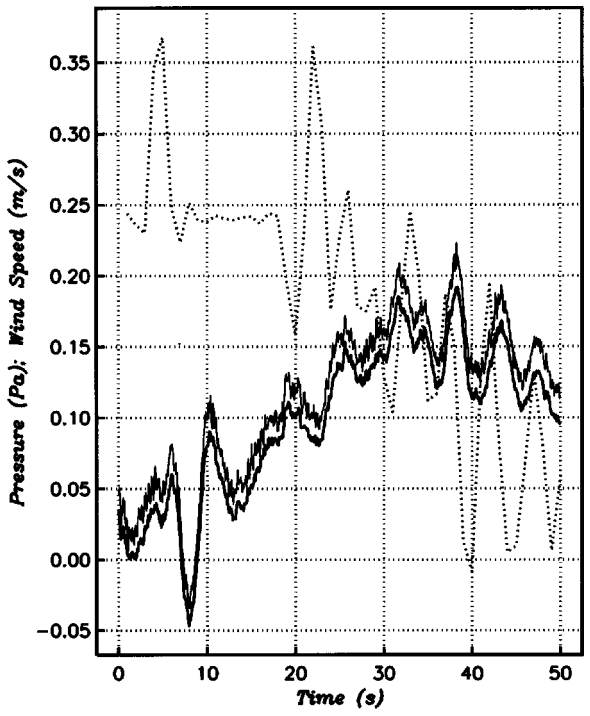


FIG. 10. The resonance occurring in the 70-m filter (upper solid curve) obscures much of the detail that is seen clearly in data from a colocated filter constructed of microporous hoses (lower solid curve beneath the black curve). The coarsely dashed curve shows the wind speed during the 50-s interval. This example shows that the resonance is significant even at times of essentially no wind.

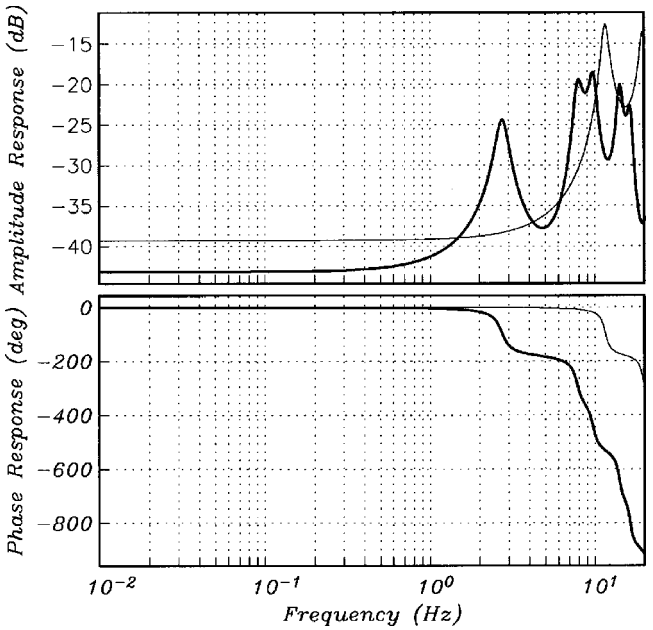


FIG. 11. Predicted amplitude and phase response of the 18-m/92-rosette filter (light curve) and the 70-m/144-rosette filter (bold curve) for one inlet. The resonance peaks coincide with significant change in the phase response of the filter. The long-period response is given by $-20\log_{10}(N)$, where N is the number of inlets.

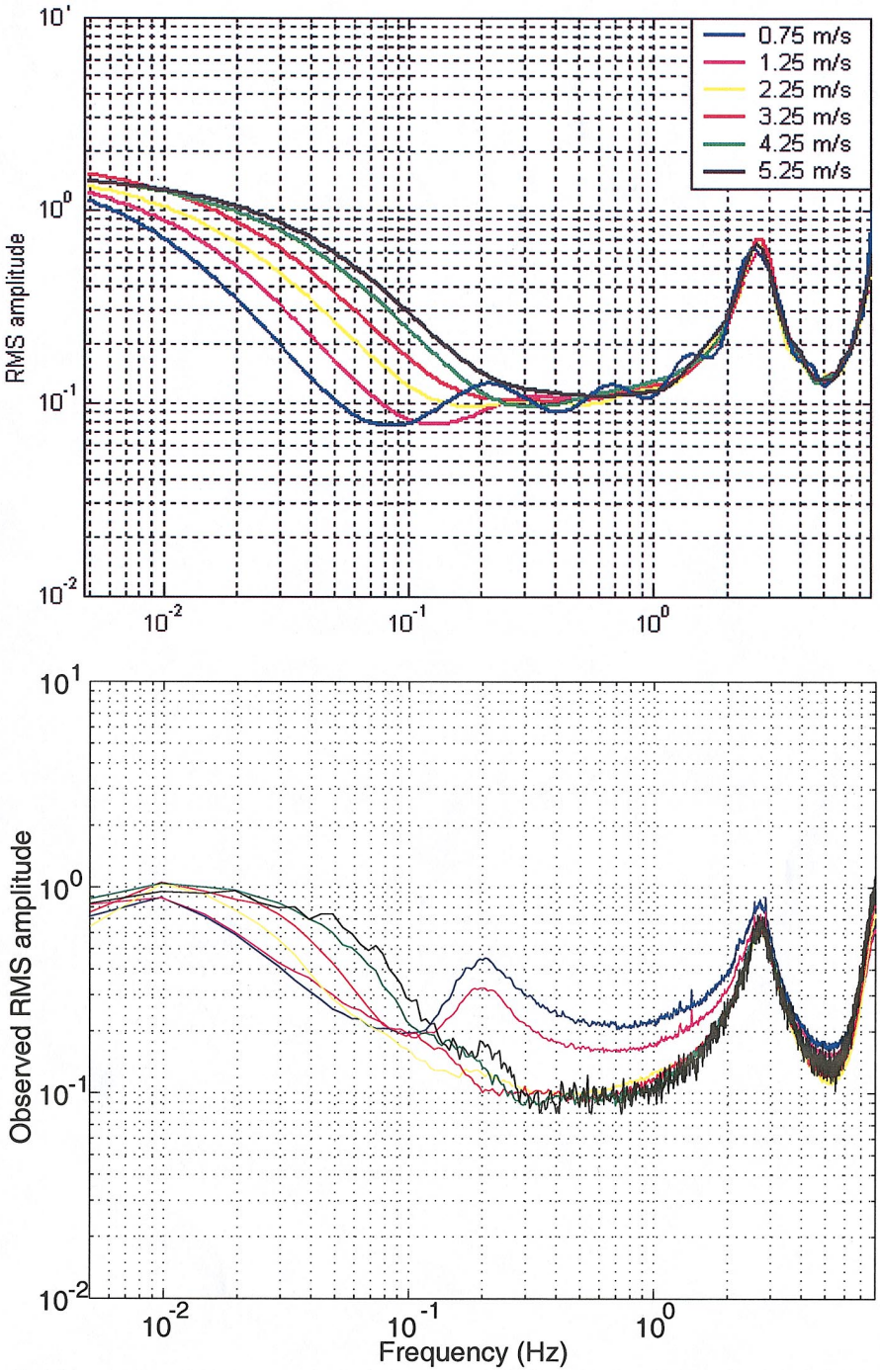


FIG. 12. Simulated noise reduction of the 70-m/144-port rosette filter for various mean wind speeds is shown in the upper panel. Each curve represents the ratio between the noise spectrum observed at the reference port and the noise spectrum of the rosette system. The corner frequency of the 70-m rosette filter is predicted to increase with increasing wind speed. Observed noise reduction is shown in the lower panel.

cause complete cancellation to occur. In effect, each space filter acts as an array of omnidirectional sensors where each port in the space filter represents a sensor. Since the propagation pathway from port to the summing manifold is identical for each port, the space filter array is mechanically beamformed so that its main lobe points in the broadside direction, i.e., in the vertical direction.

B. Analysis of a recorded signal

The theoretically derived amplitude and phase response of rosette filters that takes into account both the time delays between inlets and resonance of the energy once it has entered the pipe system can be tested by analyzing signals recorded via the filters. The ideal signal for this purpose has a long duration, a large amplitude, and arrives at a time of low ambient noise levels so there will be minimal interference with the noise the filter is designed to remove. This signal has a low phase velocity so that time delays between inlets in the filter will be large. As shown by Fig. 13, a signal arriving with a high phase velocity would only allow us to test further the theory of resonance inside the pipes as time delays between the inlets would be negligible. The ideal signal is generated by a well-documented source. Finally, to allow extraction of the amplitude and phase of a rosette filter, it is also necessary to record the ideal signal not only with the rosette filters, but by a microbarometer that is not attached to a noise reduction system, but is attached to a single inlet located at the center of the rosette filter.

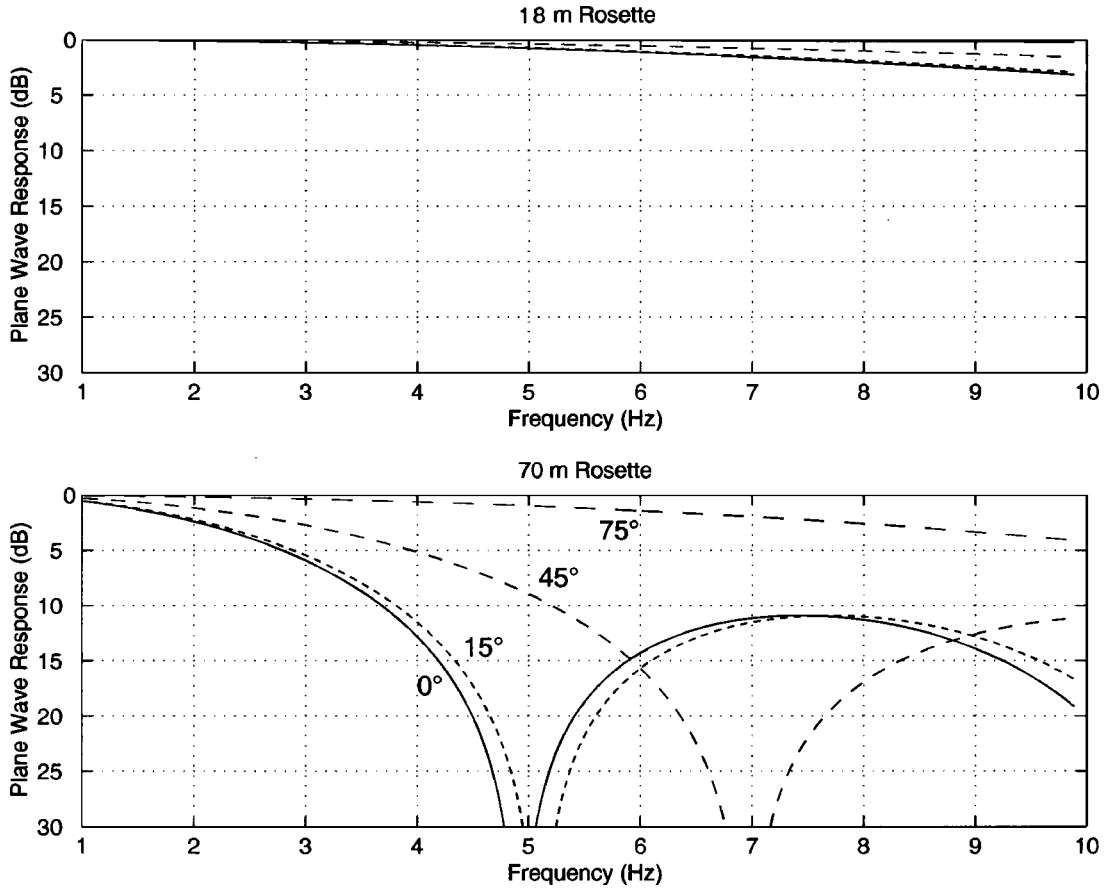


FIG. 13. In the upper and lower panels we show the plane-wave response for the 18- and 70-m rosette filters, respectively, at four arrival angles. The solid curves in each panel represent the response to horizontally propagating signals. The finely to coarsely dashed curves represent signals propagating across the two filters at 15°, 45°, and 75° above the horizontal. The elevation angles, θ_e, are calculated assuming a sound speed, c, of 347 m/s. The phase velocity, c_p, is given by $c/\cos(\theta_e)$. The phase velocities corresponding to the arrival angles at the four arrival angles are 347, 359, 491, and 1341 m/s.

Such a signal occurred on 23 April 2001. On that day, a meteor exploded off the west coast of North America with a yield of several kilotons (Garces *et al.*, 2001). Satellite data places the location of the explosion at 28.6 °N, 134.2 °W at an altitude of 30 km (Brown and Gault, 2001). This location is approximately 1800 km to the west of the Pinon Flat Observatory. This event was recorded by infrasound arrays across North America and as far away as Germany (Garces *et al.*, 2001). The explosion was recorded by the Pinon IMS array I57US and by all sites in our noise experiment. The experiment included the 18- and 70-m rosette filters that are the subject of this paper. The experiment also included noise collected by MB2000 microbarometers via single reference ports located at the center of each rosette filter. The explosion occurred at 06:09 UT and thus the signals arrived at the Pinon flat observatory near midnight local time (Fig. 14).

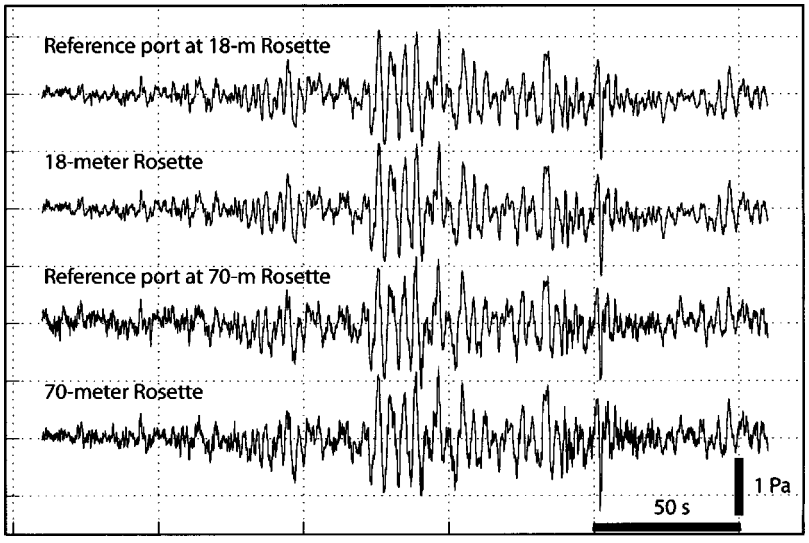


FIG. 14. The widely reported explosion of a large meteor off the coast of California on 23 April 2001 was recorded by the IMS array 157US at PFO and by sensors in our wind-noise experiment. We show recordings made by 18- and 70-m rosette filters at sites H1 and L2, respectively, in the infrasound array 157US via reference ports at the same locations. The H1 and L2 sites are ~600 m apart (Fig. 3). The recordings reveal the signal to be highly coherent. The maximum amplitude of the signal was ~2 Pa peak-to-peak. The duration of the coherent energy was ~200 s. The recordings shown in this figure are unfiltered.

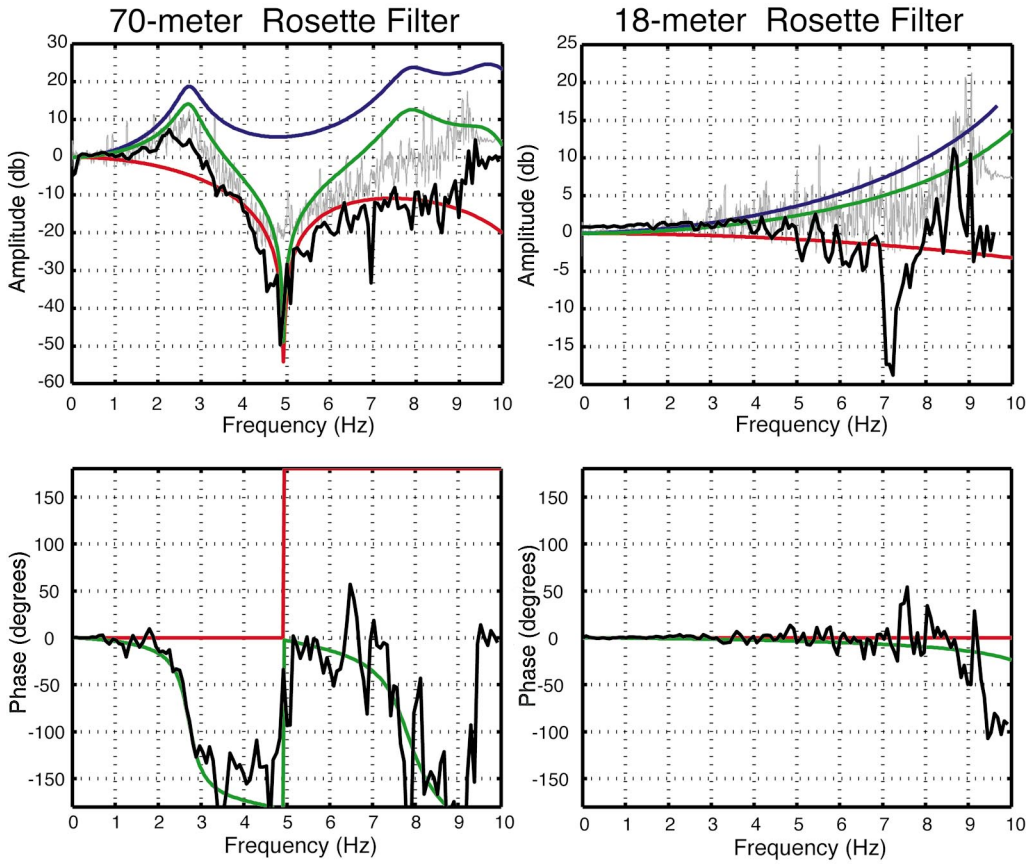


FIG. 15. Amplitude and phase response of 18- and 70-m rosette filters are shown in this figure. The blue curves in all panels represent the response due to internal resonance inside the pipe systems. The red curves represent the plane-wave response. The plane-wave response was calculated assuming a phase velocity of 330 m/s. This is the phase velocity of the energy from the 23 April bolide as determined by processing data from the 157US array. The green curves represent the total response due to time delays between inlets and to resonance inside the filters. Amplitude and phase from a phase-coherent cross-spectral analysis of data from the 23 April bolide are shown in black. The theory accurately predicts the phase of the signal but underpredicts the amplitude at all frequencies. A spectral ratio of the filtered to unfiltered data (gray curves in the upper panels) closely matches the theoretical predictions. The green curves in the lower panels plot on top of the blue curves except in the lower left panel above 5 Hz. At those frequencies the phase of the internal resonance in the 70-meter filter is less than −180 (Fig. 11).

Due to the strong diurnal cycle in infrasonic noise levels, local noise levels at that time were minimal. The wave train of arrivals remained above noise for >200 s and reached a maximum amplitude of 2 Pa peak-to-peak. Brown and Gault (2001) observed the signal was highly coherent across the Pinon array. The signal propagated across the array with a very low phase velocity of 330 m/s (D'Spain *et al.*, 2001). This corresponds to an arrival angle close to 0° above the horizontal. Signal is observed clearly above noise across the broad band from 0.05 to 7.0 Hz. This signal provides an opportunity to gauge the utility of the various noise reducing devices at preserving coherent signals from remote sources while reducing noise.

Two signal-processing approaches based on linear time-invariant filter theory can be used. In a phase-coherent approach, both the amplitude and phase response of a filter as a function of frequency are estimated by normalizing the cross spectrum Sxy between the input and output time series by the autospectrum of the input Sxx (Bendat and Piersol, 1986)

$$H(f) = Sxy(f)/Sxx(f). \tag{1}$$

The microbarometer located in the center of the space filter but not connected to any mechanical filtering system provides the input time series $[x(t)]$ and the signal recorded by the microbarometer connected to the space filter provides the output time series $[y(t)]$. The results of using this approach with nearly 1 min, 45 s of time series containing the majority of the arriving energy from the bolide are plotted as a black curve in Fig. 15. For the phase responses (lower two panels), the agreement between the data from the bolide and the theory (plotted as green curves) is excellent, particularly for the 70-m space filter. In the case of the amplitude responses, the overall spectral shape of the measured and predicted responses are in general agreement. However, the measured amplitude response is lower than predicted and appears to diverge with increasing frequency. This difference suggests that a frequency-dependent attenuation mechanism is not being accounted for in the theory.

A second approach is phase incoherent and so only provides information on the amplitude of the filter response. It is obtained simply by taking the square root of the ratio of the output autospectrum to the input autospectrum (Bendat and Piersol, 1986)

$$|H(f)| = (Syy(f)/Sxx(f))^{0.5}. \tag{2}$$

The results of this technique using exactly the same segment

of data as in the phase-coherent calculation are shown as gray curves in the upper panels of Fig. 15. They are generally in better agreement with theory. These transfer-function amplitude estimates from Eq. (2) are larger because of the addition of a component in the time series of the microbarometer with space filter that is phase uncorrelated with the reference microbarometer time series.

VIII. CONCLUDING REMARKS

To monitor Earth's atmosphere for weak signals from small nuclear tests anywhere on the globe requires that we develop an effective means to filter out noise due to atmospheric turbulence. For the International Monitoring System network under current development, rosette spatial filters are the first choice for new sites. In this study, we have examined the utility of these filters for removing incoherent noise across a broad band while preserving coherent signals.

A. Noise reduction

With minor exceptions, our observations of noise reduction provided by the 18- and 70-m filters are predicted by the theoretical method of Alcoverro (1998). The theory underpredicts coherence near the microbarom peak as this energy is propagated acoustically, not by turbulent wind. In some cases noise reduction provided by the 18-m filter exceeded the limit predicted assuming perfectly incoherent noise. This anomaly does not reflect an error with the theory but is due to the 18-m filter receiving some shelter from prevailing winds by nearby vegetation.

Data from our experiment, and from analytical simulations, attest to the utility of these filters for reducing noise at frequencies of interest to the nuclear monitoring community. We have observed, however, that the performance of the filters is severely degraded at all wind speeds by resonance. The most significant resonance occurs within the long pipes that connect the secondary summing manifolds with the microbarometer, or the primary summing manifold at the center of the filter. The resonance in the 18-m filter is observed above 3.0 Hz and is less important as nuclear monitoring relies mainly on lower frequencies. Resonance in the 70-m filters is significant at frequencies above 0.7 Hz.

Experiments indicate that the low-frequency resonance can be removed by placing impedance-matching capillaries in the long pipes connecting the summing manifolds (Hedlin, 2001). The results of this work will be reported in detail in a future paper.

Temperature changes at the recording site will cause sound-speed-sensitive changes in filter response characteristics. However, these changes are found to be insignificant if the pipes are buried and thus protected from large temperature swings.

B. Frequency-dependent directionality of the space filters

Rosette filters are designed for maximum effectiveness for vertically arriving signals. The directionality is not significant for low-frequency signals regardless of the arrival direction. Signals above 1 Hz arriving at grazing angles ($<15°$ from the horizontal) are severely attenuated by the 70-m rosette filters.

A cross-spectral analysis of signals from the explosion of bolide over the Pacific provided an opportunity to further test the resonance theory, to validate an extended theory that takes into account both the resonance and the plane-wave response of the rosette filters, and to calibrate the filters. The combined theory accurately predicts the phase response of the filters; however, the measured amplitude response is lower than predicted, suggesting that attenuation is not fully accounted for in the theory. The good agreement of predicted and observed phase shifts, as well as the general agreement with the phase-incoherent derived amplitudes with those derived from the theory, indicate that the theoretical method is sound and can be used to predict the response of the filter to signals that propagate across the filters at other velocities.

ACKNOWLEDGMENTS

The authors are indebted to Doug Christie (CTBTO) for ongoing assistance with this project. Frank Vernon, Jennifer Eakins, and Glen Offield provided the real-time data link. Clint Coon provided field assistance. The authors would like to thank Lou Sutherland and Milton Garces for many helpful comments on this paper. Funding was provided by the Defense Threat Reduction Agency under Contract No. DTRA01-00-C-0085. Funding for the rosette filters used in this study was provided by the Defense Threat Reduction Agency, the Provisional Technical Secretariat (PTS) of the UN Comprehensive Test Ban Treaty Office in Vienna, and the US Army Space and Missile Defense Command (SMDC) University Research Initiative (URI).

Alcoverro, B. (**1998**). ''Proposition d'un systeme de filtrage acoustique pour une station infrason IMS,'' CEA-DASE Scientific Report, No. 241.

Benade, A. H. (**1968**). ''On the propagation of sound waves in a cylindrical conduit,'' J. Acoust. Soc. Am. **44**, 616–623.

Bendat, J. S., and Piersol, A. G. (**1986**). *Random Data Analysis and Measurement Procedures* (Wiley, New York).

Blanc, E. (**1985**). ''Observations in the upper atmosphere of infrasonic waves from natural or artificial sources: A summary,'' Ann. Geophys. (France) **3**, 673–688.

Broche, P. (**1977**). ''Propagation des ondes acoustico-gravitationnelles excitees par des explosions,'' Ann. Geophys. (C.N.R.S.) **33**, 3, 281–288.

Brown, D. J., and Gault, A. K. (**2001**). ''Observations and Preliminary analysis of the northern Pacific infrasonic event,'' 23 April, 2001, Center for Monitoring Research technical report CMR-01/09, 21 May, 2001.

Burridge, R. (**1971**). ''The acoustics of pipe arrays'' Geophys. J. R. astr. Soc. **26**, 53–69.

Calais, E., Minster, J. B., Hofton, M. A., and Hedlin, M. A. H. (**1998**). ''Ionospheric signature of surface mine blasts from Global Positioning System measurements,'' Geophys. J. Int. **132**, 191–202.

D'Spain, G., Hedlin, M. A. H., Orcutt, J. A., Kuperman, B., DeGroot-Hedlin, C. D., Berger, L., Rovner, G., and Hudak, H. (**2001**). ''Bolide, Rocket Launch, and Background Infrasonic Noise Recordings at I57US and Anza Temporary Stations,'' Infrasound Technology Workshop, 12–15 November 2001, Kailua-Kona, Hawaii.

Daniels, F. B. (**1950**). ''On the propagation of sound waves in a cylindrical conduit,'' J. Acoust. Soc. Am. **22**, 563–564.

Daniels, F. B. (**1959**). ''Noise reducing line microphone for frequencies below 1 c/s,'' J. Acoust. Soc. Am. **31**, 529.

Davidson, M., and Whitaker, R. W. (**1992**). ''Miser's Gold,'' Los Alamos National Laboratory Technical Report: LA-12074-MS.

Francis, S. H. (**1973**). ''Acoustic-gravity modes and large-scale traveling ionospheric disturbances of a realistic, dissipative atmosphere,'' J. Geophys. Res. **78**, 13, 2278–2301.

Garces, M., Hetzer, C., Lindquist, K., Hansen, R., Drob, D., and Picone, M. (**2001**). ''Stratospheric arrivals in the upstream wind direction: Case study of the 23 April, 2001 bolide,'' Infrasound Technology Workshop, 12–15 November, 2001, Kailua-Kona, Hawaii.

Grover, F. H. (**1971**). ''Experimental noise reducers for an active microbarograph array'' Geophys. J. R. Soc., **26**, 41–52.

Hedlin, M. A. H. (**2001**). ''Recent experiments in infrasonic noise reduction: The search for that elusive broadband filter,'' Infrasound Technology Workshop, 12–15 November 2001, Kailua-Kona, Hawaii.

Hedlin, M. A. H., Berger, J., and Vernon, F. (**2002**). ''Surveying infrasonic noise on oceanic islands,'' Pure Appl. Geophys. **159**, 1127–1152.

Herrin, G., Golden, P., and Hedlin, M. A. H. (**2001**). ''Investigation of wind noise reducing filters,'' Infrasound Technology Workshop, 12–15 November 2001, Kailua-Kona, Hawaii.

Kaimal, J. C., and Finnigan, J. J. (**1994**). *Atmospheric Boundary Layer Flows: Their Structure and Measurement* (Oxford University Press, Oxford).

Keefe, D. H. (**1984**). ''Acoustical wave propagation in cylindrical ducts; Transmission line parameter approximations for isothermal and nonisothermal boundary conditions,'' J. Acoust. Soc. Am. **75**(1), 58–62.

Lamb, H. (**1932**). *Hydrodynamics*, 6th ed. (London).

Landau, L. D., and Lifshitz, E. M. (**1959**). *Fluid Mechanics* (Pergamon, Oxford).

McKisic, J. M. (**1997**). ''Infrasound and the infrasonic monitoring of atmospheric nuclear explosions: A literature review,'' Final report submitted to the DOE and Phillips Lab, PL-TR-97-2123.

Priestley, J. T. (**1966**). ''Correlation studies of pressure fluctuations on the ground beneath a turbulent boundary layer,'' Washington D.C., National Bureau of Standards Report No. 8942, U.S. Dept. of Commerce, National Bureau of Standards.

Sereno, T. J., and Orcutt, J. A. (**1985**). ''Synthesis of realistic oceanic Pn wave trains,'' J. Geophys. Res. **90**, 12755–12776.

Simons, D. J. (**1995**). ''Atmospheric methods for nuclear test monitoring. Monitoring a comprehensive test ban treaty,'' NATO ASI series E, Vol. 303, pp. 135–141, edited by E. S. Husebye and A. M. Dainty.

Welch, P. D. (**1967**). ''The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms,'' IEEE Trans. Audio Electroacoust. **AU-15**, 70–73.

Zumberge, M. A., Berger, J., Hedlin, M. A. H., Husmann, E., Nooner, S., Hilt, R., and Widmer-Schnidrig, R. (**2003**). ''An optical fiber infrasound sensor: A new lower limit on atmospheric noise between 1 and 10 Hz,'' J. Acoust. Soc. Am. **113**(5), 2474–2479.

A modified Khokhlov–Zabolotskaya equation governing shear waves in a prestrained hyperelastic solid

M. S. Cramer and M. F. Andrews
Department of Engineering Science and Mechanics, Virginia Polytechnic Institute & State University, Blacksburg, Virginia 24061-0219

(Received 27 January 2003; revised 15 July 2003; accepted 28 July 2003)

Weakly nonlinear, weakly diffracting, two-dimensional shear waves propagating in a prestrained hyperelastic solid are examined. A modification of the classical Khokhlov–Zabolotskaya equation is derived using a systematic perturbation scheme. Both dissipative and nondissipative materials were considered. The principal effect of the prestrain was seen to be the inclusion of a quadratic nonlinearity to the cubic nonlinearity found in the case of zero prestrain. Further new results include the shock jump relations and the prediction of shocks having a speed which is identical to the nonlinear wave speed ahead of or behind the shock. Explicit expressions for the nonlinearity coefficients for the special case of a Blatz–Ko material were provided. [DOI: 10.1121/1.1610460]

PACS numbers: 43.28.Kt, 43.28.Lv, 43.28.Js [MFH]

I. INTRODUCTION

The canonical equation for weakly nonlinear and weakly diffracting waves is the Khokhlov–Zabolotskaya (KZ) equation, which can be written in the general form

$$(U_\tau+\beta' UU_\chi)_\chi+\tfrac{1}{2}U_{\eta\eta}=0, \tag{1}$$

where $U=U(\chi,\eta,\tau)$ is a measure of the wave disturbance, χ is a spatial variable measured in a frame moving with the wave, η is a transverse spatial variable, and τ is a time-like variable. The order-one constant β' is a measure of the nonlinearity of the system. The KZ equation is recognized as being quadratically nonlinear, and we therefore refer to β' and similar parameters as quadratic nonlinearity parameters. The first derivation of this equation was given by Zabolotskaya and Khokhlov (1969) in the context of ordinary nonlinear acoustics.

When modified to account for dispersion and dissipation, the KZ equation (1) provides an accurate description of many commonly encountered physical systems including those associated with the nonlinear acoustics of low-pressure gases and liquids, the nonlinear acoustics of longitudinal waves in solids, magnetosonic and acoustic waves in magnetohydrodynamics, and surface and internal waves in the ocean. However, in many systems of practical interest, the quadratic nonlinearity coefficient β' may be small or vanish altogether. In such cases, (1) becomes

$$U_{\tau\chi}+\tfrac{1}{2}U_{\eta\eta}\approx 0, \tag{2}$$

and nonlinearity no longer plays a role in the wave evolution, at least over the length and time scales originally associated with (1). Examples where β' is identically zero are nonlinear optics, Alfvén waves in magnetohydrodynamics, and shear waves in an isotropic nonlinear solid; it is the latter application which is of central interest in the present paper.

While (2) may give a reasonable description of the wave evolution in some applications, any complete understanding of nonlinear diffraction in systems having $\beta'=0$ will require further study. It may also be shown that waves governed by (2) may focus if the initial wavefront is curved. Near the focal region (2) must be modified to include nonlinear effects; the resultant evolution near the focal region is also expected to be governed by a nonlinear equation of the KZ type.

The evolution equation governing systems having small or zero β' will necessarily involve an inclusion and analysis of cubically nonlinear terms in addition to the quadratically nonlinear terms normally leading to (1). Zabolotskaya (1986) has derived the explicit form of the extension of (1) for the case of shear waves propagating in a nonlinear solid having no strain in the undisturbed state. The result of Zabolotskaya's study was that the extension of (1) for shear waves can be written as

$$(U_\tau+\beta'' U^2U_\chi)_\chi+\tfrac{1}{2}U_{\eta\eta}=0, \tag{3}$$

at least when the displacements are confined to the $\chi-\eta$ plane and dissipation is ignored. Here, β'' is a cubic nonlinearity coefficient which depends on the particular material model. Zabolotskaya's result (3) is intuitively satisfying because the net effect of the detailed perturbation analysis carried out by Zabolotskaya is to replace the $\beta' UU_\chi$ term in (1) by the cubic term $\beta'' U^2U_\chi$ when $\beta'=0$. Under the conditions implicit in Zabolotskaya's derivation it can be shown that the parameter β'' can be computed by an analysis of the one-dimensional problem, i.e., a problem where all variations with respect to the transverse direction vanish, thus suggesting a simplified derivation or a check on the value of β''. Of course, similar remarks can be made for the coefficient β' seen in (1).

More general studies of cubically nonlinear systems have been carried out by Kluwick and Cox (1998) and Cramer and Webb (1998). The first set of authors developed a general multiple-scales scheme for the treatment of cubically nonlinear systems, i.e., systems in which the relevant β' is either zero or small. Kluwick and Cox's results included those for ordinary nonlinear acoustics on rays as well as the extension for diffracting systems. Cramer and Webb

(1998) presented a more restricted version of Kluwick and Cox's scheme focused on nonlinear diffraction. The primary extension of Cramer and Webb's work over that of Kluwick and Cox was the inclusion of the effects of dispersion, dissipation, and relaxation.

A remarkable feature of the general schemes presented by Kluwick and Cox (1998) and Cramer and Webb (1998) is that the general form of the KZ equation valid for small or zero β' is not of the form (3), but must be supplemented with quadratic terms representing an interaction of diffraction and nonlinearity. In the notation used for Eqs. (1)–(3), the terms which must be added to (3) can be written

$$\left\{ \beta''' U U_\eta + \beta^{(iv)} U_\chi \int U_\eta \, d\chi \right\}_\chi , \tag{4}$$

where β''' and $\beta^{(iv)}$ are system-dependent constants. Both U_η and $\int U_\eta \, d\chi$ can be shown to be zero when diffraction is negligible.

For a general system, there is no *a priori* reason why the material constants β''' and $\beta^{(iv)}$ should vanish. However, previous investigations have shown that β''' and $\beta^{(iv)}$ do in fact vanish for many cases of practical interest; an example of direct interest here is the case studied by Zabolotskaya (1986). Andrews (1999) has shown that these coefficients also vanish for the case of nonlinear optics. Kluwick and Cox (1998) and Cramer and Webb (1998) have shown that β''' and $\beta^{(iv)}$ are zero for the case of acoustic waves in a Bethe–Zel'dovich–Thompson fluid; the latter class of fluid has a quadratic nonlinearity parameter β' which can be small or even change sign at certain thermodynamic states. However, Cramer and Webb (1998) have also shown that the coefficients β''' and $\beta^{(iv)}$ are nonzero for nonlinear Alfvén waves in a plasma. Thus, although the general form of the KZ equation for cubically nonlinear problems is known from the work of Kluwick and Cox (1998) and Cramer and Webb (1998), the status of the additional terms proportional to β''' and $\beta^{(iv)}$ will depend on the details of the specific application. Furthermore, the values of β''' and $\beta^{(iv)}$ cannot be determined by or carried over from a simpler case, e.g., by an analysis of the one-dimensional theory.

In the present investigation we examine the effect of a weak prestrain on the nonlinear diffraction of shear waves. In particular, we will determine the modification to the cubic KZ equation, e.g., (3), required when the undisturbed state has a small shear strain. Such prestrained states can be of interest in acoustoelasticity (Norris, 1998) and may also occur naturally. In the present study we show that the effect of the prestrain will be to modify the linear wave speed and to introduce nonvanishing quadratic nonlinearity in the final KZ equation. A further question of interest is whether the prestrain will result in terms similar to those proportional to β''', $\beta^{(iv)}$ which are nonzero. We also account for the shock waves which will inevitably form and record the shock speed for diffracting shear waves in a prestrained material. A final extension of Zabolotskaya's results is to determine the form of the nonlinearity constants for the case of a specific material model; this model is based on the well-known Blatz–Ko strain energy.

II. FORMULATION

The stress–strain relation for a general hyperelastic solid undergoing isothermal deformations is

$$\sigma_{ij} = \alpha_0 \delta_{ij} + \alpha_1 B_{ij} + \alpha_2 B_{im} B_{mj} , \tag{5}$$

where σ_{ij} is the Cauchy stress tensor, δ_{ij} is the Kronecker delta, and B_{ij} is the left Cauchy–Green tensor. In all that follows, lower-case Latin indices will range over 1, 2, 3 and the Einstein summation convention will be employed. The left Cauchy–Green tensor is defined as

$$B_{ij} \equiv F_{i\ell} F_{j\ell} = B_{ji} , \tag{6}$$

where

$$F_{ij} \equiv \frac{\partial x_i}{\partial X_j} \tag{7}$$

is the deformation matrix, X_i are the Cartesian coordinates of a material particle in the undeformed (natural) configuration, and x_i are the coordinates of its position in the deformed configuration. We will refer to X_i as the Lagrangian coordinates and x_i as the Eulerian coordinates.

The scalar coefficients α_0, α_1, α_2 are related to the strain energy per unit mass by

$$\alpha_0 = 2\rho_0 III^{1/2} \frac{\partial U}{\partial III}, \tag{8}$$

$$\alpha_1 = \frac{2\rho_0}{III^{1/2}} \left(\frac{\partial U}{\partial I} + I \frac{\partial U}{\partial II} \right), \tag{9}$$

$$\alpha_2 = -\frac{2\rho_0}{III^{1/2}} \frac{\partial U}{\partial II}, \tag{10}$$

where $U \equiv U(I, II, III)$ is the aforementioned strain energy and the constant ρ_0 is the material density in the unstrained state. The quantities I, II, III are the invariants of B_{ij} given by

$$\begin{aligned} &I \equiv B_{\ell\ell} , \quad II \equiv \tfrac{1}{2}(I^2 - B_{im} B_{mi}), \\ &III \equiv \det(B_{ij}), \end{aligned} \tag{11}$$

where $\det(B_{ij})$ refers to the determinant of B_{ij}.

Motions of the solid are governed by Cauchy's equation

$$\rho \ddot{u}_i = \frac{\partial \sigma_{ji}}{\partial x_j}, \tag{12}$$

where the dots denote differentiation with respect to time following the particle, $u_i(X_\ell, t)$ is the displacement vector defined by

$$u_i \equiv x_i - X_i , \tag{13}$$

and ρ is the instantaneous material density related to the density in the unstrained configuration (ρ_0) by

$$\rho = \rho_0 III^{-1/2} = \rho_0 F^{-1}, \tag{14}$$

where $F \equiv \det(F_{ij})$. A useful alternate form of (12) is obtained by recasting it in terms of the Lagrangian coordinates X_i as follows:

$$\rho_0 \frac{\partial^2 u_i}{\partial t^2} = F \frac{\partial X_p}{\partial x_j} \frac{\partial \sigma_{ji}}{\partial X_p}, \tag{15}$$

where the time differentials are taken with X_i held fixed.

At this stage we note that the exact relation (5) can be reduced to the usual linear stress–strain relation by taking u_i to be small and retaining only terms linear in $\partial u_i / \partial X_j$. The resultant linearized form of (5) is found to be

$$\sigma_{ij} \approx \lambda' \frac{\partial u_\ell}{\partial X_\ell} \delta_{ij} + \mu' \left(\frac{\partial u_i}{\partial X_j} + \frac{\partial u_j}{\partial X_i} \right),$$

where

$$\mu' \equiv \tilde{\alpha}_1|_0, \quad \lambda' \equiv 2\tilde{\beta}_0, \tag{16}$$

are the usual Lamé constants, and we have required that

$$\tilde{\alpha}_0 = \tilde{\alpha}_0(I, II, III) \equiv \alpha_1 + \alpha_2 + \alpha_3 \tag{17}$$

vanishes at the undeformed state $I = II = 3$, $III = 1$; the latter condition ensures that the stress vanishes when the strain does. In the remainder of this paper we will find it convenient to define the following quantities:

$$\tilde{\alpha}_1 \equiv \alpha_1 + 2\alpha_2, \tag{18}$$

$$\tilde{\alpha}_2 \equiv \alpha_2, \tag{19}$$

and

$$\tilde{\beta}_0 \equiv \left. \frac{\partial \tilde{\alpha}_0}{\partial I} \right|_0 + 2 \left. \frac{\partial \tilde{\alpha}_0}{\partial II} \right|_0 + \left. \frac{\partial \tilde{\alpha}_0}{\partial III} \right|_0, \tag{20}$$

where the notation $(\,)|_0$ will denote quantities evaluated at the unstrained state $I = II = 3$, $III = 1$.

The perturbations to the undisturbed state will be taken to be two-dimensional, i.e., $u_i = u_i(X_1, X_2, t)$ only and

$$\begin{aligned} x_1 &= X_1 + u_1(X_1, X_2, t), \\ x_2 &= X_2 + u_2(X_1, X_2, t), \\ x_3 &= X_3. \end{aligned} \tag{21}$$

Under these conditions, the left Cauchy–Green tensor (6) can be written

$$\begin{aligned} B_{11} &= 1 + 2e_{11} + (e_{11})^2 + (e_{12})^2, \\ B_{12} &= B_{21} = (1 + e_{11})e_{21} + (1 + e_{22})e_{12}, \\ B_{22} &= 1 + 2e_{22} + (e_{22})^2 + (e_{21})^2, \\ B_{33} &= 1, \end{aligned} \tag{22}$$

with all other components equal to zero. Here

$$e_{ij} \equiv F_{ij} - \delta_{ij} = \frac{\partial u_i}{\partial X_j} \neq e_{ji} \tag{23}$$

represents the deviation of the deformation matrix from the undisturbed state and should not be confused with the strain; in the linear approximation, the latter can be written as $(e_{ij} + e_{ji})/2$. The invariants (11) now reduce to

$$\begin{aligned} I &= 1 + B_{11} + B_{22}, \\ II &= B_{11} + B_{22} + B_{11}B_{22} - B_{12}^2, \\ III &= B_{11}B_{22} - B_{12}^2, \end{aligned} \tag{24}$$

with B_{ij} given by (22). When (22) is substituted in (5), we find that the only nonzero components of the stress are

$$\sigma_{11} = \tilde{\alpha}_0 + \tilde{\alpha}_1(2e_{11} + e_{11}^2 + e_{12}^2) + \tilde{\alpha}_2(Q_{112} + Q_{113} + Q_{114}), \tag{25}$$

$$\sigma_{22} = \tilde{\alpha}_0 + \tilde{\alpha}_1(2e_{22} + e_{22}^2 + e_{21}^2) + \tilde{\alpha}_2(Q_{222} + Q_{223} + Q_{224}), \tag{26}$$

$$\sigma_{12} = \sigma_{21} = \tilde{\alpha}_1(e_{21} + e_{12} + e_{11}e_{21} + e_{22}e_{12}) + \tilde{\alpha}_2(Q_{122} + Q_{123} + Q_{124}), \tag{27}$$

$$\sigma_{33} = \tilde{\alpha}_0, \tag{28}$$

where

$$Q_{112} \equiv 4e_{11}^2 + e_{21}^2 + 2e_{21}e_{12} + e_{12}^2, \tag{29}$$

$$Q_{113} \equiv 4e_{11}^3 + 4e_{11}e_{12}^2 + 2e_{11}e_{21}^2 + 2(e_{11} + e_{22})e_{21}e_{12} + 2e_{22}e_{12}^2, \tag{30}$$

$$Q_{114} \equiv e_{11}^4 + 2e_{11}^2e_{12}^2 + e_{12}^4 + e_{11}^2e_{21}^2 + 2e_{22}e_{11}e_{21}e_{12} + e_{22}^2e_{12}^2, \tag{31}$$

$$Q_{222} = 4e_{22}^2 + e_{21}^2 + 2e_{21}e_{12} + e_{12}^2, \tag{32}$$

$$Q_{223} = 4e_{22}^3 + 2e_{11}e_{21}^2 + 2e_{11}e_{21}e_{12} + 2e_{22}e_{21}e_{12} + 2e_{22}e_{12}^2 + 4e_{22}e_{21}^2, \tag{33}$$

$$Q_{224} = e_{22}^4 + 2e_{22}^2e_{21}^2 + e_{21}^4 + e_{11}^2e_{21}^2 + 2e_{22}e_{11}e_{12}e_{21} + e_{22}^2e_{12}^2, \tag{34}$$

$$Q_{122} \equiv 2(e_{11} + e_{22})(e_{21} + e_{12}), \tag{35}$$

$$Q_{123} \equiv (e_{21} + e_{12})(e_{12}^2 + 2e_{11}e_{22} + e_{21}^2) + 3e_{11}^2e_{21} + 3e_{22}^2e_{12} + e_{11}^2e_{12} + e_{22}^2e_{21}, \tag{36}$$

$$Q_{124} \equiv e_{11}^3e_{21} + e_{11}e_{22}^2e_{21} + e_{11}^2e_{22}e_{12} + e_{22}^3e_{12} + e_{11}e_{12}^2e_{21} + e_{11}e_{21}^3 + e_{12}^3e_{22} + e_{21}^2e_{22}e_{12}. \tag{37}$$

Expressions (25)–(37) are exact. However, the quantities Q_{114}, Q_{224}, Q_{124} are of fourth order in e_{ij} and will turn out to be negligible for the purposes of the present study.

We now reduce Cauchy's equation to its simplified form. We first note that it will be convenient to convert the system of second-order (in time) equations (15) into an equivalent system of first-order equations. To do this we will define the particle velocities

$$v_i \equiv \frac{\partial u_i}{\partial t}, \tag{38}$$

where $v_3 = 0$ due to the fact that $u_3 = 0$. If we now combine the condition of two-dimensional motions (21), the definition (38), and Cauchy's equation (15), we find that (15) can be rewritten in the following vector form:

$$\frac{\partial \mathbf{w}}{\partial t}+\underline{\underline{A}}\frac{\partial \mathbf{w}}{\partial X_1}+\underline{\underline{B}}\frac{\partial \mathbf{w}}{\partial X_2}=\mathbf{0}, \tag{39}$$

where $\mathbf{w}$ is a 6×1 solution vector defined as

$$\mathbf{w}\equiv\begin{pmatrix} e_{11}\\ e_{12}\\ e_{21}\\ e_{22}\\ v_1\\ v_2\end{pmatrix}. \tag{40}$$

Here, we note that the last two equations in (39) are the two-dimensional form of Cauchy's equation and the first four arise from the compatibility conditions

$$\frac{\partial e_{ij}}{\partial t}=\frac{\partial v_i}{\partial X_j} \tag{41}$$

obtained by combining the definition (38) with the time differential of (23). The elements of the 6×6 coefficient matrices $\underline{\underline{A}}$ and $\underline{\underline{B}}$ were found to be

$$A_{15}=A_{36}=-1, \tag{42}$$

$$A_{51}=-\frac{1}{\rho_0}\left[(1+e_{22})\frac{\partial\sigma_{11}}{\partial e_{11}}-e_{12}\frac{\partial\sigma_{21}}{\partial e_{11}}\right], \tag{43}$$

$$A_{53}=-\frac{1}{\rho_0}\left[(1+e_{22})\frac{\partial\sigma_{11}}{\partial e_{21}}-e_{12}\frac{\partial\sigma_{21}}{\partial e_{21}}\right], \tag{44}$$

$$A_{61}=-\frac{1}{\rho_0}\left[(1+e_{22})\frac{\partial\sigma_{12}}{\partial e_{11}}-e_{12}\frac{\partial\sigma_{22}}{\partial e_{11}}\right], \tag{45}$$

$$A_{63}=-\frac{1}{\rho_0}\left[(1+e_{22})\frac{\partial\sigma_{12}}{\partial e_{21}}-e_{12}\frac{\partial\sigma_{22}}{\partial e_{21}}\right], \tag{46}$$

$$B_{51}=-\frac{1}{\rho_0}\left[(1+e_{22})\frac{\partial\sigma_{11}}{\partial e_{12}}-e_{12}\frac{\partial\sigma_{21}}{\partial e_{12}}+(1+e_{11})\frac{\partial\sigma_{21}}{\partial e_{11}}-e_{21}\frac{\partial\sigma_{11}}{\partial e_{11}}\right], \tag{47}$$

$$B_{25}=B_{46}=-1, \tag{48}$$

$$B_{52}=-\frac{1}{\rho_0}\left[-e_{21}\frac{\partial\sigma_{11}}{\partial e_{12}}+(1+e_{11})\frac{\partial\sigma_{21}}{\partial e_{12}}\right], \tag{49}$$

$$B_{53}=-\frac{1}{\rho_0}\left[(1+e_{22})\frac{\partial\sigma_{11}}{\partial e_{22}}-e_{12}\frac{\partial\sigma_{21}}{\partial e_{22}}+(1+e_{11})\frac{\partial\sigma_{21}}{\partial e_{21}}-e_{21}\frac{\partial\sigma_{11}}{\partial e_{21}}\right], \tag{50}$$

$$B_{54}=-\frac{1}{\rho_0}\left[-e_{21}\frac{\partial\sigma_{11}}{\partial e_{22}}+(1+e_{11})\frac{\partial\sigma_{21}}{\partial e_{22}}\right], \tag{51}$$

$$B_{61}\equiv-\frac{1}{\rho_0}\left[(1+e_{22})\frac{\partial\sigma_{12}}{\partial e_{12}}-e_{12}\frac{\partial\sigma_{22}}{\partial e_{12}}+(1+e_{11})\frac{\partial\sigma_{22}}{\partial e_{11}}-e_{21}\frac{\partial\sigma_{12}}{\partial e_{11}}\right], \tag{52}$$

$$B_{62}\equiv-\frac{1}{\rho_0}\left[-e_{21}\frac{\partial\sigma_{12}}{\partial e_{12}}+(1+e_{11})\frac{\partial\sigma_{22}}{\partial e_{12}}\right], \tag{53}$$

$$B_{63}\equiv-\frac{1}{\rho_0}\left[(1+e_{22})\frac{\partial\sigma_{12}}{\partial e_{22}}-e_{12}\frac{\partial\sigma_{22}}{\partial e_{22}}+(1+e_{11})\frac{\partial\sigma_{22}}{\partial e_{21}}-e_{21}\frac{\partial\sigma_{12}}{\partial e_{21}}\right], \tag{54}$$

$$B_{64}=-\frac{1}{\rho_0}\left[-e_{21}\frac{\partial\sigma_{12}}{\partial e_{22}}+(1+e_{11})\frac{\partial\sigma_{22}}{\partial e_{22}}\right], \tag{55}$$

with all other elements equal to zero. Because the material is isotropic we would naturally expect that the $\underline{\underline{A}}$ and $\underline{\underline{B}}$ coefficient matrices would be very similar to each other. However, we have anticipated the fact that we will consider only quasi-plane waves which propagate primarily in the X_1 direction. Most of our calculations will then involve the $\underline{\underline{A}}$ matrix and we have simplified this matrix by using the definition (23) to replace

$$\frac{\partial e_{12}}{\partial X_1}\ \text{by}\ \frac{\partial e_{11}}{\partial X_2}\quad\text{and}\quad\frac{\partial e_{22}}{\partial X_1}\ \text{by}\ \frac{\partial e_{21}}{\partial X_2},$$

wherever they occur.

The undisturbed state of the solid will be taken to be initially at rest and to have at most a small constant shear strain. As a result, we will take the undisturbed state to be

$$v_i=0,\quad u_1=u_3=0,\quad u_2=e_{21}^{(0)}X_1, \tag{56}$$

for all X_i. Here, $e_{21}^{(0)}$ will be taken to be small and constant. The value of the vector (40) in the undisturbed configuration may therefore be written

$$\mathbf{w}_0=\begin{pmatrix} 0\\ 0\\ e_{21}^{(0)}\\ 0\\ 0\\ 0\end{pmatrix}. \tag{57}$$

The main goal of this paper is to derive the evolution equation governing weakly nonlinear, quasi-plane waves. It is well known that longitudinal (or *p*-) waves will have non-negligible quadratic nonlinearity and will therefore be governed by the classical Khokhlov–Zabolotskaya equation. On the other hand, shear (or *s*-) waves will have zero quadratic nonlinearity in the absence of prestrain. Because the shear mode will be governed by the nonclassical form of the Khokhlov–Zabolotskaya equation described by Kluwick and Cox (1998) and Cramer and Webb (1998), we will focus on the shear mode only.

The wave evolution will be governed by a form of the Khokhlov–Zabolotskaya equation whenever the effects of nonlinearity and diffraction occur at the same rate. For the shear mode this will require that the initial pulse shape should be such that the length scale associated with the transverse variation, i.e., the variations in the X_2 direction, is of order L/ϵ, where ϵ is a small, nondimensional parameter measuring the amplitude of the pulse and L is the disturbance length scale in the X_1 (or propagation) direction. As dis-

cussed by Kluwick and Cox (1998) and Cramer and Webb (1998), these waves will undergo significant distortion over time scales of order $L/(\epsilon^2 c_s)$, where c_s is the linearized shear wave speed defined more precisely in the next section.

One of the contributions of the present study is to extend the work of Zabolotskaya (1986) and Andrews (1999) to include the effects of prestrain. Here, we will take the prestrain to be small and, in particular, we will take $e_{21}^{(0)} = 0(\epsilon)$. With this restriction on the prestrain, the two constraints on the matrices $\underline{\underline{A}}$ and $\underline{\underline{B}}$ which were imposed by Kluwick and Cox (1998) and Cramer and Webb (1998) are satisfied, at least to the accuracy required here. The first constraint requires that the appropriate measure of the quadratic nonlinearity be of order ϵ at most; the fact that this constraint is satisfied is discussed explicitly in Sec. III of the present study. The second constraint is recorded as Eq. (8) of Cramer and Webb (1998) and ensures that the primary propagation direction is in the X_1 direction. Although not discussed explicitly in the present paper, it was shown that the appropriate quantity is of $0(\epsilon^2)$ whenever $e_{12}^{(0)} = 0(\epsilon)$, and therefore is sufficiently small to apply the theory of Cramer and Webb (1998).

III. THE MODIFIED KZ EQUATION

In this section we present the specific form of the modified KZ equation corresponding to the hyperelastic solid described in Sec. II. Detailed expressions for the solution vector (40) will also be provided. The general form of (39) and the restrictions on the relative sizes of the amplitude and the length and time scales are identical to those considered by Cramer and Webb (1998). Thus, we can simply evaluate the expressions developed by Cramer and Webb (1998) for the specific system of interest here. Because we will frequently need to refer to the article by Cramer and Webb (1998), we will use the shorthand notation CW for that study.

We should point out that the notation used here will differ slightly from that found in CW. For example, the spatial variables x and y and the dependent variable $\mathbf{u}$ introduced in Eq. (2) of CW will need to be replaced by X_1, X_2, and $\mathbf{w}$, respectively, in the present study.

Under the conditions described in Sec. II of the present investigation and, equivalently, CW, the general solution for (40) is

$$\mathbf{w} = \mathbf{w}_0 + \epsilon \mathbf{r} W + \epsilon^2 \left(\boldsymbol{\delta} V + \boldsymbol{\gamma} \frac{W^2}{2} + \mathbf{r} h \right) + 0(\epsilon^3), \tag{58}$$

where $W \equiv W(X, X_2, t)$ is a scalar shape function related to the main disturbance, $V \equiv V(X, X_2, t)$ is a second scalar function related to the disturbances induced by diffraction, and $X \equiv X_1 - \lambda t$ is a wave coordinate measured in a frame moving at the linear plane-wave speed λ. The 6×1 vectors $\mathbf{r}$, $\boldsymbol{\delta}$, and $\boldsymbol{\gamma}$ provide the relative sizes of the components of $\mathbf{w}$ for the various perturbations of the solution vector $\mathbf{w}$. The quantity $\mathbf{r}h$ seen in (58) represents an arbitrary homogeneous solution parallel to $\mathbf{r}$.

As discussed by CW, $\mathbf{r}$ and λ are the eigenvectors and eigenvalues of $\underline{\underline{A}}^{(0)} \equiv \underline{\underline{A}}(\mathbf{w}_0) =$ speed matrix $\underline{\underline{A}}$ evaluated at the undisturbed state. As in the zero prestrain case studied by Andrews (1999), a total of six eigenvalues were found for the $\underline{\underline{A}}^{(0)}$ matrix associated with (42)–(55). Two of these eigenvalues were zero and correspond to nonpropagating modes. The squares of the eigenvalues corresponding to the propagating modes were found to be

$$\lambda_s^2 = c_{so}^2 + \frac{e_{21}^{(0)2}}{2\rho_0} \left\{ \left. \frac{\partial^3 \sigma_{12}}{\partial e_{21}^3} \right|_0 + \frac{2}{\rho_0 (c_{so}^2 - c_{po}^2)} \left. \frac{\partial^2 \sigma_{11}}{\partial e_{21}^2} \right|_0 \left. \frac{\partial^2 \sigma_{12}}{\partial e_{11} \partial e_{21}} \right|_0 \right\} + 0(e_{21}^{(0)3}), \tag{59}$$

$$\lambda_p^2 = c_{po}^2 + \frac{e_{21}^{(0)2}}{2\rho_0} \left\{ \left. \frac{\partial^3 \sigma_{11}}{\partial e_{11} \partial e_{21}^2} \right|_0 + \frac{2}{\rho_0 (c_{po}^2 - c_{so}^2)} \left. \frac{\partial^2 \sigma_{11}}{\partial e_{21}^2} \right|_0 \left. \frac{\partial^2 \sigma_{12}}{\partial e_{11} \partial e_{21}} \right|_0 \right\} + 0(e_{21}^{(0)3}), \tag{60}$$

where

$$c_{po} \equiv \left(\frac{1}{\rho_0} \left. \frac{\partial \sigma_{11}}{\partial e_{11}} \right|_0 \right)^{1/2} = \left(2 \frac{\tilde{\beta}_0 + \tilde{\alpha}_1|_0}{\rho_0} \right)^{1/2}, \tag{61}$$

$$c_{so} \equiv \left(\frac{1}{\rho_0} \left. \frac{\partial \sigma_{12}}{\partial e_{21}} \right|_0 \right)^{1/2} = \left(\frac{\tilde{\alpha}_1|_0}{\rho_0} \right)^{1/2} \tag{62}$$

are the p- and s-mode wave speeds in the unstrained state. When we recall (16) the speeds (61)–(62) are recognized as the classical results for the body-wave speeds in linear elasticity.

In the course of the derivation of (59) and (60) it was necessary to show that

$$\left. \frac{\partial \tilde{\alpha}_0}{\partial e_{11}} \right|_0 = 2\tilde{\beta}_0,$$

$$\left. \frac{\partial \tilde{\alpha}_0}{\partial e_{21}} \right|_0 = \left. \frac{\partial \tilde{\alpha}_1}{\partial e_{21}} \right|_0 = \left. \frac{\partial \tilde{\alpha}_2}{\partial e_{21}} \right|_0 = 0,$$

$$\left. \frac{\partial^2 \tilde{\alpha}_0}{\partial e_{21} \partial e_{11}} \right|_0 = \left. \frac{\partial^2 \tilde{\alpha}_1}{\partial e_{21} \partial e_{11}} \right|_0 = \left. \frac{\partial^2 \tilde{\alpha}_2}{\partial e_{21} \partial e_{11}} \right|_0 = 0.$$

These results as well as others not recorded here will be used without explicit comment in the remainder of this paper. Examples of the details of such calculations can be found in the thesis of Andrews (1999).

In all that follows we will focus on the shear or s-mode only. When this is done we find that the components of the eigenvector $\mathbf{r}$ corresponding to (59) are

$$r_1 = r_3 \frac{e_{21}^{(0)}}{\rho_0 (c_{so}^2 - c_{po}^2)} \left. \frac{\partial^2 \sigma_{11}}{\partial e_{21}^2} \right|_0 + 0(e_{21}^{(0)2}), \tag{63}$$

$$r_2 = r_4 = 0, \tag{64}$$

$$r_5 = -\lambda_s r_1 = -\lambda_s r_3 \frac{e_{21}^{(0)}}{\rho_0(c_{so}^2 - c_{po}^2)} \left.\frac{\partial^2 \sigma_{11}}{\partial e_{21}^2}\right|_0 + 0(e_{21}^{(0)2}), \tag{65}$$

$$r_6 = -\lambda_s r_3, \tag{66}$$

respectively. Throughout this paper, we will take r_3 to be the nonzero but otherwise arbitrary component of the eigenvector $\mathbf{r}$. Following the procedure of CW, we find that the vectors $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ can be written

$$\boldsymbol{\delta} = r_3 \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \\ \lambda_s \\ 0 \end{pmatrix} + 0(\epsilon), \tag{67}$$

and

$$\boldsymbol{\gamma} = \gamma_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ -\lambda_s \\ 0 \end{pmatrix} + 0(\epsilon), \tag{68}$$

where

$$\gamma_1 \equiv \frac{r_3^2}{\rho_0(c_{so}^2 - c_{po}^2)} \left.\frac{\partial^2 \sigma_{11}}{\partial e_{21}^2}\right|_0 + 0(\epsilon). \tag{69}$$

In (67)–(68) any portions of the vectors $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ which are parallel to $\mathbf{r}$ can be shown to make no contribution to the resultant evolution equations and, in (58), can be absorbed in the $\mathbf{r}h$ term. Thus, any part of (67)–(68) which is parallel to $\mathbf{r}$ can be ignored with no loss in generality.

An explicit form of the solution for $\mathbf{w}$ can be obtained by combining (58), (63)–(69) with (40) to obtain

$$e_{11} = \epsilon r_1 W + \epsilon^2 \left[-r_3 V + \gamma_1 \frac{W^2}{2} \right] + 0(\epsilon^3), \tag{70}$$

$$e_{12} = 0(\epsilon^3), \tag{71}$$

$$e_{21} = e_{21}^{(0)} + \epsilon r_3 W + 0(\epsilon^3), \tag{72}$$

$$e_{22} = \epsilon^2 r_3 V + 0(\epsilon^3), \tag{73}$$

$$v_1 = -\lambda_s e_{11} + 0(\epsilon^3), \tag{74}$$

$$v_2 = -\lambda_s e_{21} + 0(\epsilon^3). \tag{75}$$

When we recall that we are taking $e_{21}^{(0)} = 0(\epsilon)$, we see that both e_{11} and e_{22} are of second order in the shear pulse amplitude. If we combine (23) and (38) with (74), we can show that the disturbance to e_{21} gives rise to a disturbance in the longitudinal displacement u_1 which, to lowest order, travels with the main pulse at speed λ_s. The contributions to e_{11} are from three sources. The first contribution can be written as

$$\epsilon r_1 W \approx (e_{21} - e_{21}^{(0)}) \frac{e_{21}^{(0)}}{\rho_0(c_{so}^2 - c_{po}^2)} \left.\frac{\partial^2 \sigma_{11}}{\partial e_{21}^2}\right|_0 \approx (e_{21} - e_{21}^{(0)}) e_{21}^{(0)} \frac{\gamma_1}{r_3^2},$$

where we have used (63), (69), and (72). This term represents the part of e_{11} generated by the interaction of the shear wave and the prestrain. The second contribution $-\epsilon^2 r_3 V \approx -e_{22}$ is due to the diffraction, i.e., the transverse variations. The final contribution is due to the nonlinearity of the stress–strain relation (5) and can be written

$$\epsilon^2 \gamma_1 \frac{W^2}{2} \approx \frac{\gamma_1}{2r_3^2}(e_{21} - e_{21}^{(0)})^2,$$

where we have used (72).

The fact that the passage of the shear wave causes a density perturbation of second order can be seen by adding (70) to (73). We then use (63), (69), and (72) to obtain

$$e_{11} + e_{22} \approx \frac{1}{2} \frac{\gamma_1}{r_3^2}(e_{21}^2 - e_{21}^{(0)2}) + 0(\epsilon^3). \tag{76}$$

If we further expand (14) and the last of (24) in a manner which is consistent with (70)–(75), we find that

$$\frac{\rho}{\rho_0} \approx 1 - \frac{1}{2}(III - 1) + 0(\epsilon^4) \approx 1 - (e_{11} + e_{22}) + 0(\epsilon^4),$$

which, when combined with (76), yields

$$\frac{\rho}{\rho_0} \approx 1 - \frac{1}{2} \frac{\gamma_1}{r_3^2}(e_{21}^2 - e_{21}^{(0)2}) + 0(\epsilon^3), \tag{77}$$

where γ_1 / r_3^2 can be obtained from (69). Result (77) is consistent with the well-known result that material nonlinearity can generate density perturbations under shear loads.

We now turn to the determination of the specific form of the KZ equation corresponding to the hyperelastic material described in Sec. II. In CW the general form of the KZ equation in terms of the shape functions W and V was given. If we use (72) and (73) to replace W and V by e_{21} and e_{22}, we find that the evolution equations governing weakly nonlinear, weakly diffracting waves in hyperelastic solids are

$$\frac{\partial e_{21}}{\partial t} \pm c_{so}\left[\tilde{\Gamma} + \frac{\tilde{\Lambda}}{2}(e_{21} - e_{21}^{(0)})\right](e_{21} - e_{21}^{(0)}) \frac{\partial e_{21}}{\partial X} \pm \frac{c_{so}}{2} \frac{\partial e_{22}}{\partial X_2} = 0, \tag{78}$$

$$\frac{\partial e_{22}}{\partial X} = \frac{\partial e_{21}}{\partial X_2}, \tag{79}$$

to lowest order. Here, $X \equiv X_1 - \lambda_s t$

$$\tilde{\Lambda}\equiv\frac{1}{2\rho_0 c_{so}^2}\left[\left.\frac{\partial^3\sigma_{12}}{\partial e_{21}^3}\right|_0 + \frac{3}{\rho_0(c_{so}^2-c_{po}^2)}\left.\frac{\partial^2\sigma_{11}}{\partial e_{21}^2}\right|_0\left.\frac{\partial^2\sigma_{12}}{\partial e_{21}\partial e_{11}}\right|_0\right]+0(\epsilon), \tag{80}$$

$$\tilde{\Gamma}\equiv\frac{e_{21}^{(0)}}{2\rho_0 c_{so}^2}\left[\left.\frac{\partial^3\sigma_{12}}{\partial e_{21}^3}\right|_0 + \frac{3}{\rho_0(c_{so}^2-c_{po}^2)}\left.\frac{\partial^2\sigma_{11}}{\partial e_{21}^2}\right|_0\left.\frac{\partial^2\sigma_{12}}{\partial e_{21}\partial e_{11}}\right|_0\right] + 0(e_{21}^{(0)2})\approx e_{21}^{(0)}\tilde{\Lambda}. \tag{81}$$

The $\pm$ signs in (78) correspond to right and left running waves, and the differential with respect to time seen in (78) is taken holding X and X_2 constant.

Once (78)–(79) are solved for e_{21} and e_{22}, e_{11} can be determined from (76) and u_1, u_2 can be determined by straightforward integration once it is recognized that (71) implies that $u_1=u_1(X_1,t)$ to lowest order.

We note that (79) can be further simplified through use of (23) and conversion back to X_1, X_2, t coordinates to yield

$$\frac{\partial^2 u_2}{\partial X_1\partial X_2}=\frac{\partial^2 u_2}{\partial X_2\partial X_1}. \tag{82}$$

Thus, (79) can be recognized as a compatibility condition on the shear displacement $u_2(X_1,X_2,t)$.

The term proportional to $\tilde{\Gamma}$ introduces quadratic nonlinearity into the evolution equation (78). Inspection of (81) reveals that $\tilde{\Gamma}\to 0$ as the prestrain $e_{21}^{(0)}\to 0$. Thus, our restriction that $e_{21}^{(0)}=O(\epsilon)$ guarantees that the quadratic nonlinearity is weak and therefore that condition (12) of CW holds.

The term proportional to $\tilde{\Lambda}$ in (78) represents cubic nonlinearity and is seen to be consistent with that obtained by Zabolotskaya (1986) for transverse waves in a nonlinear isotropic solid once differences in coordinate systems are taken into account. We also note that the one-dimensional version of (78) is consistent with the work of Carman and Cramer (1992) when their result is simplified using the small disturbance approximation.

In the course of the derivation of (78)–(79), it was found that the terms A, B defined in Eqs. (28) and (29) of CW were of order ϵ and are therefore negligible in the physical system considered here. The coefficients A and B are proportional to the coefficients β''' and $\beta^{(iv)}$ in (4) of the present investigation and represent terms generated by an interaction of the weak nonlinearity and the weak diffraction. Thus, in spite of the complexity of the original physical system (39)–(40) and (42)–(55), the primary effect of the small prestrain is seen to be the modification of the linear wave speed (59) and the addition of the term representing quadratic nonlinearity to the KZ equations of Zabolotskaya (1986) and Andrews (1999).

Because the scaled value of the quadratic nonlinearity coefficient $\tilde{\Gamma}$ is proportional to $\tilde{\Lambda}$, it can be shown that we can recast (78) in an alternate form which reads

$$\frac{\partial e_{21}}{\partial t}\pm c_{so}\frac{\tilde{\Lambda}}{2}(e_{21}^2-e_{21}^{(0)2})\frac{\partial e_{21}}{\partial X}\pm\frac{c_{so}}{2}\frac{\partial e_{22}}{\partial X_2}=0. \tag{83}$$

This form of the evolution equation is particularly useful when considering the effect of the sign of the prestrain $e_{21}^{(0)}$ on the wave motion. Inspection of (83) reveals that the evolution is invariant with respect to the sign of $e_{21}^{(0)}$. This conclusion may not be immediately obvious when we recall that the quadratic nonlinearity parameter (81) changes sign as the prestrain changes sign.

Finally, we note that the evolution equation (78) can be reduced to a modified or cubic KZ equation by differentiation with respect to X and use of (79); the result is found to be

$$\left[e_t\pm c_{so}\left(\tilde{\Gamma}e+\frac{\tilde{\Lambda}}{2}e^2\right)e_X\right]_X\pm\frac{c_{so}}{2}e_{X_2X_2}=0, \tag{84}$$

where we have replaced $e_{21}-e_{21}^{(0)}$ by e for convenience and the subscripts X, X_2, t again denote partial derivatives. Equation (84) is of the same general form as (3) with an additional term representing quadratic nonlinearity.

A second reduction of (78)–(79) to a single equation can be obtained by noting that (79) implies the existence of a potential function. In fact, (82) implies that the appropriate potential is the shear displacement u_2. It will be convenient to define

$$u\equiv u_2-e_{21}^{(0)}X_1, \tag{85}$$

which is a measure of the actual shear displacement from the undisturbed state. When (85) is combined with (23) and (78), we find that

$$u_{Xt}\pm c_{so}\left[\tilde{\Gamma}u_X+\frac{\tilde{\Lambda}}{2}(u_X)^2\right]u_{XX}\pm\frac{c_{so}}{2}u_{X_2X_2}=0, \tag{86}$$

where we have used the fact that $e_{22}=u_{X_2}$ and

$$e_{21}\equiv\frac{\partial u_2}{\partial X_1}=\frac{\partial u_2}{\partial X}\approx\frac{\partial u}{\partial X}+e_{21}^{(0)}.$$

We note that (86) is a cubically nonlinear version of the low-frequency transonic equation derived by Cramer and Seebass (1978) in the context of focusing shock waves in perfect gases.

In closing this section we note that Eqs. (78)–(79), (84), and (86) are written in a coordinate system which translates at speed λ_s in the X_1 direction. When we transform back to a frame which is stationary with respect to the undisturbed solid, we find that the counterpart of (78)–(79) can be obtained simply by replacing $X(\equiv X_1-\lambda_x t)$, X_2, t by X_1, X_2, t in (78)–(79) and by adding the term $\lambda_s\partial e_{21}/\partial X_1$ to the result of replacing X, X_2, t by X_1, X_2, t. As a final step, it can also be shown that no additional terms arise when our equations are converted from the Lagrangian coordinates X_i to the Eulerian coordinates x_i. That is, it was found that the KZ equation was essentially the same form regardless of whether Lagrangian or Eulerian coordinates are employed, at least to the accuracy needed here.

IV. DISSIPATIVE WAVES

In this section we show how weak dissipation modifies the evolution equation (78). Simple dissipation will be included through use of a Kelvin–Voigt model. To incorporate this model we write the stress (5) as

$$\sigma_{ij}=\sigma_{ij}^{(e)}+\sigma_{ij}^{(v)}, \tag{87}$$

where $\sigma_{ij}^{(e)}$ is the elastic part of the stress identical to (5). The second term is the viscous part of the stress given by

$$\sigma_{ij}^{(v)}\equiv\lambda_v\frac{\partial v_\ell}{\partial X_\ell}\delta_{ij}+\mu_v\left(\frac{\partial v_i}{\partial X_j}+\frac{\partial v_j}{\partial X_i}\right), \tag{88}$$

where μ_v and λ_v are shear and second viscosities, respectively. As discussed by CW, only the lowest-order contributions to the dissipative terms will be required. As a result, the linear stress–rate of strain relation (88) will be adequate for our purposes. Furthermore, μ_v and λ_v may be taken to be constants and the distinction between Lagrangian and Eulerian coordinates can be ignored in (88).

When (87)–(88) are substituted in (15), we find that the system of equations governing the wave propagation is

$$\frac{\partial\mathbf{w}}{\partial t}+\underline{\underline{A}}\frac{\partial\mathbf{w}}{\partial X_1}+\underline{\underline{B}}\frac{\partial\mathbf{w}}{\partial X_2}\approx\underline{\underline{D}}^{(xx)}\frac{\partial^2\mathbf{w}}{\partial X_1^2}+\underline{\underline{D}}^{(xy)}\frac{\partial^2\mathbf{w}}{\partial X_1\partial X_2}+\underline{\underline{D}}^{(yy)}\frac{\partial^2\mathbf{w}}{\partial X_2^2}, \tag{89}$$

where $\mathbf{w}$, $\mathbf{w}_0$, $\underline{\underline{A}}$, $\underline{\underline{B}}$ are identical to those given in Sec. II. The only difference between (89) and (39) is the presence of the dissipative terms having the $N\times N$ matrices $\underline{\underline{D}}^{(xx)}$, $\underline{\underline{D}}^{(xy)}$, $\underline{\underline{D}}^{(yy)}$ as coefficients. The only nonzero components of the latter are found to be

$$D_{55}^{(xx)}=\frac{\lambda_v+2\mu_v}{\rho_0},\quad D_{66}^{(xx)}=\frac{\mu_v}{\rho_0}, \tag{90}$$

$$D_{56}^{(xy)}=D_{65}^{(xy)}=\frac{\lambda_v+\mu_v}{\rho_0}, \tag{91}$$

$$D_{55}^{(yy)}=\frac{\mu_v}{\rho_0},\quad D_{66}^{(yy)}=\frac{\lambda_v+2\mu_v}{\rho_0}. \tag{92}$$

The effects of weak dissipation on the shear waves will become noticeable at times of order ρ_0L^2/μ_v. If the dissipative effects are much larger than the effects of nonlinearity and diffraction, the shear pulse will damp out long before the latter effects will be noticeable. On the other hand, pulses for which the dissipation is much weaker than the nonlinearity will evolve with no noticeable dissipation, except in the interior of shock waves, over the time scales considered here. We will therefore take the length scale, L, and the pulse amplitude to be such that dissipation, nonlinearity, and diffraction all occur at the same rate. As a result, we require that $\rho_0L^2/\mu_v=0(L/(\epsilon^2c_{so}))$, or that

$$\mathrm{Re}_s\equiv\frac{c_{so}\rho_0L}{\mu_v}=0\left(\frac{1}{\epsilon^2}\right), \tag{93}$$

where Re_s is recognized as the Reynolds number based on the linearized shear wave speed.

An extension of the Khokhlov–Zabolotskaya equation to include weak relaxation, dissipation, and dispersion has been derived by CW for the case where these effects are of the same strength as the nonlinearity and diffraction. When Cramer and Webb's scheme is applied to (89), it is found that the dissipative form of the evolution equation (78) may now be written

$$\frac{\partial e_{21}}{\partial t}\pm c_{so}\left[\tilde{\Gamma}+\frac{\tilde{\Lambda}}{2}(e_{21}-e_{21}^{(0)})\right](e_{21}-e_{21}^{(0)})\frac{\partial e_{21}}{\partial X}\pm\frac{c_{so}}{2}\frac{\partial e_{22}}{\partial X_2}=\frac{\mu_v}{2\rho_0}\frac{\partial^2e_{21}}{\partial X^2}, \tag{94}$$

with (79) unchanged. We note further that the inclusion of material dissipation leaves the basic solutions (70)–(75) unchanged. As a result, the procedure for finding u_1 and u_2 from e_{21}, e_{22} is exactly the same as in the nondissipative case. The only difference between the dissipative and nondissipative cases is the difference in the governing equations (78) and (94). The evolution equation (94) is again seen to be completely consistent with previous investigations as well as the dispersion relation for linear Kelvin–Voigt solids.

Because (79) is unchanged when dissipation is included, we can immediately write down the extensions of (83), (84), and (86); these read

$$\frac{\partial e_{21}}{\partial t}\pm c_{so}\frac{\tilde{\Lambda}}{2}(e_{21}^2-e_{21}^{(0)2})\frac{\partial e_{21}}{\partial X}\pm\frac{c_{so}}{2}\frac{\partial e_{22}}{\partial X_2}=\frac{\mu_v}{2\rho_0}\frac{\partial^2e_{21}}{\partial X^2}, \tag{95}$$

$$\left[e_t\pm c_{so}\left(\tilde{\Gamma}e+\frac{\tilde{\Lambda}}{2}e^2\right)e_X\right]_X\pm\frac{c_{so}}{2}e_{X_2X_2}=\frac{\mu_v}{2\rho_0}e_{XXX}, \tag{96}$$

and

$$u_{Xt}\pm c_{so}\left[\tilde{\Gamma}u_X+\frac{\tilde{\Lambda}}{2}(u_X)^2\right]u_{XX}\pm\frac{c_{so}}{2}u_{X_2X_2}=\frac{\mu_v}{2\rho_0}u_{XXX}, \tag{97}$$

where e and u have the same definitions as in Sec. III and the subscripts X, X_2, t in (96)–(97) denote partial differentiation.

V. SHOCK JUMP CONDITIONS

It is expected that the presence of nonlinearity and dissipation will generally give rise to the formation of shock waves. The shock jump relations can be derived in a standard way by first converting (78)–(79) to a conservative form and then by integrating over a volume enclosing the possible discontinuity. In terms of the original X_1, X_2, t system, we find that the jump conditions for a right-running shock wave can be written

$$\frac{s}{\lambda_s}\approx1+\frac{\tilde{\Gamma}}{2}\frac{[\![e^2]\!]}{[\![e]\!]}+\frac{\tilde{\Lambda}}{6}\frac{[\![e^3]\!]}{[\![e]\!]}+o(\epsilon^2), \tag{98}$$

$$[\![e_{22}]\!]\approx\theta_s[\![e]\!]+o(\epsilon^2), \tag{99}$$

where e is again simply $e_{21}-e_{21}^{(0)}$, the double brackets denote the jump in the indicated quantity across the shock, and

$\theta_s = 0(\epsilon)$ is the acute angle between the shock surface and the positive X_2 axis, measured in a counterclockwise direction. The quantity s is the dimensional shock speed relative to the original X_1, X_2, t frame.

We note that the lowest-order shock speed (98) is identical to that derived in the one-dimensional theory of Carman and Cramer (1992). From the point of view of the shock, the primary way in which the two-dimensional character of the motion is incorporated is through the jump condition (99).

An important feature of shock waves in nonlinear solids can be revealed by an inspection of the characteristic surfaces associated with (78)–(79). If the equation of the characteristic surface is $F(X_1, X_2, t) = 0$, then the unit normal ($\mathbf{n}$) and speed (s_c) of this surface can be written

$$\mathbf{n} \equiv \frac{\nabla F}{|\nabla F|}, \quad s_c \equiv -\frac{F_t}{|\nabla F|}, \tag{100}$$

respectively. If we apply the method of characteristics to (78)–(79) and recognize that $n_{X_2} = 0(\epsilon n_{X_1}) = o(n_{X_1})$, we find that the speed of a right-running characteristic surface can be approximated by

$$\frac{s_c}{\lambda_s} = 1 + \tilde{\Gamma} e + \frac{\tilde{\Lambda}}{2} e^2 + o(\epsilon^2). \tag{101}$$

Thus, the speed of the characteristic surface is only dependent on $e \equiv e_{21} - e_{21}^{(0)}$ and is identical to that of the one-dimensional theory described by Carman and Cramer (1992).

We close this section by noting that the necessary and sufficient conditions for the local existence of shear shocks can be written

$$s_{c1} \leqslant s \leqslant s_{c2}; \tag{102}$$

that is, the shock speed is required to be larger than the wave speed immediately ahead of the shock (defined as s_{c1}) and smaller than the wave speed immediately behind the shock (defined as s_{c2}). Condition (102) is recognized as Lax's (1971) speed-ordering condition and holds in a trivial way even for cases having strictly quadratic nonlinearity. The derivation of (102) is based either on an analysis of the dissipative structure of the shock or on the basis of a casualty argument, i.e., an argument denying the possiblity that events in the present or past can be influenced by events which occur in the future.

A graphical interpretation of (102) can be developed by defining the flux function

$$Q = Q(e) \equiv \frac{\tilde{\Gamma}}{2} e^2 + \frac{\tilde{\Lambda}}{6} e^3. \tag{103}$$

A sketch of the Q vs e curve for $\tilde{\Gamma} > 0$, $\tilde{\Lambda} > 0$ has been provided in Fig. 1. Straightforward differentiation can be used to show that

$$Q' = \tilde{\Gamma} e + \frac{\tilde{\Lambda}}{2} e^2 \approx \frac{s_c}{\lambda_s} - 1, \tag{104}$$

where $Q' \equiv dQ/de$. Futhermore

$$\frac{[\![Q]\!]}{[\![e]\!]} = \frac{\tilde{\Gamma}}{2} \frac{[\![e^2]\!]}{[\![e]\!]} + \frac{\tilde{\Lambda}}{6} \frac{[\![e^3]\!]}{[\![e]\!]} = \frac{s}{\lambda_s} - 1. \tag{105}$$

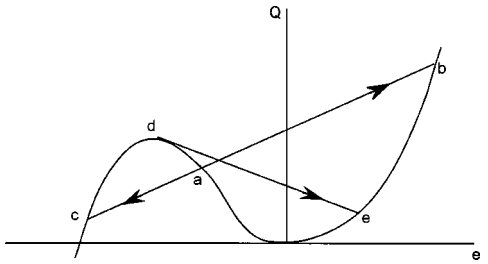


FIG. 1. Sketch of the flux function (103) for the case $\tilde{\Gamma}$, $\tilde{\Lambda} > 0$. Arrows denote the direction of jumps for admissible shocks and d–e denotes a sonic shock.

Thus, the slope of the Q vs e curve is closely related to the speed of the characteristic surfaces, and the slope of the straight line connecting the states ahead of and behind the shock is closely related to the shock speed. Furthermore, (102) can be recast as

$$Q_1' \leqslant \frac{[\![Q]\!]}{[\![e]\!]} \leqslant Q_2', \tag{106}$$

where $Q_i' \equiv Q'(e_i)$ for $i = 1,2$. As in (102), the subscripts 1 and 2 will denote states ahead of and behind the shock, respectively. Thus, the admissibility of a proposed shear discontinuity can be determined by a relatively simple inspection of the corresponding Q vs e curve. As an example, we consider the states *a*, *b*, *c* indicated on the Q curve in Fig. 1. The states *a* and *b* will correspond to states on either side of an admissible shock, i.e., one satisfying (102) or (106), if and only if *a* is the state ahead of the shock and *b* is the state behind the shear shock. To emphasize this fact we have added an arrow in Fig. 1 to indicate the direction of the admissible jump. If we consider a discontinuity between the points *a* and *c* in Fig. 1, we can conclude that the only admissible shock is one for which the solid is taken from state *a* to state *c*. Thus, both *e*-lowering and *e*-raising shocks are possible if both $\tilde{\Lambda}$ and $\tilde{\Gamma}$ are positive; it can be shown that the same conclusion will be valid if $\tilde{\Lambda}$ and $\tilde{\Gamma}$ are simply nonzero.

An example of an inadmissible discontinuity is one having end states given by *c* and *b*. Because the slopes of the Q curve at *c* and *b* are greater than the slope of the chord $c-b$, (106) will always be violated for this discontintuity.

Tangency points are also clearly possible. One such shock is represented by the states $d-e$ in Fig. 1. Such tangency points correspond to the equalities in (102) and (106) and represent a shear shock for which the shock speed is identical to the characteristic speed immediately ahead of or behind the shock. The resultant shock wave is referred to as a sonic shock and is not possible in the classical KZ theory. A detailed analysis of the Q curve in Fig. 1 shows that the sonic shocks will correspond to the maximum or minimum shock strength when such an extremum is not zero or infinite. Furthermore, experience with the one-dimensional theory strongly suggests that sonic shocks will be the result of partial disintegrations of inadmissible discontinuities.

Finally, we consider the unstrained undisturbed state discussed by Zabolotskaya (1986). In this case $e_{21}^{(0)} = 0$, $\tilde{\Gamma} = 0$, and $e = e_{21}$. Thus, (103) reduces to the pure cubic relation

$$Q=\frac{\tilde{\Lambda}}{6}e^3. \tag{107}$$

Sonic shocks as well as shocks carrying both increases and decreases of e are possible. An interesting special case is that where $\tilde{\Lambda}<0$. In this case, no shocks are possible which have $e=0$ ahead of the shock. Thus, when the prestrain is zero, a disturbance propagating into an undisturbed media, i.e., one having $e=0$, cannot be headed by a shock regardless of the degree of focusing or attenuation.

VI. RESULTS FOR A BLATZ–KO MATERIAL

In this section we consider a specific material model and provide explicit formulas for the nonlinearity coefficients derived in Sec. III. The model chosen is that developed by Blatz and Ko (1962) for compressible foam rubbers. To specify any particular material we simply need to give the dependence of the strain energy on the invariants I, II, III. For the Blatz–Ko model this relation is

$$U=\frac{\mu_0 b}{2\rho_0}\left\{I-3+\frac{1-2\nu}{\nu}(III^{-\nu/(1-2\nu)}-1)\right\} + \frac{\mu_0(1-b)}{2\rho_0}\left\{\frac{II}{III}-3+\frac{1-2\nu}{\nu}(III^{\nu/(1-2\nu)}-1)\right\}, \tag{108}$$

where μ_0, ν, and b are material constants.

It can be shown that the Lamé constants (16) for the Blatz–Ko material can be written

$$\mu' \equiv \tilde{\alpha}_1|_0=\mu_0, \tag{109}$$

$$\lambda' \equiv 2\tilde{\beta}_0=2\mu_0\frac{\nu}{1-2\nu}. \tag{110}$$

Thus, the Blatz–Ko parameters μ_0 and ν are just the shear modulus and Poisson ratios of the linear theory. As a result, these parameters may be determined by examining the material response in the small strain limit. The parameter b cannot be determined in the limit of linear elasticity. Blatz and Ko (1962) performed experiments showing that $\mu_0 \approx 32$ psi≈ 221 kPa, $b\approx 0$, and $\nu=1/4$ for the foamed polyurethene rubbers generated in their lab. Further discussion of the Blatz–Ko model and its limitations can be found in the original article and those by Beatty and Stalnaker (1986) and Beatty (1987).

When (108) is substituted in (8)–(10), we find that

$$\alpha_0=\mu_0\left\{(1-b)\left[III^{(4\nu-1)/[2(1-2\nu)]}-\frac{II}{III^{3/2}}\right] - bIII^{-1/[2(1-2\nu)]}\right\}, \tag{111}$$

$$\alpha_1=\frac{\mu_0}{III^{1/2}}\left\{b+\frac{I}{III}(1-b)\right\}, \tag{112}$$

$$\alpha_2=-\mu_0\frac{1-b}{III^{3/2}}, \tag{113}$$

exactly. We note in passing that the values of α_0 and α_1 given by Carman and Cramer (1992) contained typographical errors in the authors' original manuscript. However, those typographical errors had no effect on their reported results.

Results (111)–(113) may now be combined to yield the specific forms of (17)–(19); these read

$$\tilde{\alpha}_0=\mu_0(1-b)\left[III^{(4\nu-1)/[2(1-2\nu)]}+\frac{I-II-1}{III^{3/2}}\right] + \mu_0 b[III^{-1/2}-III^{-1/[2(1-2\nu)]}], \tag{114}$$

$$\tilde{\alpha}_1=\mu_0\left[\frac{b}{III^{1/2}}+(1-b)\frac{I-2}{III^{3/2}}\right], \tag{115}$$

$$\tilde{\alpha}_2=\alpha_2=-\mu_0\frac{1-b}{III^{3/2}}, \tag{116}$$

exactly. Results (114)–(115) were used to generate the expressions for the Lamé constants (109)–(110) for the Blatz–Ko material. The results for $\tilde{\beta}_0$, $\tilde{\alpha}_1|_0$ may be combined with (61)–(62) to obtain the linearized shear and longitudinal wave speeds for the unstrained material. In the remainder of this section we use (114)–(116) to compute the parameters associated with the nonlinear aspects of the wave propagation.

We begin by noting that the only derivatives of σ_{ij} which are required can be written

$$\left.\frac{\partial^2\sigma_{11}}{\partial e_{21}^2}\right|_0=2\left[\left.\frac{\partial\tilde{\alpha}_0}{\partial I}\right|_0+\left.\frac{\partial\tilde{\alpha}_0}{\partial II}\right|_0+\tilde{\alpha}_2|_0\right], \tag{117}$$

$$\left.\frac{\partial^2\sigma_{12}}{\partial e_{21}\partial e_{11}}\right|_0=2\tilde{\beta}_1+\tilde{\alpha}_1|_0+2\tilde{\alpha}_2|_0, \tag{118}$$

$$\left.\frac{\partial^3\sigma_{12}}{\partial e_{21}^3}\right|_0=6\left[\left.\frac{\partial\tilde{\alpha}_1}{\partial I}\right|_0+\left.\frac{\partial\tilde{\alpha}_1}{\partial II}\right|_0+\tilde{\alpha}_2|_0\right], \tag{119}$$

where

$$\tilde{\beta}_1\equiv\left.\frac{\partial\tilde{\alpha}_1}{\partial I}\right|_0+2\left.\frac{\partial\tilde{\alpha}_1}{\partial II}\right|_0+\left.\frac{\partial\tilde{\alpha}_1}{\partial III}\right|_0. \tag{120}$$

Details of the derivations leading to (117)–(120) can be found in the thesis by Andrews (1999). When (114)–(116) are substituted in (117)–(120), we find that

$$\left.\frac{\partial^2\sigma_{11}}{\partial e_{21}^2}\right|_0=\left.\frac{\partial^2\sigma_{12}}{\partial e_{21}\partial e_{11}}\right|_0=-2\mu_0(1-b), \tag{121}$$

$$\left.\frac{\partial^3\sigma_{12}}{\partial e_{21}^3}\right|_0=0. \tag{122}$$

We may now determine the parameters (69) and (80)–(81) for the Blatz–Ko material. The lowest-order values of these quantities read

$$\frac{\gamma_1}{r_3^2}=\frac{2c_{so}^2}{c_{po}^2-c_{so}^2}(1-b), \tag{123}$$

$$\tilde{\Lambda}=-\frac{6c_{so}^2}{c_{po}^2-c_{so}^2}(1-b)^2, \quad \tilde{\Gamma}=\tilde{\Lambda}e_{21}^{(0)}. \tag{124}$$

In like manner, the perturbation to the linear wave speed and density may be computed by substituting (121)–(123) in (59) and (77) to yield

$$\lambda_s \approx \pm c_{so}\left[1-e_{21}^{(0)2}\frac{2c_{so}^2}{c_{po}^2-c_{so}^2}(1-b)^2+0(\epsilon^3)\right], \tag{125}$$

$$\frac{\rho}{\rho_0}=1-\frac{c_{so}^2}{c_{po}^2-c_{so}^2}(1-b)[e_{21}^2-e_{21}^{(0)2}]+0(\epsilon^3). \tag{126}$$

If we make the usual assumptions that $\mu'>0$ and $0<\nu<1/2$ for any linear material, then $c_{po}^2>c_{so}^2$. Thus, (124)–(125) may be used to show that $\tilde{\Lambda}<0$ and that the magnitude of the linear wave speed is decreased by any nonzero prestrain $e_{21}^{(0)}$. If we further take $b<1$ we conclude that shear waves decrease the material density from ρ_0 whenever $|e_{21}|>|e_{21}^{(0)}|$. In particular, $\rho<\rho_0$ for a $b<1$ Blatz–Ko material having no prestrain, at least in the context of the small disturbance theory presented here.

Inspection of (123)–(126) reveals that the effects of nonlinearity are related in a simple way to the material constant b. Blatz and Ko (1962) found that $b\approx 0$, $\nu=1/4$ for their polyurethene foams, whereas they found that $b\approx 1$, $\nu\approx 0.463$ for solid polyurethene rubber. If we note that

$$\frac{c_{so}^2}{c_{po}^2-c_{so}^2}=1-2\nu,$$

we find that

$$\frac{\gamma_1}{r_3^2}\approx 1, \quad \tilde{\Lambda}\approx -3, \quad \tilde{\Gamma}\approx -3e_{21}^{(0)},$$

for Blatz and Ko's polyurethene forms. On the other hand, the nonlinearity appears to vanish in the limit of $b\to 1$. That is, the Blatz–Ko model predicts that the propagation of quasi-plane shear waves is at most quartic in the $b\to 1$ limit. Experience has shown that the evolution equation must be modified even when $1-b=0(\epsilon)$. Under these conditions, the time scale over which nonlinearity becomes noticeable is proportional to ϵ^{-3} rather than the ϵ^{-2} used here. Furthermore, diffraction can balance the weakened nonlinearity only if the transverse gradients are also correspondingly weaker.

Such a reduction in the nonlinearity should not be surprising when we note that (108) reduces to

$$U\approx\frac{\mu_0}{2\rho_0}(I-3), \tag{127}$$

when $b\approx 1$; here, we have noted that (126) implies that the density perturbations induced by the shear wave are of order ϵ^3 at most. The strain energy (127) is recognized as the energy of a neo-Hookean solid which in turn is just a simplification of the classical Mooney–Rivlin strain enegy. It is well known that the Mooney–Rivlin model yields a linear response in a simple (or homogeneous) shear.

It should be noted that other, more accurate, models of solid rubbers have nonzero cubic nonlinearity parameters. One such example is Ogden's (1972) model. The behavior of the nonlinearity coefficients as $b\to 1$ seen here should be regarded as a characteristic of the Blatz–Ko and Mooney–Rivlin family of models rather that as a general statement of the behavior of all solid rubbers.

VII. SUMMARY

We have examined weakly nonlinear, weakly diffracting shear waves in a hyperelastic solid. The extension of the Khokhlov–Zabolotskaya equation which is valid for small shear prestrains has been derived. Two alternate forms of our extended KZ equation were presented; the nondissipative forms are given by (78)–(79), (84), and (86) and the dissipative forms are given by (95)–(97).

The primary effect of the prestrain was seen to be to modify the linear wave speed and to change the nonlinearity from a pure cubic nonlinearity to a mixed cubic and quadratic nonlinearity. Surprisingly, the terms representing interactions between nonlinearity and diffraction found in the general modified KZ equation derived by Kluwick and Cox (1998) and CW were found to be negligible even when the prestrain is nonzero. To the authors' knowledge, the present case is the largest and most complicated physical system for which such nonlinear diffraction terms play no role in the lowest order evolution.

A second new result is that for the shock jump conditions (98)–(99). The shock existence conditions were also given in terms of the speed-ordering condition (102). We have also provided a short discussion of a graphical approach to the analysis of the admissibility of proposed discontinuities. A new type of shock, not possible in the classical KZ theory and referred to as a sonic shock, was briefly described. Experience with the one-dimensional theory suggests that such shocks will play a key role in the nonlinear diffraction of shear shocks.

Finally, we have given examples of the nonlinearity constants for a specific material model. The model chosen was based on the well-known Blatz–Ko formula (108) for the strain energy. The explicit form of the relevant constants is given by (124). If we take $\mu'>0$, $0<\nu<1/2$, and $b\neq 1$, (124) reveals that $\tilde{\Lambda}<0$ for all Blatz–Ko materials. If we further restrict attention to the case of no prestrain studied by Zabolotskaya (1986), a disturbance propagating into an undisturbed solid will never be headed by a shock wave. In this $e_{12}^{(0)}=0$, $\tilde{\Lambda}<0$ case, the resultant wavefront will always be smooth and will travel with the linearized wave speed.

Andrews, M. F. (**1999**). "Evolution equations for weakly nonlinear, quasi-planar waves in isotropic dielectrics and elastomers," Masters thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Beatty, M. F. (**1987**). "Topics in finite elasticity: Hyperelasticity of rubber, elastomers, and biological tissues—with examples," Appl. Mech. Rev. **40**(12), 1699–1734.

Beatty, M. F., and Stalnaker, D. O. (**1986**). "The Poisson function of finite elasticity," J. Appl. Mech. **53**, 807–813.

Blatz, P. J., and Ko, W. L. (**1962**). "Application of finite elastic theory to the deformation of rubbery materials," Trans. Soc. Rheol. **VI**, 223–251.

Carman, G. P., and Cramer, M. S. (**1992**). ''Nonlinearity parameters for pulse propagation in isotropic elastomers,'' J. Acoust. Soc. Am. **91**(1), 39–51.
Cramer, M. S., and Seebass, A. R. (**1978**). ''Focusing of weak shock waves at an arête,'' J. Fluid Mech. **142**, 9–37.
Cramer, M. S., and Webb, C. (**1998**). ''A Zabolotskaya–Khokhlov equation for cubic and near-cubic nonlinearity,'' AIAA Paper AIAA-98-2957.
Kluwick, A., and Cox, E. A. (**1998**). ''Nonlinear waves in materials with mixed nonlinearity,'' Wave Motion **27**, 23–41.
Lax, P. D. (**1971**). ''Shock waves and entropy,'' in *Contributions to Nonlinear Functional Analysis*, edited by E. H. Zarantonello (Academic, New York).
Norris, A. N. (**1998**). ''Finite-amplitude waves in solids,'' in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, New York), pp. 263–277.
Ogden, R. W. (**1972**). ''Large deformation isotropic elasticity—on the correlation of theory and experiment for incompressible rubberlike solids,'' Proc. R. Soc. London, Ser. A **326**, 565–584.
Zabolotskaya, E. A. (**1986**). ''Sound beams in a nonlinear isotropic solid,'' Sov. Phys. Acoust. **32**(4), 296–299.
Zabolotskaya, E. A., and Khokhlov, R. V. (**1969**). ''Quasi-plane waves in the nonlinear acoustics of confined beams,'' Sov. Phys. Acoust. **15**(1), 35–40.

Numerical simulation of the excitation of a Helmholtz resonator by a grazing flow

S. Mallick and R. Shock
EXA Corporation, Lexington, Massachusetts 02420

V. Yakhot
Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

(Received 11 February 2003; revised 2 July 2003; accepted 7 July 2003)

The process of noise generation in a flow-excited Helmholtz resonator involves strong interaction between a time-dependent fluid flow and acoustic resonance. Quantitative prediction of this effect, requiring accurate prediction of time-dependent features of a flow over complex three-dimensional bodies, turbulence modeling, compressibility and Mach number effects, is one of the major challenges to computational fluid dynamics. In this paper a numerical procedure based on the lattice kinetic equation, combined with the RNG turbulence model, is applied to describe a well-controlled experiment on acoustic resonance excitation by a grazing flow [Nelson *et al.*, J. Sound Vib. **78**, 15–27 (1981)]. The achieved agreement between numerical and physical experiments is very good. The simulations reveal a universality transformation enabling comparison of the data for different inlet conditions. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1606464]

PACS numbers: 43.28.Ra [MSH]

I. INTRODUCTION

The basic physics of sound generation by a resonator exposed to an external flow field has been understood for over a hundred years.[1] According to the early studies, the process involves a time-dependent, quasiperiodic, velocity field strongly interacting with the acoustic resonance of a cavity. The two separate phenomena of quasiperiodic vortex shedding by an incompressible (low mach number) flow past bluff bodies, and excitation of acoustic resonances in cavities, pipes, etc. can be accurately predicted numerically. When the frequency of the flow fluctuations is far from that of acoustic resonance, the two phenomena are weakly coupled and can be treated independently. However, when these two frequencies are close an intense energy exchange between the driving flow oscillations and fluid inside the cavity can occur. The phenomenon occurs in a wide range of engineering applications ranging from musical instruments to engines and combustors. The noise, produced by an open sunroof or window of a moving car or train, is a familiar and annoying manifestation of this effect.

A quantitative description of the excitation of acoustic resonance by a time-dependent flows involves knowledge of fine details which cannot be achieved without extremely accurate prediction of various complex flow features like vortices, time-dependent vortex shedding and convection, compressibility, and Mach number effects, etc. When the Reynolds number of the flow is high the description must also involve accurate turbulence models. All this makes quantitative prediction of an acoustic resonance–flow interaction one of the major challenges to computational fluid dynamics (CFD).

In this paper we are concerned with the problem of a grazing flow excitation of acoustic resonance in a three-dimensional Helmholtz resonator. In addition to the basic interest, this problem has many applications like sunroof buffeting in automobile manufacturing, design and construction of combustors and burners, where the flow and combustion-induced instabilities, coupled with acoustic resonances, often limit performance. The numerical procedure used in this paper has recently been successfully applied to quantitative prediction of the sunroof buffeting.[2,3] However, experimental data on the cavities inside a production car are influenced by various uncertainties, such as panel vibrations or leaks (through seals or into heating/air-conditioning systems), or into the trunk). It is desirable to test the methods and ideas on well-controlled experimental situations. In this paper we present the results of numerical investigation of the noise generation in the Helmholtz resonator experimentally investigated in the early 1980's by Nelson, Halliwell, and Doak (NHD).[4,5]

The problem of the noise generation by a low Mach number flow over a slot on top of the cavity was theoretically analyzed by Howe,[6,7] who considered an instability of a shear layer, modeled as an infinitesimally thin vortex sheet formed in the resonator neck. The linear stability analysis yielded the transverse displacement of the sheet leading to excitation of acoustic modes. The quantitative information was deduced by imposing a Kutta condition on the upstream lip of the slot. Howe's theory was used by Nelson *et al.* to explain their experimental data.[4,5] Despite shedding light on some important aspects of the problem, the theory was unable to predict important nonlinear features such as the observed shear layer growth, the magnitude and variation of the speed of perturbation transport across the resonator neck, amplitudes of the spectral lines of pressure fluctuations, etc. These features were successfully simulated in Ref. 2, studying a two-dimensional version of the NHD experiment. While the magnitude of the resonance frequency was correctly predicted,[2] the calculated amplitude of the correspond-

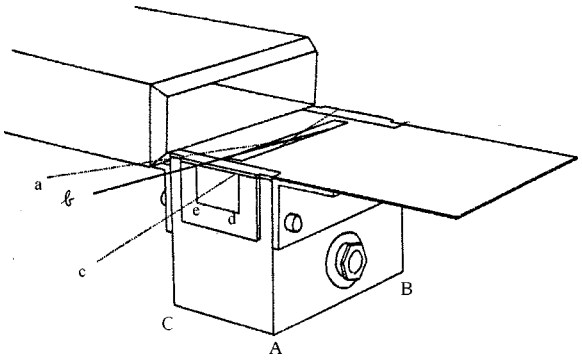


FIG. 1. NHD experimental setup. Lines a; b; c show the LDV laser beam arrangement. The depth of the cavity $AB=9.5$ cm; the width $de=5.8$ cm. The length of the resonator neck is $L=1$ cm. CA is the direction of the flow.

ing mode was somewhat lower than observed. Three-dimensional numerical simulations of grazing flow-excited Helmholtz resonator[8] are also based on a model derived in a low-order expansion in powers of Mach number (Ma). The compressibility effects were accounted for by an $O(\mathrm{Ma}^2[\partial\rho/\partial t]$ contribution to the continuity equation, where ρ is density of a fluid. For simplicity and without any justification, the same order correction to the Navier–Stokes equation was neglected. Simulations were performed at the fixed Reynolds number $\mathrm{Re}=2\times10^5$ and the Ma-dependent resonance was studied by variation of the Mach number of the flow. Since the Reynolds number (Re)-dependent structure of the shear layer in the resonator neck is of crucial importance, it is hard to estimate how well this interesting approach[8] represents experimental reality.

II. NHD's EXPERIMENT

In this section we briefly review the setup and main results of the experimental study.[4] The body of the resonator, made of the thick (1.25-cm) aluminum, is shown in Fig. 1. The coordinated axes $(x;y;z)$ were chosen in a stream-wise, span-wise, and vertical direction, respectively. The internal dimensions of the cavity were: $AB=L_x=9.5$ cm; $de=L_y=8.0$ cm; $h=L_z=5.8$ cm. The neck of the resonator was produced by cutting a slot of length $a=10$ cm and width $L=1.0$ cm across the top of the box made of $h=0.6$-mm-thick aluminum sheet. The rectangular nozzle of the height $H=3.4$ cm, placed at a distance of 3–4 cm upstream, delivered a steady grazing flow across the slot. The mean velocity in the jet 15 m$<U_\infty<$30 m was achieved at approximately $y_b=1.7$ cm.

The design of the experiment was limited by the LDV (laser Doppler velocimetry) requirements. Since the vortex shedding frequency f in the flow considered in Ref. 4 satisfied the formula

$$fL/U_\infty=\mathrm{St}, \tag{1}$$

with the Strouhal number $\mathrm{St}=1/4$ determined in a separate experiment,[9] the experiment was supposed to generate the convenient for the Ref. 4 instrumentation value of the Helmholtz frequency $f\approx600$ Hz achieved in accord with (1) at the jet speed $U_\infty\approx20$ m/s. This indeed was found experimentally: varying the jet velocity and accurately measuring the spectra of pressure fluctuations $p'(f,U_\infty)$ inside the resonator just below the top plate, the global maximum of the multivariable function $p'(f,U_\infty)$ was observed at the frequency $f\approx600$ Hz when the mean jet velocity was $U_\infty=22$ m/s.

The universal applicability of expression (1) with the velocity scale U_∞ as a relevant dynamic characteristic of the process was analyzed by Nelson *et al.*[4,5] The authors concluded that, since the flow was generated by a very wide nozzle placed not too far from the *narrow* slot, the x component of velocity vector $u(y)=U_\infty\approx22$m/s, reached at $y_b\approx1.7$ cm, far above the resonator surface, was dynamically irrelevant, and to understand the physics one needed the detailed data for the flow field within both the resonator neck and the cavity. The main results for the flow field reported in Ref. 4 are summarized as follows: (1) The flow is close to two-dimensional; (2) The relatively slow large-scale motion within the cavity $(y<0)$ does not influence the sound generation; (3) Shear layer formation, leading to vortex generation, starts at the upstream lip of the slot; (4) The Kutta condition there is well satisfied. (5) The shed vortex is fully formed at approximately the mean point of the resonator neck $(x=0,y=0)$. (6) The vortex is advected downstream with the mean velocity $V\approx6$ m/s, reaching a maximum velocity $u_{\max}\approx12$ m/s just after being ejected from the resonator neck in the vicinity of the downstream edge; (7) The shear layer thickness varies more or less linearly with the distance from the upstream lip.

As a result of the qualitative flow analysis, the relevant velocity scale, describing convection of the shear layer perturbation, was identified as $u_{\max}\approx12$ m/s and not as $U_\infty=22$ m/s. The importance of this picture, well understood by Nelson *et al.*,[4,5] can hardly be overestimated: it basically unifies the data on the noise generation in Helmholtz cavities by grazing flows. Indeed, with the relevant velocity $u_{\max}\approx12$ m/s, the relation (1) gives the Strouhal number $\mathrm{St}=0.5$, corresponding to the experimentally observed frequency $f\approx600$ Hz. This value is very close to the one obtained in the recent experiments on the sunroof buffeting.[3]

A simple model, based on Howe's theory,[6,7] gave qualitative explanations of the experimental findings.[5] Below, we report full three-dimensional simulations of the noise generated by the grazing flow over the NHD cavity.

III. NUMERICAL SETUP AND PROCEDURE

The kinetic (Boltzmann) equation is an intermediate (between molecular dynamics and hydrodynamics) method to describe statistical and kinetic systems. The classical equations of hydrodynamics are contained in this description since they can be derived in the low-order expansion in powers of the so-called Knudsen number $\mathrm{Kn}=\lambda/L\ll1$ where λ is the mean-free path in a fluid (gas) and L is a characteristic (hydrodynamic) length scale of the flow.

When the Knudsen number is small, the expansion, leading to the Navier–Stokes equations is accurate. Sometimes, in strongly strained or time-dependent flows, the hydrodynamic equations break down and one has to account for the high-order terms of the expansions in powers of Kn. The derivation of equations of the so-called nonlinear hydrody-

namics are prohibitively difficult and, to simulate the flow, one has to use molecular dynamics (MD) or kinetic theory approaches.

Continuous kinetic equations are to be solved in the seven-dimensional space (three coordinates, three velocity components, plus time), which is prohibitively expensive. This is the main reason why the method was useless for the flow calculations. Recently, the kinetic equation method was formulated for idealized particles (numbers) occupying nods of simple lattices (sets of integers). ''Collisions'' between the particles at the nearest nod were defined to satisfy all basic conservation laws, and that is why the equations of hydrodynamics, which are basically the expressions of conservation laws for the fluid elements, are contained in this simplified kinetic description. Since the ''gas of lattice particles'' obeys the ideal gas equation of state, the hydrodynamic equations contained in the lattice Boltzmann equation are those of compressible hydrodynamics. As such, the method is well suited for aero-acoustic applications. Details of the LBM are presented in various publications.[10–14] Below we describe the physical ideas behind this simple and fast numerical algorithm [lattice Boltzmann method (LBM)].

Introducing a ''single particle'' distribution function $f(\mathbf{x},\mathbf{c},t)$ in the configuration space Γ, where $\mathbf{c}$ denotes velocities of ''molecules,'' one can, in principle, write down an exact kinetic equation which involves an unspecified collision integral C

$$\partial_t f + c_i \partial_i f = C. \tag{2}$$

Knowing the distribution, we can calculate all moments of the particle velocities c_i and obtain the large-scale (averaged over many lattice sites) hydrodynamic velocity field, stresses, and all other parameters of the flow. The collision integral C can be written in a relatively simple way[15] in the case of a rarefied gas where only the binary collisions are important

$$C(f) = \int w'(f'f_1' - ff_1)d\Gamma_1\, d\Gamma' d\Gamma_1' = \int u_{\mathrm{rel}}(f'f_1' - ff_1)d\sigma\, d^3p, \tag{3}$$

where $\partial_t = \partial/\partial t$; $\partial_x = \partial/\partial x_i$, and collision integral $C(f)du\, d\Gamma$ is rate of change of the number of molecules in the phase volume $d\Gamma = d^3x\, d^3u$, u_{rel} is the relative velocity of colliding molecules, and $d\sigma$ and $\mathbf{p}$ are the differential scattering cross section and momentum of the molecule, respectively. The state variable Γ describes all degrees of freedom of a molecule and w' stands for the probability of a transition of two molecules initially in a state Γ and Γ_1 to a state Γ_1', Γ' as a result of collision. The kinetic equation (2) with collision integral (3) is the celebrated Boltzmann equation. The local equilibrium distribution function is determined from the relation $C(f^{\mathrm{eq}}) = 0$. If the deviations from thermodynamic equilibrium are small, we can write $f = f^{\mathrm{eq}} + \delta f$ and

$$C = -\frac{f}{\tau_0} + \int u_{\mathrm{rel}} f' f'\, d\sigma\, d^3p \approx -\frac{f - f^{\mathrm{eq}}}{\tau_0}, \tag{4}$$

with $\tau_0 \approx \int u_{\mathrm{rel}} f^{\mathrm{eq}} d\sigma\, d^3p$. This equation contains the hydrodynamic description of a simple fluid. The equation (4) for the single-particle distribution function, often called the ''BGK'' (mean-field) approximation,[16] reflects an important principle: a weakly perturbed gas tends to relax to thermodynamic equilibrium. If, in accord with Bogolubov,[17] we accept that the return-to-local-equilibrium assumption is plausible even when the deviations from equilibrium δf are not small, we can, to develop a model for strongly nonlinear (turbulent) flow, use (4) with *the relaxation time reflecting the relevant complex physics of a nonlinear (turbulent) fluid far from equilibrium*.

To begin, we have to choose the parameter characterizing nonequilibrium properties of the fluid. The most natural parameter having the dimensionality of inverse time is the rate of strain $S_{ij} = \frac{1}{2}(\partial_i u_j + \partial_j u_i)$.

To come up with the modification of the relaxation time in Eq. (4), let us write the most general definition, the second-rank tensor leading to expression for viscosity coefficient

$$\tau_{ij} = \varphi(|S_{ij}|)S_{ij} + \sum \psi_n(|S_{ij}|)S_{ia_1}S_{a_1a_2}\cdots S_{a_nj}, \tag{5}$$

where the first term in the right side corresponds to the ''isotropic viscosity'' while the sum in (5) accounts for the anisotropic effects. Our goal is to write a kinetic equation (2)–(4) with the relaxation time leading in the lowest order of the Chapman–Enskog expansion[15] to the (turbulent) viscosity defined by Eq. (5). Below, to simplify the theory we will neglect the anisotropic contributions to ''viscosity'' in (5).

In the low-order expansion in powers of dimensionless strain $\tau_0 S_{ij}$, the correct effective equation we are attempting to derive should generate the Navier–Stokes equation with the generalized viscosity given by (5). In the hydrodynamic approximation, viscosity includes the information about two parameters: speed of sound and the mean-free path or relaxation time. Since our kinetic theory is based on a single governing parameter, the relaxation time, the simplest Galilean invariant model for the collision integral is

$$C = -\frac{f - f^{\mathrm{eq}}}{\tau}, \tag{6}$$

where

$$\tau \approx \frac{K}{\varepsilon} \approx \frac{1}{2}\frac{\overline{u^2}}{\overline{\nu(\partial u_i/\partial x_j)^2}}, \tag{7}$$

with

$$\partial_i u_j = \partial_i \int c_j f(\mathbf{c},\mathbf{x})d\Gamma. \tag{8}$$

In this work we use the differential equation for turbulent kinetic energy K and dissipation rate ε (RNG model).[18–20] It has been shown[12] that in the low order of expansion in powers of the dimensionless shear, a set of equations similar to (2), (6)–(8), generate the Reynolds stress (5) with

$$\varphi(|S_{ij}|) \approx -\nu_T \propto \frac{K^2}{\varepsilon}\Psi(|S_{ij}|, \eta, K, \varepsilon), \tag{9}$$

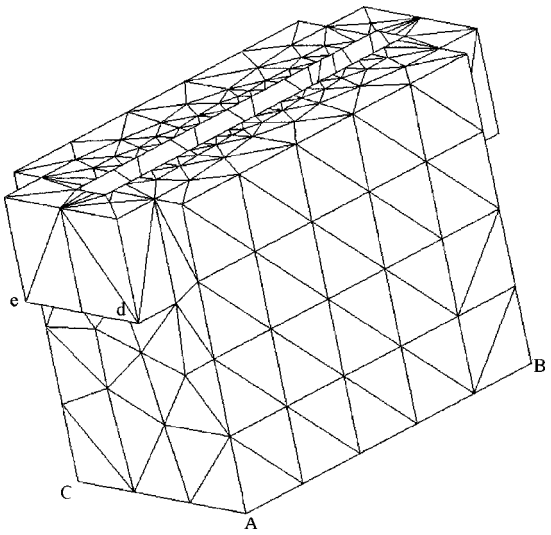


FIG. 2. Resonator geometry and surface mesh used in simulations. Labels are those of Fig. 1.

with the dimensionless scaling function $\Psi(|S_{ij}|,\eta,K,\varepsilon)$ accounting for strong shear, vorticity, and helicity. *In the case of very strong shear, the high-order contributions to the hydrodynamic (turbulence) model, which are impossible to derive in terms of an expansion in powers of a large parameter, are contained in a simplified kinetic equation.*

The above ideas have been realized in the commercially available numerical code POWERFLOW™, which has been used for the simulations described below. This fast and accurate enables one to generate automatically computational meshes over bodies of complex geometry. The geometry of the numerical experiment and computational mesh in the resonator neck are shown in Figs. 2 and 3. High-resolution requirements in the vicinity of the resonator neck are obtained by using nested regions of increased resolution. Resolution tests have shown that it is necessary to use at least 50 volume cells across the length of the opening. The highest resolution needs to be applied some distance upstream of the opening in order to capture the growth of the boundary layer,

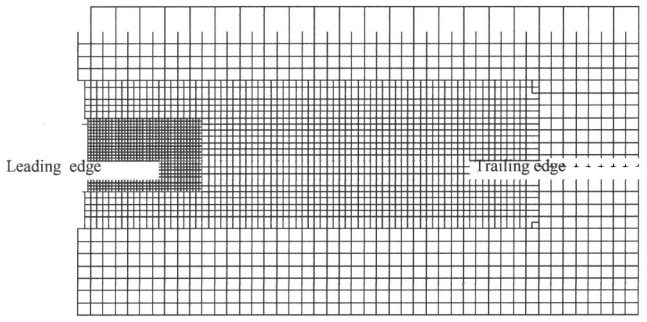


FIG. 3. Volumetric mesh in the neck of the resonator. The resolution is highest at the leading edge.

as well as at the upstream side of the opening. However, it is not necessary to have high resolution along the length of the slot. The total number of the mesh points was 3.8×10^6.

IV. MEAN FLOW FEATURES

The inlet flow velocity profile $u(y)=U=\text{const}$ was prescribed at the distance $l=0.8$ cm from the upstream lip of the resonator slot. This distance was sufficient for the development of a small boundary layer prior to the formation of the shear layer within the resonator neck. The principal features of the shear layer for $U=15$ m/s are shown in Fig. 4, where the velocity profiles $u(y)$ are presented for three stations ($x=-0.049$ cm, 0), (0,0), and (0.49,0), corresponding to the upstream lip, midpoint, and the downstream lip of the resonator neck, respectively. We can see that, as in the experimental study,[4] the large-scale flow inside the cavity ($y<0.06$ cm) where velocity $u(y)\leqslant 2$ m/s, is weak. The shear layer widths at these three stations are: $\delta_1\approx0.1$ cm, $\delta_2\approx0.2$ cm, and $\delta_3\approx0.3$ cm, in very good agreement with the experimentally found close-to-linear boundary layer thickness as a function of the distance from the upstream lip.[4]

The mean flow stream lines are shown in Fig. 5. The instantaneous stream lines, not shown here, indicate formation of a strong vortex, convected downstream toward the edge, where it is ejected from the resonator neck. In fact, the

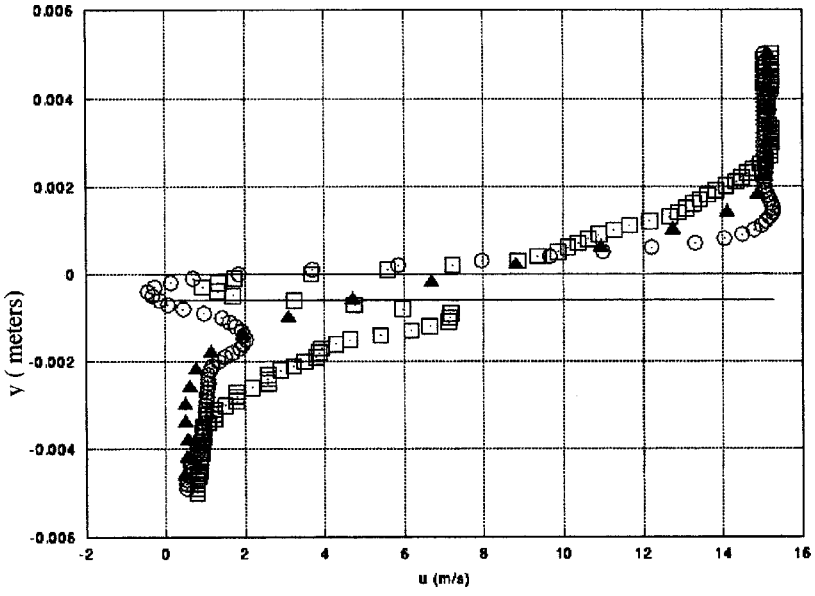


FIG. 4. Velocity profiles, $u(y,x)$, in resonator neck at three stations: $x_1=-0.49$ cm (○); $x_2=0.0$ cm (▲); $x_3=0.49$ cm (□). One can see increase of the width of the shear layer with the distance from the leading edge at $x=-0.5$ cm. For the details see the text.

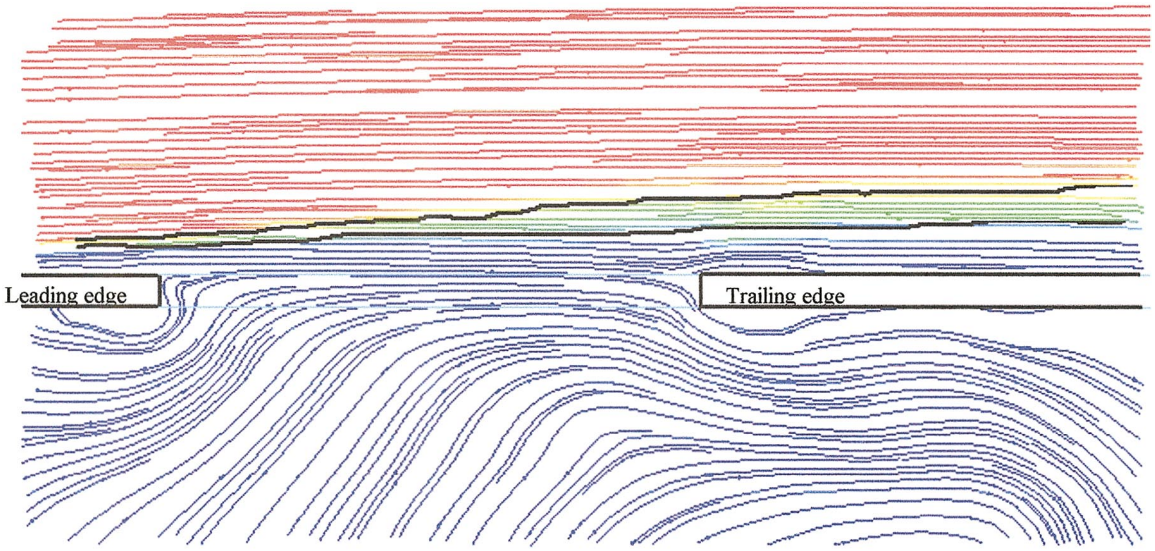


FIG. 5. Streamlines of long time-averaged flow field at the neck of resonator. The region in which the x velocity is bounded by 12 ± 0.5 m/s is highlighted. One observes separation bubbles near leading and trailing edges.

instantaneous picture was much more complex, showing formation of additional vortexes somewhat weaker than the main one. These structures are responsible for the high-frequency component of the pressure spectrum fluctuations, which were experimentally observed and successfully simulated in the present work (see below). The velocity at the center and slightly above the strongest vortex $u \approx 12$ m/s can be interpreted as a convection velocity, very close to that experimentally observed.[4]

V. SPECTRA OF PRESSURE FLUCTUATIONS

As in the experiment,[4] the data on the pressure fluctuations were taken just below the top plate of the resonator, left of the resonator neck. A typical time history of the calculated pressure signal evolution is presented in Fig. 6. The fast Fourier transform (FFT) applied to the signal enables one to compute the pressure fluctuation spectra for a given inlet velocity U_∞

$$\mathrm{SPL} \equiv p(f, U_\infty) = 20 \log\left(\sqrt{\frac{1}{N} \sum_{i=1}^{N} p_i'(f, U_\infty) p_i'(-f, U_\infty)} \right). \quad (10)$$

Here, SPL stands for the sound-pressure level. In what follows the scalar denoted by the letter p stands for pressure, which is not to be mistaken for the momentum vector described in Sec. III. To obtain a well-converged mean, the time series was subdivided into realizations, each one containing 2000 time steps. The FFT was applied to each realization generating the sets $p_i(f, U_\infty)$ and, to achieve good convergence, a large number of realizations $N \approx 10^3$ was used. The spectrum $p(f,15)$, corresponding to the inlet velocity $U_\infty = 15$ m/s or $u_{\max} \approx 12$ m/s, is shown in Fig. 7 together with the experimentally observed[4] pressure spectrum for $U_\infty = 22$ m/s ($u_{\max} \approx 12$ m/s). As we see, our numerical simulations show a well-pronounced main peak of intensity $p_{\max}(570,15) \approx 130$ dB at the frequency $f(15) \approx 570 \pm 4$ Hz.

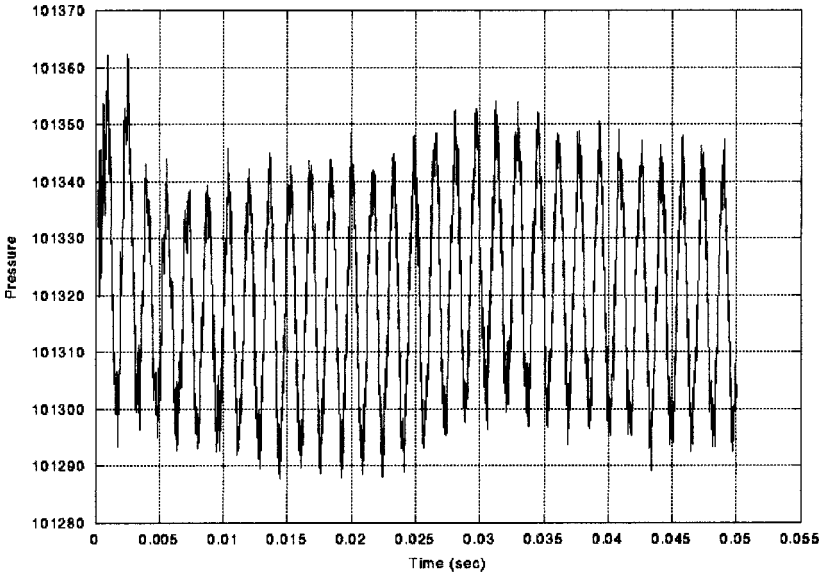


FIG. 6. Calculated typical time history of pressure signal at a point within the cavity for $U_\infty = 18$ m/s.

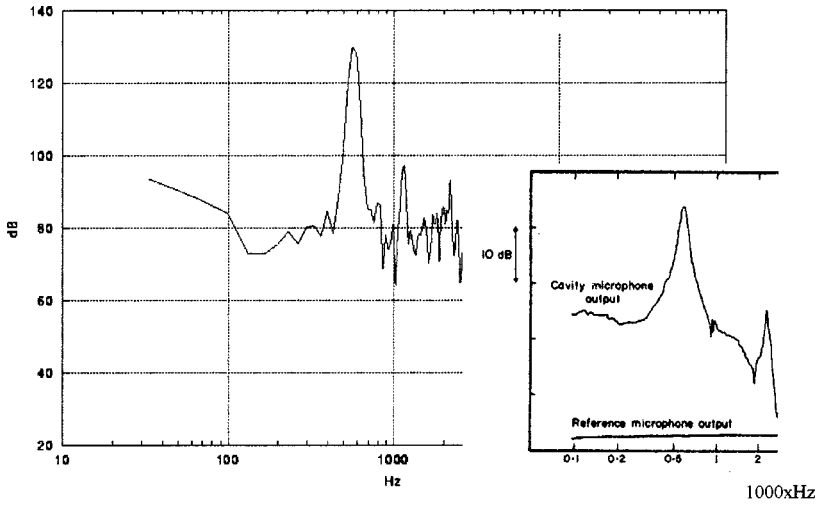


FIG. 7. Computed spectrum of pressure fluctuations. Insert: experimentally observed spectrum (Ref. 4) at $U_\infty = 22$ m/s.

This spike in the computed spectrum is to be compared with the experimentally observed peak of intensity $p_{\max}(603,22) \approx 134 \pm 3$ dB.[4] It is interesting that not only the experimental and numerical peak frequencies and intensities compare favorably, but the line shapes and the spectral features in the vicinities of the main peak are very similar. The important peak at $f \approx 1100$ Hz corresponds to the flow pattern involving more than one (maybe two) vortexes generated within the resonator opening.

Repeating the simulation for different magnitudes of free-stream velocity ($12 < U_\infty < 20$ m/s), we found that each spectrum had a dominant spectral line. By analyzing the free-stream velocity dependence of the dominant peak, we found that maximum intensity $p_{\max}(U_\infty)$ was attained at $U_\infty \approx 15$ m/s, which corresponds to the Helmholtz resonance frequency $f(15) \approx 570 \pm 4$ Hz. The amplitude and frequency of the calculated pressure spectrum peak as a function of the free-stream velocity is plotted in Fig. 8 (lower curve).

VI. UNIVERSALITY TRANSFORMATION

One feature of Fig. 8 deserves attention. According to general principles of acoustics, the resonance frequency is a unique property of a resonator geometry, while grazing flow is one of the many ways to excite the resonant acoustic mode. Due to substantial differences in the inlet velocity

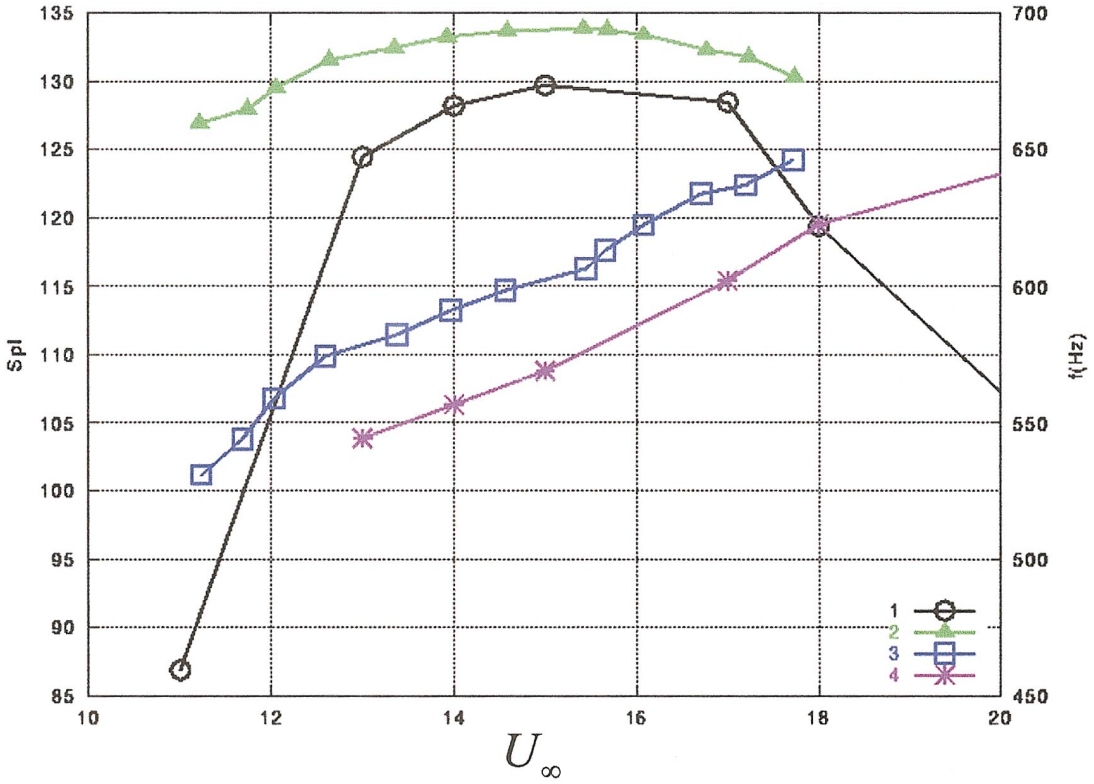


FIG. 8. Evidence for universality of convection velocity relative to mean velocity. The maximum observed in experiment of Ref. 4 is at $U_\infty = 22$ m/s. The universality transformation makes experimental and calculated curves very close. Peak intensity: (1) Present work (○); (2) experiment (▲). Resonant frequency: (3) Present work (□); (4) experiment (*). Error bar of experimental data ±3 dB not shown on graphs.

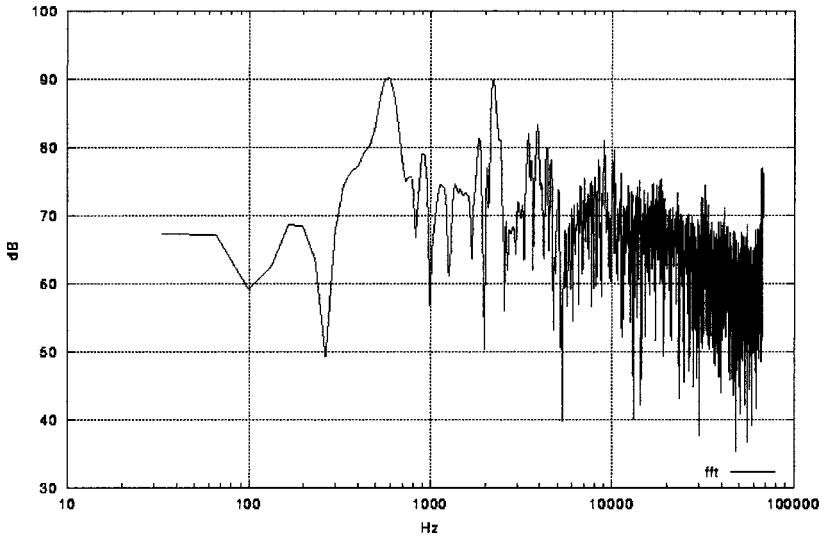


FIG. 9. No-flow case. Calculated power spectrum for the pressure perturbed NHD cavity. The initial high-pressure region is generated at the center of the cavity.

profile, the same resonance frequency $f \approx 600$ Hz corresponds to experimental and numerical values of the inlet velocity $U_\infty \approx 22$ m/s and 15 m/s. Detailed comparison of experimental data with numerical results revealed substantial difference in the properties of the boundary layer (velocity profile, width) in the vicinity of the upstream lip. The experimental boundary layer[4] was very thick, with velocity $u(y) \approx U_\infty = 22$ m/s reached at $y \approx 1.7$ cm. Thus, the convection velocity $u_{\max} \approx 12$ m/s of the experimentally observed shed vortex across the narrow ($L = 1$ cm) resonator neck, was unaffected by the large value of the mean jet velocity U_∞, but rather corresponded to a value $u(y)$ somewhere deep within the boundary (shear) layer. On the other hand, because of the flat inlet velocity profile imposed 0.8 cm upstream of the edge of the slit, the boundary layer in numerical simulation was narrower and, as a result, the same convection velocity $u_{\max} \approx 12$ m/s was much closer to the corresponding free-stream velocity $U_\infty = 15$ m/s.

To understand the role of the boundary layer in noise generation, we conducted a numerical simulation with the same uniform inlet velocity profile $u(y) = 15$ m/s, imposed much farther from the resonator neck: $L = 5.3$ cm instead of $L = 0.8$ cm tested above. In this case we expect a much thicker boundary layer. This was indeed observed: the boundary layer was 2–3 times wider than before. The most interesting result of this test was the excitation of the low-intensity off-resonance (hydrodynamic) peak at the frequency $f \approx 480$ Hz, at a convection velocity $u_{\max} \approx 9.6$ m/s. In this case, to excite the resonance at the frequency $f \approx 570-600$ Hz, we had to substantially increase the inlet velocity. Below, we present a simple similarity transformation, unifying experimental data corresponding to different boundary layers at the entrance to the resonator neck.

It is clear from the physics of the noise generation that in the vicinity of the maximum (resonance), both the shape of the curve $p(f_r, U_\infty)$ and the spectrum are entirely defined by the function $u_{\max}(U_\infty)$, where $u_{\max}$ is a convection velocity of the shear layer perturbation (vortex) as a function of free-stream velocity. Assuming that all velocities in the problem scale with U_∞, we conclude that two different experimental curves $p_{\max}(U_\infty(1))$ and $p_{\max}(U_\infty(2))$ should collapse as a result of the mapping $p_{\max}(U_\infty(2)) = p_{\max}(U_\infty(1) \times [U_{\max}(2)/U_{\max}(1)])$, where $U_{\max}(i)$ is an inlet velocity (U_∞) corresponding to the resonance frequency measured in the ith experiment. For the case considered in this paper, the results of numerical simulations (ns) are mapped onto the NHD experimental data by

$$p_{\max}(U_\infty(ns)) = p_{\max}\left(U_\infty(NHD) \frac{U_{\max}(ns)}{U_{\max}(NHD)} \right)$$

$$\approx p_{\max}(U_\infty(NHD)\, \tfrac{15}{22}). \qquad (11)$$

The result of the mapping is presented in Fig. 8. Without the coordinate transformation the maximum of the experimentally observed curve is shifted to $U_\infty = 22$ m/s.

It is clear that, apart from slight differences in the peak intensities [the error bars of the NHD result $p_{\max}(603,22) \approx 134 \pm 3$ dB are not shown in Fig. (8)], the similarity of the curves in the vicinity of their maxima is very good.

VII. THE RESONANT FREQUENCY REVISITED

To test the consistency of the above results, the resonance frequency of the NHD cavity was also determined numerically by introducing a perturbation to equilibrium pressure field. A very small high-pressure region was created at time $t = 0$ at the center of the cavity. The time evolution of this pressure perturbation showed effective generation of acoustic excitations in the cavity. The FFT of the resulting time-dependent pressure field yielded the spectrum of pressure fluctuations shown in Fig. 9. The computed frequency of the dominant mode is $f = 577 \pm 2$ Hz, which is to be compared to $f = 570 \pm 4$ Hz, obtained above for a grazing flow excitation of the acoustic resonance. Since this simulation is very simple and quite reliable, the calculated resonant frequency serves as a consistency test for the much more involved numerical simulations presented in this paper.

VIII. DISCUSSION AND CONCLUSIONS

The modified lattice kinetic (Boltzmann) equation was applied to the computation of the pressure fluctuations spectrum in a grazing flow-excited Helmholtz resonator studied

experimentally. The calculated spectral line of acoustic resonance $p_{\max}(570,15) \approx 130$ dB is to be compared with the measured one: $p_{\max}(603,22) \approx 134 \pm 3$ dB. The computed convection velocity of the shear layer perturbation $u_{\max} \approx 12$ m/s was extremely close to that experimentally observed.[4] This is the dynamically relevant velocity, and we have proposed a universality transformation (11) that collapses the data obtained from different experimental (physical and/or numerical) setups. The deviation of numerical results from experimental data for both the intensity and the peak frequency is not greater than a few percent. Moreover, the computed shape of the spectral line is very close to the observed one.

Accurate two-dimensional ($L_x = 9.5$ cm; $L_z = 5.8$ cm) lattice Boltzmann simulations of the NHD cavity using the same numerical procedure as the one described in this paper were conducted.[2] These 2D simulations produced the resonance frequency $f \approx 570$ Hz, very close to ours, but the peak intensity was lower by some 10 dB than both experimentally observed[4] and computed in this paper. The experimental data clearly demonstrated the two-dimensionality of the mean flow in the resonator neck[4] and the fact that the 2D and 3D simulations of the same system, based on the same method, gave such different peak intensities, indicate the possibility that three-dimensional perturbations of the two-dimensional mean shear layer flow contribute a lot to the overall peak intensity. Another, in our opinion less probable, scenario may be related to the differences between intensities of the 2D and 3D sound radiation from the cavity.[21]

Often, disagreements between experiments and simulations are attributed to poor resolution, accuracy of numerical method, simplified geometry, uncertainties at the inlet boundary conditions, turbulence modeling (favorite culprit), etc. In our work the geometry of the simulations was extremely close to experimental. Numerical resolution used in our was not too high and the RNG turbulence model ($K-\varepsilon$ equations) was the one widely used in engineering modeling. Thus, the natural question arises: why the simulations of a complex system, presented in this paper, gave such accurate results. It is too early to attribute the success of the above simulations to superiority of the lattice Boltzmann equations over the hydrodynamic approximations for description of strongly time-dependent and sheared flows. This possibility cannot be ruled out though, since the equations of hydrodynamics (known turbulence models) are contained in the kinetic equations (2),(6)–(8) as the low-order result of the Chapman–Enskog expansion in powers of dimensionless rate of strain. If the high-order nonlinear corrections to the model are important, then the kinetic equations (2), (6)–(8), which contain all of them, can indeed be more accurate. To shed light on the issue, it would be useful to conduct numerical simulations of the well-controlled NHD experiment using more conventional methods of turbulence modeling (both RANS and LES).

ACKNOWLEDGMENTS

We are indebted to D. Ricot for his help in setting up the computational domain and for helpful discussions. The correspondence with P.A. Nelson, related to interpretation of some of the experimental features, was really illuminating. Very interesting conversations with H. Chen, M.S. Howe, R. Zhang, S. Remondi, and I. Staroselsky are gratefully acknowledged.

[1] Excellent review of the history of the subject can be found in: J. W. S Rayleigh, *The Theory of Sound* (Dover, New York, 1945).
[2] D. Ricot, V. Maillard, and C. Bailly, ''Numerical simulation of the unsteady flow past a cavity and application to the sun-roof buffeting,'' AIAA J. ■, 2001–2112 (2001).
[3] D. Ricot, V. Maillard, and C. Bailly, private communication.
[4] P.A. Nelson, N.A. Halliway, and P.E. Doak, ''Fluid dynamics of flow excited resonance. I. Experiment,'' J. Sound Vib. **78**, 15–27 (1981).
[5] P.A. Nelson, N.A. Halliway, and P.E. Doak, ''Fluid dynamics of flow excited resonance. II. Flow excited resonance,'' J. Sound Vib. **91**, 375–389 (1983).
[6] M.S. Howe, Philos. Trans. R. Soc. London, Ser. A **303**, 151–164 (1981).
[7] M.S. Howe, J. Sound Vib. **70**, 407–418 (1980).
[8] M. Inagaki, O. Murata, T. Kondoh, and K. Abe, ''A flow excitation of acoustic resonance,'' AIAA J. **40(9)**, 1823–1830 (2002).
[9] J.S. Bolton, ''The excitation of an acoustic resonance by pipe flow,'' MSc dissertation, Institute of Sound and Vibrations, University of Southhampton, 1976.
[10] S. Chen and G. Doolen, ''Lattice Boltzmann method for fluid flows,'' Annu. Rev. Fluid Mech. **30**, 329–364 (1998).
[11] S. Hou, Q. Zou, S. Chen, G. Doolen, and A. Cogley, J. Comput. Phys. **47**, 118–126 (1995).
[12] C. Teixeira, ''Incorporating turbulence models into the lattice Boltzmann method,'' Int. J. Mod. Phys. C **9**, 1159–1171 (1998).
[13] N. Martys and H. Chen, ''Simulations of multi-component fluids in complex three-dimensional geometries by the lattice Boltzmann method,'' Phys. Rev. E **53**, 743–750 (1996).
[14] H. Chen, C. Teixeira, and K. Molvig, ''Realization of fluid boundary conditions via discrete Boltzmann dynamics,'' Int. J. Mod. Phys. C **9**, 1281–1288 (1998).
[15] L.D. Landau and E.M. Lifshitz, *Physical Kinetics* (Butterworths/Heinemann, London, 1995).
[16] P. Bhatnagar, E. Gross, and M. Krook, Phys. Rev. **94**, 511–528 (1954).
[17] N.N. Bogolubov, ''Problemy dinamicheskoii teorii v statisticheskoi phisike,'' (in Russian), Moscow, 1946.
[18] V. Yakhot and S.A. Orszag, ''Renormalization group analysis of turbulence,'' J. Sci. Comput. **1**, 3–51 (1986).
[19] V. Yakhot and L.M. Smith, ''The renormalization group, the ε-expansion and derivation of turbulence models,'' J. Sci. Comput. **3**, 35–47 (1992).
[20] V. Yakhot, S.A. Orszag, S. Thangam, T.B. Gatski, and C.G. Speciale, ''Development of turbulence models for shear flows using double expansion technique,'' Phys. Fluids A **4**, 1510–1520 (1992).
[21] We are grateful to M.S. Howe for bringing our attention to this possibility.

Sound absorption by suspensions of nonspherical particles: Measurements compared with predictions using various particle sizing techniques

Simon D. Richards[a)]
Marine & Acoustics Centre, QinetiQ, Winfrith Technology Centre, Dorchester Dorset, DT2 8XJ, United Kingdom

Timothy G. Leighton and Niven R. Brown[b)]
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

(Received 27 January 2003; revised 15 July 2003; accepted 28 July 2003)

Knowledge of the particle size distribution is required in order to predict ultrasonic absorption in polydisperse particulate suspensions. This paper shows that the method used to measure the particle size distribution can lead to important differences in the predicted absorption. A reverberation technique developed for measuring ultrasonic absorption by suspended particles is used to measure the absorption in suspensions of nonspherical particles. Two types of particulates are studied: (*i*) kaolin (china clay) particles which are platelike in form; and (*ii*) calcium carbonate particles which are more granular. Results are compared to theoretical predictions of visco-inertial absorption by suspensions of spherical particles. The particle size distributions, which are required for these predictions, are measured by laser diffraction, gravitational sedimentation and centrifugal sedimentation, all of which assume spherical particles. For a given sample, each sizing technique yields a different size distribution, leading to differences in the predicted absorption. The particle size distributions obtained by gravitational and centrifugal sedimentation are reinterpreted to yield a representative size distribution of oblate spheroids, and predictions for absorption by these spheroids are compared with the measurements. Good agreement between theory and measurement for the flat kaolin particles is obtained, demonstrating that these particles can be adequately represented by oblate spheroids. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1610449]

PACS numbers: 43.30.Es, 43.35.Bf [DKW]

I. INTRODUCTION

The propagation of high frequency (tens to hundreds of kHz) sound in shallow, littoral seas is complicated by many phenomena, including variable bathymetry and sound speed profiles, seabed and sea surface interactions, tides, currents, turbulence, ambient noise and the effects of bubbles and suspended mineral particles throughout the water column.

The last of these, the effect of suspended mineral particles, has been shown[1,2] to be important for high frequency sonar performance in turbid environments. In order to account for this effect in sonar performance predictions it is necessary to calculate the absorption coefficient in particulate suspensions. A reverberation time technique has been developed for the laboratory measurement of absorption in dilute suspensions and measurements made with spherical glass particles have been shown[3] to be in good agreement with predictions which assume the particles to be homogenous spheres.

Natural particles are not generally spherical, and there have been a number of studies of attenuation by irregular particles. A coupled-phase theory[4] has been modified[5] to include the effects of particle shape using an effective radius approach. Another approach[6] treats the suspension as a fractal medium, with the Reynolds number as the fractal dimension, and this has been shown[7] to give good agreement with measured data for airborne suspensions. This approach is of limited use since it requires that the fractal dimension be determined by fitting the model to observations at a given frequency. In this paper Urick's[8] formula for visco-inertial absorption by spherical particles has been used, along with an extension[9] to that model which accounts for spheroidal particles.

In order to characterize absorption by natural marine particles measurements in representative suspensions are required. This paper addresses that issue by comparing the measured absorption in suspensions of nonspherical particles with predictions for both spheres and spheroids. Particular attention is paid to the interpretation of the measurements of particle size distribution, which are required as inputs to the theoretical predictions.

II. EXPERIMENT

A. Protocol

The experimental method used in this study employs measurements of reverberation time to infer the acoustic ab-

[a)]Electronic mail: sdrichards@qinetiq.com
[b)]Present address: Industrial Research Limited, PO Box 2225, Auckland 1015, New Zealand.

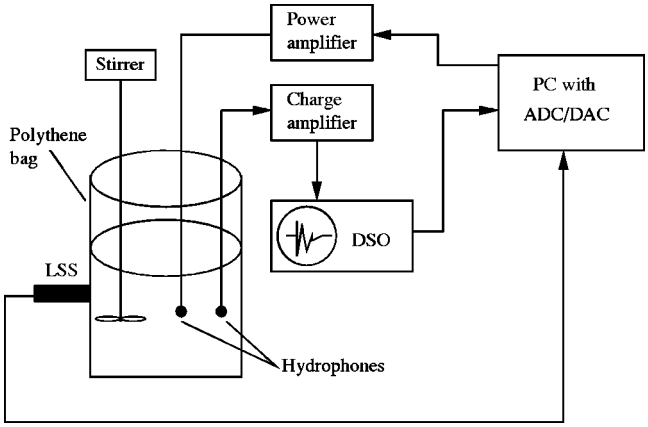


FIG. 1. A schematic of the experimental apparatus used to measure absorption in dilute suspensions over the frequency range 50 to 150 kHz. The stirrer is removed during acoustic measurements.

sorption in the system. The use of reverberation time measurements has been credited[10] to the work of Meyer and Skudrzyk. Relative measurements of the reverberation decay rate between clear water and water containing suspended particles are used to determine the absorption due to the particles. Initial experiments showed the importance of reducing boundary losses in the system.[3,11] This resulted in the development of the current system in which 0.016 m^3 of water is contained within a thin-walled plastic membrane, suspended in air by means of fine wires. This configuration approximates well to a pressure-release boundary condition surrounding the volume, thus minimizing the losses at the boundary. It may be noted that, even with the system optimized for minimum boundary losses in this way, the absorption due to a dilute suspension of particles is a small contribution to the total attenuation in the system. For example, a small increase (O(mm)) in the depth of the hydrophones was observed to produce a significant change in the reverberation time of the system. Experiments conducted with a third hydrophone[11] demonstrated that this was caused by additional attenuation due to the increased length of submerged hydrophone cable, thus illustrating the challenging nature of the measurements. The apparatus is shown schematically in Fig. 1, and Fig. 2 shows a photograph of the apparatus.

The signal generation, data acquisition and signal processing were all performed under the LabVIEW laboratory instrument management system, running on a personal computer (PC). The signal from the PC's digital-to-analog converter (DAC) board was fed to a power amplifier driving a Brüel and Kjær (B&K) 8103 hydrophone. A second B&K 8103 hydrophone was used for the receiving transducer and the signal from this was amplified and fed to a LeCroy digital storage oscilloscope (DSO). The DSO was connected to the PC via a GPIB interface, and the data were transferred to the computer for storage and post processing. A mechanical stirring device (propeller) was used to resuspend particles that had settled out of suspension. This was removed from the water when the acoustic measurements were performed, as additional absorbing/reflecting surfaces complicate the acoustic system. The temporal variation in the concentration of the suspension was monitored using a Sea Tech light scattering sensor (LSS). The LSS was mounted horizontally on the outside of the bag, with its window in close contact with the side of the bag. The mating faces of the LSS window and the side of the bag were wetted to ensure good optical coupling. This configuration ensured that any spatial integrating inherent in the LSS measurements would be in the horizontal direction, not in the vertical where the concentration would be expected to be spatially varying after a finite period of settling.

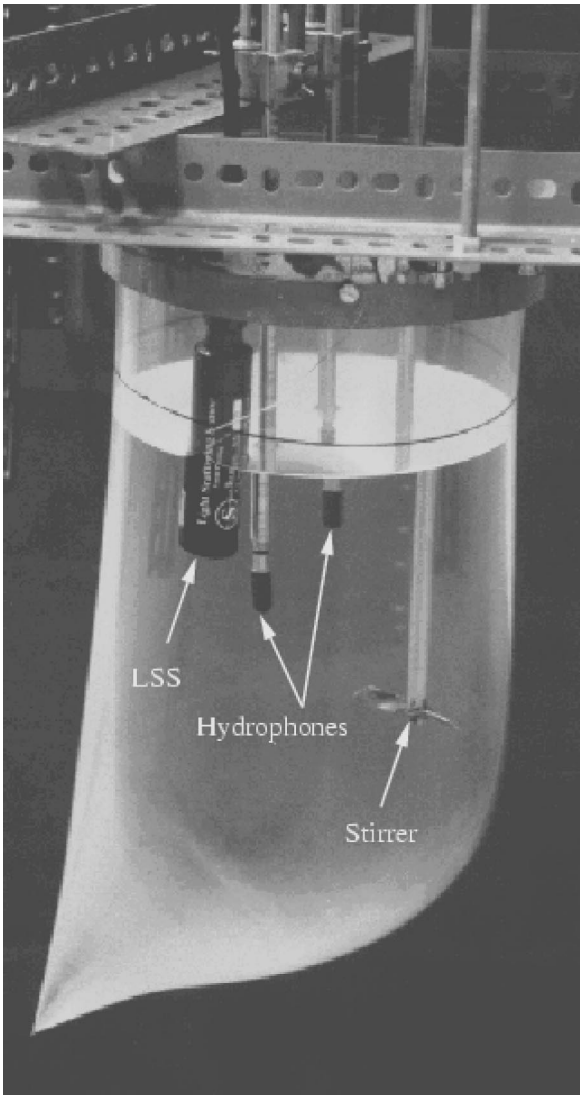


FIG. 2. Photograph of the experimental apparatus. The bag containing the test fluid is surrounded by air, resulting in a boundary condition which is close to pressure-release. This photograph is illustrative only, as the LSS and stirrer were removed during acoustic measurements. The majority of experiments were carried out with the LSS mounted horizontally on the outside of the bag (see text). Other instrumentation included temperature and dissolved oxygen probes (not shown). The hydrophones are mounted in rigid tubes to prevent movement.

The photograph (Fig. 2) shows the polythene bag suspended from its mechanical support by means of fine wires attached to a metal hoop. The items dipping into the water are, from left to right, the LSS used to monitor the concentration of the suspended particles (shown here mounted vertically, rather than in the more usual horizontal configuration as described above), the transmitting and receiving hydro-

phones and the propeller used to stir the water in order to suspend the particles. The diameter of the bag at the water surface is about 235 mm.

Prior to acoustic measurements the water was passed through a reverse osmosis system and then filtered to remove particulate matter. This ensured that no particles of diameter greater than 0.22 μm remained in the water. The water was then degassed under partial vacuum in order to avoid the presence of bubbles which may be acoustically significant. The temperature of the water was measured using a Jenway 3071 *p*H and temperature meter and this measurement was used to calculate the speed of sound in the ambient water. The level of dissolved oxygen was monitored throughout the experiments (between acoustic runs) using a Jenway 9010 dissolved oxygen probe attached to the 3071 meter. This was to ensure that the addition of the particles or the stirring process did not cause significant gas entrainment. The dissolved oxygen content was found to vary from 51% to 65% (expressed as a percentage of the saturation level) over the course of a measurement period, and no bubbles were observed.

In each series of experiments, reverberation time measurements were first made in particulate-free water in order to provide a reference measurement. Particles were then added in stages, in known quantities by mass, to enable a series of measurements to be made at varying concentrations. The concentration was taken to be the spatial average over the volume, i.e., the mass of particles added divided by the volume of water.

Before each series of acoustic measurements, the suspension was stirred by the mechanical stirrer until the particles were evenly distributed in the suspension. The time taken for this to occur was shown, by measurements made with the LSS, to be of the order of a few seconds. Care was taken to ensure that the particles that collect in the corners formed by the bottom seam of the bag were resuspended by the stirring. As the stirring takes place just before each set of acoustic measurements, the clear water reference measurements were also made on stirred water. This ensured that the reference signal used was obtained under conditions which most closely represented those experienced during the measurements made with the particles in suspension. It may be noted that stirring induces turbulence which can, in principle, lead to absorption. The mechanism for this absorption is the interaction between two phenomena: the perturbation of the turblence field by the acoustic wave, leading to anisotropic Reynolds stresses; and the redistribution of turbulent kinetic energy as it cascades from the large scale to the dissipation scale. As a result of these two effects the Reynolds stress is not in phase with the acoustic field and there is a net transfer of energy from the acoustic field to turbulent kinetic energy. This absorption mechanism has been shown[12] to be negligible for the current measurements.

The acoustic measurement itself involved first generating the sound field, then switching off the sound source and recording the decay of the reverberant sound field.

Two techniques for generating the sound field were investigated; an impulse and a long burst (20 ms) of uniform white noise. Both of these methods produce a broadband sound field. The advantage of using the long burst technique is that it gives sufficient time for a steady-state sound field to build up before the source is switched off, and the decaying sound field is also less prone to large perturbations resulting from direct reflections and particular modes of the volume. For these reasons the long burst generation method was used for the results presented in this paper.

Typically a series of measurements consisted of emitting ten noise bursts from the transmitting transducer and recording their responses via the receiver hydrophone, with the entire sequence lasting approximately 35 s. During this period there was no observable decrease in the concentration according to the estimates obtained using the LSS.

The decaying, reverberant sound field measured by the receiving hydrophone demonstrates fluctuations resulting from the interference between the many modes within the reverberating volume. These fluctuations limit the accuracy with which the decay rate of the reverberant field may be determined from the measured decay curve. In order to improve the accuracy of the estimate of the reverberation time of a reverberating volume, it is usual practice to repeat the measurement many times and average the decay rates obtained from the individual measurements. This method is, however, inefficient owing to the large number of measurements which must be made in order to obtain an accurate estimate of the decay rate or reverberation time. This averaging also obscures any ping-to-ping variations in the decay rate and any temporal variation within the integration time, which can be long due to the requirement for a large number of measurements.

The decay rates were therefore determined by applying the method of integrated impulse response[13] (IIR) to the sound field from the time that the driving signal was switched off. This method was used, even for signals derived from the noise burst, as it gives a smooth decay curve.

The raw data were filtered using second-order Butterworth bandpass filters in 10-kHz bands over the frequency range 50 to 150 kHz, and the IIR analysis was performed in each band. A second-order Butterworth filter has a roll-off of 12 dB per octave outside of the pass band, and it is therefore recognized that there will be some out of band data in each of the 10-kHz bands. The data were also reduced into time bins representing the rms value of the signal for a user-defined number of samples, typically 100. The sampling rate of the DSO was 500 kHz.

Errors in the determination of the decay rate in such an experiment can arise as a result of spatial variations in the sound field. There are two approaches to mitigating against this problem: (*i*) the measurements should be made at many points and the spatial average determined; or (*ii*) the measurements must be made under diffuse field conditions, in which the average energy density is the same throughout the entire volume and all directions of propagation are equally probable. The diffuse field approach was adopted for this study, and this is discussed later.

It will be noted that while these experiments are broadband, the theory referred to Sec. III is obtained for a single insonifying frequency. However, under the assumption that the individual frequency components are linearly indepen-

dent, the single-frequency models may be applied to the interpretation of the broadband measurements in each narrow band. The results of frequency-binned measurements made with spherical particles were found to be in very good agreement with single-frequency predictions,[3] thus validating the assumption that the frequency components may be considered as linearly independent.

B. Interpretation

In the experiments described in this paper the total acoustic loss in a volume of fluid containing suspended particles was inferred from measurements of the reverberation time of the system. The attenuation due to the suspended particles may thus be estimated by comparing the reverberation times of the system with and without the particles present. Making measurements relative to the clear water case in this way partially corrects for the effects of the boundary losses and other contributions to the total attenuation in the system, such as absorption by the hydrophone mounts and the hydrophones themselves. However, it should be noted that when the suspended particles' contribution to the total attenuation is small (i.e., at low concentration) the effect of the relatively large losses due to the boundary and the hydrophones will be to give large errors in the relative measurement of the influence of the suspended particles on the reverberation time.

The reverberation time of a reverberating volume is defined as the time taken for the sound pressure level to fall by 60 dB after the sound source is removed and may be given by[14]

$$t_{60}=\frac{55.26\ V_{\mathrm{enc}}}{c(S_{\mathrm{enc}}\bar{a}+8\zeta V_{\mathrm{enc}})}, \tag{1}$$

where V_{enc} is the volume of the enclosure, S_{enc} is the surface area, c is the compression wave speed, $\bar{a}$ is the average Sabine absorptivity and ζ is the absorption coefficient of the fluid in Nepers per meter. The average Sabine absorptivity is defined

$$\bar{a}=\frac{A_{\mathrm{enc}}}{S_{\mathrm{enc}}}, \tag{2}$$

where A_{enc} is the total sound absorption of the enclosure, expressed in units of m^2, such that $\bar{a}$ is dimensionless.

If t_{60} is the reverberation time of a volume of particulate-free water, then t'_{60} is the corresponding reverberation time in water containing suspended particles, differing only from the definition given in Eq. (1) in that ζ is replaced by ζ'. It is then clear from Eq. (1), and the relation $\alpha=20\log(e)\zeta$, that the difference in the attenuation coefficients of the two fluids, expressed in dB m^{-1}, is given by

$$\Delta\alpha=20\log(\mathrm{e})\frac{55.26}{8c}\left(\frac{1}{t'_{60}}-\frac{1}{t_{60}}\right)=\frac{60}{c}\left(\frac{1}{t'_{60}}-\frac{1}{t_{60}}\right), \tag{3}$$

provided that the addition of the particles does not significantly affect (i) the sound speed in the medium; (ii) the volume of fluid; or (iii) the absorption at the boundaries. It should be noted that while $\Delta\alpha$ is defined in terms of the reverberation times t_{60} and t'_{60}, in practice it is determined from the rate of decay of the sound field rather than from a direct measurement of t_{60}.

The change in sound speed resulting from the introduction of the particles at concentrations used in the experiments is sufficiently small[12] (less than a few tens of $\mathrm{cm\,s}^{-1}$ even at the highest concentrations) not to invalidate Eq. (3). Similarly, the volume change due to the addition of the particles at the concentrations used in these experiments is sufficiently small (volume fractions are typically much smaller than 10^{-3}), and it may be assumed that the particles in suspension have little effect on the absorption properties of the walls. Particles that settle to the bottom of the volume may, however, affect the boundary losses, but attempts were made to make measurements with the great majority of the particles in suspension.

As discussed above, the reverberation time should be determined from the decay of a diffuse sound field. The onset of a diffuse sound field in an enclosure can be described by the Schroeder cut-off frequency, which gives an indication of the lowest frequency at which the modal density, i.e., the number of modes per unit bandwidth, is sufficient to constitute a diffuse field. The Schroeder cut-off frequency, f_s, may be written[15]

$$f_{\mathrm{s}}=\left(\frac{c^3}{4\ln 10}\right)^{1/2}\left(\frac{t_{60}}{V_{\mathrm{enc}}}\right)^{1/2}. \tag{4}$$

Ideally then, all experiments should be carried out at frequencies well above the Schroeder cut-off in order to ensure that the measurements are not influenced by modal structure in the sound field. For the experimental system described in this paper the Schroeder cut-off frequency was typically around 50 to 75 kHz. To test the assumption that the field is diffuse, measurements of reverberation time were made with the hydrophones in different positions.[11] It was found that the change in reverberation time occurring when the hydrophones were moved vertically was dominated by the change in absorption owing to the difference in length of submerged hydrophone cables. It was therefore not possible to investigate experimentally the variation in reverberation time in the vertical direction. However, measurements made in a number of hydrophone positions in the horizontal plane, i.e., with the hydrophones at a constant depth such that the amount of submerged cable is unchanging, showed a 4% standard deviation in the reverberation time. This is the dominant source of experimental error in the system and is taken into account in the error analysis.

C. Particle concentration and size distributions

The attenuation depends on both the absolute concentration of particles in suspension, and the size distribution of the particles. As described previously, the concentration is determined by carefully weighing the dry particles before they are added to the volume and ensuring that the acoustic measurements are made with all of the particles in suspension.

Because the particles settle out of suspension over time, the LSS was used to monitor the concentration to ensure that

it did not change significantly over the time taken for a series of acoustic measurements to be carried out.

The LSS only provides a relative measurement of the total concentration of suspended particles in the measurement volume, and does not provide any information on the sizes of the particles present. Since the particle size is an important parameter influencing the acoustic absorption by suspended particles, it is important to know the size distribution of the particles in suspension. While this is possible with spherical particles, it is not a simple matter where nonspherical particles are used. This is discussed in the following sections.

The particle size distribution was determined using three standard techniques: (*i*) laser diffraction analysis; (*ii*) gravitational sedimentation; and (*iii*) centrifugal sedimentation. These measurements were carried out independently and it was not possible to measure the particle size distribution dynamically during the acoustic measurements.

1. Light scattering sensor

It is important here to note the limitations of the LSS device for quantitative measurements of suspended particle concentration. The LSS device emits light in the forward direction and detects the light which is backscattered from the medium. While the intensity of the backscattered light depends upon the concentration of particles in suspension, it also depends upon the properties of the particles, such as their size relative to the wavelength of the laser light, shape and complex refractive index. The LSS may be calibrated using a known concentration of the suspension of interest. However, in cases where the suspension contains particles with a range of sizes, larger particles will settle out of suspension more quickly. The size distribution will therefore be time-varying, thus invalidating the calibration. It should be noted that the LSS was not used to obtain quantitative measurements of particle concentration for normalization of the attenuation measurements. This was done by weight, as previously described.

2. Laser diffraction analysis

Laser diffraction analysis uses the diffraction pattern of laser light scattered by a sample of particulate in suspension to infer the particle size distribution. The instrument used in this study measured particle diameters in the range 0.4 to 1000 μm.

It is instructive to describe briefly the laser diffraction method of particle sizing, which exploits the fact that small particles in a laser beam scatter light in a characteristic pattern, i.e., the diffraction pattern. The details of the diffraction pattern depend upon the distribution of particle sizes contributing to the light scattering. Information about the particle size distribution can thus be inferred from the details of the light flux pattern.

The simplest diffraction pattern due to a collection of spheres is that from a monomodal dispersion. It is the familiar Airy pattern consisting of a central bright spot surrounded by concentric light and dark rings, the intensity of which diminish further from the center of the diffraction pattern, that is to say at higher scattering angles. The scattering angles at which the diffraction maxima and minima occur depend on the size of the particles relative to the laser wavelength, with smaller particles leading to higher scattering angles. The particle size in such a monomodal dispersion of spheres can therefore be simply inferred from the scattering angles.

These diffraction patterns obey the principle of linear superposition, meaning that the total scattering pattern for a mixture of two or more monomodal dispersions is obtained by adding the intensities of the scattering patterns from each constituent monomodal dispersion. This allows the possibility of inferring the particle size distribution from sufficiently accurate measurements of the scattering pattern due to a multimodal dispersion of spherical particles.

The interpretation of laser diffraction measurements of particle size distribution becomes complicated when nonspherical particles are involved. The standard analysis yields the size distribution of spherical particles which would give the observed diffraction pattern. In principle it would be possible to infer the size distribution of particles of a different shape by inverting a model for the diffraction pattern obtained from forward scattering by such particles. However, for highly irregular particles and suspensions containing many different particles, this becomes impractical. A more fundamental question is what exactly we mean by the size when discussing nonspherical particles. In general nonspherical particles cannot be described by a single number, such as the diameter of a sphere. The laser diffraction technique may be considered to yield an effective spherical diameter for optical scattering. Other particle sizing techniques are available which also yield effective spherical diameters for nonspherical particles. In order to compare with the measurements made by laser diffraction, measurements have also been made using two alternative techniques: (*i*) gravitational sedimentation and (*ii*) centrifugal sedimentation.

3. Gravitational sedimentation

The gravitational sedimentation technique is based on the measurement of the steady-state settling velocity of the particles as they fall through a fluid under the influence of gravity. This method yields the Stokes diameter of the particle, defined as the diameter of a sphere which has the same density and the same free falling velocity in a given fluid as the particle, within the viscous flow regime. From Stokes' equation the Stokes diameter may be written

$$d_{\mathrm{st}}=\left(\frac{18\eta h_{\mathrm{s}}}{(\rho'-\rho)gt}\right)^{1/2}, \tag{5}$$

where η is the molecular viscosity of the suspending fluid, h_{s} is the distance the particle falls in time t, ρ' and ρ are the densities of the particle and fluid, and g is the acceleration due to gravity. Typically, in gravitational sedimentation measurements, the time taken for the particle to reach its terminal velocity is negligible.[16] The free fall velocity is therefore taken to be $v_{\mathrm{s}}=h_{\mathrm{s}}/t$.

Stokes' equation is only valid in the region of viscous flow, which sets an upper limit on the particle size which may be determined by this technique. This limit is determined by the Reynolds number

$$\mathrm{Re}=\frac{\rho v_s d_{st}}{\eta}. \tag{6}$$

The Reynolds number should not exceed 0.25 if the error in the Stokes diameter is not to exceed 3%.[16] Setting $\mathrm{Re}=0.25$ and equating Eq. (5) and Eq. (6) yields the following expression for the limiting Stokes diameter:

$$d_{st}|_{\max}=\left(\frac{4.5\eta^2}{(\rho'-\rho)\rho g}\right)^{1/3}. \tag{7}$$

For example, for silica ($\rho'=2650\ \mathrm{kg\,m^{-3}}$) settling in water ($\rho=1000\ \mathrm{kg\,m^{-3}}$; $\eta=0.001$ Pa s), the limiting Stokes diameter is 65.3 μm.

The lower size limit is determined partly by the long settling times experienced by small particles, and partly by other motions which may be significant compared to the small settling velocities, such as Brownian motion, diffusion and convection currents which may be set up over long integration periods. For these reasons the use of gravitational sedimentation is not usually recommended for particles smaller than about 1 μm.

4. Centrifugal sedimentation

Some of the difficulties associated with the use of gravitational sedimentation for fine particles may be reduced by speeding up the settling time. This can be achieved through the use of centrifugal sedimentation techniques.

As in the case of gravitational sedimentation the Stokes diameter is determined from Stokes' law, but now the acceleration due to gravity is replaced by the centrifugal acceleration and the free fall velocity is replaced by the radial settling velocity to give

$$d_{st}=\left(\frac{18\eta}{(\rho'-\rho)\omega_c^2 r_c}\frac{\mathrm{d}r_c}{\mathrm{d}t}\right)^{1/2}, \tag{8}$$

where r_c is the radial distance of the particle from the axis of the centrifuge, $\mathrm{d}r_c/\mathrm{d}t$ is the radial settling velocity and ω_c is the rotation speed of the centrifuge in radian $\mathrm{s^{-1}}$.

Centrifugal sedimentation is usually used for particles up to a few microns in diameter, although this can easily be extended using more viscous suspending fluids and longer settling times. The lower limit on particle size is determined by the consideration that the radial displacement of the particles by Brownian motion during sedimentation should be much smaller than the displacement due to centrifugal motion. For a typical example[17] the minimum Stokes diameter is less than 0.01 μm.

For the measurements of the size distributions of clay-like particles presented in this paper, all samples were dispersed in the suspension media and insonified with ultrasound (normally for one minute) to assist in dispersal and breaking up of agglomerates. Additionally, the sample portions used for the centrifugal sedimentation were dispersed using more prolonged ultrasonics in an attempt to further disseminate the aggregates and obtain a particle size distribution more representative of the discrete particles. This additional deflocculation was only used for the centrifugal sedimentation samples, in order to exploit the ability of that technique to characterize smaller particles.

III. THEORY

Urick[8] obtained the following expression for the visco-inertial attenuation coefficient by employing the expression for the viscous drag on an oscillating sphere developed by Stokes:[18]

$$\alpha_v=20\log(\mathrm{e})\frac{\phi k(\sigma-1)^2}{2}\left[\frac{s}{s^2+(\sigma+\tau)^2}\right]\quad \mathrm{dBm^{-1}}, \tag{9}$$

where

$$\tau=\frac{1}{2}+\frac{9}{4}\left(\frac{\delta_v}{a}\right) \tag{10}$$

and

$$s=\frac{9}{4}\left(\frac{\delta_v}{a}\right)\left(1+\frac{\delta_v}{a}\right). \tag{11}$$

Here ϕ is the volume fraction of suspended particles, k is the acoustic wavenumber, σ is the ratio of the densities of the solid and fluid phases, a is the particle radius and $\delta_v=\sqrt{2\eta/(\omega\rho)}$ is the skin depth for shear waves, where ω is the acoustic frequency. Polydisperse suspensions may be accounted for by summing the contributions from each size bin.

Implicit in Eq. (9) is the assumption that the absorption coefficient in a suspension of similar particles is linearly proportional to the volume fraction, ϕ, i.e., the process is linearly additive. This assumption is valid for dilute suspensions, in which interparticle interactions may be neglected. Urick[8] showed experimentally that this linear relationship between the viscous absorption coefficient and concentration holds for volume fractions of up to 8% to 9% for kaolin particles at MHz frequencies. Note that a volume fraction of 8% corresponds to a mass concentration of approximately 200 $\mathrm{kg\,m^{-3}}$ for kaolin particles. This is far higher than concentrations found in the natural environment, except perhaps in the boundary region near the seabed.

Equation (9) may be extended to account for spheroidal particles by employing suitable expressions[9] for τ and s, employing a shape factor[19] and inertia coefficient[20] for the drag on spheroidal bodies. The application of these equations to the calculation of attenuation by the suspensions under investigation in this paper was first described by the current authors in another paper[21] and the equations will not therefore be reproduced here. The only additional parameters introduced in this analysis are the semimajor and semiminor axes of the spheroids.

The absorption due to spheroidal bodies in suspension depends on the orientation of the body with respect to the insonifying wave. The model used in this work accounts for

particles having the two orthogonal orientations: with the axis of symmetry orientated parallel to the direction of motion, and with the axis perpendicular to the direction of motion. For comparison with the measurements made using the reverberation technique it is appropriate to assume that the particles are orientated randomly with respect to the acoustic field. That is not to say that they are randomly orientated with respect to the gravitational field. Indeed, there may well be a preferred orientation for particles settling slowly under gravity. However, under diffuse field conditions all propagation directions are equally probable and so it may be assumed that there is no preferred particle orientation with respect to the insonifying sound waves.

When one is considering attenuation by suspended particles in a directional sonar beam, the dependence on orientation should be taken into account if the particles are expected to have a preferred orientation. This will be dealt with in a forthcoming paper.

IV. RESULTS

The results of measurements made with spherical glass particles were found to be in very good agreement with the predictions of Eqs. (9)–(11).[3] In this paper results are presented for two types of nonspherical particle. The first is a form of kaolin (china clay) with the trade name Speswhite, and the second is a form of calcium carbonate with the trade name Polcarb. Both are well characterized industrial samples obtained from English China Clays International (ECCI, now Imerys). These particulates were chosen because they are nonspherical particles representative of some of the types of particles found in the marine environment, but with well known properties, avoiding the problems of working with poorly characterized natural samples. Results from natural marine sediment particles will be presented in a future paper.

Samples of both types of particles were examined using a scanning electron microscope.[12,21] The Speswhite particles were found to take the form of very thin, flat plates, whilst the Polcarb particles were more angular in form. The densities of the Speswhite and Polcarb particles are, respectively, 2600 and 2700 $\mathrm{kg\,m^{-3}}$.

A. Particle size distributions

Figures 3 and 4 show particle size distributions for the Speswhite and Polcarb particles, respectively, measured using laser diffraction analysis, gravitational sedimentation and centrifugal sedimentation.

Clearly there are very significant differences between the size distributions obtained using these different methods. This serves as an illustration of the fundamental difficulties of characterizing irregular particles. As described in Sec. II C, the laser diffraction technique may be considered to yield an effective optical scatterer dimension, since it gives the size distribution of spherical particles which would give the observed optical diffraction pattern. Both the gravitational and centrifugal sedimentation techniques use Stokes' law to determine particle size, and hence yield an effective Stokes diameter.

In the case of Speswhite in particular the centrifugal sedimentation measurement shows a strong bias towards smaller particles. Since samples used for the centrifugal sedimentation measurements were treated with prolonged ultrasound to break up aggregates this might possibly be an indication that there was some degree of aggregation of the Speswhite particles. This is consistent with scanning electron microscope images of the Speswhite particles,[12] which do show some clumping of particles.

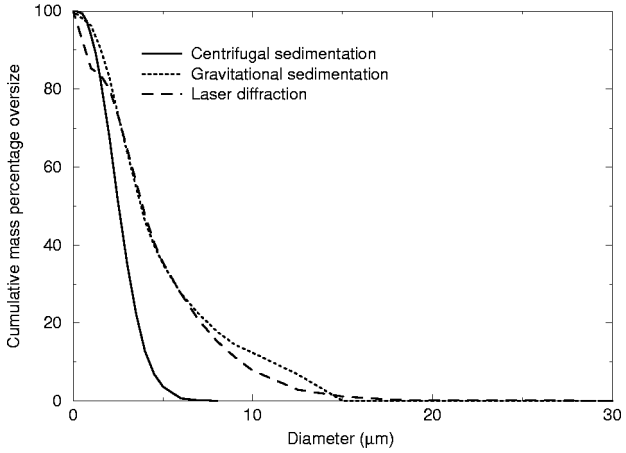


FIG. 3. Size distribution of Speswhite particles measured by centrifugal sedimentation, gravitational sedimentation and laser diffraction.

B. Absorption measurements

Figure 5 shows the measured attenuation coefficient for the Speswhite (kaolin) particles as a function of frequency, normalized with respect to concentration. The symbols indicate measurements made at different concentrations. The measurements are binned at 10 kHz intervals over the range 50–150 kHz, but are shown offset slightly so that the individual error bars can be resolved.

Also shown on this graph is the attenuation predicted by Eqs. (9)–(11) for spherical particles using the three size distributions shown in Fig. 3. These three predictions are in surprisingly close agreement with each other, given the apparent differences displayed by the size distributions measured by the different techniques. However, it must be remembered that the absorption is dominated by those particles

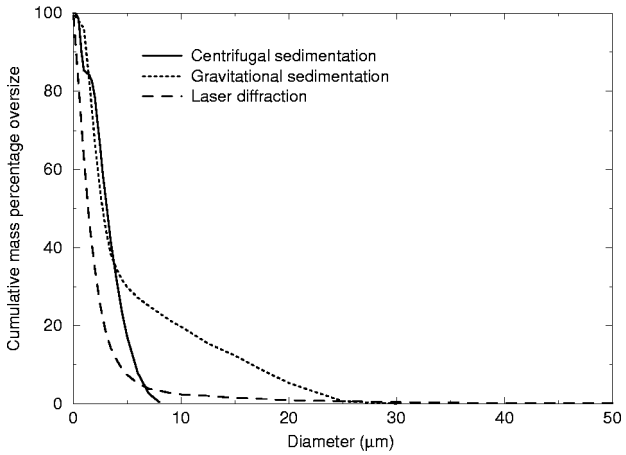


FIG. 4. Size distribution of Polcarb particles measured by centrifugal sedimentation, gravitational sedimentation and laser diffraction.

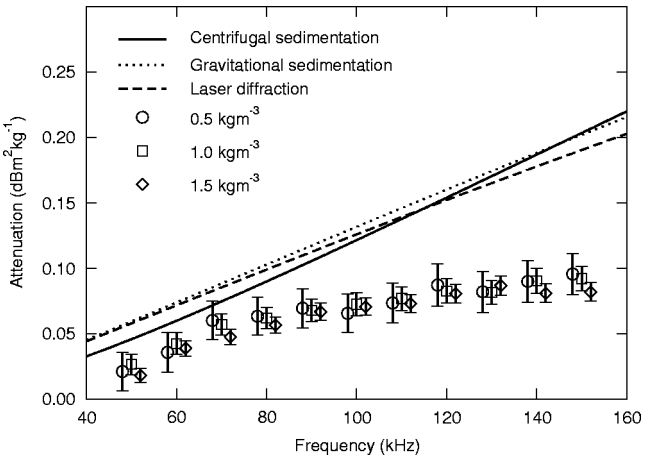


FIG. 5. Normalized attenuation coefficient for Speswhite particles: experimental data and theoretical predictions assuming spherical particles. Data points have been offset in frequency to show individual error bars (see text).

whose size is of the order of the skin depth for shear waves. At 100 kHz, for example, the skin depth in water is approximately 2 μm, and 80% of the absorption is due to particles smaller than 6.5 μm. As expected for these highly nonspherical particles, the prediction of the theory for spherical particles does not agree well with the measured attenuation.

Figure 6 shows the results of the measurements of attenuation with Polcarb (calcium carbonate) particles. Again, the symbols indicate measurements made at different concentrations, and the data have been normalized with respect to concentration. In this case the theoretical predictions using the size distributions yielded by the different sizing techniques show greater differences than in the case of the Speswhite. This is because the size distributions have greater differences in the particle size range contributing most to the absorption, although this is not immediately apparent from visual inspection of the curves showing cumulative mass percentage oversize.

The measurements for Polcarb show much better agreement with the theoretical predictions than was the case for Speswhite. Although the theoretical curves do not have the same form for the frequency dependence as suggested by the

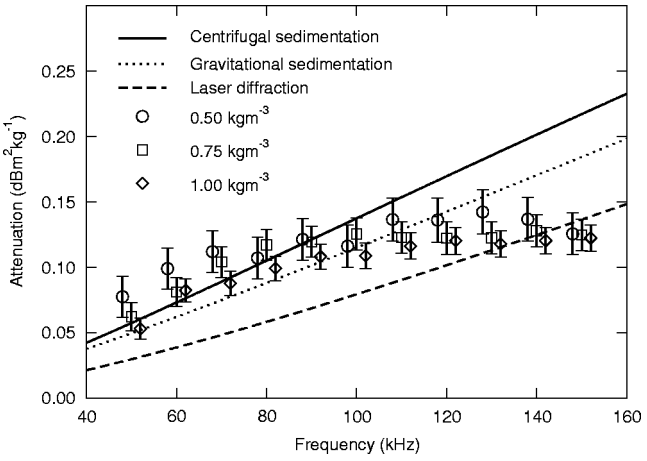


FIG. 6. Normalized attenuation coefficient for Polcarb particles: experimental data and theoretical predictions assuming spherical particles. Data points have been offset in frequency to show individual error bars (see text).

data, the magnitude of the predicted absorption is close to the measured absorption over much of the frequency range. The fact that the Polcarb measurements agree better with the predictions of the spherical particle theory than the Speswhite measurements is not surprising, since inspection of electron microscope images shows the Speswhite particles to be highly platelike whereas the Polcarb particles, although certainly irregular and nonspherical, are closer to spheres in aspect ratio.

These figures show the comparison between the predictions of Eq. (9) with Eqs. (10) and (11), which assumes spherical particles, and measurements made with nonspherical particles. These results demonstrate that the spherical model is of some limited use in providing estimates of the attenuation from the Polcarb particles. In the case of the highly nonspherical Speswhite particles, the spherical theory significantly overpredicts the attenuation.

A method for calculating the attenuation due to suspensions of oblate spheroids was referred to in Sec. III. Since the degenerate form of an oblate spheroid is a thin circular disk it is appropriate to approximate the platelike Speswhite particles as oblate spheroids.

In order to apply this method it is necessary to know the size distribution of the spheroids representing the particles. It is not appropriate to use the size distributions shown in Figs. 3 and 4 as these are distributions of equivalent spheres. Therefore these distributions have been used to derive new distributions for spheroids.

The starting point for this process is the distribution of Stokes diameters obtained from gravitational sedimentation measurements. These were originally derived by applying Eq. (5) to measurements of the fraction of particles which settle out of suspension as a function of time. Although the original time-domain data were not available, it was possible to recalculate them by inverting Eq. (5) and applying it to the distributions of Stokes diameters. This gives the fraction of particles which settle out of suspension as a function of time. Now, if the settling velocity of spheroids is known as a function of their major and minor radii, these data can be used to derive a particle size distribution for the spheroids.

The steady-state settling velocity, v_s, for spheroids with shape factor[19] K_{sf} (which depends on the aspect ratio) may be obtained by equating the drag force on a spheroid:

$$F_0 = 6\pi\eta K_{sf} a' v_s \tag{12}$$

with the gravitational force and rearranging to give

$$v_s = \frac{4\pi a' b' (\rho' - \rho) g}{18\pi\eta K_{sf}}, \tag{13}$$

where a' and b' are, respectively, the semi-major and semi-minor axes for oblate spheroids and the semi-minor and semi-major axes for prolate spheroids.

In this way the size distribution for Speswhite particles derived by gravitational sedimentation, shown in Fig. 3, was used to calculate the size distribution of spheroids of given aspect ratio, $h = b'/a'$, which would give the same measured settling time history. A similar analysis is used for the centrifugal sedimentation measurements.

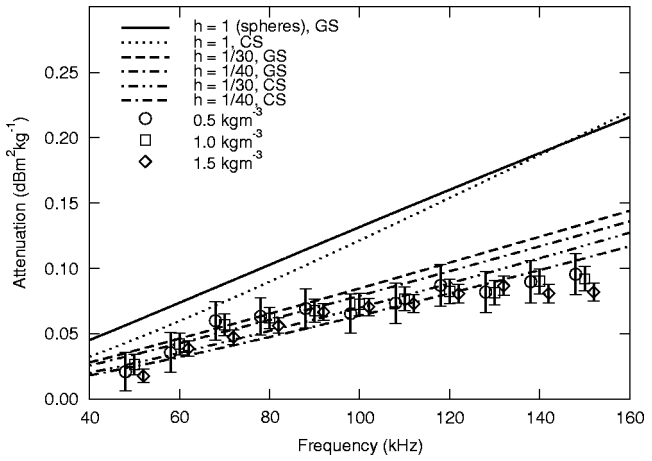


FIG. 7. Normalized attenuation coefficient for Speswhite particles: experimental data and theoretical predictions using the model for spheroidal particles. The predictions are based on size distributions derived from gravitational sedimentation (GS) and centrifugal sedimentation (CS) measurements.

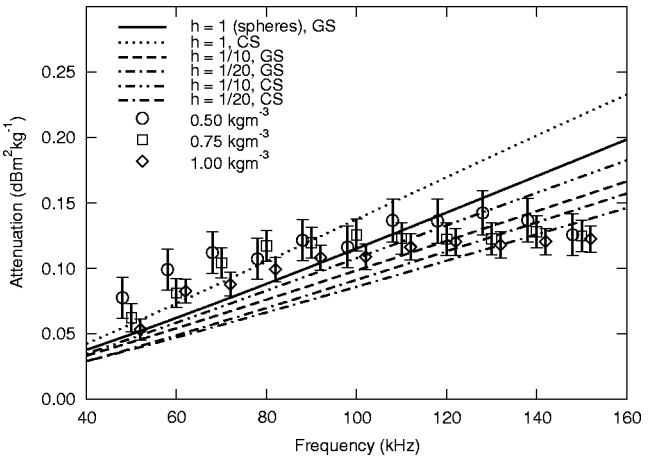


FIG. 8. Normalized attenuation coefficient for Polcarb particles: experimental data and theoretical predictions using the model for spheroidal particles. The predictions are based on size distributions derived from gravitational sedimentation (GS) and centrifugal sedimentation (CS) measurements.

These particle size distributions were then used to calculate the attenuation spectrum due to the suspension of spheroids, using the approach referred to in Sec. III. The attenuation measurements, as described in Sec. II, employ an approximately diffuse field in which all directions of propagation are equally probable. Ideally, then, these calculations would be integrated over all particle orientations. However, the method only yields solutions for the two orthogonal cases. In the case of oblate spheroids the three independent spatial directions may be resolved into the broadside direction and two edgewise directions, and the results presented below were therefore obtained assuming that two-thirds of the particles were orientated edgewise to the sound field and one-third orientated broadside.

Figure 7 shows the comparison between the attenuation predicted using the oblate spheroid model and the measured attenuation for Speswhite particles. The range of typical aspect ratios of the Speswhite particles as quoted by the suppliers is $h=\frac{1}{30}$ to $h=\frac{1}{40}$. The figure therefore, shows predictions for these values of aspect ratio, together with the curve for $h=1$, i.e., for spherical particles. The predictions for spheres, using the spheroidal model with $h=1$, are in agreement with the predictions of the spherical model shown in Fig. 5, and thus overestimate the attenuation. The predictions for the aspect ratios which are representative of the Speswhite particles, however, show much better agreement with the measured attenuation. It is notable that this agreement is achieved without using any *a priori* knowledge of the attenuation measurements. While the predictions of the spheroid model show significantly improved agreement with the data, it is noted that the exact form of the frequency dependence suggested by the data is not predicted by the model. In particular there is a prominent increase in attenuation over the range 50 to 70 kHz which is not predicted. This is in the frequency range of the Schroeder cut-off and this is suggested as a possible explanation for this feature of the data. It is also noted that the predictions based on the centrifugal sedimentation measurements of particle size distribution appear to be slightly closer to the measurement than those based on gravitational sedimentation, but the data do not really support any definitive conclusions on this matter.

The spheroidal approach has also been applied to the Polcarb particles in exactly the same manner. However, inspection of the scanning electron microscope images[12,21] showed that they are very angular and irregular. The spheroidal approach was therefore not expected to yield significantly improved agreement over the spherical model. Furthermore, no information was available on the aspect ratio, other than perhaps what could be inferred from the microscope images. Therefore calculations were performed for a number of different aspect ratios and those which gave predictions which were reasonably close to the measured data are shown in Fig. 8. All that can really be concluded from this result is that the predictions of the spheroidal model for $0.05 \leqslant h \leqslant 1$ are neither better nor worse than the predictions of the spherical model. This is consistent with expectations, given the nature of the particles.

Particle size data based on laser diffraction analysis were also presented in Sec. IV A. It is not, however, appropriate to use these data in the spheroidal analysis, as was done using the gravitational and centrifugal sedimentation data. That is because the analysis relies on equating the steady-state settling velocities of spheres and spheroids. To do that with the laser diffraction data would require that the effective optical scatterer diameter be equated to the Stokes diameter, a step which would be both arbitrary and nonphysical.

V. DISCUSSION

The experimental method described in this paper has previously been shown[3] to give results for absorption by suspensions of spherical particles which are in good agreement with theory. In this paper the method has been applied to the measurement of absorption in suspensions of highly nonspherical particles of Speswhite (kaolin) and Polcarb (calcium carbonate). With these nonspherical particles the method of obtaining the particle size distribution became an important issue, and three different particle sizing techniques

were employed. Significant differences were observed in the particle size distributions yielded by the three techniques, leading to differences in the predicted attenuation.

The predictions of Urick's equation for visco-inertial absorption by spherical particles, integrated over the size distributions obtained using the three particle sizing techniques, significantly overestimate the attenuation by the Speswhite particles. This poor agreement is to be expected since these platelike particles are far from spherical. Agreement between the spherical particle model and the experimental data for Polcarb particles is much better, although the model fails to predict the observed frequency dependence in the attenuation. These particles, while still highly irregular, have aspect ratios closer to unity than the kaolin particles do. The closer agreement is therefore consistent with expectation.

A model for acoustic attenuation by dilute suspensions of spheroidal particles has been used to predict the attenuation in suspensions of the two different particles. This model is most appropriate for the platelike Speswhite particles as the degenerate form of an oblate spheroid is a circular disk. The predictions of the model for oblate spheroids, using particle size data based on gravitational sedimentation measurements, were found to be in good agreement with the measurements of attenuation by Speswhite particles. This agreement was obtained without the need for any *a priori* information about the attenuation measurements. Agreement between this model and the measurements made with Polcarb particles was less good, as expected due to the shape of these particles.

It may be noted that instruments based on the standard, spherical analysis of laser diffraction measurements are commonly used for *in situ* particle sizing.[22] The results given in this paper suggest that such measurements may be of limited use when applying the derived size distributions to the prediction of acoustic attenuation spectra.

It is concluded that the model for visco-inertial absorption by oblate spheroids used in this paper is appropriate for estimating the absorption in dilute suspensions of platelike clay particles.

ACKNOWLEDGMENTS

The authors are grateful to S. G. Foran, QinetiQ Bridgwater, for performing the particle size measurements, and to Gary Heald, QinetiQ Winfrith, for valuable discussions. TGL would like to thank the Royal Society and the Leverhulme Trust for a Senior Research Fellowship. This work was funded by Technology Group 01 of the MoD Corporate Research Program.

[1] S. D. Richards, A. D. Heathershaw, and P. D. Thorne, ''The effect of suspended particulate matter on sound attenuation in seawater,'' J. Acoust. Soc. Am. **100**, 1447–1450 (1996).
[2] S. D. Richards and T. G. Leighton, ''Acoustic sensor performance in coastal waters: Solid suspensions and bubbles,'' in *Acoustical Oceanography*, Vol. 23 of *Proceedings of the Institute of Acoustics*, edited by T. G. Leighton, G. J. Heald, H. D. Griffiths, and G. Griffiths (Bath U. P., Bath, 2001), pp. 399–406.
[3] N. R. Brown, T. G. Leighton, S. D. Richards, and A. D. Heathershaw, ''Measurement of viscous sound absorption at 50–150 kHz in a model turbid environment,'' J. Acoust. Soc. Am. **104**, 2114–2120 (1998).
[4] S. Temkin and R. A. Dobbins, ''Attenuation and dispersion of sound by particulate-relaxation processes,'' J. Acoust. Soc. Am. **40**, 317–324 (1966).
[5] J. M. Evans, ''Models for sound propagation in suspensions and emulsions,'' Ph.D. thesis, Open University, 1996.
[6] Z. W. Qian, ''Fractal dimensions of sediments in nature,'' Phys. Rev. E **53**, 2304–2306 (1996).
[7] Q. Wang, K. Attenborough, and S. Woodhead, ''Particle irregularity and aggregation effects in airborne suspensions at audio and low ultrasonic frequencies,'' J. Sound Vib. **236**, 781–800 (2000).
[8] R. J. Urick, ''The absorption of sound in suspensions of irregular particles,'' J. Acoust. Soc. Am. **20**, 283–289 (1948).
[9] A. S. Ahuja and W. R. Hendee, ''Effects of particle shape and orientation on propagation of sound in suspensions,'' J. Acoust. Soc. Am. **63**, 1074–1080 (1978).
[10] G. Kurtze and K. Tamm, ''Measurements of sound absorption in water and aqueous solutions of electrolytes,'' Acustica **3**, 34–48 (1953).
[11] T. G. Leighton, N. R. Brown, and S. D. Richards, ''Effect of acoustic absorption by hydrophone and cable on a reverberation technique for measuring sound absorption coefficient of particulate suspensions,'' ISVR Technical Report No. 299, University of Southampton, 2002.
[12] S. D. Richards, ''Ultrasonic visco-inertial dissipation in dilute particulate suspensions,'' Ph.D. thesis, University of Southampton, 2002.
[13] M. R. Schroeder, ''New method of measuring reverberation time,'' J. Acoust. Soc. Am. **37**, 409–412 (1965).
[14] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982).
[15] A. D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications* (McGraw-Hill, New York, 1981), reprinted by the Acoustical Society of America, 1989 and 1994.
[16] BS3406 Part 2, *British Standard Methods for Determination of Particle Size Distribution. Part 2: Recommendations for Gravitational Liquid Sedimentation Methods for Powders and Suspensions*, 1984.
[17] BS3406 Part 6, *British Standard Methods for Determination of Particle Size Distribution. Part 6: Recommendations for Centrifugal Liquid Sedimentation Methods for Powders and Suspensions*, 1985.
[18] G. G. Stokes, ''On the effect of internal friction of fluids on the motion of pendulums,'' Trans. Cambridge Philos. Soc. **IX**, 8–106 (1851).
[19] J. Happel and H. Brenner, *Low Reynolds Number Hydrodynamics* (Prentice–Hall, Englewood Cliffs, NJ, 1965).
[20] H. Lamb, *Hydrodynamics*, 6th ed. (Dover, New York, 1945).
[21] S. D. Richards, T. G. Leighton, and N. R. Brown, ''Visco-inertial absorption in dilute suspensions of irregular particles,'' Proc. R. Soc. London, Ser. A **459**, 1–15 (2003).
[22] Y. C. Agrawal and H. C. Pottsmith, ''Instruments for particle size and settling velocity observations in sediment transport,'' Mar. Geol. **168**, 89–114 (2000).

Consistency and accuracy of perturbative inversion methods for group travel time data

Brian J. Sperry
Ocean Sciences Group, SAIC, McLean, Virginia 22102

B. Edward McDonald[a)]
Naval Research Laboratory, Washington, DC 20375

Arthur B. Baggeroer
Department of Ocean Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 28 June 2002; accepted for publication 14 July 2003)

Perturbation theory for ocean acoustic modal group speed responses to small environmental changes is investigated with regard to its applicability to ocean acoustic tomography. Assuming adiabaticity, the inverse problem for each vertical eigenmode is an integral equation whose kernel involves the eigenfunction and its frequency derivative. A proof is given for the equivalence of two dissimilar forms of the integral equation. Numerical examples are given for the inversion kernel for four types of sound-speed profiles, and then the parameter range (amplitude and scale size) in which perturbation theory is accurate is examined. It is found that the range of validity is determined not only by the amplitude of the perturbations, but also by their vertical scale size.
[DOI: 10.1121/1.1605953]

PACS numbers: 43.30.Pc, 43.30.Bp [WLS]

I. INTRODUCTION

The theoretical framework for ocean acoustic tomography gives rise to some mathematical questions concerning the consistency and accuracy of perturbative inversion methods. This work explores these questions theoretically and numerically. The current paper grew out of earlier, mostly theoretical work[1,2] on the linear perturbation of modal group speeds which did not consider the limits in which the results are accurate. The earlier investigations assumed implicitly that the accuracy of the results is determined by the amplitude of the input perturbation parameters (in this case the sound-speed perturbations). Numerical work by one of the authors [B.J.S.] later produced the unexpected result that, with the amplitude of the sound-speed perturbation fixed at a value of order 2 m/s, the linear perturbation results were sometimes accurate, and sometimes inaccurate. The accuracy of perturbation theory turned out to depend on the vertical scale size of the imposed perturbations and on their amplitude. This finding was first noted in an unpublished internal report[3] by the present authors. The current paper presents both the theory and its domain of accuracy as determined by numerical simulation.

Perturbative approaches to ocean acoustic tomography[4–9] have sought to estimate corrections to an assumed basic ocean state by measuring perturbations in acoustic transmission properties. Large-scale ocean experiments (e.g., HIFT,[10] ATOC,[11] and SLICE89[12]) have recorded pulse arrivals on vertical hydrophone arrays, yielding large amounts of path-integrated acoustic information concerning the state of the ocean. A physically revealing way to display and/or process this information is to employ modal filters[13,14] and modal group travel time tomography.

Modal group travel time tomography relates arrival time information for individual acoustic eigenmodes to details of sound-speed structure in the water column[5] or ocean bottom[15] within a vertical slice taken along a presumed horizontal path. In the current work we investigate the adequacy of perturbative methods to calculate modal group slowness changes resulting from imposed perturbations in the water column. We assume sufficiently weak range dependence so as to allow the use of adiabatic normal-mode theory. While the use of adiabatic normal-mode theory leads to many helpful theoretical results, it is often the case in the real ocean that strong range dependence introduces mode coupling effects which are not considered here.

We show that two independent derivations of perturbed modal group slowness which appear to give different results[1–3] are actually consistent, as a result of a reciprocity principle in Sturm–Liouville systems. We refer to this apparent difference of results as the ''third term problem.'' We also address the evaluation of the frequency derivative of the acoustic eigenmode $\partial\psi_n(z,\omega)/\partial\omega$. This quantity plays a central role in perturbative inversion. Previous approaches[5,6,15,16] solved the appropriate eigenvalue problem at two neighboring frequencies and evaluated the frequency derivative by finite differences.[17] Finite differences may be sensitive to iteration errors at high mode numbers or in situations where the two eigenvalue solutions converge slowly.

Summarizing earlier work[2] by one of the present authors, we give (a) an independent method in which an equation for $\partial\psi_n/\partial\omega$ is solved directly, and (b) a least-squares tomographic inversion scheme which is consistent with the range-integrated formulation of Shang,[6] allowing for different methods of evaluating $\partial\psi_n/\partial\omega$. Rather than giving

a)Electronic mail: mcdonald@sonar.nrl.navy.mil

sample inversions of synthetic data for a single ''canonical'' sound-speed profile as in Ref. 6, we give numerical examples for the perturbative inversion kernel for the canonical profile and three other types of sound-speed profiles (SSP): a smooth midlatitude SSP, a mostly upward-refracting high-latitude SSP with strong surface interaction, and a double-ducted high-latitude SSP.

Next, we investigate numerically the parameter range in which linear perturbation theory yields accurate results for modal group slowness. In particular, we give numerical examples for the canonical profile in which the error in the perturbation method is calculated as a function of the amplitude and vertical extent of assumed perturbations in the water column.

II. ADIABATIC MODE THEORY

Linear perturbation of modal group slowness as calculated from adiabatic mode theory is summarized from earlier work[2] as follows. Vertical eigenmodes $\psi_n(z;x,y)$ (where the vertical coordinate is z, and the dependence of the environment on (x,y) is weak) satisfy

$$\rho\frac{\partial}{\partial z}\left(\rho^{-1}\frac{\partial\psi_n}{\partial z}\right)+\frac{\omega^2}{c^2(z;x,y)}\psi_n=\kappa_n^2(x,y)\psi_n, \tag{1}$$

with orthonormality

$$(\psi_n,\psi_m)=\begin{cases}1, & m=n\\ 0, & m\neq n,\end{cases} \tag{2}$$

where the inner product $(*,*)$ is

$$(A,B)\equiv\int\rho^{-1}AB\,dz. \tag{3}$$

Boundary conditions for (1) are $\psi_n=0$ at the pressure release surface $z=0$, and $\psi_n\to 0$ for $z\to-\infty$, corresponding to mode trapping. In Eq. (1) ρ is density, allowing for stratification in the ocean bottom.

For notational convenience we define slowness as the reciprocal of sound speed

$$s(z;x,y)\equiv c(z;x,y)^{-1}, \tag{4}$$

so that for a time-harmonic signal of radian frequency ω the wave number field is

$$k(z;x,y)=\omega s(z;x,y). \tag{5}$$

We represent (1) symbolically as

$$L\psi_n=\kappa_n^2(x,y)\psi_n, \tag{6}$$

where the linear operator L is

$$L=\rho\partial_z\rho^{-1}\partial_z+\omega^2 s^2(z;x,y). \tag{7}$$

With the stated boundary conditions and real sound speed (no attenuation in the basic state) the operator L is self-adjoint.

From (1) one has for the square of the modal horizontal wave number

$$\kappa_n^2(x,y)=(\psi_n,L\psi_n). \tag{8}$$

III. CONSISTENCY OF PERTURBATION RESULTS

From here on, weak range dependence is suppressed, so that $f(z)$ really represents $f(z;x,y)$. From Eq. (8) we will derive expressions for changes in modal group slowness due to perturbations $\delta s(z)$ in the water column sound slowness. Inversion schemes may then be constructed to relate variations in modal travel times (and thus modal group slowness) to the water column variations $\delta s(z)$.

One can either perturb (8) with respect to δs and then take a frequency derivative, or take the frequency derivative of (8) and then perturb with respect to δs. Following these different paths leads to very different-looking results. In this section we will show that they are indeed the same using a more fundamental approach than was used previously.[1,2]

A. Frequency derivative of the perturbed system

Upon perturbing the sound slowness profile while keeping the ocean depth constant, $s\to s+\delta s$ one has from (8)

$$\begin{aligned}2\kappa_n\delta\kappa_n&=(\psi_n,L\,\delta\psi_n)+(\delta\psi_n,L\psi_n)+2\omega^2(\psi_n,s\,\delta s\psi_n)\\&=2\kappa_n^2(\psi_n,\delta\psi_n)+2\omega^2(\psi_n,s\,\delta s\psi_n),\end{aligned} \tag{9}$$

where we have invoked (6) in the second expression. Orthonormality (2) requires that the first term on the right-hand side of the second expression be zero, so that the perturbation in the horizontal wave number is

$$\delta\kappa_n=\frac{\omega^2}{\kappa_n}(\psi_n,s\,\delta s\psi_n). \tag{10}$$

The frequency derivative of (10) gives

$$\begin{aligned}\frac{\partial}{\partial\omega}\delta\kappa_n\equiv\delta s_n^g=&\left(2\frac{\omega}{\kappa_n}-\frac{\omega^2}{\kappa_n^2}\frac{\partial\kappa_n}{\partial\omega}\right)(\psi_n,s\,\delta s\psi_n)\\&+2\frac{\omega^2}{\kappa_n}\left(\psi_n,s\,\delta s\frac{\partial\psi_n}{\partial\omega}\right),\end{aligned} \tag{11}$$

where δs_n^g refers to the perturbation in the modal group slowness of mode n. Equation (11) gives the following integral for the perturbed group slowness:

$$\delta s_n^g=\frac{\omega}{\kappa_n}\int\rho^{-1}\left(\left(2-\frac{\omega}{\kappa_n}s_n^g\right)\psi_n^2+2\omega\psi_n\frac{\partial\psi_n}{\partial\omega}\right)s\,\delta s\,dz. \tag{12}$$

Equation (12) relates the perturbation in modal group slowness to the change δs in the slowness distribution in an expression which requires knowledge of the eigenfunction, its frequency derivative, and unperturbed modal parameters κ_n and $s_n^g\equiv\partial\kappa_n/\partial\omega$. The perturbation $\delta\psi_n$ in the eigenfunction does not appear in (12), having been eliminated by orthonormality.

B. Perturbation of the frequency derivative

Returning to (8), a frequency derivative gives

$$2\kappa_n s_n^g=2\left(\frac{\partial\psi_n}{\partial\omega},L\psi_n\right)+2\omega(\psi_n^2,s^2). \tag{13}$$

The first term on the right is again zero due to (1) and orthonormality. The modal group slowness is thus

$$s_n^g = \frac{\omega}{\kappa_n}(\psi_n^2, s^2). \tag{14}$$

Perturbing (14) with respect to δs now gives

$$\delta s_n^g = 2\frac{\omega}{\kappa_n}[(\psi_n, s^2\delta\psi_n) + (\psi_n, s\,\delta s\,\psi_n)] - \delta\kappa_n\frac{\omega}{\kappa_n^2}(\psi_n^2, s^2)$$

$$= \frac{\omega}{\kappa_n}\int \rho^{-1}dz\left(\left(2 - s_n^g\frac{\omega}{\kappa_n}\right)\psi_n^2 s\,\delta s + 2s^2\psi_n\delta\psi_n\right). \tag{15}$$

The second expression here results from (10) and (14). Equation (15) gives the change in the modal group slowness due to δs in an expression requiring knowledge of eigenfunction, its response $\delta\psi_n$, and unperturbed modal parameters κ_n and s_n^g. The frequency derivative term in (13) is eliminated by orthonormality.

Equations (12) and (15) must give identical results for physical consistency, but they differ in each of their third terms. Consistency of the third terms of (12) and (15) requires that the following general relation hold among the eigenfunction, its frequency derivative, the imposed sound slowness perturbation, and the eigenfunction's response to the perturbation:

$$\int \rho^{-1}s(z)^2\psi_n\delta\psi_n\,dz \stackrel{?}{=} \omega\int \rho^{-1}s(z)\psi_n\frac{\partial\psi_n}{\partial\omega}\delta s(z)dz. \tag{16}$$

Equation (16) is a curious and unexpected relation we refer to as the ''third term problem.'' It points to a fundamental but not obvious role played by the eigenmode's frequency derivative in relating the water column perturbation δs to the eigenmode's response $\delta\psi_n$.

C. Reciprocity in perturbed Sturm–Liouville systems

We will show that Eq. (16) is not a constraint on the perturbations, but an example of a general property of Sturm–Liouville eigenvalue problems of the form $L\psi_n = \lambda_n\psi_n$ with orthonormality as in (2). Let us represent the system and its linear perturbation as

$$L\psi_n = \lambda_n\psi_n, \tag{17a}$$

$$L\,\delta\psi_n + \delta L\,\psi_n = \lambda_n\delta\psi_n + \delta\lambda_n\psi_n, \tag{17b}$$

and thus

$$(L-\lambda_n)\,\delta\psi_n = (\delta\lambda_n - \delta L)\,\psi_n. \tag{18}$$

The left-hand side of this equation is orthogonal to ψ_n, so that the right-hand side must be also. As a result

$$\delta\lambda_n = (\psi_n, \delta L\,\psi_n). \tag{19}$$

Since the Sturm–Liouville eigenvalues λ_n are distinct, the operator $(L-\lambda_n)$ has no zeros on the vector space $\mathbf{R}_{\perp n} \equiv \{\psi_j, j\neq n\}$ orthogonal to ψ_n. In other words, $(L-\lambda_n)$ is nonsingular on $\mathbf{R}_{\perp n}$ as a result of the Fredholm alternative.[18] Thus, we may solve (18) to yield

$$\delta\psi_n = (L-\lambda_n)^{-1}(\delta\lambda_n - \delta L)\,\psi_n$$

$$= -(L-\lambda_n)^{-1}(\delta L\,\psi_n)_{\perp n}, \tag{20}$$

where the subscript $\perp_n$ refers to projection onto $\mathbf{R}_{\perp n}$ (or equivalently orthogonalization to ψ_n).

Now, consider two linearly independent perturbations of the system $\delta L^{(j)}$, $j=1, 2$ with responses $\delta\psi_n^{(j)}$, $j=1, 2$ in the eigenfunctions. Consider the following cross term between perturbations and responses resulting from (20):

$$(\delta\psi_n^{(1)}, \delta L^{(2)}\psi_n) = -((L-\lambda_n)^{-1}(\delta L^{(1)}\psi_n)_{\perp n}, \delta L^{(2)}\psi_n)$$

$$= -(\delta L^{(1)}\psi_n, (L-\lambda_n)^{-1}(\delta L^{(2)}\psi_n)_{\perp n})$$

$$= (\delta L^{(1)}\psi_n, \delta\psi_n^{(2)}). \tag{21}$$

Thus

$$(\delta\psi_n^{(1)}, \delta L^{(2)}\psi_n) = (\delta\psi_n^{(2)}, \delta L^{(1)}\psi_n). \tag{22}$$

The left and right sides of (22) differ only by the interchange of indices. This statement of index interchangeability between independent perturbations δL and responses $\delta\psi_n$ in (22) brings to mind the forcing-response reciprocity[19] in many reversible physical systems.

If we let $\delta L^{(1)}$ refer to a frequency perturbation $\delta\omega$ and $\delta L^{(2)}$ refer to a perturbation $\delta s(z)$ in the sound slowness profile, we have from (22) and (7)

$$(\delta\psi_n^{(1)}, 2\omega^2 s(z)\,\delta s(z)\,\psi_n) = (\delta\psi_n^{(2)}, 2\omega\,\delta\omega s^2(z)\,\psi_n). \tag{23}$$

Dividing both sides by $\delta\omega$, taking its limit to zero, and suppressing superscript (2) for the eigenfunctions' response to the sound slowness perturbation, we have

$$\left(\omega\frac{\partial\psi_n}{\partial\omega}, s(z)\,\delta s(z)\,\psi_n\right) = (\delta\psi_n, s(z)^2\psi_n), \tag{24}$$

which proves Eq. (16). We have thus established the equivalence of (12) and (15). To summarize, we have established that the perturbed group slowness may be calculated from either of the following equivalent expressions:

$$\delta s_n^g = \frac{\omega}{\kappa_n}\int \rho^{-1}\left(\left(2 - \frac{\omega}{\kappa_n}s_n^g\right)\psi_n^2 + 2\omega\psi_n\frac{\partial\psi_n}{\partial\omega}\right)s\,\delta s\,dz$$

$$= \frac{\omega}{\kappa_n}\int \rho^{-1}\left(\left(2 - \frac{\omega}{\kappa_n}s_n^g\right)\psi_n^2\delta s + 2s\psi_n\delta\psi_n\right)s\,dz. \tag{25}$$

IV. THE INVERSION PROBLEM

A. The group slowness kernel

One can derive from (20) an equation for the eigenfunction's frequency derivative. The result is[2]

$$\frac{\partial\psi_n}{\partial\omega} = -2\omega(L-\kappa_n^2)^{-1}(s^2\psi_n)_{\perp n}. \tag{26}$$

Equation (26) may have utility in areas outside the current investigation. For example, LePage[20] has developed a narrow-band approximation to synthesize time-domain pulses by Taylor expanding the modal eigenfunctions in frequency to two or three terms, and carrying out frequency integration under a Gaussian window. In order to obtain physically reasonable arrival patterns in deep water, it is essential to include the frequency derivative of the

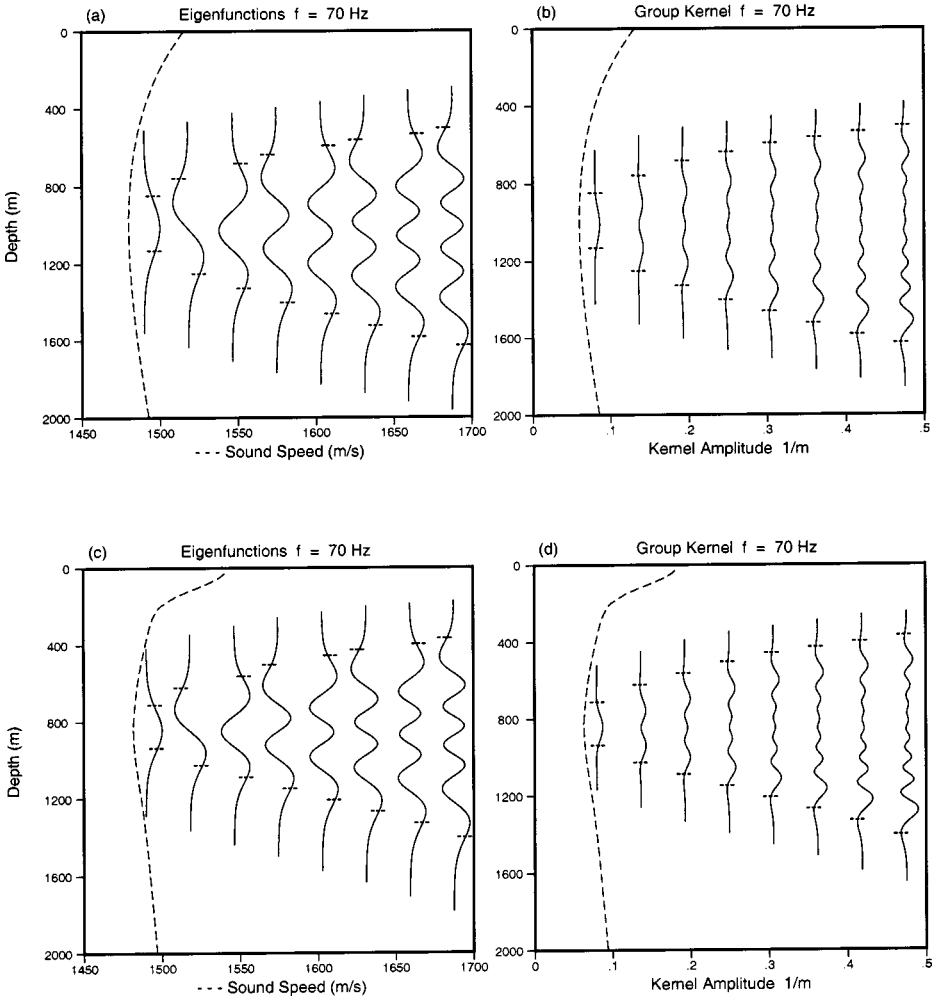


FIG. 1. (a) The first eight eigenmodes and (b) group slowness kernels for the canonical profile of Eq. (34). (c) Eigenmodes and (d) group slowness kernels for an SSP taken near Ascension Island in January. The horizontal marks on each plot give the upper and lower ray turning points. Each plot is terminated where the amplitude of the function permanently falls below 0.001 times its maximum amplitude.

eigenfunctions.[21] From Eqs. (20), (26), and (12), we find

$$\delta s_n^g = \frac{\omega}{\kappa_n}\left(2-\frac{s_n^g}{s_n^p}\right)(\psi_n, s(z)\delta s(z)\psi_n) - 4\frac{\omega^3}{\kappa_n} \times ((L-\kappa_n^2)^{-1}(s^2(z)\psi_n)_{\perp n}, s(z)\delta s(z)\psi_n), \quad (27)$$

where $s_n^p = \kappa_n/\omega$ is the modal phase slowness. Equation (27) may be written as in integral equation with kernel K

$$\delta s_n^g = \int dz\, \delta s(z) K_n(z), \quad (28)$$

where

$$K_n(z) = \rho^{-1}\frac{\omega}{\kappa_n}\psi_n s(z)\left\{\left(2-\frac{s_n^g}{s_n^p}\right)\psi_n - 4\omega^2(L-\kappa_n^2)^{-1} \times (s^2(z)\psi_n)_{\perp n}\right\}. \quad (29)$$

Equation (28) gives the inverse problem for changes in the ocean sound slowness structure $\delta s(z)$. Arrival time data from a vertical array of sufficient resolution can be processed to yield changes δs_n^g in modal group slowness. Knowledge of the functions $K_n(z)$ determined from the unperturbed environment may be used in a least-squares inversion of (28) as follows.

B. Least-squares inversion

The form of (28) leads to a convenient approximate inversion scheme, providing that the K_n can be obtained numerically. While the K_n in general are neither orthogonal or complete, they are linearly independent so that one can carry out a least-squares approximation for $\delta s(z)$ when the δs_n^g are known. Other parametrizations may have advantages, but we wish to explore theoretical results arising naturally from (28). We define

$$\delta s(z)|_K = \sum_n A_n K_n(z), \quad (30)$$

where the quantity on the left-hand side is the projection of the slowness perturbation $\delta s(z)$ onto the space of functions $K_n(z)$. Coefficients A_n are then determined by minimizing the vertically integrated squared difference between $\delta s(z)$ and the right-hand side of (30). Invoking (28), the result is

$$\mathbf{A} = \mathbf{C}^{-1}\delta\mathbf{s}^{\mathbf{g}}, \quad (31)$$

where the kernel correlation matrix $\mathbf{C}$ is given by

$$C_{nm} \equiv \int K_n(z) K_m(z) dz. \quad (32)$$

Previous numerical results[2] involving the lowest eight eigenmodes found C_{nm} to be diagonally dominant for three of the

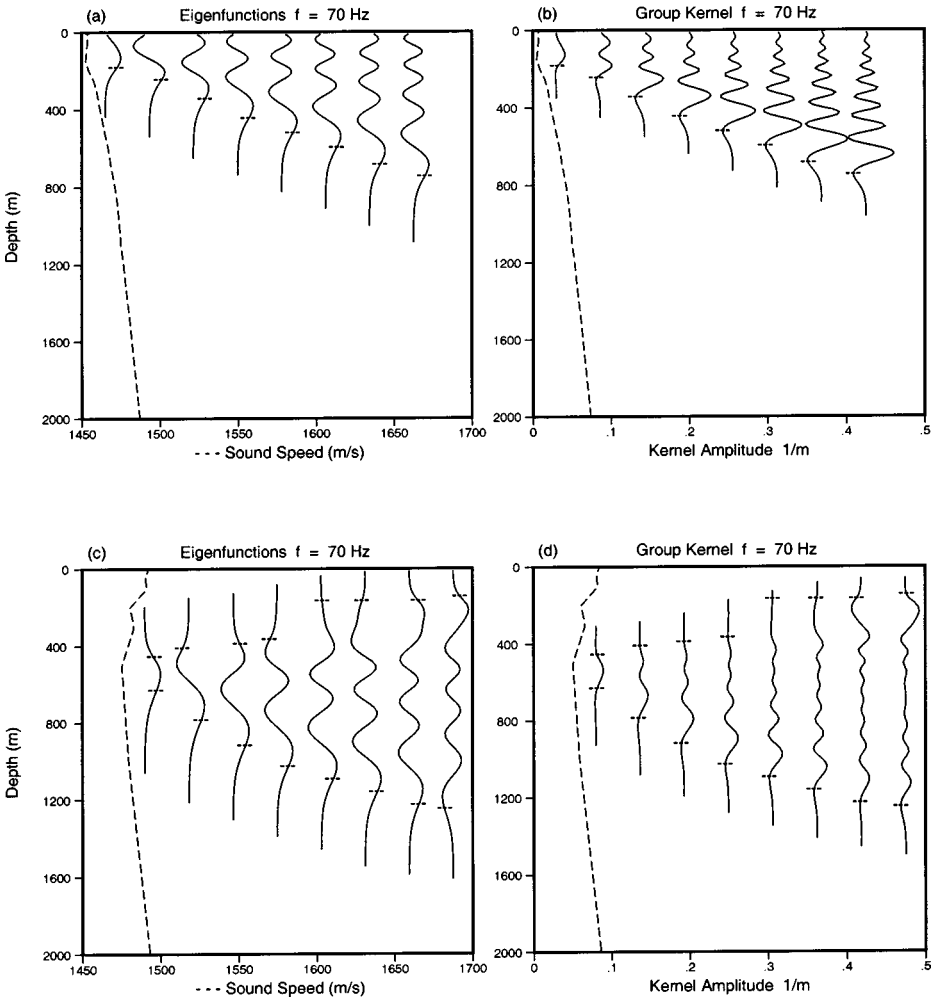


FIG. 2. (a) and (b) Results for a near-polar SSP with strong surface interaction. (c) and (d) Results for a double ducted SSP. Other details are as in Fig. 1.

environments to be considered below (the fourth was not considered in that work). These results suggest that $K_n(z)$ are an appropriate set of functions for least squares. The inversion of (28) is now summarized as

$$\delta s(z) \simeq \mathbf{K}(z)^T \mathbf{C}^{-1} \delta \mathbf{s}^{\mathbf{g}}. \tag{33}$$

Equations (29)–(33) give the approximate inversion of group slowness data to yield the slowness perturbation in the water column, and are consistent with a previously derived range-integrated formulation.[6]

V. NUMERICAL RESULTS

A. Group slowness kernels for various environments

Figure 1(a) shows the first eight eigenmodes at a frequency of 70 Hz for a ''canonical'' sound-speed profile (also known as a Munk profile) with parameters $z_o = 1000\,\mathrm{m}$, $B = 1300\,\mathrm{m}$, $\epsilon = 0.0113$, $\eta = (z - z_o)/(B/2)$, $c_o = 1480\,\mathrm{m/s}$:

$$c(z) = c_o[1 + \epsilon(\eta + e^{-\eta} - 1)], \tag{34}$$

where z is taken to be positive downward. Figure 1(b) gives the corresponding inversion kernels $K_n(z)$ as calculated from (29). The equivalent ray turning points, where the modal phase speed equals the local sound speed, are marked on the eigenfunctions and on the $K_n(z)$ for comparison. The $K_n(z)$ in Fig. 1(b) have dimensions of inverse meters, with the full width of the abscissa representing $0.5\ \mathrm{m}^{-1}$.

For comparison with the canonical profile results, we present results for three real ocean profiles adapted from Ref. 2 to conform to the style of Figs. 1(a) and (b). Figures 1(c) and (d) show results for a smoothly varying deep-water sound channel typical of January near Ascension Island,[22] where some HIFT data were taken. The two smooth deep-water sound-speed profiles of Figs. 1(a) and (c) result in quite similar eigenmodes and inversion kernels. For this reason we use the Munk profile in estimating the accuracy of the perturbative approach to ocean tomography in Sec. VI.

Figures 2(a) and (b) illustrate a sound speed typical of the near-polar oceans with strongly surface-interacting modes. The SSP is from Levitus[23] for a point near Heard Island (53S,74E), where the ocean is 1500 m deep. Below 1500 m we have extended the profile by assuming a hydrostatic sound-speed gradient. The modes shown are not bottom interacting, so this extension of the SSP does not affect the results. The amplitude of K_n for this surface-limited SSP is larger than either of the two deep ocean SSPs illustrated, indicative of high dispersion.

Figures 2(c) and (d) show an SSP with a double duct typical of a point (44S,41E) northwest of Crozet Island in January.[24] Modes 7 and 8 are partially trapped in the duct, and show increased group slowness sensitivity near the secondary duct.

For all of the results in this section, the $K_n(z)$ occupy

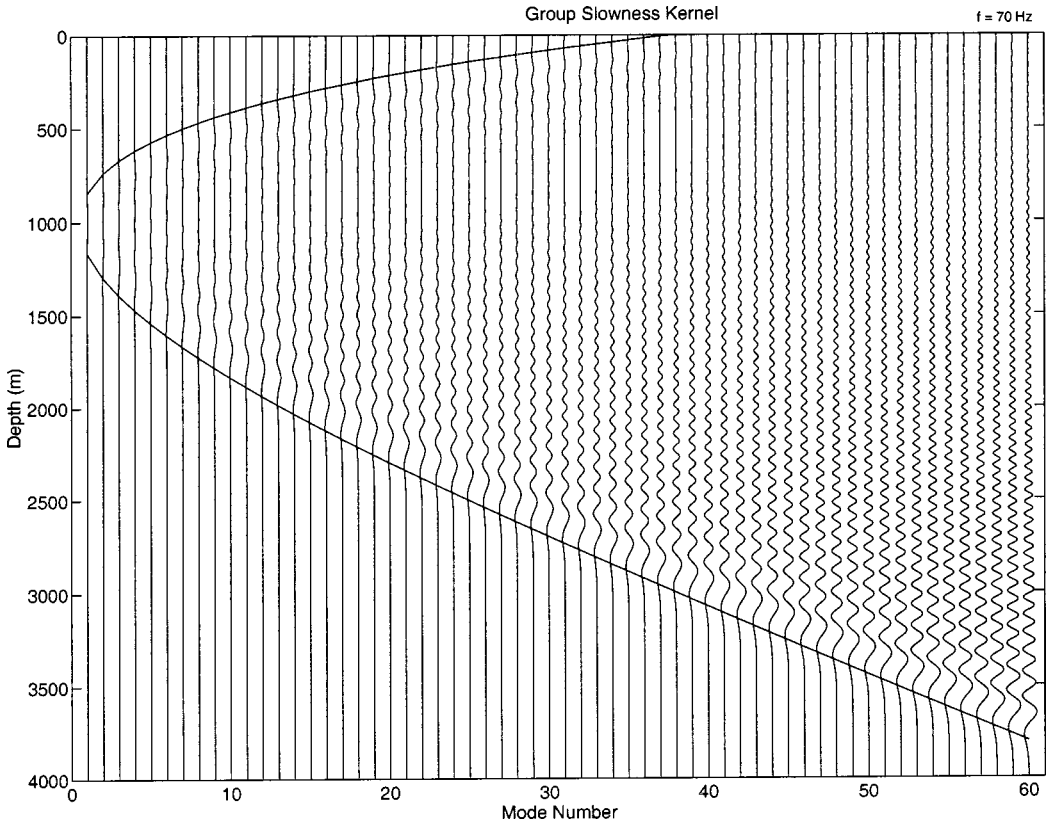


FIG. 3. Group slowness kernel K_n for the first 60 modes resulting from the Munk SSP of Eq. (34). Upper and lower turning depths for each mode are indicated by curved lines.

approximately the same vertical extent as the eigenfunctions $\psi_n(z)$, but oscillate roughly twice as many times, a result of the quadratic dependence of K_n upon ψ_n in (29). The group kernel has its first minimum (coming up from the bottom) just below the lower ray turning point for the mode under consideration (i.e., just below the lowest inflection point in ψ_n), and then has a large maximum just above the first extremum in ψ_n. Strong surface interaction and double ducting in Figs. 2(a) and (c) complicate the near- surface behavior of K_n in Figs. 2(b) and (d).

For smooth deep-water SSPs [Figs. 1(a) and (c)], modal group slowness sensitivity to water column slowness is greatest near outer inflection points (corresponding to ray turning points). The highly dispersive SSP of Figs. 2(a) and (b) shows about three times the sensitivity as the smooth deep-water case [cf. kernel amplitudes in Figs. 1(b) and (d) and 2(b)]. But again, maximum sensitivity occurs just inside the ray turning point locations. Double ducting [Figs. 2(c) and (d)] complicates the relation between water column perturbations and group slowness response. There is only weak sensitivity in the upper water column for modes 2–6 [Fig. 2(d)], but modal resonances with the secondary duct produce large responses of opposite sign for modes 7 and 8.

The results of Figs. 1 and 2 support the unsurprising conclusion that group slowness perturbations are most sensitive to sound-speed changes near turning points in a ray description, but with the added complication that the sign of the inversion kernel near the turning points may be positive or negative in a double-ducted SSP.

The formalism provided here gives information about the projection of slowness perturbations $\delta s(z)$ on the nonorthogonal oscillatory functions $K_n(z)$, whose behavior is illustrated in the right-hand panels of Figs. 1 and 2. For a midlatitude SSP such as that of Fig. 1(c), the response of the modal group slowness to slowness changes in the water column is weak, and peaks at upper and lower turning points. The more restricted polar SSP of Fig. 2(a) reveals increased sensitivity of the group slowness to $\delta s(z)$ due to the vertical confinement of the eigenmodes. The double-ducted SSP of Fig. 2(c) shows increased sensitivity in modes which are partially trapped in the near-surface ducts.

As one considers higher and higher mode numbers, one finds a transition from mode trapping near the sound axis to strong surface interaction. This is illustrated for the Munk SSP in Fig. 3, where the K_n for the first 60 modes are given. One sees a monotonic tendency of the group slowness kernel to become strong approaching the lower turning point, and weak near the surface. In a ray description this would correspond to a surface-interacting ray spending more time near the turning point than near the surface.

VI. LIMITATIONS ON THE LINEAR PERTURBATION METHOD

We now give some numerical examples in which perturbation theory results are accurate, followed by some in which the accuracy is degraded. The first example compares δs_n^g as determined from (28) with that found by solving Eqs. (1) and (14) for two different environments whose sound slowness differs by $\delta s(z)$. The eigenmodes were computed

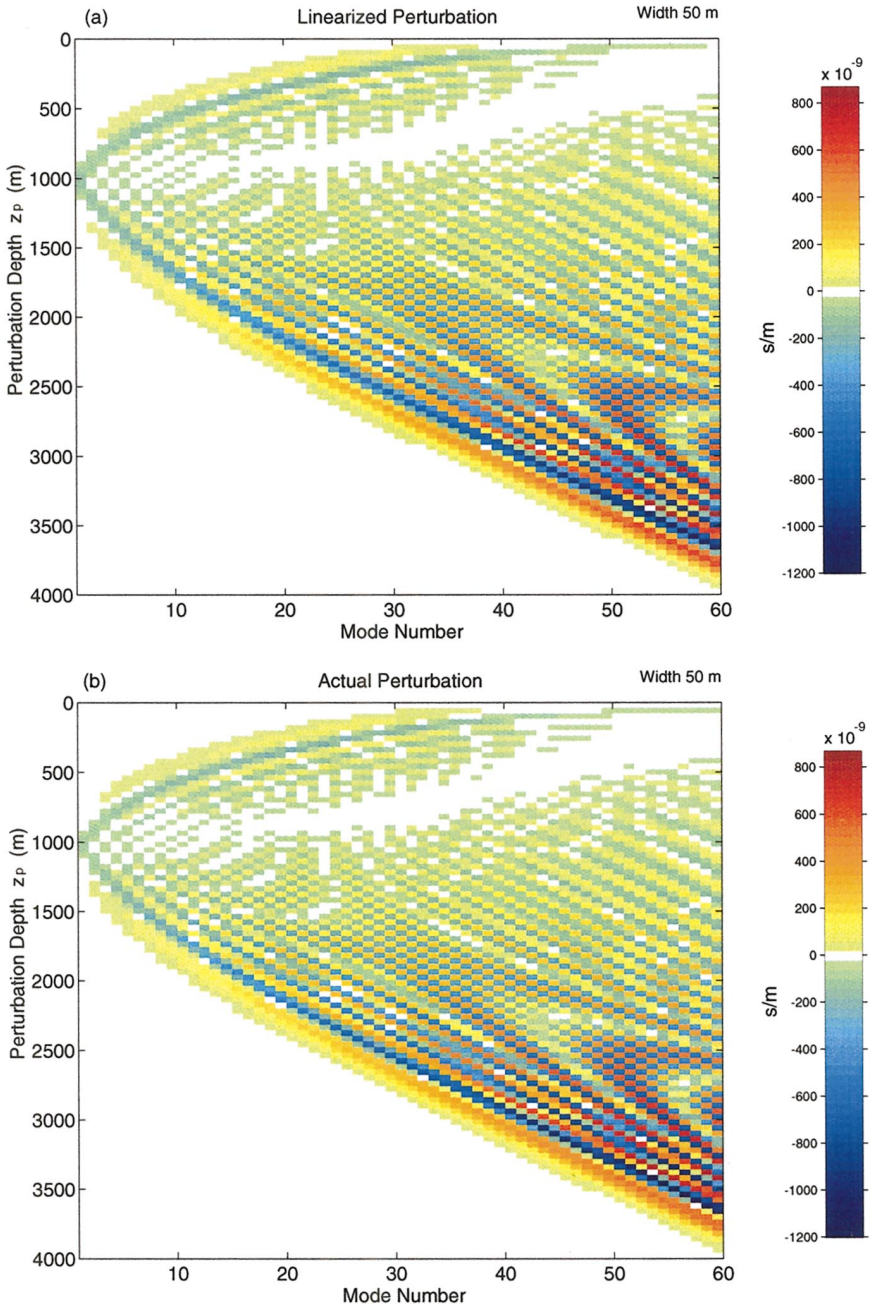


FIG. 4. (a) Linearized group slowness perturbation at 70 Hz from (28) for a 50-m Hanning window perturbation (35) of the Munk SSP; (b) Actual group slowness change calculated directly from the two different SSPs.

using KRAKEN.[25] The unperturbed case is taken to be that of the Munk SSP of Eq. (34). A Hanning window was used to create the perturbation signal with a given amplitude, extent and depth location

$$\delta s(z) = \begin{cases} \alpha \cos^2\left(\dfrac{\pi(z-z_p)}{\Delta z}\right), & |z-z_p| < \Delta z/2 \\ 0, & |z-z_p| > \Delta z/2. \end{cases} \tag{35}$$

Figure 4 shows δs_n^g for $\Delta z = 50\,\mathrm{m}$ and amplitude $\alpha = -1 \times 10^{-6}$ s/m, leading to a peak sound-speed perturbation of about 2.25 m/s. The 50-m width is smaller than most vertical correlation scales encountered in the deep ocean and is used mainly to illustrate a case of good qualitative agreement between the linear and exact methods. The linear perturbation result of Fig. 4(a) accurately captures the high-frequency structure seen in the higher modes. In fact, one can match the oscillations in Fig. 4(a) to those seen in Fig. 3. The smallest oscillation has a vertical wavelength of about 60 m, on par with the 50-m perturbation scale. The actual perturbation seen in Fig. 4(b) as computed from Eqs. (1) and (14) for the two different environments is almost indistinguishable from the linear perturbation result (28).

Next, we give an example in which the linearized and exact group slowness perturbations are considerably different. Figure 5 gives results for slowness perturbations in the water column as in Fig. 4, but with the Hanning width increased to 200 m. We show linear versus actual results δs_n^g for mode 9 in Fig. 5(a) and for mode 21 in Fig. 5(b). One sees that the linear result for these modes is off by approximately 30% to 50%, with the highest errors occurring when the perturbation is near the modal turning points. The error is

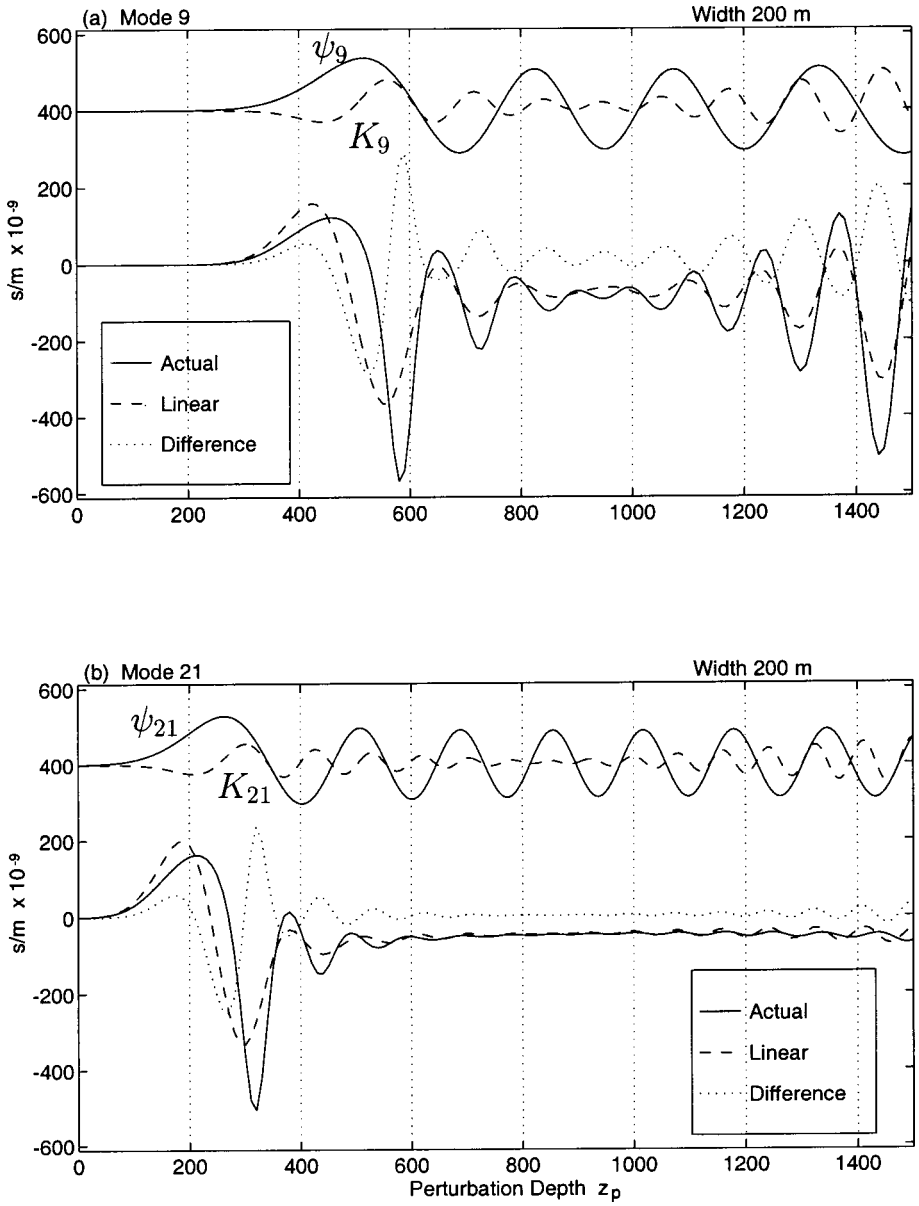


FIG. 5. Linear and actual group slowness perturbations for a 200-m-width Hanning perturbation pulse for modes 9 (a) and 21 (b) at 70 Hz. The upper two curves in each plot are the unperturbed mode shape $\psi_n(z)$ (solid line) and the group slowness kernel $K_n(z)$ (dashed). Below those are three lines, giving the actual δs_n^g (solid), linearized version (dashed) and their difference (dotted).

oscillatory, with a vertical wavelength comparable to that of the group slowness kernel.

If a perturbation width of 50 m gives good agreement between linear and exact results, we must ask why an increase to 200 m produces such poor agreement. After all, one could approximate the 200-m width window by a linear superposition of 50-m windows properly spaced. The answer is that δs_n^g is not linear with respect to changes in the environment, so that a linear approximation must begin to fail somewhere. By enlarging the width from 50 to 200 m we have perturbed more of the water column, even if the pointwise changes in the water column are of the same magnitude.

To illustrate the onset of nonlinear dependence of δs_n^g on perturbation amplitude and width, we give in Fig. 6(a) the difference between linear and actual δs_n^g for mode 9 as a function of both these parameters, with the depth of the perturbation kept constant at 250 m. This depth is above the upper turning point of mode 9 by approximately 250 m. One sees that as the perturbation amplitude increases from zero, the error in the linear expression (28) develops a strong maximum for $\Delta z = 500\,\mathrm{m}$ and $\alpha = 3\times10^{-6}$ s/m. Then, for $\Delta > 1000$ m, the error is negative for both positive and negative α. A contour is drawn where the amplitude of the group slowness error is 10^{-8} s/m, corresponding to travel time accuracy of 10 ms per 1000 km. This level is a nominal figure for accurate ocean basin scale acoustic tomography. For mode 9 at the stated perturbation depth of 250 m, the slowness perturbation amplitude needs to be smaller than approximately 10^{-6} s/m to achieve this nominal level of accuracy for the linear expression (28).

Figure 6(b) repeats the calculation of Fig. 6(a) for mode 21. Here, the error in the linear expression is always negative, and the amplitude limit for the nominal accuracy level is even more restrictive on α for perturbation widths $\Delta z < 800\,\mathrm{m}$. For $\Delta z > 800\,\mathrm{m}$, the nominal amplitude limit $\alpha < 10^{-6}$ s/m still applies for the linear method to give group slowness accuracy within 10^{-8} s/m. The consistent negative bias for the linear method at large vertical scale lengths Δz implies, at least for the modes under consideration, that the

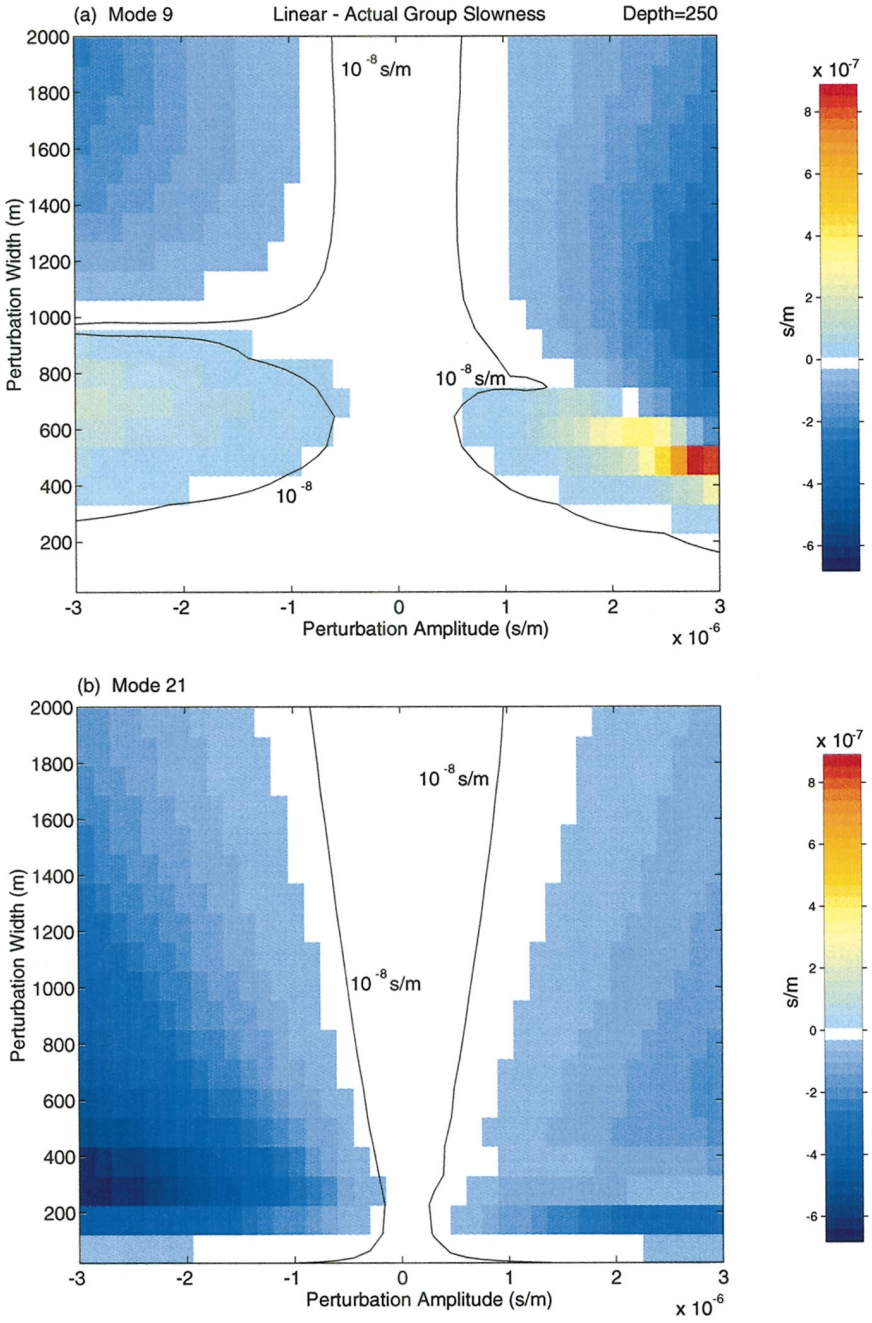


FIG. 6. Error in the linear result for group slowness perturbation at 70 Hz for modes 9 (top) and 21 (bottom) as a function of perturbation width Δz and amplitude α. The perturbation to the sound slowness profile is centered at a depth of 250 m. The black contour shows the region in which the error is smaller than the nominal value 10^{-8} s/m, considered acceptable for ocean basin scale tomography.

linear method may give biased group slowness perturbations even for a horizontally zero-mean field of water column perturbations.

VII. SUMMARY

We have presented perturbation theory for modal group slowness within the adiabatic approximation. The form of the result depends on the order in which one does frequency derivative and sound-speed perturbation. We showed that the results are equivalent by showing that (16) emerges as an inherent property of Sturm–Liouville systems such as Eq. (1) for ocean acoustic eigenmodes. This result emphasizes a fundamental but not obvious role played by the eigenmode's frequency derivative in relating a water column perturbation to the eigenmode's response. We have given a direct method for evaluating the frequency derivative of the eigenmode which may be a useful alternative to finite differences.

We have expressed modal group slowness perturbations as an integral equation (28) in $\delta s(z)$, the water column sound slowness perturbation. Equations (29)–(33) give an approximate least-squares inversion for $\delta s(z)$ in terms of the modal group slowness perturbations.

The kernel of the integral equation has been given numerically in Figs. 1 and 2 for four sound-speed profiles: The Munk SSP, a smooth midlatitude SSP, a mostly upward-refracting high-latitude SSP with strong surface interaction, and a double-ducted high-latitude SSP. The first two SSPs show similar response to water column perturbations, but the last two show increased sensitivity at the lower turning point for the high-latitude SSP, and in the surface ducts for the last SSP considered.

We then examined the accuracy of the linear perturbation method and found decreased accuracy at large perturbation amplitude ($|\delta s(z)| \gtrsim 10^{-6}$ s/m for which $|\delta c(z)| \gtrsim 2.25$ m/s) *and* at large vertical perturbation width ($\Delta z \gtrsim 50$ m) for the modes considered. We also found that for large perturbation width the linear perturbation method gave biased results (predominantly negative error in Fig. 6 regardless of the sign of the water column perturbation). It appears from other work (not presented here) that the sign of the bias may depend on whether the water column perturbation is inside or outside the modal turning points.

ACKNOWLEDGMENTS

Work supported by the Office of Naval Research. Part of the work[3] was performed while one of the authors (B.E.M.) was at Saclantcen.

[1] B. E. McDonald and A. B. Baggeroer, "Modal eigenfunction perturbations and group speed tomography," J. Acoust. Soc. Am. **94**, 1794 (1993).
[2] B. E. McDonald, "The inversion kernel in group travel time tomography," in *Full Field Inversion Methods in Ocean and Seismo Acoustics, Proceedings NATO Conference, Saclant Center, La Spezia Italy, 27 June–1 July 1994*, edited by O. Diachok, A. Caiti, P. Gerstoft, and H. Schmidt (Kluwer Academic, Dordrecht, 1995).
[3] B. E. McDonald, B. Sperry, and Arthur B. Baggeroer, "Theoretical and Numerical Issues in Travel Time Tomography," Saclantcen Report SR-333 (US postal address: Saclantcen Library, APO AE 09613), October, 2000.
[4] W. H. Munk and C. Wunsch, "Ocean acoustic tomography: A scheme for large-scale monitoring," Deep-Sea Res., Part A **26**, 123–161 (1979).
[5] W. Munk, P. Worcester, and C. Wunsch, *Ocean Acoustic Tomography* (Cambridge University Press, London, 1995).
[6] E. C. Shang, "Ocean acoustic tomography based on adiabatic mode theory," J. Acoust. Soc. Am. **85**, 1531–1537 (1989).
[7] E. C. Shang and Y. Y. Wang, "On the possibility of monitoring El Niño by using modal ocean acoustic tomography," J. Acoust. Soc. Am. **91**, 136–140 (1992).
[8] J. F. Lynch, S. D. Rajan, and G. V. Frisk, "A comparison of broadband and narrow-band modal inversions for bottom geoacoustic properties at a site near Corpus Christi, Texas," J. Acoust. Soc. Am. **89**, 648–665 (1991).
[9] P. Sutton, W. Morawitz, B. D. Cornuelle, G. Masters, and P. Worcester, "Incorporation of acoustic normal mode data into tomographic inversions in the Greenland Sea," J. Geophys. Res. **99**, 12487–12502 (1994).
[10] A. B. Baggeroer, J. Miller, C. S. Chiu, G. Froger, P. N. Mikhalevsky, and K. von der Heydt, "Vertical array resolution of the normal modes from the Heard Island signals," J. Acoust. Soc. Am. **90**, 2330 (1991).
[11] W. H. Munk, "Acoustic thermometry of ocean climate," J. Acoust. Soc. Am. **100**, 2580 (1996).
[12] B. M. Howe, J. A. Mercer, R. C. Spindel, P. F. Worcester, J. A. Hildebrand, W. S. Hodgkiss, Jr., T. F. Duda, and S. M. Flattè, "SLICE89: A single slice tomography experiment," in *Ocean Variability and Propagation*, edited by J. Potter and A. Warn-Varnas (Kluwer, Dordrecht, 1991), pp. 81–86.
[13] B. E. McDonald, M. D. Collins, W. A. Kuperman, and K. D. Heaney, "Comparison of data and model predictions for Heard Island acoustic transmissions," J. Acoust. Soc. Am. **96**, 2357–2370 (1994).
[14] K. D. Heaney, J. Colosi, and W. A. Kuperman, "Estimation of internal wave strength from mode space source localization," J. Acoust. Soc. Am. **100**, 2582 (1996).
[15] S. Rajan, J. F. Lynch, and G. V. Frisk, "Perturbative inversion methods for obtaining bottom geoacoustic parameters in shallow water," J. Acoust. Soc. Am. **82**, 998–1017 (1987).
[16] E. C. Shang and Y. Y. Wang, "On the calculation of modal travel time perturbation," Sov. Phys. Acoust. **37**, 411–413 (1991).
[17] W. L. Rodi, P. Glover, T. M. C. Li, and S. S. Alexander, "A fast, accurate method for computing group velocity partial derivatives for Rayleigh and Love modes," Bull. Seismol. Soc. Am. **65**, 1105–1114 (1975).
[18] C. T. H. Baker, *The Numerical Treatment of Integral Equations* (Clarendon, Oxford, 1977).
[19] A. D. Pierce, *Acoustics* (Acoustic Society of America, Woodbury, NY, 1989), Sec. 4.9.
[20] K. D. LePage, "Acoustic time series variability and time reversal mirror defocusing due to cumulative effects of water column variability," J. Comput. Acoust. **9**, 1455–1474 (2001).
[21] K. D. LePage, Personal communication (2000).
[22] D. Palmer, Personal communication (1993).
[23] S. Levitus *Climatological Atlas of the World Ocean*, NOAA Prof. paper 13 (U.S. Government Printing Office, Washington DC, 1982).
[24] *World Ocean Atlas*, edited by S. Gorshov (Pergamon, New York, 1979).
[25] M. B. Porter and E. L. Reiss, "A numerical method for ocean acoustic normal modes," J. Acoust. Soc. Am. **76**, 244–252 (1984).

Seabed reflection measurement uncertainty

Charles W. Holland[a)]
Applied Research Laboratory, The Pennsylvania State University, State College, Pennsylvania 16804

(Received 31 December 2002; accepted for publication 14 July 2003)

The seabed reflection coefficient is a fundamental property of the ocean waveguide. Measurements of the frequency and angular dependence of the reflection coefficient can provide information about the geoacoustic properties of the seabed or can be used as an input to propagation models. The uncertainty of the measurements must be known in order to determine prediction uncertainties for the acoustic field and/or the geoacoustic properties. Analysis indicates that the reflection measurements have a standard deviation from ± 0.5–1 dB at full angular resolution depending on frequency and experiment geometry. The dominant contribution to the error is source amplitude variability, and a new processing approach was developed that reduces the error for frequencies above a few hundred Hz. Further reduction in the uncertainty can be obtained by averaging in angle, for example, a $\pm 1°$ angle averaging leads to a standard deviation of less than ± 0.5 dB. Errors in the angle estimate are a few tenths of a degree from 0–34° grazing angle: the crucial angular range for predicting long-range propagation or for geoacoustic property inversion. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1605388]

PACS numbers: 43.30.Ma, 43.20.El, 43.30.Xm [WLS]

I. INTRODUCTION

Acoustic interaction with the seafloor often dominates and controls propagation and reverberation in continental shelf environments (e.g., Urick, 1970; Jensen and Kuperman, 1983; Eller and Gershfeld, 1985). Despite the availability of high-fidelity models (e.g., Porter, 1991; Collins, 1993; Weinberg and Keenan, 1996), acoustic predictions may have large error bars because the seafloor geoacoustic data required to drive the models have large uncertainties (e.g., Ferla and Jensen, 2002). The impact of seabed variability on acoustic predictions can be studied by analyzing fundamental acoustic measures that control seabed interaction: the seabed reflection coefficient and scattering strength. These two quantities, in principle, permit study of the seafloor independent of oceanographic variability. In this study, uncertainties associated with the reflection coefficient are examined.

Many existing techniques to estimate shallow-water seabed properties, or equivalently the seabed reflection coefficient, measure long-range propagation (e.g., Rubano, 1980; Frisk and Lynch, 1984; Beebe and Holland, 1986; Collins *et al.*, 1992; Chapman *et al.*, 2001). These techniques and similar generally spatially average over km to tens of km. A recently developed local measurement technique for reflection (Holland and Osler, 2000) averages over a much smaller footprint, ~100 m, permitting the spatial variability in the vertical and horizontal to be probed at much higher resolution. In addition, problems of intermingled geoacoustic variability with spatial-temporal oceanographic variability in long-range measurements (e.g., Siderius *et al.*, 2001) are greatly diminished because of the short distances and the short time interval over which the local measurements occur.

The reflection measurements can be used as ''ground truth'' for long-range methods or to extract sediment geoacoustic properties, such as sound speed, attenuation, and density. In both cases, errors associated with the reflection data are required. The objective of this research is to quantify the uncertainty associated with the direct path reflection technique.

Uncertainty or error analysis is important for several reasons. First, it is necessary in order to generally interpret the results. For example, if the data are employed to estimate seabed geoacoustic parameters, the uncertainty of the reflection data must be known in order to predict the uncertainty in the resulting geoacoustics. Second, a careful error analysis can sometimes (as it does here) point to ways in which the error can be reduced.

In any experiment the total error is a combination of the random or precision error and the systematic or bias error. The approach to error estimation taken here is two-pronged. First, the error is treated analytically and the uncertainty is derived where possible from known and estimated distributions of the measurement error. This analysis gives the random or precision error. Second, the measurement error is analyzed experimentally, i.e., by comparing multiple measurements made at nearly the same location. These two complementary approaches give a consistent picture of the uncertainty in the reflection measurements and roughly correspond to Secs. III and IV, respectively.

II. THE MEASUREMENT TECHNIQUE

The measurement technique is reviewed first, followed by pertinent details regarding the data processing. Measurements have been conducted at a wide variety of seabed types in the Mediterranean Sea and the North Atlantic Ocean. Measurements referred to in this paper were conducted on the North Tuscany Shelf, NTS, and the Malta Plateau, MP (see Fig. 1).

a)Electronic mail: holland-cw@psu.edu

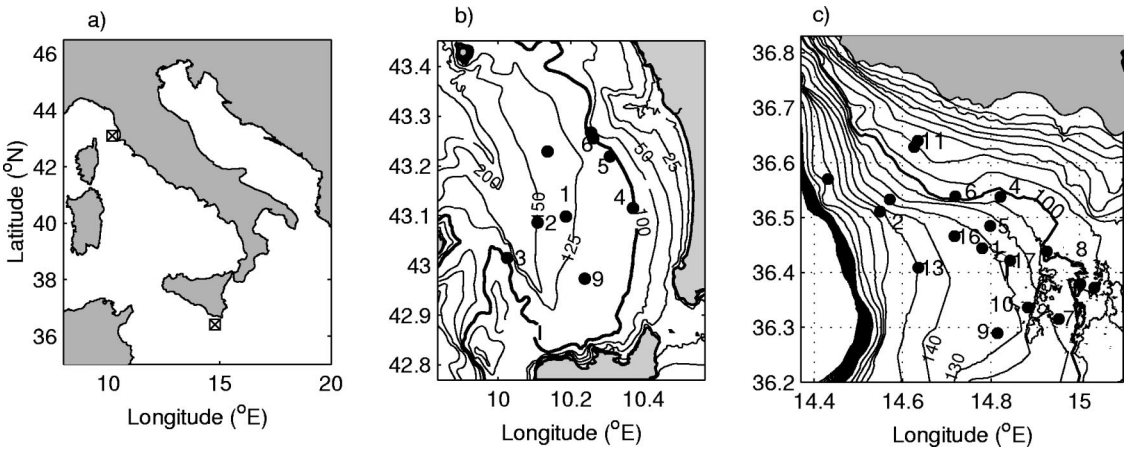


FIG. 1. Location of measurements employed to examine uncertainty in the reflection loss measurements (a) Italian littoral region, the two sites are marked by ⊗; (b) North Tuscany shelf; and (c) Malta Plateau.

A. Measurement geometry

The measurement geometry was specifically designed for shallow water, but could also be used in deep water (see Fig. 2). A broadband source is towed close to the surface, typically at a depth of a few tens of cm, and a fixed single receiver is placed far enough away from the seabed to prevent interference from the direct- and bottom-reflected paths. One of the challenges in making single bounce reflection measurements in shallow water is the presence of multipath. Towing the source close to the surface has the advantage that the surface-reflected multipaths arrive very close in time and angle and thus do not degrade the ability to associate time, range, and angle. At least two disadvantages of a near-surface source are that the surface is always changing, and that small changes in source depth can have relatively large impact on source spectrum.

Two geophysical sources have been used in the experiments, an EG&G model 265 Uniboom prior to 1999 and a GeoAcoustics model 5813B Geopulse boomer after 1999 (see Table I). The sources are metal plates driven with a high-energy impulsive signal resulting in a beam pattern similar to piston source. Pulse repetition rate was generally 1 pulse per second. Both source trigger and data acquisition were controlled by the same GPS clock to eliminate synchronization problems. The tow speed is typically about 4 knots.

The receiver and acquisition configuration has also varied over the experiments (see Table I). Prior to May 1999, a 12-bit A/D converter was used that required manual gain changes as a function of source–receiver offset in order to avoid clipping of the direct path. Subsequently, a 20-bit A/D converter was employed, eliminating the need for gain changes. The data in all experiments were telemetered to the R/V ALLIANCE via radio link. Other details about the experiment geometry can be found in Holland and Osler (2000).

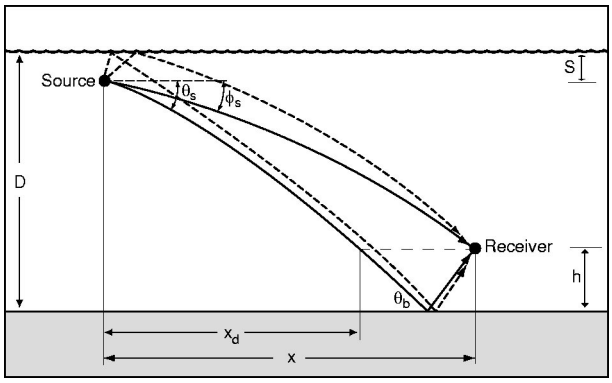


FIG. 2. Measurement geometry showing the direct and bottom bounce path ray paths along with the associated surface reflections.

B. Reflection processing

The magnitude of the pressure reflection coefficient can be defined as

$$|R(\theta_b,f)| = \left|\frac{p_r(x,f)}{p_o(x,f)}\right|, \tag{1}$$

where p_r is the received pressure from the bottom-reflected path at range x and p_o is the received level of a bottom-reflected path at x for a perfectly reflecting half-space (see Fig. 2 for measurement geometry). The denominator is written as

$$|p_o(x,f)| = |p_s(\theta_s,f)|\,\gamma_o, \tag{2}$$

where p_s is the source pressure amplitude at 1 m from the source, γ_o is the transmission factor from source to receiver along the bottom-reflecting path including the effects of spreading, refraction, and absorption, multiple paths (i.e., surface path), and bottom reflection (which for a perfectly reflecting bottom is unity). If the source were omnidirectional, p_s could simply be obtained measuring the pressure amplitude of the direct path at ϕ_s and correcting it back to the source. However, since the source is directional, the direct path amplitude at the same angle must come from a different range, x_d

$$|p_s(\theta_s,f)| = q_d(x_d,f)\,\gamma_d^{-1}, \tag{3}$$

where q_d is the amplitude of the direct path at launch angle θ_s and height h. The finite range interval of the measurements (approximately 2 m) means that in practice q_d is interpolated rather than measured directly. γ_d is the transmission factor from source to receiver along the direct path

TABLE I. List of reflection experiments with source and receiver characteristics. NTS is the northern Tuscany shelf; MP is the Malta Plateau; NJS is the New Jersey Shelf and SS in the Scotian Shelf. *El* is number of elements in the receiver string and *fs* is sampling frequency.

Cruise	Location	Date	Source	Receiver
SCARAB97	NTS	June 97	EG&G	16 *el*, 12 bit, $fs=24$ kHz
SCARAB98	MP	April 98	EG&G	16 *el*, 12 bit, $fs=24$ kHz
Boomer99	NTS	January 99	GeoAcoustics	4 *el*, 12 bit, $fs=48$ kHz
Malta99	MP	May 99	GeoAcoustics	4 *el*, 20 bit, $fs=48$ kHz
Geoscat99	NTS	October 99	GeoAcoustics	4 *el*, 20 bit, $fs=48$ kHz
Boundary00	MP	April 00	GeoAcoustics	4 *el*, 20 bit, $fs=48$ kHz
GeoClutter	NJS	April 01	GeoAcoustics	4 *el*, 20 bit, $fs=48$ kHz
Boundary01	NJS, SS	May 01	GeoAcoustics	4 *el*, 20 bit, $fs=48$ kHz
Boundary02	MP	May 02	GeoAcoustics	4 *el*, 20 bit, $fs=48$ kHz

including the effects of spreading, refraction, absorption, and multipath.

Thus, the expression for the reflection coefficient is

$$|R(\theta_b,f)|=\frac{|p_r(x,f)|}{q_d(x_d,f)}\frac{\gamma_d}{\gamma_o}, \tag{4}$$

where the first factor is a ratio of measured and interpolated pressures and the second factor is a ratio of modeled transmission factors. It will be convenient to refer to these two ratios in the following analysis as the ''pressure ratio'' and ''transmission ratio,'' respectively.

III. UNCERTAINTY ANALYSIS

The propagation of errors in the data processing and in the geometric factors that produce uncertainty in angle are computed using a Taylor series expansion around the mean values (Bevington and Robinson, 1992).

A. Uncertainties associated with transmission factor

The transmission ratio of Eq. (4) has uncertainties associated with the modeling of the transmission factors for the direct- and the bottom-reflected paths, which are a function of the experiment geometry. The uncertainties in the experiment geometry come from uncertainties in source depth, receiver depth, water depth, and range. We assume that errors in the model itself (Westwood, 1987) are small with respect to the errors in the geometry and thus can be neglected.

In the data processing, the transmission ratio is computed using the measured ocean sound profile and attenuation. However, in order to estimate errors, the analysis is simplified using a lossless isovelocity profile, where the transmission ratio can be written independent of range (or angle) as

$$\frac{\gamma_d}{\gamma_o}=(D-S-h)^{-1}\frac{[(D-S+h)^2+x^2]^{1/2}}{[1+x^2(D-S+h)^{-2}]^{1/2}}=\frac{D-S+h}{D-S-h}, \tag{5}$$

where D is water depth, S is source depth, and h is receiver height above the seafloor (see Fig. 2). The attenuation is ignored for the error analysis because its contribution is small (less than 3%) even at the highest frequency, 10 kHz, and longest range, 1 km.

Water depth is measured acoustically using a 12-kHz fathometer, a swath mapping system, and/or the near-normal

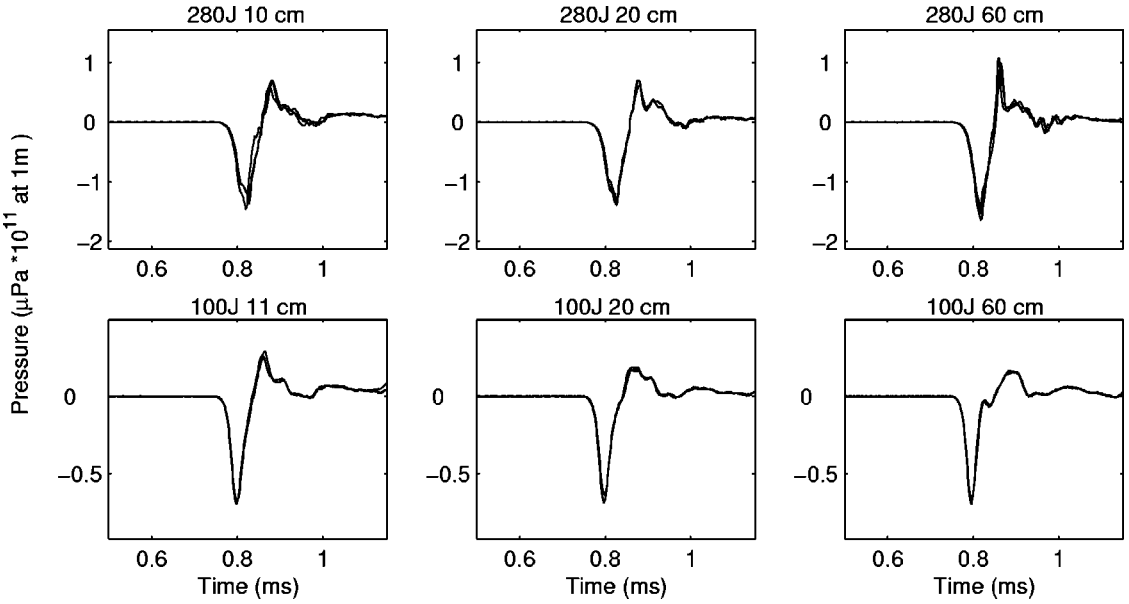


FIG. 3. Tank measurements of GeoAcoustics boomer at various energy levels (280 and 100 joules) and source depths at normal incidence. Each panel has three pulses.

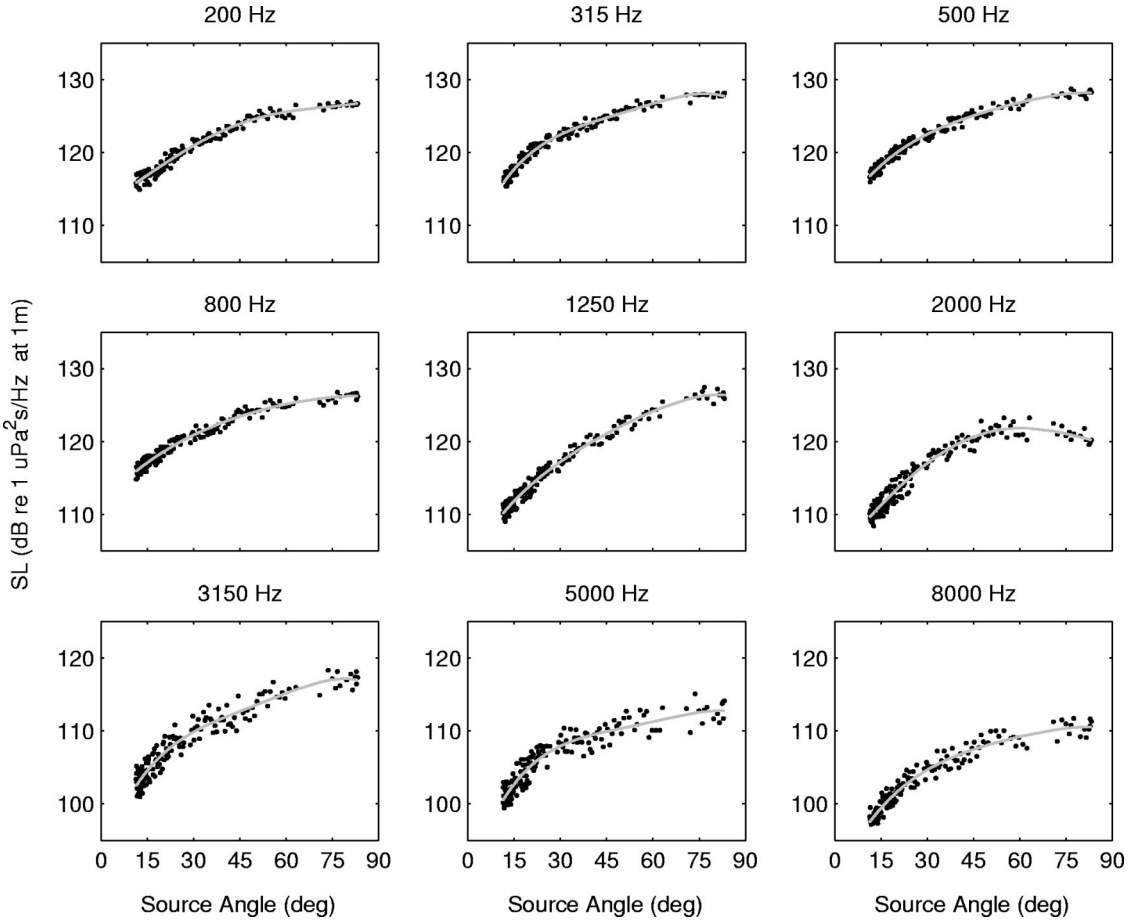


FIG. 4. Energy spectral density source level measurements (●) with polynomial fits (gray line) for various 1/3-octave frequency bands at MP site 7.

incidence bottom reflection on an array towed near the source. This estimate has a standard deviation of order a few tens of cm. The greater source of error for D occurs because the source is mounted on a surface-towed catamaran, so that the effective water depth varies with the passing waves, which may be up to of order 1 meter. The variability associated with S, which has a mean depth of 0.35 m, is associated with the nonconstant drag forces on the catamaran including small changes in instantaneous ship speed, wake turbulence, unsteady surface currents, and surface waves. It is believed that instantaneous changes in ship speed and wake effects dominate the variability in S. Receiver height is measured onboard before deployment; the length of the Kevlar rope attached to the sea anchor fixes its height. Receiver height errors occur because of rope stretching (presumed very small for Kevlar), displacement due to currents, and sinking of the sea anchor into the seabed.

Assuming that the errors in water depth, source depth, and receiver height are independent (which seems reasonable given that the processes that govern them are independent), the standard deviation, σ_t, associated with the transmission ratio [Eq. (5)] can be written

$$\sigma_t = ((\sigma_D^2 + \sigma_S^2)(1-\mu_t)^2 + \sigma_h^2(1+\mu_t)^2)^{1/2} \times (\mu_D - \mu_S - \mu_h)^{-1}, \quad (6)$$

where μ_t is the mean transmission factor. For reasonable mean and standard deviations ($\mu_D = 120$ m, $\mu_S = 0.35$ m, $\mu_h = 15$ m, $\sigma_D = 1$ m, $\sigma_S = 0.25$ m, $\sigma_h = 0.3$ m) the resulting relative errors are less than 1%, or 0.05-dB absolute error.

B. Uncertainties associated with the source level

The pressure ratio of (4) has no uncertainty associated with the calibration, since the bottom-reflected signal and the direct path signal are influenced by the hydrophone calibration and receiving electronics in precisely the same way.

However, the pressure ratio has uncertainty associated with the source amplitude and is the major contribution to the error budget. The source amplitude variability comes from two factors: (1) inherent variability in the drive voltage and the source plate response, and (2) nonconstant drag forces on the catamaran (caused by small changes in ship speed/direction, wake turbulence, and passing waves), which result in variability of the source plate depth and angle from ping to ping. Tank measurements of the source show it to be highly repeatable (see Fig. 3). Therefore, the largest contribution to the error is source variability due to motion. While small variations in source plate depth (a few tens of cm) played a nearly insignificant role for the transmission factor, they can lead to significant variations in amplitude (a few dB) both because of the sensitivity of plate response to hydrostatic pressure and sensitivity of the received pressure to the sea surface multipath.

Figure 4 shows 1/3-octave averaged source measurements at various frequencies along with the fitted polynomi-

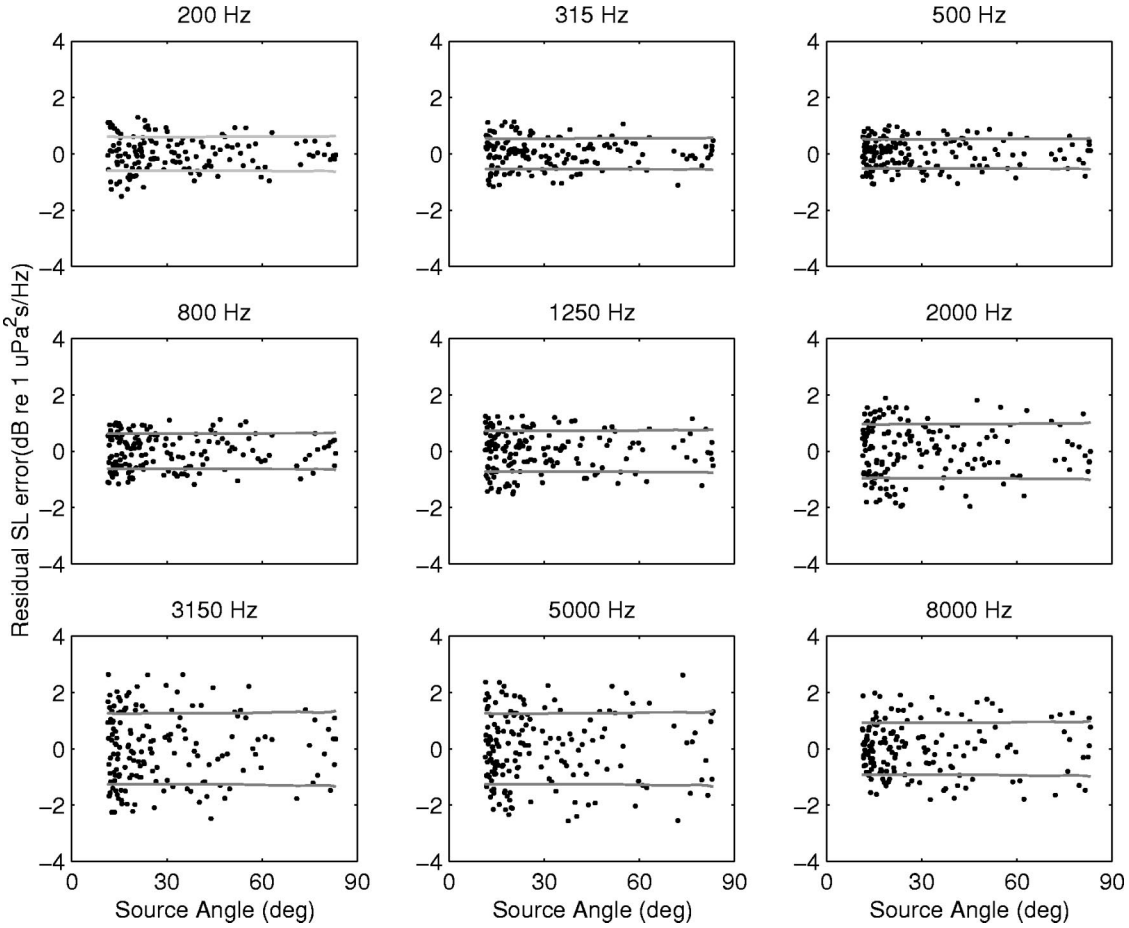


FIG. 5. Residual error (●) associated with polynomial fits of Fig. 3 with standard deviation (gray line) at MP site 7.

als. The data have been screened for signal-to-noise ratios greater than 6 dB, and outliers (greater than 2σ) have been removed. The residual errors of the polynomial fits are shown in Fig. 5. Note that the errors appear to be random, which means that the polynomial fit is appropriate. The distribution of the errors is approximately Gaussian. The standard deviations of the fits are relatively independent of angle but are dependent upon frequency ranging from a minimum of 0.5 dB at low frequencies to a maximum of 1.4 dB at 5000 Hz (see Fig. 6). Although these are reasonably typical, the frequency dependence of the standard deviation varies from run to run. As a general rule, the standard deviation is proportional to sea state: the higher the sea state, the larger the deviation. Measurements are generally conducted in sea state 3 or less. Assuming that the variability at ranges x and x_d are uncorrelated, the standard deviation in the reflection coefficient ranges from about 1–2 dB, depending on frequency.

C. Source variability normalization

While the errors associated with the data processed by Eq. (4) are reasonably small, error analysis suggested a novel way to reduce the error associated with source variability. If the source variability is predominantly due to amplitude fluctuations rather than beam-pattern fluctuations (e.g., caused by source plate tilt), then the variability can be reduced by normalizing by the same ping for the direct path as the bottom-reflected path and correcting by the ratios of the fitted average

$$|p_s(\theta_s,f)| = |p_d(x,f)| \frac{q_d(x_d,f)}{q_d(x,f)} \gamma_d^{-1}. \tag{7}$$

Now the reflection coefficient becomes

$$|R(\theta_b,f)| = \left|\frac{p_r(x,f)}{p_d(x,f)}\right| \frac{q_d(x,f)}{q_d(x_d,f)} \frac{\gamma_d}{\gamma_o}, \tag{8}$$

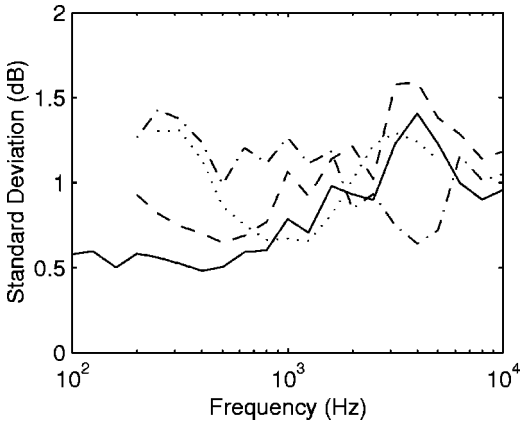


FIG. 6. Standard deviation of source level fits: MP site 7 (solid), an example from a single run in 1998 (dotted), 2000 (dashed), 2002 (chain dashed).

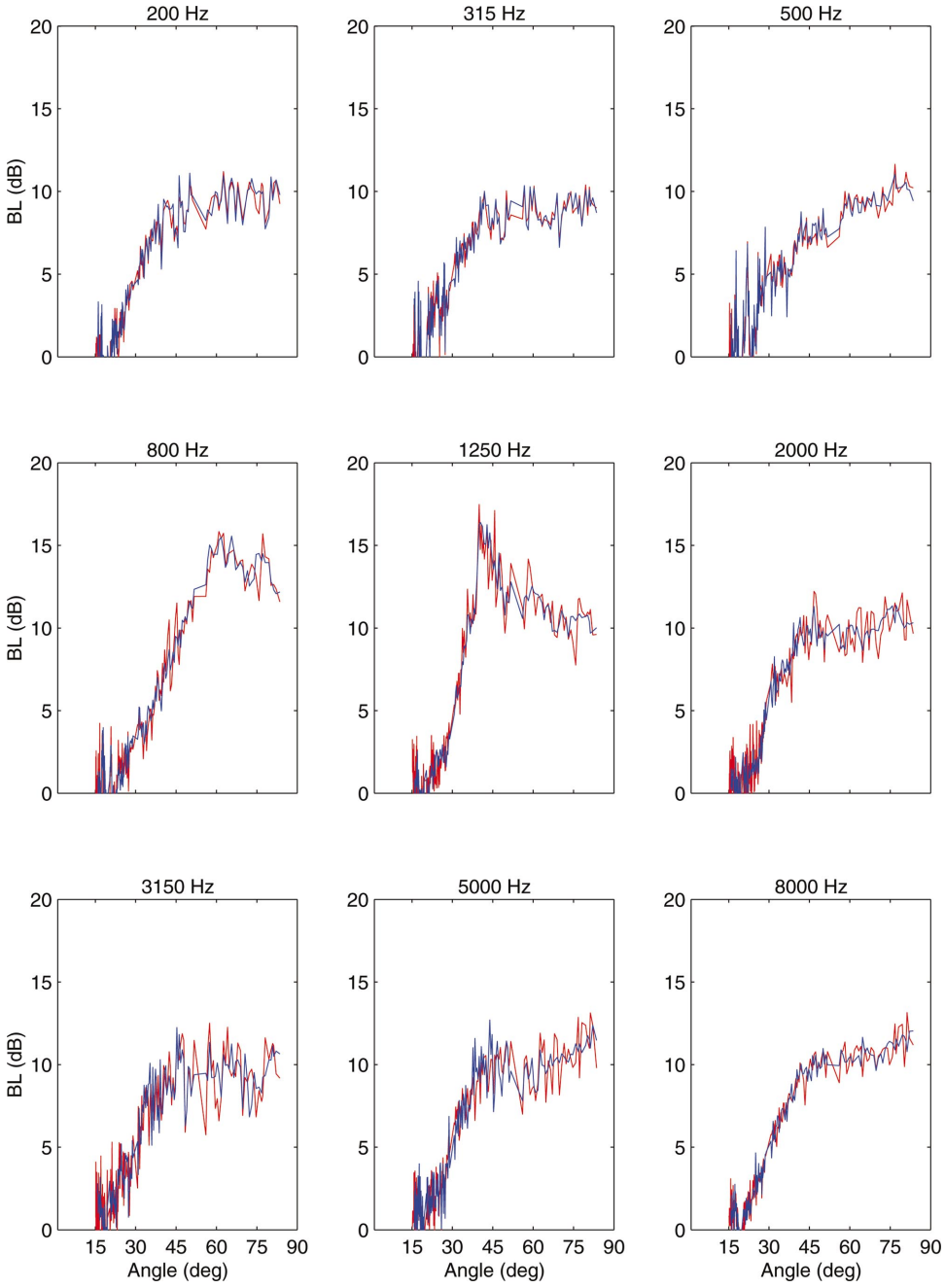


FIG. 7. Results of processing with Eq. (4) (red) and Eq. (8) (blue) at MP site 7; note the decrease in the fluctuations, especially above 800 Hz. The source level data are from Fig. 3.

where the first factor is a ratio of measured pressures for the same ping, so that ping-to-ping amplitude variability is completely normalized. In this form, the variability of the reflection coefficient only depends upon the variability in the fitted source level and thus should be about 0.5–1.5 dB (see Fig. 6).

The form of Eq. (8) requires no more computational load, and provides a fully normalized reflection coefficient. If there is significant ping-to-ping rotation of the source plate, then the errors associated with (8) actually are larger than (4). However, for most of the measurements conducted to date, (8) appears to reduce the error budget. As an example, Fig. 7 shows reflection data processed using Eq. (4) and Eq. (8) at MP site 7 (same site as Fig. 4). The fluctuations do appear to be smaller for (8). In order to obtain the quantitative difference in the standard deviation, the exact reflection coefficient must be known. That is, the apparent fluctuations in the reflection coefficient could be real, for example due to resonant interaction between layers, or at the higher frequencies, due to scattering. In order to estimate the change of (8), we make the assumption that the exact reflection coefficient at this site is perfectly smooth. Then, the standard deviations show that while some of the lower frequencies are degraded slightly, the midfrequency deviations are reduced (see Fig. 8). These standard deviations are an upper bound, given the foregoing assumption.

D. Angle uncertainty and resolution

The first step in determining the angle uncertainty is to determine the uncertainty in range. In the data processing, range is estimated by fitting modeled (Westwood, 1987) to

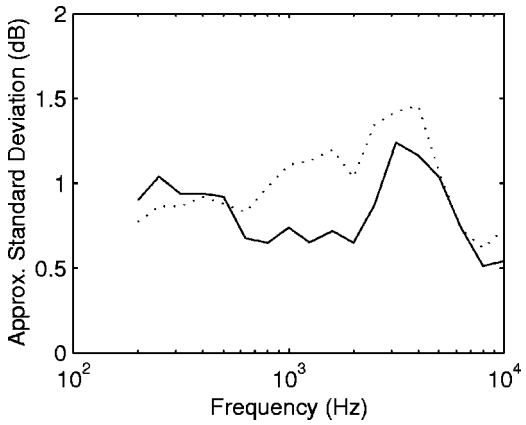


FIG. 8. Effect of source amplitude normalization; Eq. (4) (dotted) and Eq. (8) (solid). Bottom angles from 25–90° are included in the statistics.

measured arrival times, which are obtained by amplitude thresholding the direct path arrival. Although the measured sound-speed profile is used in the data processing, an isovelocity profile is used to simplify the error analysis. Thus, range, x, and its standard deviation are given by

$$x=\left|(c\tau)^2-(D-S-h)^2\right|^{1/2}, \tag{9}$$

$$\sigma_x=\mu_x^{-1}[\mu_c^2\mu_\tau^2(\sigma_\tau^2\mu_c^2+\sigma_c^2\mu_\tau^2)+(\mu_D-\mu_S-\mu_h)^2 \times(\sigma_D^2+\sigma_S^2+\sigma_h^2)]^{1/2}, \tag{10}$$

where c is sound speed and τ is travel time of the direct path. Sound speed is measured generally with a CTD and XBT; a reasonable estimate of the standard deviation from these instruments is $\sigma_c=0.25$ m/s. An estimate of the standard deviation in the direct arrival is $\sigma_\tau=0.1$ ms or about 1 pulse width (see Fig. 3). Assuming the same means and standard deviations as in Sec. III A, the resulting mean and standard deviation of the range are shown in Fig. 9. Note that near normal incidence, there is a bias error in the mean range (9) and that the standard deviation (10) becomes infinite. Near normal incidence ($x\to 0$), the mean and standard deviation can be estimated as

$$\mu_{x=0}\approx 2^{1/4}[\mu_c^2\mu_\tau^2(\sigma_\tau^2\mu_c^2+\sigma_c^2\mu_\tau^2)+(\mu_D-\mu_S-\mu_h)^2 \times(\sigma_D^2+\sigma_S^2+\sigma_h^2)]^{1/4}, \tag{11}$$

$$\sigma_{x=0}\approx \tfrac{1}{2}[\mu_c^2\mu_\tau^2(\sigma_\tau^2\mu_c^2+\sigma_c^2\mu_\tau^2)+(\mu_D-\mu_S-\mu_h)^2 \times(\sigma_D^2+\sigma_S^2+\sigma_h^2)]^{1/4}. \tag{12}$$

The angle at the seabed, θ, and its associated standard deviation (in radians) are given by

$$\theta=\tan^{-1}((D-S+h)/x), \tag{13}$$

$$\sigma_\theta=\mu_x^{-1}\cos^2\mu_\vartheta(\sigma_D^2+\sigma_S^2+\sigma_h^2+\sigma_x^2\tan^2\mu_\vartheta)^{1/2}. \tag{14}$$

At normal incidence (i.e., $\theta=\pi/2$) the mean angle and the standard deviation (13)–(14) have bias errors, but can be estimated as

$$\mu_{\vartheta=\pi/2}=\tan^{-1}((\mu_D-\mu_S+\mu_h)/\mu_{x=0}), \tag{15}$$

$$\sigma_{\vartheta=\pi/2}=\sigma_{x=0}(\mu_D-\mu_S+\mu_h)^{-1}. \tag{16}$$

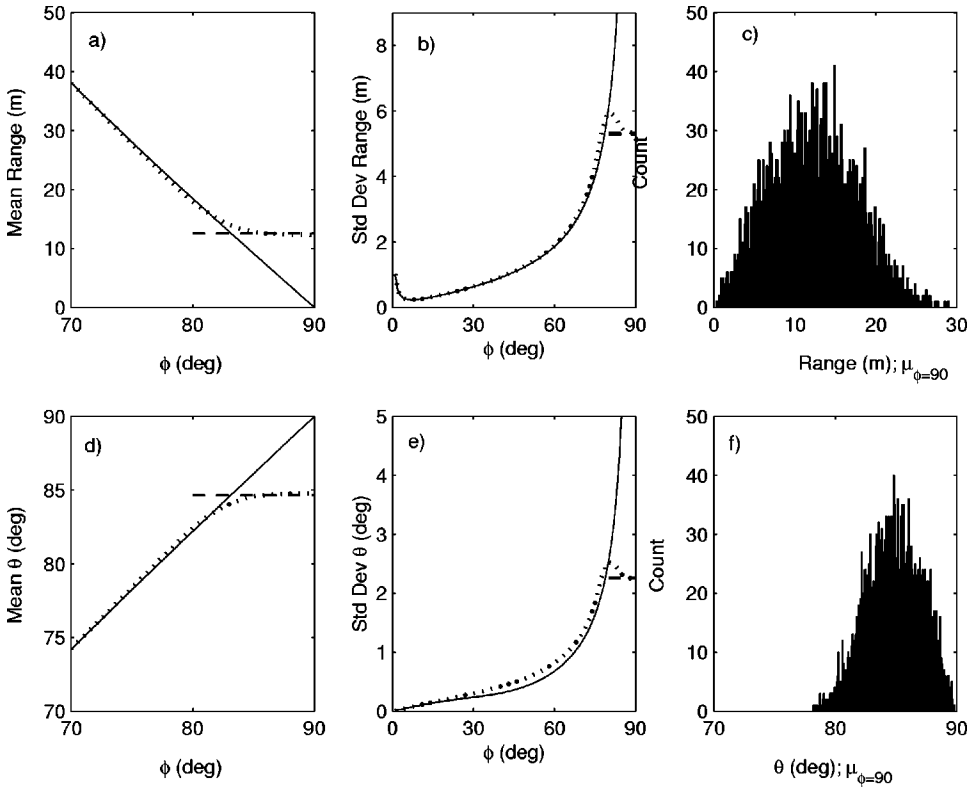


FIG. 9. Theoretical (solid), simulated (dotted line) and approximate (dashed line) solutions for: (a) mean of range estimate, Eq. (9) is solid line, Eq. (11) is dashed line; (b) standard deviation of range estimate, Eq. (10) is solid line, Eq. (12) is dahsed line; (d) mean of angle estimate, Eq. (15) is dashed line; (e) standard deviation of angle estimate, Eq. (14) is solid line, Eq. (16) is dashed line. The histograms show simulation results at normal incidence for (c) range and (f) angle estimates.

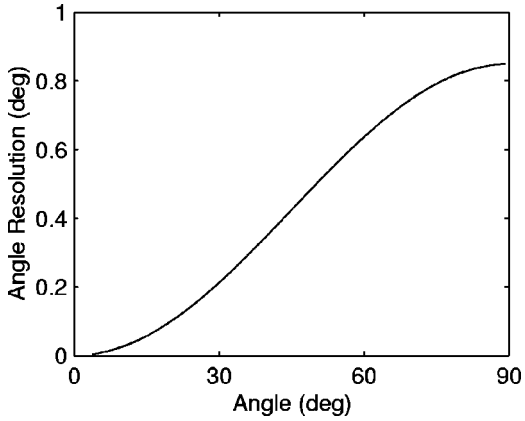


FIG. 10. Typical angular resolution of the reflection measurements.

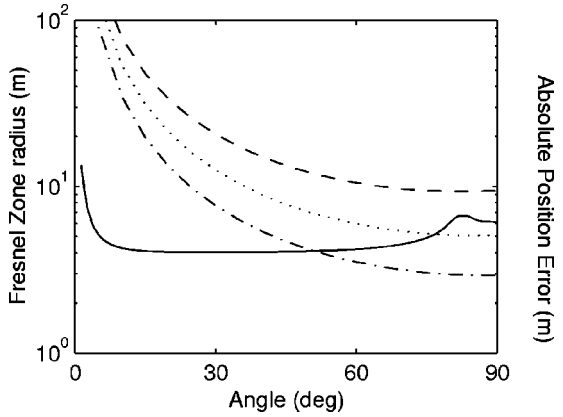


FIG. 11. Absolute position error (solid line) for a source, receiver 120 and 15 m above the seabed, respectively, and the Fresnel zone radius 300 Hz (dashed), 1000 Hz (dotted), and 3000 Hz (chain dashed).

The part of the angular range that is the most crucial for minimizing errors is dictated by where the critical angle or angle of intromission is expected, since these angles control long-range propagation. Hamilton (1980) indicates that for unconsolidated sediments on the continental shelf, critical angles vary from 0–34°. In that range the standard deviation of the angle uncertainty is quite small, from about 0.01–0.3°. At normal incidence, the errors are considerably larger; however, the increase in errors is mitigated, in part, by the fact that the reflection coefficient itself is often nearly constant between 70–90° (e.g., Fig. 7).

The angular resolution (difference in angle between two adjacent measurements) is a function of the ship speed ν, pulse repetition rate ϖ, and the geometry. Since the pulse repetition rate is constant, the resolution δ_θ is a function of angle. For an isovelocity sound-speed profile, the angle resolution is

$$\delta_\vartheta = \nu\varpi\,\mu_x^{-1}\cos^2\mu_\vartheta((\mu_D-\mu_S+\mu_h)/\mu_x). \tag{17}$$

Given typical parameters, $\nu=2$ m/s and $\varpi=1$ pulse per s, with the geometry as above, the angular resolution is less than 1° (see Fig. 10).

E. Absolute position uncertainty

A reflection experiment yields four measurements around the fixed receiver, typically two aligned with the bathymetric contours (i.e., an incoming and an outgoing), and two perpendicular. The region of the bottom that is sampled by each measurement depends on water depth, sound-speed profile, and receiver depth, but generally ranges from about 70–150 m in length. The uncertainty of the absolute position of each measurement is due to uncertainty in the location of the fixed receiver and location of the source.

The location of the receiver is determined by a least-squares fit to echolocation data. Transponders operating at 9–11 kHz mounted on the hull of the ship and on the fixed hydrophone string a few meters above the bottom provide the echolocation data. A typical echolocation run consists of an ''x'' pattern, attempting to place the center of the ''x'' as close as possible to the location of the drop position of the array. The least-squares fit to a model assuming straight-line ray paths gives fits of ±2–7 m depending on how close the estimated drop position was to the actual hydrophone position. In addition, there is a ±2–3-m uncertainty in the differential GPS (DGPS) estimate, so that the receiver position is generally known to within ±3–8 m. The isovelocity approximation for the echolocation data is reasonable since the transponder data are generally at grazing angles greater than about 30 deg.

The source is towed from a crane. The position uncertainty of the source relative to the ship is about ±2 m, and with the DGPS uncertainty of ±2–3 m, the overall source position accuracy is about ±4 m. Since range is estimated acoustically and is more precise than the positions, and source position standard deviation, σ_{sp}, is less than that of the receiver, the source position and the range is used to estimate absolute position. The standard deviation of the absolute position is

$$\sigma_L = (\sigma_{sp}^2 + (\sigma_D^2+\sigma_S^2)\tan^{-2}\theta + \sigma_\theta^2(\mu_D-\mu_S)^2\sin^{-4}\theta)^{1/2}. \tag{18}$$

The absolute positional uncertainty should be considered in light of the size of the region on the seabed illuminated by the acoustic field. The radii of the first Fresnel zone (see Fig. 11) are useful metrics for estimating that size. In Fig. 11, the larger radius of the Fresnel ellipse was used (in the plane of the source–receiver). Note that the positional errors are of the same order or smaller than the Fresnel zone radius except at high frequencies and high angles.

F. Other errors

There is another potential source of error that may contribute to the reflection coefficient related to assumptions inherent in the measurement and data processing. It is implicitly assumed that the bathymetry, sediment layer geometry, and sound velocity profile are independent of space (several hundred meters) and time (5–10 min) during the measurement. Bathymetry and normal incidence seismic reflection data are always collected and analyzed to ensure that the first two assumptions are met. Generally a conductivity–temperature–depth (CTD) cast is taken before the experiment commences and an expendable bathythermograph (XBT) is collected during the actual measurement evolution near the hydrophone position. The temperature data from the

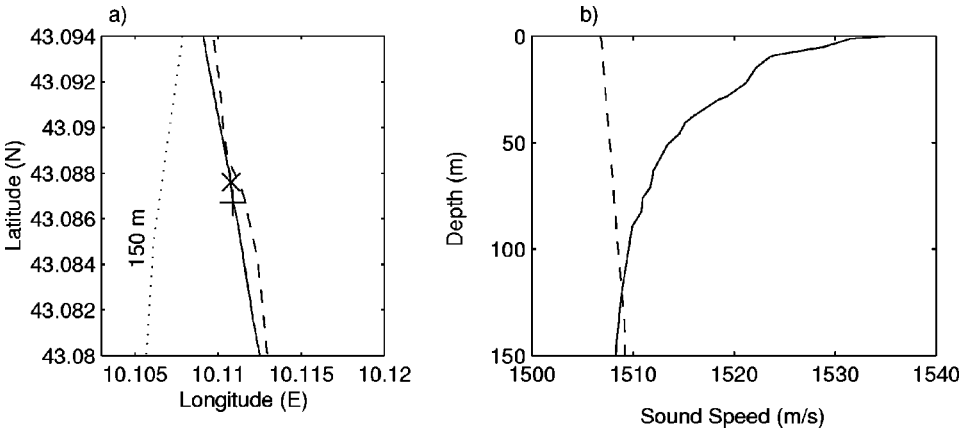


FIG. 12. NTS site 2(a) array position, for 1997 (+), 1999 (×), and reflection loss tracks 1997 (solid), 1999 (dashed) with bathymetry contour; (b) sound-speed profiles for June 1997 (solid), January 1999 (dashed).

XBT cast with the salinity data from the CTD are employed to represent the sound-speed structure over the minutes and hundreds of meters of the experiment. If sound-speed variability exists within this time/space scale, the effect would probably be a small shift in grazing angles.

IV. OBSERVED UNCERTAINTY

Multiple reflection loss measurements in the same area provide the opportunity to analyze the uncertainty from the measurement standpoint. Since the measurements are conducted under varying conditions, it is desirable to define in what sense the measurements are repeated.

Coleman and Steele (1999) draw a useful distinction between the words ''repetition'' and ''replication'' of measurements that is adopted here. Replication implies that the measurements are repeated in a particular fashion. Zeroth-order replication means that the measurements are repeated with identical instrumentation and perfectly constant experimental conditions. Only changes inherent in the source–receiver over time-space contribute to variation in the results. First-order replication means repetition with identical instrumentation but with variation in the experimental conditions, so that the observed variability in the data would be a combination of variability of the instrumentation and the conditions. *N*th-order replication is when both the instrumentation and the experimental conditions change. By completely changing both the instrumentation and the experimental conditions, *n*th-order replication permits an estimate of the total error.

In the suite of reflection measurements (Table I), there are never instances of pure replication of any order, i.e., where the same seabed is measured multiple times. There are instances, however, when nearly the same portion is measured multiple times. For example, seismic reflection data often indicate that the seabed on the ingoing and outgoing legs has a similar character. This roughly corresponds to the

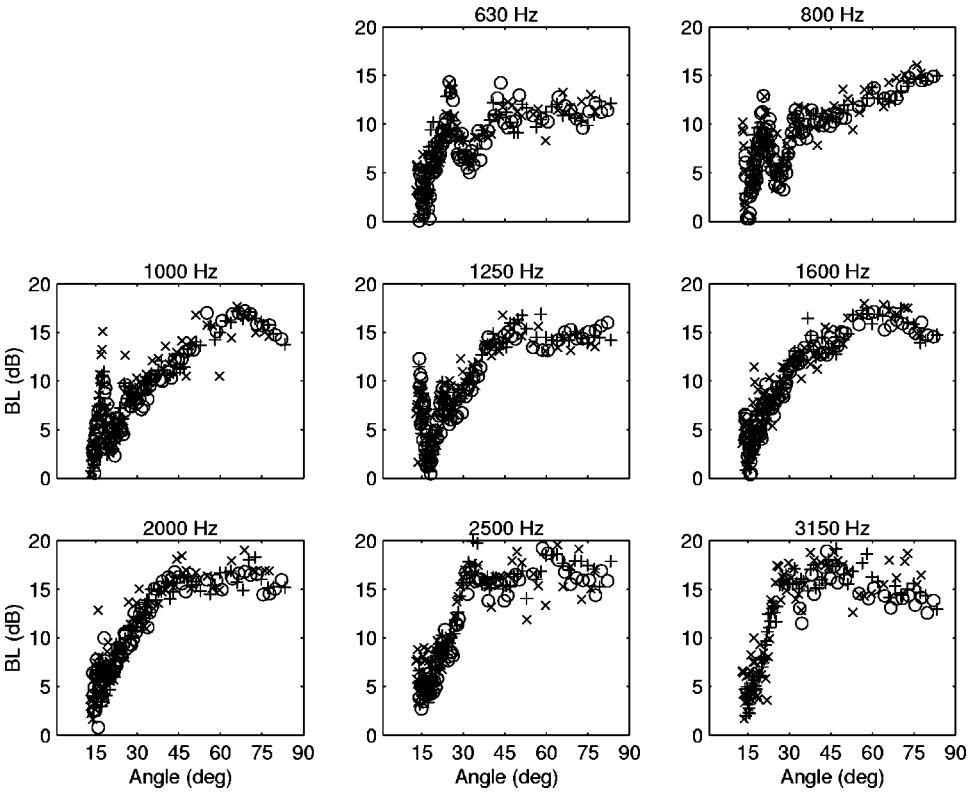


FIG. 13. Measured reflection loss at NTS site 2: south leg June 1997 (+), north leg June 1997 (○), south leg January 1999 (×).

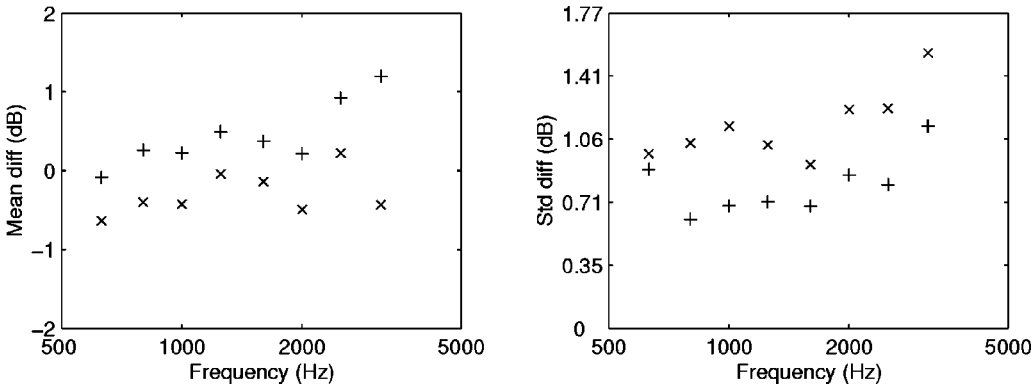


FIG. 14. NTS site 2 mean and standard deviation of the reflection loss difference: between south and north legs June 1997 (+); south leg June 1997, and south leg January 1999 (×).

first-order replication, where the source and receiver are identical, but with the passage of time between the experiments, the conditions (e.g., sound-speed profile) are allowed to vary. Actually, though the source itself is the same between the incoming and outgoing legs, the source characteristics change measurably because of an inherent tilt in the source plate, so it could be argued that this replication is *n*th order. Another replication is where the source and receiver are completely different, and the sound-speed profile is completely different. This roughly corresponds to the *n*th-order replication. In both cases the observed variance will be an outer bound because the region of sampled seabed is (slightly) different.

In order to determine if the replication results were a function of seabed type, repeated measurements were analyzed in two very distinct seabed environments: Site 2 in 150-m water depth in the North Tuscan shelf in the northern Tyrrhenian Sea (silty-clay host with random thin shelly layers), and site 7 in 107-m water depth on the Malta Plateau, Straits of Sicily (fine sand over limestone; see Fig. 1 for locations).

A. North Tuscany observations

Multiple measurements at site 2 permit comparison of reflection results under various conditions. During the SCARAB97 experiment in June 1997, measurements were conducted on a nearly N–S track. Data from the southern leg (i.e., south of the array position) were analyzed and reported in Holland and Osler (2000). Since seismic reflection data show that the layering structure changes little over the northern leg, data from the two legs can be compared. In addition, measurements were made in the Boomer99 experiment in January 1999 with the receive array 100 ± 7 m north of the 1997 array position [see Fig. 12(a)]. The source and receiver depths in 1997 were 0.35 and 138 m, and in January 1999 0.11 and 136 m. The source used in the 1997 experiment was an EG&G model 265 Uniboom and in 1999 a GeoAcoustics Uniboom with higher source level, but the shallower tow depth meant higher source level variability.

These three data sets provide the opportunity to examine the uncertainty (including the variability) over quite different experimental conditions; the sound-speed profiles for the two seasons are shown in Fig. 12(b). The reflection data in Fig. 13 show strong similarities. A statistical comparison was performed by forming the difference between data sets (after interpolation) and computing the mean and standard deviation of the difference (see Fig. 14). The mean of the data difference indicates whether or not there is a bias; the standard deviation gives an indication of the variance between the data sets. Frequently, at low angles, the data rapidly oscillate in angular increments that are at or below the uncertainty of the measurements. Out-of-phase oscillations can produce very large variances; this effect was reduced by per-

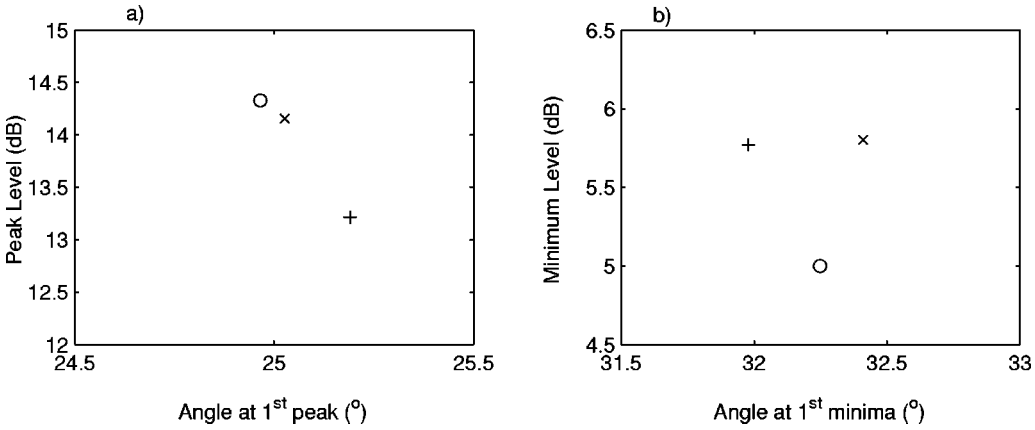


FIG. 15. Comparison of three data sets at 630 Hz for the (a) first peak and (b) first minima.

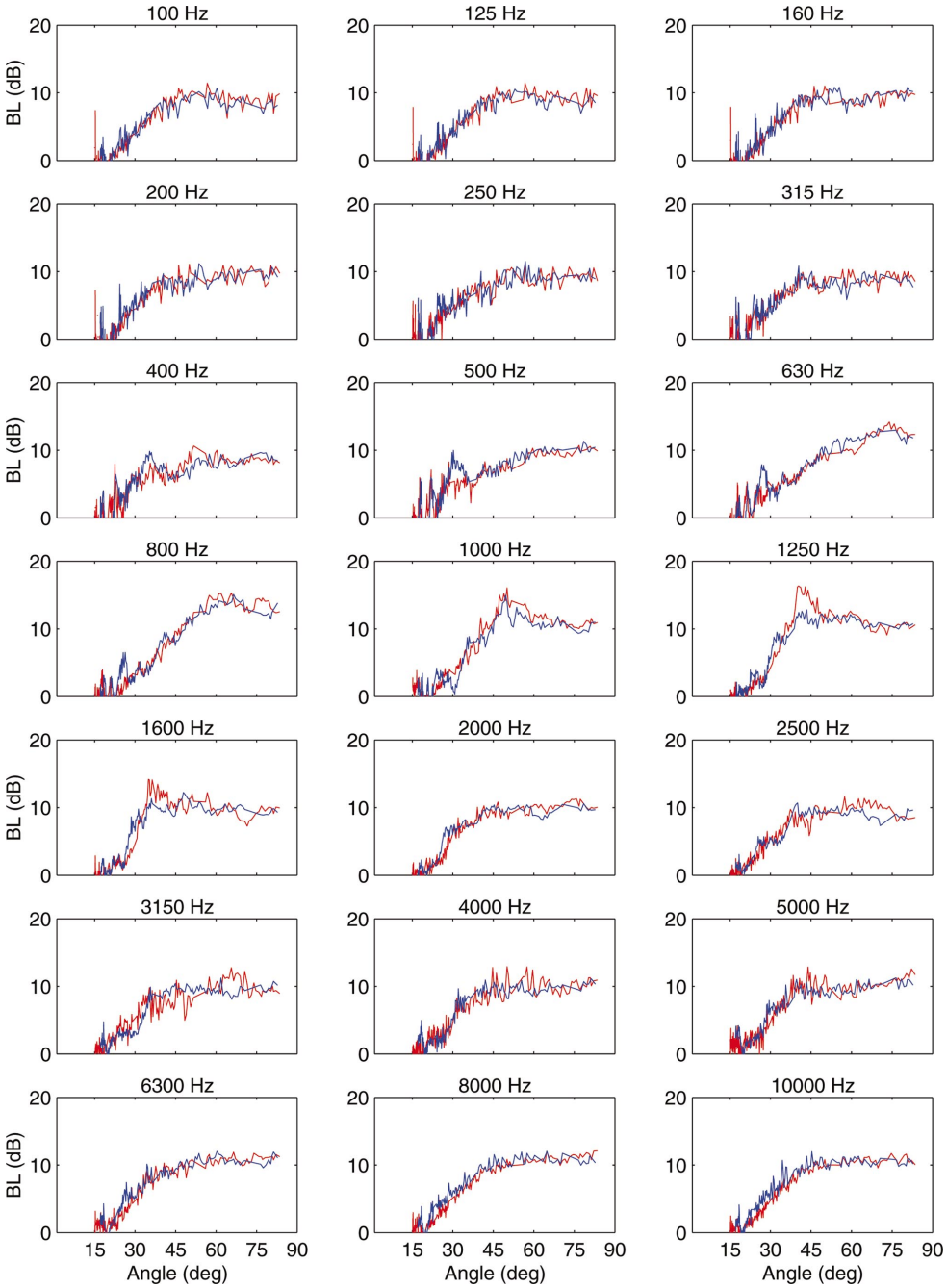


FIG. 16. MP site 7 comparison of reflection loss: south leg (red) and north leg (blue).

forming the statistics on data greater than 25°.

The mean difference between the 1997 measurements (Δ_{sn}) is about 1 dB or less (Fig. 14), with the greatest difference being at the highest frequencies. The mean difference between the 1997 and the 1999 (Δ_{ss}) measurements is also less than 1 dB. Δ_{sn} is somewhat smaller than Δ_{ss} below 2 kHz but larger above 2 kHz. The reason for this is unknown, but is probably related to the fact that the signal-to-noise ratio of the northern path was considerably smaller than that for the incoming. The important point of this figure is that the measurements are repeatable within 1 dB or less and there is no apparent consistent bias in the measurements.

The angle error inherent in the measurement predicted in Eq. (14) can be compared against the measurements by examining the angular offset between maxima (or minima) in the reflection data. The 1997 data were processed with a decimation factor of 3 (1 pulse every 3 s); the 1999 data have a decimation of 4 s. For a strict comparison, the decimation should be equal and as low as possible; however, even the data as they are permit an upper bound estimate of the error. In order to try and minimize possible differences due to spatial variability (the measurements sampled different parts of the seabed, of order 100 m apart), the lowest frequency, 630 Hz, is used for this comparison.

At 630 Hz, the first peak and minima are convenient points at which to compare the angles from the various experiments. A plot of the resulting first peak angle and first minima is shown in Fig. 15. The theory [Eq. (14)] indicates

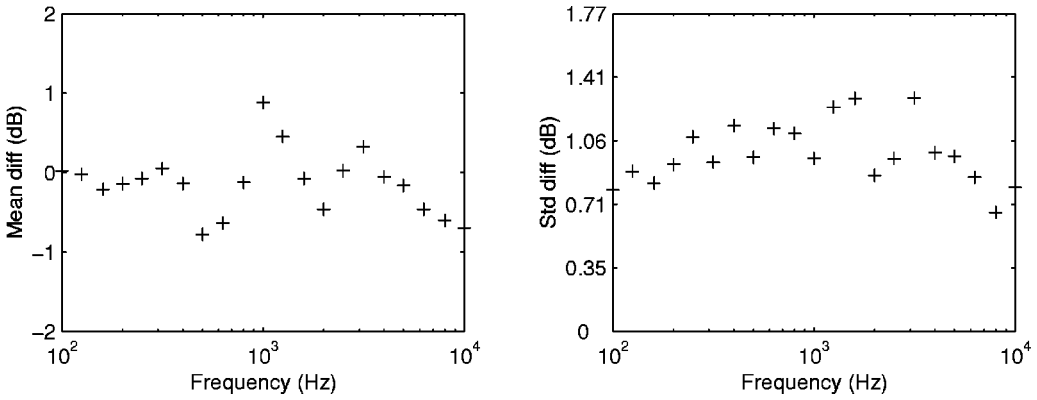


FIG. 17. MP site 7 mean and standard deviation of the reflection loss difference between the south and north legs.

that at 25°, the standard deviation $\sigma_\theta = 0.15°$ for the first peak, the observed $\sigma_\theta = 0.12°$. For the reflection loss minimum at about 32°, the theory indicates a $\sigma_\theta = 0.18°$ and the observed $\sigma_\theta = 0.22°$. Thus, the predicted angle errors seem to be quite reasonable. The site specific parameters used for the predictions were: $\mu_D = 150$ m, $\mu_S = 0.35$ m, $\mu_h = 13$ m, $\sigma_D = 1$ m, $\sigma_S = 0.25$ m, $\sigma_h = 0.3$ m, $\sigma_c = 0.25$ m/s, and $\sigma_\tau = 0.1$ ms.

The data can also be used to compare the uncertainty of the peak levels; however, this is a rather stringent test given that the amplitude of a narrow peak will not be well estimated for various sampling intervals as is the case here. Nevertheless, a comparison is useful as a guide; the standard deviation of the peaks at 25° and 32° are 0.6 and 0.5 dB, respectively, which are well within the estimated standard deviation.

B. Malta Plateau observations

Multiple reflection experiments at site 7 permit comparison at a different location. These experiments were conducted during Boundary2000 in May 2000. A combination of the source (GeoAcoustics Uniboomer) and the seabed type, thick sand over consolidated limestone, allowed a comparison over a broader frequency band than possible on the north Tuscany shelf. Figure 16 shows the reflection measurements for the southern and northern legs of the experiment. The agreement is generally good, although there are some clear differences that are believed to be due to slight differences in sediment fabric. Even including the variability due to sediment inhomogeneity, the mean difference between the two runs is quite small (Fig. 17), less than 1 dB.

V. REDUCING UNCERTAINTY

Although the uncertainty in the reflection data is modest, there are several ways in which the variance could be further reduced: by averaging in angle or by modifying the source.

A. Reducing uncertainty by averaging

Given that resolution can be traded for variance (e.g., Menke, 1989), averaging over angle space can be done to reduce the variance. As an example, the source level data of site 7 were reprocessed with a ±1° sliding window. The resulting standard deviation, compared with the full angular resolution (see Fig. 18) is significantly reduced to ~0.1–0.5 dB.

The advantage of decreased variance may or may not offset the loss in resolution depending on the particular problem. However, given the very high resolution in angle, especially at low angles (see Fig. 10), angle averaging should provide a useful reduction in variance for many situations. Some situations suggest a strategy of a variable window size, i.e., a window that is a function of angle. For example, in obtaining geoacoustic properties in fine-grained sediments (see Holland, 2002) a low variance near 90° with a high angle resolution near the angle of intromission (in that case 15°) would yield the highest precision in the resulting velocity and density estimates. That could be easily accomplished with a large window size near 90° and a small (or zero) window near the angle of intromission. Obtaining high-precision velocity and density estimates from sandy sediments (with a critical angle) suggests a similar strategy.

B. Reducing uncertainty by modifying the source

Another way to reduce the variance would be to use an omnidirectional source, which would eliminate sensitivity to source rotation. Practically speaking this is difficult, since a large bandwidth is desirable. A sparker source was investi-

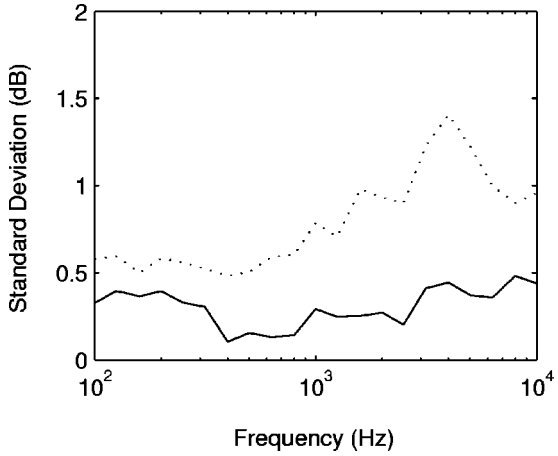


FIG. 18. Reduction in variance achieved by angle averaging: no angle averaging (dotted line) and ±1-deg angle averaging (solid line).

gated, which has a more omnidirectional beam pattern, but the source amplitude variability was much larger than that of the Boomer.

Yet another way to diminish the errors would be to tow the source at a much greater depth. In an experiment planned in 2004, the source will be mounted on an autonomous undersea vehicle (AUV) and flown a few tens of meters above the bottom. In addition to eliminating the uncertainties due to the air–sea interface, the deep-tow geometry will substantially reduce the time required for multiple measurements.

VI. SUMMARY AND CONCLUSIONS

Error analysis of the reflection measurements has shown that standard deviations are typically ± 0.5–1 dB at full angular resolution. Averaging in angle can reduce that substantially, for example $\pm 1°$ angle averaging leads to a standard deviation of ± 0.1–0.5 dB. Specific error estimates depend upon frequency, geometry, and location. The dominant contribution to the error is source amplitude variability, and a new processing approach was developed that reduces the error for frequencies above a few hundred Hz.

Errors in the angle estimate are a few tenths of a degree below 35 deg, which is the crucial angular range for predicting long-range propagation or inverting for geoacoustic properties. The largest contribution to the error in the angle estimates comes from wave motion that induces variability in source height. Absolute position errors of the measurements are about 3–8 m.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of the NATO SACLANT Undersea Research Center and the Office of Naval Research. Enzo Michelozzi conducted the tank measurements.

Beebe, J. H., and Holland, C. W. (**1986**). ''Shallow-water propagation effects over a complex, high-velocity bottom,'' J. Acoust. Soc. Am. **80**, 244–250.

Bevington, P. R., and Robinson, D. K. (**1992**). *Data Reduction and Error Analysis for the Physical Sciences*, 2nd ed. (McGraw-Hill, New York).

Chapman, R., Chin-Bing, S., King, D., and Evans, R. (**2001**). ''Benchmark workshop for geoacoustic inversion techniques in range-dependent waveguides,'' J. Acoust. Soc. Am. **110**, 2660.

Collins, M. D., Kuperman, W. A., and Schmidt, H. (**1992**). ''Nonlinear inversion of sediment parameters in a range-dependent region of the Gulf of Mexico,'' J. Acoust. Soc. Am. **92**, 2770–2783.

Coleman, H. W., and Steele, W. G. (**1999**). *Experimentation and Uncertainty Analysis for Engineers* (Wiley, New York).

Collins, M. D. (**1993**). ''A split-step Padé solution for the parabolic equation method,'' J. Acoust. Soc. Am. **93**, 1736–1742.

Eller, A. I., and Gershfeld, D. A. (**1985**). ''Low frequency acoustic response of shallow water ducts,'' J. Acoust. Soc. Am. **78**, 622–631.

Ferla, C., and Jensen, F. B. (**2002**). ''Are current environmental databases adequate for sonar predictions in shallow water?'' in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. Pace and F. Jensen (Lerici, Italy) (Kluwer Academic Publishers, Dordecht, The Netherlands).

Frisk, G. V., and Lynch, J. F. (**1984**). ''Shallow water waveguide characterization using the Hankel transform,'' J. Acoust. Soc. Am. **76**, 205–216.

Hamilton, E. L. (**1980**). ''Geoacoustic modeling of the seafloor,'' J. Acoust. Soc. Am. **68**, 313–339.

Holland, C. W., and Osler, J. (**2000**). ''High resolution geoacoustic inversion in shallow water: A joint time and frequency domain technique,'' J. Acoust. Soc. Am. **107**, 1263–1279.

Holland, C. W. (**2002**). ''Geoacoustic inversion for fine-grained sediments,'' J. Acoust. Soc. Am. **111**, 1560–1564.

Jensen, F. B., and Kuperman, W. A. (**1983**). ''Optimum frequency of propagation in shallow water environments,'' J. Acoust. Soc. Am. **73**, 813–819.

Menke, W. (**1989**). *Geophysical Data Analysis: Discrete Inverse Theory* (Academic, New York).

Porter, M. B. (**1991**). ''The KRAKEN normal mode program,'' SACLANTCEN Memorandum SM-245, SACLANT Undersea Research Centre, La Spezia.

Rubano, L. A. (**1980**). ''Acoustic propagation in shallow water over a low-velocity bottom,'' J. Acoust. Soc. Am. **67**, 1608–1613.

Siderius, M., Nielsen, P. L., Sellschopp, J., Snellen, S., and Simons, D. (**2001**). ''Experimental study of geo-acoustic inversion uncertainty due to ocean sound-speed fluctuations,'' J. Acoust. Soc. Am. **110**, 769–781.

Urick, R. J. (**1970**). ''Reverberation-derived scattering strength of the shallow sea bed,'' J. Acoust. Soc. Am. **48**, 392.

Weinberg, H., and Keenan, R. E. (**1996**). ''Gaussian ray bundles for modeling high-frequency propagation loss under shallow-water conditions,'' J. Acoust. Soc. Am. **100**, 1421–1431.

Westwood, E. K., and Vidmar, P. J. (**1987**). ''Eigenray finding and time series simulation in layered bottom ocean,'' J. Acoust. Soc. Am. **81**, 912–924.

Inversion for sediment geoacoustic properties at the New England Bight

Gopu R. Potty[a] and James H. Miller
Department of Ocean Engineering, University of Rhode Island, Narragansett, Rhode Island 02882

James F. Lynch
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 15 May 2002; revised 25 June 2003; accepted 14 July 2003)

This article discusses inversions for bottom geoacoustic properties using broadband acoustic signals obtained from explosive sources. Two different inversion schemes for estimating the compressional wave speeds and attenuation are presented in this paper. In addition to these sediment parameters, source–receiver range is also estimated using the arrival time data. The experimental data used for the inversions are SUS charge explosions acquired on a vertical hydrophone array during the Shelf Break Primer Experiment conducted south of New England in the Middle Atlantic Bight in August 1996. The modal arrival times are extracted using a wavelet analysis. In the first inversion scheme, arrival times corresponding to various modes and frequencies from 10 to 200 Hz are used for the inversion of compressional wave speeds. A hybrid inversion scheme based on a genetic algorithm (GA) is used for the inversion. In an earlier study, Potty *et al.* [J. Acoust. Soc. Am. **108**(3), 973–986 (2000)] have used this hybrid scheme in a range-independent environment. In the present study results of range-dependent inversions are presented. The sound speeds in the water column and bathymetry are assumed range dependent, whereas the sediment compressional wave speeds are assumed range independent. The variations in the sound speeds in the water column are represented using empirical orthogonal functions (EOFs). The replica fields corresponding to the unknown parameters were constructed using adiabatic theory. In the second inversion scheme, modal attenuation coefficients are calculated using modal amplitude ratios. The ratios of the modal amplitudes are also calculated using time-frequency diagrams. A GA-based inversion scheme is used for this search. Finally, as a cross check, the computed compressional wave speeds along with the modal arrival times were used to estimate the source–receiver range. The inverted sediment properties and ranges are seen to compare well with *in situ* measurements and historical data.
 [DOI: 10.1121/1.1605391]

PACS numbers: 43.30.Pc, 43.30.Ma, 43.30.Bp [WLS]

I. INTRODUCTION

Acoustic propagation in shallow water is greatly influenced by the properties of the bottom. Indirect methods for the estimation of bottom properties have been given much attention in underwater acoustics as direct measurements (e.g., cores) are very hard to make. In this article we discuss two different inversion schemes for the estimation of sediment compressional wave speeds and compressional attenuation using broadband data. In inversion scheme I, sediment compressional wave speeds are estimated using a hybrid inversion scheme based on the dispersion behavior of broadband acoustic propagation. The application of this inversion scheme to a range-independent environment is discussed in detail by Potty *et al.* (2000) in a previous article. This hybrid scheme is a combination of a genetic algorithm (GA) and the Levenberg–Marquardt optimization method. Compressional wave attenuation values are estimated using inversion scheme II based on modal amplitude ratios. In addition to the sediment properties, other parameters such as bathymetry, source depth, receiver depth, range, and source level are also treated as unknowns in this inversion scheme. The relative importance of these parameters is assessed by a sensitivity study. This latter inversion scheme is also carried out using a GA.

When a broadband acoustic source is used in shallow-water waveguide, the acoustic propagation exhibits dispersion effects. The group velocities, i.e., the speeds at which energy is transported, differ for different frequencies and modes. This dispersion effect can be observed by time–frequency analysis of an acoustic signal recorded at sufficiently large distance away from the source. The times of arrivals of different modes at various frequencies can be directly extracted from these time–frequency distributions. Lynch, Rajan, and Frisk (1991) successfully used dispersion characteristics for the inversion of geoacoustic properties using linear perturbation methods. Nonlinear inverse methods for estimating bottom properties were subsequently developed by Collins *et al.* (1992), Gerstoft *et al.* (1996), and others. It should be noted that dispersion analysis of seismic interface waves has been used extensively to determine the shear properties of near-bottom ocean sediments (Jensen and Schmidt, 1985; Caiti *et al.*, 1994; Stoll *et al.*, 1994). Turning to the New England Shelf Break environment, Potty *et al.* (2000) have recently used global optimization methods for

[a]Electronic mail: potty@oce.uri.edu

 0001-4966/2003/114(4)/1874/14/$19.00

their geoacoustic inversion in the range-independent shallow-water environment case. Their genetic algorithm-based inversion gave good estimates for sediment compressional wave speeds, which matched well with deep core data [Atlantic Margin Coring (AMCOR) Project] and shallow gravity cores at the same location. In this article we extend this inversion scheme to the mildly range-dependent environment using adiabatic theory.

The recovery of range-dependent structures in the ocean environment has been a subject of interest in ocean acoustics. Tolstoy *et al.* (1991) proposed a linearized matched field processing approach to acoustic tomography, assuming adiabatic normal-mode propagation for low-frequency signals received on vertical arrays. Taroudakis and Marakaki (1997) used GAs and modal phases to invert for a range-dependent environment due to a cold eddy using adiabatic theory. Siderius, Gerstoft, and Nielsen (1998) have used GAs for geoacoustic inversion in a range-dependent environment.

Turning to the estimation of medium attenuation, Tindle (1982) and Zhou *et al.* (1987) used modal amplitude ratios to extract modal attenuation coefficients. Using this approach they were able to model the nonlinear frequency dependence of the acoustic attenuation in the upper sediment layer in a shallow-water location in the Yellow Sea. Rajan, Frisk, and Lynch (1992) estimated modal attenuation coefficients using various methods based on the pressure field or its Hankel transform. The bottom attenuation profile is obtained from these modal attenuation coefficients by solving an integral equation using linear inverse theory. They also separated contributions from other attenuating mechanisms (shear, rough surface scattering, etc.). In this article, an inverse scheme for attenuation is presented which utilizes the compressional wave-speed values determined using the dispersion-based inversion. This scheme calculates the modal attenuation coefficients based on modal amplitude ratios and transmission loss data. In addition to modal attenuation coefficients, source depth, receiver depth, range, and source levels are treated as unknowns in the inversion and checked for consistency.

Our study is arranged as follows. Sediment compressional wave speeds are first addressed in this study, using inversion scheme I based on the group speed dispersion. Section II A contains the details of this inversion scheme. Section II B begins by presenting inversion scheme II for compressional wave attenuation. This inversion scheme is applied to synthetic data as well as field data from the Shelf Break Primer Experiment. Next, the group speeds estimated using the inversion scheme and the arrival times obtained from the field data are used to estimate the source–receiver ranges assuming range independence. Section II C briefly explains this source distance estimation method.

A mode-based sensitivity study, performed to assess the relative influence of various parameters (compressional wave speed, ocean sound speed, and water depth), is discussed in Sec. III A. Sensitivity of group speeds to these parameters is discussed in Sec. III B and sensitivity of modal amplitude ratios in Sec. III C. Section IV A presents the details of the Shelf Break Primer Experiment. Section IV B contains the description of the historic geoacoustic data pertaining to the experimental region. Section V contains the major new results of this study. Section V A shows the results of compressional speed inversion, Sec. V B discusses the results of range estimation, and Sec. V C contains result of attenuation estimates. Section VI summarizes and concludes the article.

II. INVERSION SCHEMES

We use a genetic algorithm (Goldberg, 1988) as the basic search tool for our inversion of sediment compressional wave speeds and attenuation. The principle of the GA is simple and closely resembles the genetic cycle. From all the possible parameter vectors, an initial population of members is randomly selected. The ''fitness'' of each member is computed on the basis of an objective function. Based on the fitness of the members, a set of ''parents'' is selected and through a randomization procedure a set of ''children'' is produced. These children replace the least fit of the original population and the process iterates to develop an overall fitter population. A hybrid scheme is used in the inversion for sediment compressional wave speeds, where the best parameter vector obtained using the GA is further optimized using a local search. The Levenberg–Marquardt algorithm (Fletcher, 1980) was employed for this local search. By applying this method at the end of the GA search, we can both assess the quality of the GA solution locally and also search for a better solution. Error bounds on the parameters were also estimated using two different approaches. During the optimization using GA, all the population is stored and is later used to estimate *a posteriori* probabilities. In addition to the best possible estimate, moments of the *a posteriori* distributions such as mean and covariance can also be estimated. This error estimation procedure is discussed in detail by Gerstoft (1994) and Potty *et al.* (2000). These error estimates provide a measure of the convergence of the optimization procedure and can be used to make comparisons between retrieved parameters. Added to this *a posteriori* error estimate, the quality of the inversion is also examined locally by numerically calculating the standard deviation using the Hessian matrix (Sen and Stoffa, 1995). Elements of the Hessian matrix are the second partial derivatives of the objective function with respect to the model parameters. They are local estimates that only characterize the region about the model at which they are calculated and are numerically evaluated in the neighborhood of the best solution. Hessian uncertainty will represent the true uncertainty if the best solution corresponds to the true model. The ability of the global approaches to efficiently navigate the multipeaked and noisy search space and to converge to the true solution increases the effectiveness of the Hessian approach. It should be noted that this Hessian method assumes that the error surface is Gaussian.

A. Method of inversion for sediment compressional speeds: Inversion scheme I

In this section, we discuss our inversion for sediment compressional wave speed. The sound speed in the water column and bathymetry are considered range dependent, whereas the sediment compressional wave speed is treated as range independent in this inversion. This is because we have

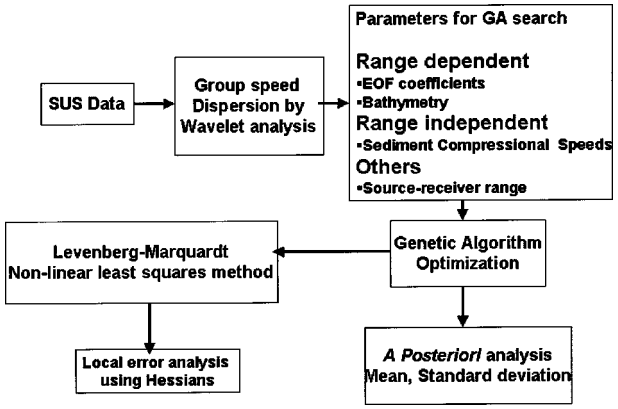


FIG. 1. Steps involved in the inversion scheme I. The source–receiver range is divided into five sections in which the bathymetry and water column sound speeds are allowed to vary.

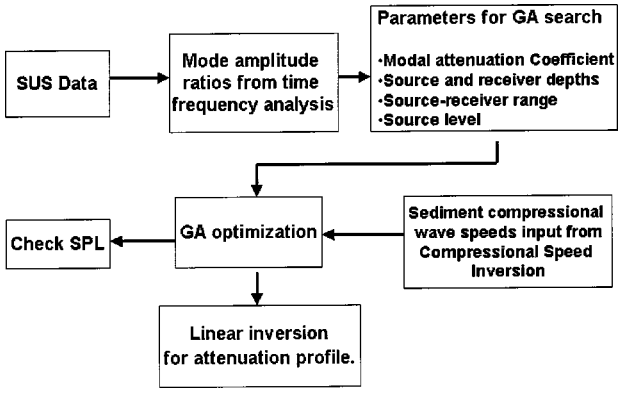


FIG. 2. Steps involved in the inversion scheme II. Sound speeds in the water column and the compressional speeds in the sediment are obtained from a previous inversion by Potty *et al.* (2000).

a priori information about the range dependence of the water column sound speed and bathymetry. Shear effects in the sediment are neglected, as shear speeds are expected to be of the order of 150 m/s (Hamilton, 1980) for the type of sediments present at the experimental location. In the inversion scheme to estimate the sediment compressional wave speeds, the parameter vectors searched for consisted of coefficients of empirical orthogonal functions (EOFs) of water column sound speed, the bathymetry, and the source to receiver range, in addition to the sediment compressional wave-speed profile.

The objective function for the inversion was based on the minimization of group speed differences, and was of the form

$$E(m)=\sum_i \frac{[d_i-F_i(m)]^2}{\sigma_i^2}. \tag{1}$$

In Eq. (1), $E(m)$ is the objective function for the parameter vector m and σ_i is the standard deviation associated with the ith data point. The numerator of this equation represents the mismatch between the observed data $(d,N\times 1)$ and the prediction $[F(m),N\times 1]$ of the forward model. A normal-mode routine is used to calculate the predictions $[F(m)]$. The major steps involved in this inversion scheme are shown in Fig. 1.

B. Method of inversion for the compressional wave attenuation coefficient: Inversion scheme II

The compressional wave attenuation is estimated in this paper assuming range-independent propagation. In a range-independent environment, the acoustic pressure at range r in the far field can be expressed as a sum of M normal modes

$$P(r,z)=\frac{e^{i\pi/4}}{\rho(z_s)\sqrt{8\pi r}}\sum_{m=1}^{M}\frac{\psi_m(z_s)\psi_m(z)(e^{(i\kappa_{rm}-\beta_m)r})}{\sqrt{\kappa_{rm}}}, \tag{2}$$

where $P(r,z)$ is the acoustic pressure at a range r and at depth z, z_s is the depth of the source, ψ_m is the mth mode function, κ_{rm} is the horizontal wave number, and β_m is the modal attenuation coefficient. At a given range r from the source, the ratio of the amplitude of the second mode to the amplitude of the first mode can be written as

$$R_{21}(f)=\sqrt{\frac{\kappa_1}{\kappa_2}}\left|\frac{\psi_2(z_s)\psi_2(z)}{\psi_1(z_s)\psi_1(z)}\right|e^{(\beta_1-\beta_2)r}. \tag{3}$$

Ratios of spectral amplitudes between other modes also can be similarly expressed.

Modal attenuation coefficients were inverted by minimizing the difference between the theoretical spectral ratios computed using Eq. (3) and the experimental spectral ratios calculated from time–frequency diagrams in a least-squares sense. The modal amplitude ratios corresponding to the first three modes $[R_{21}(f),\ R_{31}(f),$ and $R_{23}(f)]$ are obtained by time–frequency analysis using wavelet-based methods. The mode functions (ψ) and eigenvalues (κ) are obtained using a standard normal-mode routine. The sound-speed profile obtained from the inversion done previously (Potty *et al.*, 2000) is used for this purpose. The unknowns in the inversion scheme are the source depth (z_s), receiver depth (z), the range (r), and the modal attenuation coefficients $(\beta_1,\ \beta_2,$ and $\beta_3)$. As a check, transmission loss is calculated using the inverted modal attenuation coefficients and compared with experimental values. The source level of the explosion is also treated as unknown. A genetic algorithm (GA) was used to perform this inversion.

Having obtained the modal attenuation coefficients using this inversion, the compressional wave attenuation profile is determined from the integral equation

$$\kappa_{rm}\beta_m=\int_0^\infty \alpha(z)k(z)|\psi_m(z)|^2dz, \tag{4}$$

where $k(z)$ is the wave number and $\alpha(z)$ is the attenuation profile. We solve this equation using linear inverse theory (Rajan *et al.*, 1987). The ability of this method to estimate the attenuation profile depends primarily on the amplitudes of the mode functions at various depths. The mode function falls off exponentially with depth beyond the turning depth, and hence this method will not be able to estimate the attenuation coefficient reliably at depths much greater than the turning depth. The steps involved in the inversion scheme are shown in Fig. 2. This inversion scheme is first tested using synthetic data generated for a known sound-speed profile and attenuation.

C. Method of estimating the source–receiver ranges from mode arrival times

The source–receiver range (r) can be evaluated from the arrival time difference between two frequencies at a single mode or from two modes at a single frequency. This serves as a cross check for the compressional wave-speed inversions. The arrival time difference between two frequencies for any given mode i at range r is given by

$$\Delta T_{ii}(f)=\left[\frac{1}{V_g^i(f)}-\frac{1}{V_g^i(f_H)}\right]r, \tag{5}$$

where f_H is a reference frequency which is different from f. At a distance r, for the same frequency the arrival time difference between mode j and mode i is given by

$$\Delta T_{ji}(f)=\left[\frac{1}{V_g^j(f)}-\frac{1}{V_g^i(f)}\right]r, \quad i\neq j. \tag{6}$$

These equations are of the form

$$\Delta T(f)=[K_t(f)]r, \tag{7}$$

in which ΔT is the experimental travel time differences [left-hand side of Eqs. (5) and (6)] and K_t is the theoretical group slowness differences [the quantity within brackets in the right-hand sides of Eqs. (5) and (6)]. The group speeds (V_g) calculated theoretically using the compressional speed inversion and arrival time differences ΔT obtained from the experimental data are used to calculate the range r. It should be noted that these equations are applicable only to range-independent environments, so that errors will be incurred if the environment contains range dependence.

III. SENSITIVITY STUDY

A. Mode-based sensitivity study

In order to prepare for our inversion, the relative importance of sound speed in water column (cw), compressional speeds in four layers of sediments [cp1 (0–6 m), cp2 (6–12 m), cp3 (12–18 m), and cp4 (18–24 m)] each 6 m thick, and the water depth was analyzed using a mode-based sensitivity study. These values were then compared with sensitivities calculated based on group speed changes due to changes in the above parameters. The mode-based sensitivity study is described in detail by Kessel (1999) and is applicable to weakly range-dependent environments. Some of the important aspects of this study are included in this section.

A possible measure of sensitivity can be written in terms of the change in the pressure field due to some environmental changes relative to a suitable norm as $\Delta P(x,x_s)/\|P(x,x_s)\|$. Here, $P(x,x_s)$ is the original pressure and $\Delta P(x,x_s)$ is the change in pressure. If $P'(x,x_s)$ is the new field corresponding to a small change in the environment, the change in pressure can be written as contributions from modes $m=1,2,3,...,M$.

$$\Delta P(x,x_s)=\sum_{m=1}^{M}\left(P'_m(x,x_s)-P_m(x,x_s)\right). \tag{8}$$

Here, $x_s=(r_s,z_s)$ and $x=(r,z)$ define the positions of the source and receiver where r and r_s are horizontal positions of receiver and source, respectively. z and z_s represent the receiver and source depths, respectively. The overall tendency of the perturbation $|\Delta P(x,x_s)|^2$ can be represented by the incoherent mode sum

$$|\Delta P(x,x_s)|^2_{\text{incoh}}=\sum_{m=1}^{M}|P'_m|^2+|P_m|^2-2\,\text{Re}(P'_mP^*_m), \tag{9}$$

in which * denotes complex conjugation. Noting that the long-range horizontal phase $e^{i\xi_mR}$ of the modes is the most sensitive part of the field to changes in environment, we can write

$$P'_m(x,x_s)\approx P_m(x,x_s)e^{i\Delta\xi_mR}, \tag{10}$$

where

$$\Delta\xi_m=\xi'_m-\xi_m. \tag{11}$$

In order to evaluate the sensitivity for a given environmental change, the local wave numbers κ_m and κ'_m corresponding to original and modified environments, respectively, for each mode m are calculated. Corresponding changes in the modal phase ($\Delta\xi_m$) are computed using these wave numbers

$$\xi_m=\frac{1}{r-r_s}\int_{r_s}^{r}\kappa_m(r)dr, \tag{12}$$

and

$$\xi'_m=\frac{1}{r-r_s}\int_{r_s}^{r}\kappa'_m(r)dr. \tag{13}$$

It should also be noted that, if the medium absorbs energy, then ξ_m is complex

$$\xi_m=\gamma_m+i\alpha_m. \tag{14}$$

The perturbation in the field can now be written as

$$|\Delta P(x,x_s)|^2_{\text{incoh}}=\sum_{m=1}^{M}|P'_m|^2\Omega_m(r,r_s), \tag{15}$$

in which

$$\Omega_m(r,r_s)=1+e^{-2\Delta\alpha_mR}-2e^{-\Delta\alpha_mR}\cos\Delta\gamma_mR. \tag{16}$$

It should also be noted that, in order to isolate the role of long-range horizontal phase, $\Delta\alpha_m$ may be taken as zero. The mean of this perturbation over a large number of source–receiver positions will then be

$$\overline{|\Delta P(x,x_s)|^2_{\text{incoh}}}=\|p(r,r_s)\|^2\overline{\Omega(r,r_s)}, \tag{17}$$

in which $\overline{\Omega(r,r_s)}$ is a weighted average over modes as shown below (Kessel, 1999)

$$\overline{\Omega(r,r_s)}=\frac{1}{\|p(r,r_s)\|^2}\sum_{m=1}^{M}\overline{|p_m|^2}\Omega_m(r,r_s). \tag{18}$$

If we use N source–receiver positions $x_s^n=(r_s,z_s^n)$ and $x^n=(r,z^n)$ $(n=1,2,3,...,N)$, respectively, then

$$\overline{|P_m|^2}=\frac{1}{N}\sum_{n=1}^{N}|P_m(x^n,x_s^n)|^2, \tag{19}$$

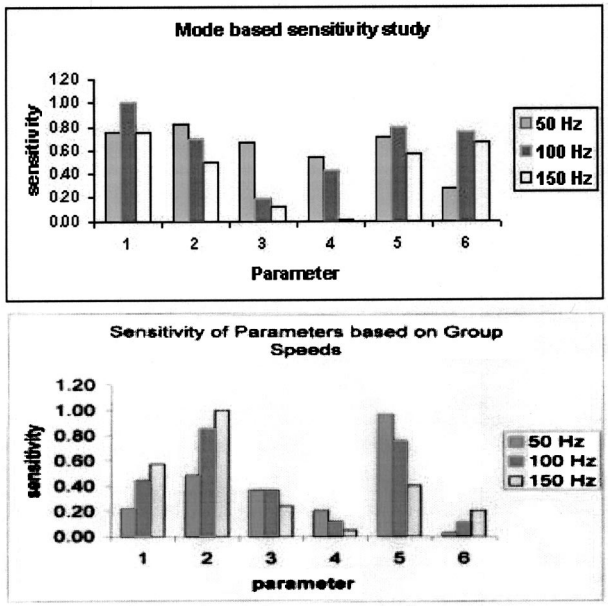


FIG. 3. Sensitivity of sediment compressional speeds at 0–6 m (parameter 1), 6–12 m (parameter 2), 12–18 m (parameter 3), 18–24 m (parameter 4). Parameter 5 corresponds to changes in water depth and parameter 6 indicates ocean sound-speed variations. Top panel shows the sensitivities calculated based on the mode-based sensitivity study. Sensitivities shown in the bottom panel were calculated based on the changes in group speeds.

and

$$\|p(r,r_s)\|^2=\overline{|p_m|^2} \tag{20}$$

(r_s,z_s) and (r,z) are the horizontal and vertical positions of the source and receiver such that the range $R=|r-r_s|$.

The top panel in Fig. 3 shows the sensitivities of these parameters for frequencies of 50, 100, and 150 Hz. The sound-speed profile obtained by the compressional wave-speed inversion scheme (Potty *et al.*, 2000) was used for this analysis. The sediment compressional speeds at depths 0–6 m (cp1), 6–12 m (cp2), 12–18 m (cp3), and 18–24 m (cp4) were changed by ±30 m/s, water depth by ±4 m, and ocean sound speed (cw) by ±3 m/s. These sensitivity values are scaled relatively, larger values indicating higher sensitivity for the changes in the parameters as mentioned above. For the given changes in compressional speeds, ocean sound speed, and water depth the sensitivities are nearly equal. Water depth variations have equal sensitivities for the three frequencies considered, which are comparable to the sensitivities of all the other parameters in magnitude. The sensitivities also show the expected dependence on frequency. At the lower frequency (50 Hz) changes in compressional speeds at deeper sediment depths are very sensitive, whereas at 150 Hz the sensitivity is very low at deeper sediment depths. At higher frequencies, the ocean sound speed and water depth are more sensitive than deep sediment compressional speeds. At 50 Hz, sound speed in the water column has comparatively less effect than the other parameters.

B. Sensitivity of group speeds

The bottom panel in Fig. 3 shows the sensitivities calculated based on the changes in group speeds corresponding to changes in parameters mentioned earlier. The magnitudes of these changes were same as in the mode-based sensitivity study. The group speeds seem to be more sensitive to changes in water depth and compressional wave speeds at depths 0–6, and 6–12, and 12–18 m. Unlike the mode-based sensitivity study, frequencies 100 and 150 Hz are more effective at depths 0–6 and 6–12 m. This may be due to the fact that more modes are included at higher frequencies compared to 50 Hz. These results are identical to the results of the sensitivity study reported by Potty *et al.* (2000).

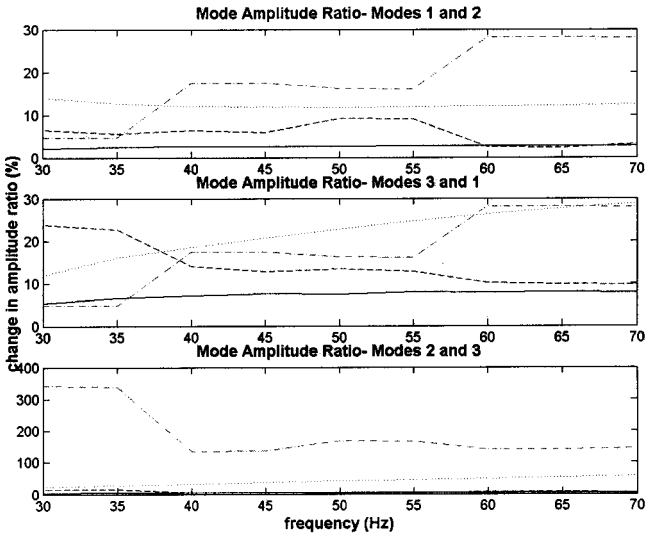


FIG. 4. Effect of variations in the source depth, receiver depth, range, and modal amplitude ratio on the spectral ratios. Solid line indicates variations due to 1.5-m source depth increase from baseline, dotted line indicates 1.5-m increase in receiver depth, dash dot line 30% increase in beta, and dashed line indicates 1.5-km increase in range. The top, middle, and bottom panels show the ratio of modes 1 and 2, modes 1 and 3, and modes 2 and 3, respectively.

C. Sensitivity of modal amplitude ratios

A sensitivity study was performed to understand the effect of model parameters on the spectral amplitude ratios. Figures 4 and 5 show the influences of source depth, receiver depth, modal attenuation coefficient, and range on modal amplitude ratios. This analysis was done using the sound-speed profile obtained from the inversion for sediment compressional speeds (Potty *et al.*, 2000). Figure 4 shows the variation of modal amplitude ratios with changes in source depth (+1.5 m), receiver depth (+1.5 m), range (+1.5 km), and modal attenuation coefficient (+30%) from the baseline model. The baseline model corresponds to a source depth of 18 m, receiver depth of 66 m, and at a range of 41 km. Figure 5 shows the variations in modal amplitude ratios with changes in source depth (−1.5 m), receiver depth (−1.5 m), range (−1.5 km), and modal attenuation coefficient (−30%) from the baseline model. The changes in the modal amplitude ratio of mode 3 to mode 1 seem to be comparatively higher than the ratios of modes 2 to 1 and modes 2 to 3. The effect of receiver depth variation is more prominent when compared to source depth variation (Figs. 4 and 5). A ±30% change in modal attenuation coefficient produces far higher variations in modal amplitude ratios, when compared to reasonable changes in range and source depth. Considering the

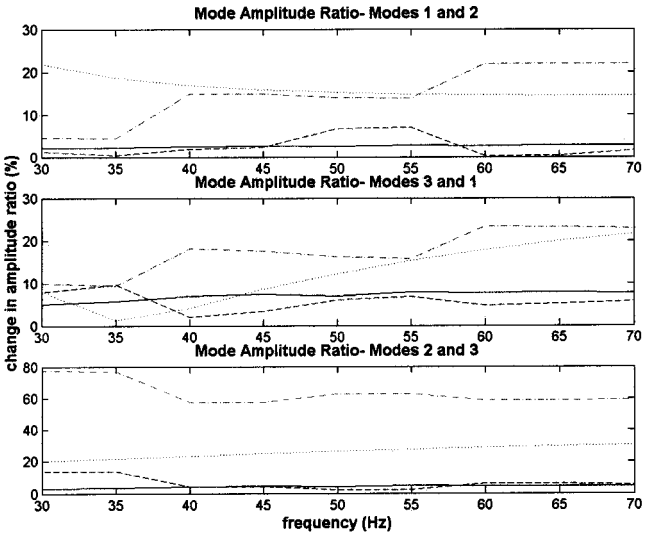


FIG. 5. Effect of variations in the source depth, receiver depth, range, and modal amplitude ratio on the spectral ratios. Solid line indicates variations due to 1.5-m source depth decrease from baseline, dotted line indicates 1.5-m decrease in receiver depth, dash dot line 30% decrease in beta, and dashed line indicates 1.5-km decrease in range. The top, middle, and bottom panels show the ratio of modes 1 and 2, modes 1 and 3, and modes 2 and 3, respectively.

influence of source depth, depth of receiver, and range, all three of these parameters were included as unknowns in the inversion scheme for determining the modal attenuation coefficients.

IV. PRIMER FIELD STUDY

A. General description of the Shelf Break PRIMER experiment

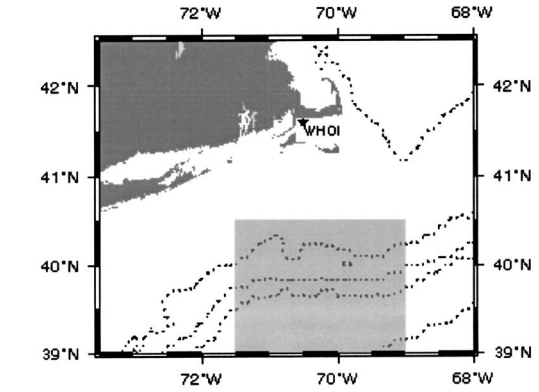
In the summer of 1996, a number of oceanographic and acoustic measurements were taken on the shelf break south of New England in the Middle Atlantic Bight (Fig. 6). Details of the experiment, with emphasis on bottom inversion, are discussed by Potty *et al.* (2000). The SUS component of the experiment involved acquisition of broadband acoustic data on two vertical line arrays (VLAs) on the northeast (NE) and northwest (NW) corners of the experimental area. The SUS charges were of type MK61 and consisted of 0.82 kg of TNT set to detonate at a depth of 18 m. About 80 charges were dropped on the continental shelf and slope in water depths varying from 85 to 300 m. In this study acoustic signals from SUS explosions in the shelf area received at the NE VLA are used to invert for geoacoustic parameters. The positions of these SUS drops and the location of the NE VLA are shown in Fig. 7. The three SUS explosions D2, D4, and D6 are part of the down-slope run, whereas C6 and C9 are part of the cross-slope run. Signals received from these shots on the NE vertical array were analyzed in the present study for compressional wave speed. The NE VLA consisted of 16 hydrophones spanning the water column from a depth of 45.42 to 92.72 m. Data were acquired on the receivers at a sampling frequency of 1395.1 samples/second. Signals from D1 and D2 are used to calculate the modal attenuation coefficients.

The acoustic signal received at the VLA is analyzed using wavelet-based methods to produce time–frequency dispersion diagrams. The advantage of analyzing the signal with wavelets as analyzing kernels is that it enables us to study features of the signal locally with a detail matched to their scale, i.e., broad features on a large scale and fine features on a small scale. This enables us to get the time of arrivals corresponding to higher modes with good resolution. This becomes important since the higher modes penetrate farther into the sediment and hence enable us to invert for deeper sediment properties. The arrival times corresponding to spectral peaks for the various modes at different frequencies are picked from the time–frequency scalograms. It should be noted that there might be multiple peaks corresponding to bubble pulses generated by the explosion. This sometimes makes the identification of the peaks and their corresponding arrival times difficult for various modes. This is a serious problem for the early arrivals above 70 Hz where the arrivals corresponding to various modes are closer.

Figure 8 shows the modal dispersion for four SUS signals (D2, D4, C6, and C9) received at 45.42 m at the NE VLA. The SUS signals C6 and C9 are closer to the VLA compared to the other two, and hence the arrivals are not as well spread out in time. From these wavelet scalograms arrival time data corresponding to individual modes can be extracted, especially those corresponding to late arrivals. The arrival pattern is similar at other depths, except that the relative strengths of the modes differ at these depths. Modes 1 to 9 can be identified from the time–frequency diagram for SUS signal D2. It can also be noted that overall the individual modes are identifiable and well separated, which indicates no major coupling of energy between modes while they propagate. This feature facilitated good quality inversions using this data.

B. Geoacoustic data at the PRIMER site

During the Shelfbreak PRIMER experiment oceanographic parameters such as temperature and salinity were monitored using various methods including SEASOAR (Gawarkiewicz *et al.*, 2001) measurements. These measurements were useful in generating background sound-speed profiles for the ocean and also the empirical orthogonal functions (EOFs) used to represent the sound-speed variations in the ocean. The region adjacent to the PRIMER experimental site has been investigated for sediment properties extensively by many investigators. A detailed review of these studies is presented by Potty *et al.* (2000). The top layer of sediment in the experimental site consists of fine-grained sediments below which sands of fine or medium grain size are found. Only one deeper core (AMCOR Project site 6012) is available within the experimental area. This AMCOR site is down-slope in the southwest corner of the experimental site. Compressional wave-speed profiles have been computed using this core data based on the Biot–Stoll model. Another nearby site is AMCOR 6010, which is on the shelf at a shallower water depth southwest of the experimental site. Trevorrow and Yamamoto (1991) have computed the compressional wave-speed profile at this location. Comparing these two sites, it can be observed that the surface fine-



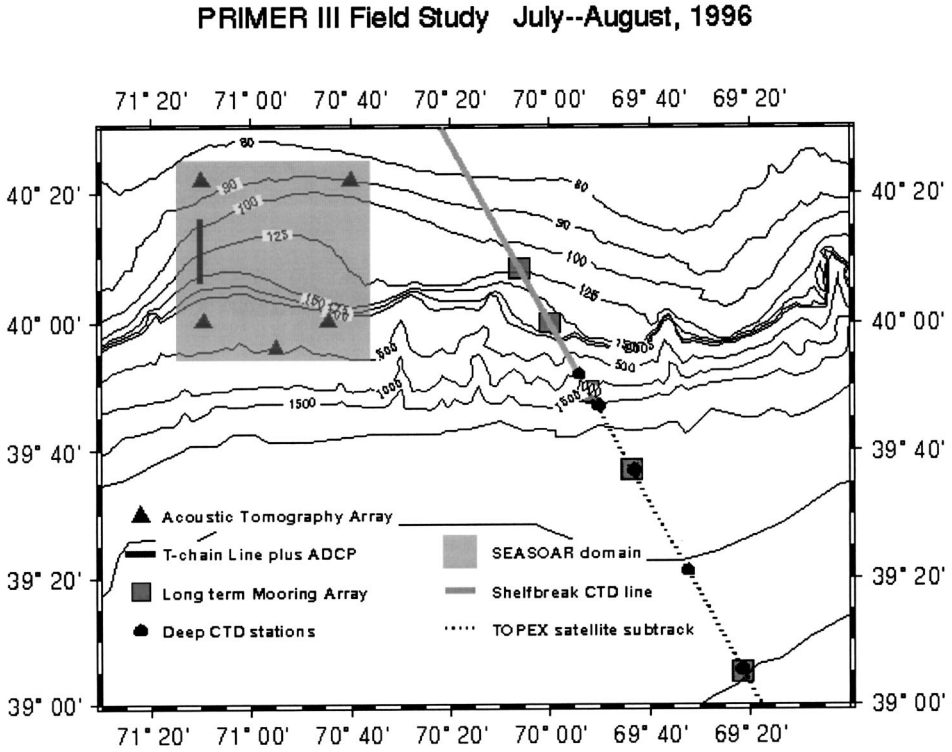


FIG. 6. Location of the PRIMER experimental study.

grained layer becomes thinner along the slope. The mean velocity in the top 10 m of the sediment measured using seismic methods was approximately 1650–1675 m/s (McGinnis and Otis, 1979).

In order to get more data for the top few meters of the sediment, gravity cores were taken in three locations on the shelf at water depths of approximately 90 m, and two more in deeper water depths on the slope (Fig. 7). These cores penetrated down to a maximum depth of 1.5 m in the shelf locations. Beyond this depth the core encountered sandy sediments and refused to penetrate further. For the slope locations the penetration was less due to the presence of the sandy layer at shallower depths. The cores were logged at the Marine Geomechanics Laboratory at the University of Rhode Island to obtain the compressional wave speed, bulk density, and attenuation profiles. It should be noted that these gravity cores give geoacoustic parameters for the top 1–2 m of the sediment only. The present inversion gives compressional speed values down to 25 m depth with reasonable quality. Hence, the gravity cores are useful only to a limited extent for comparison and validation of the present inversion. However, the gravity core data will be much more important for propagation at higher acoustic frequencies.

V. RESULTS AND DISCUSSION

A. Compressional wave-speed inversion: Inversion scheme I

We now pursue the inverse problem. To begin with, the acoustic signals from the SUS charge explosions received at the VLA are now analyzed to evaluate the time–frequency distribution. A signal recorded on a single hydrophone from SUS explosion D2 is shown in Fig. 9. Frequency dispersion is clearly observable even in the raw time series, especially for the later arrivals. We can observe high-frequency late arrivals at 3.5 s after the onset and some low frequencies arriving earlier at around 2.5 s. A Morlet wavelet, commonly used in geophysics and acoustical analysis (Badiey *et al.*,

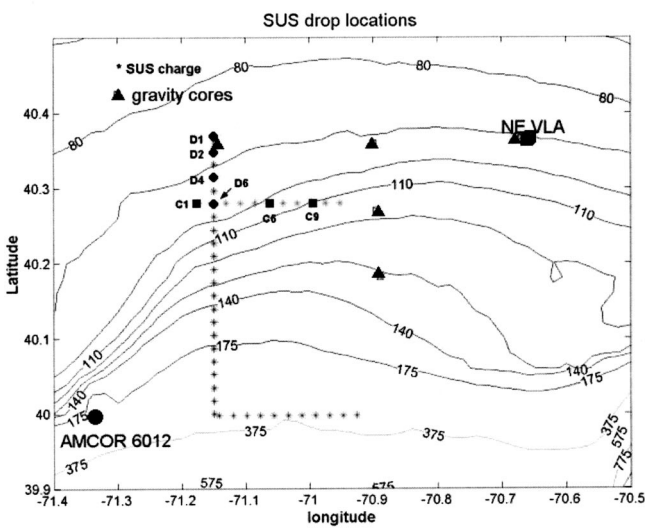


FIG. 7. Location of the SUS charges analyzed in this study. D1, D2, D4, and D6 are part of the down-slope run, whereas C1, C6, and C9 are from the cross-slope run. Locations of the gravity cores and the AMCOR drill site are also shown.

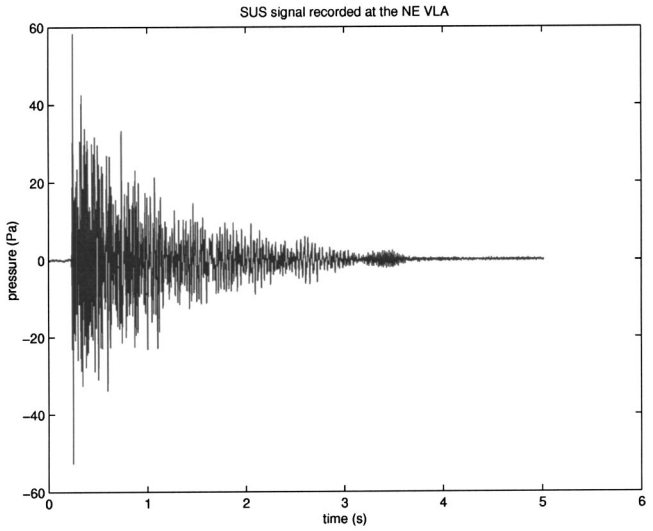


FIG. 9. SUS signal received at the top hydrophone at 45.42-m depth. The source to receiver range is 41 km.

1994), is then used to produce the time–frequency diagram of the signal as shown in Fig. 10. It is seen that the arrivals corresponding to 3.5 and 2.5 s observed in the time series belong to mode 4 and mode 2, respectively. The continuous lines shown in this figure correspond to the theoretical group speed values obtained by the inversion scheme. It should also be noted that, by looking at the time–frequency diagrams for all the receiving hydrophones at various depths along the array, we can extract arrivals for most of the modes. That is, we do not lose modes due to the receiver being in a modal null. The times of arrivals corresponding to the various modes over our range of frequencies were peak picked and used as data for the inversion. The widths of the spectral peaks corresponding to 95% of peak value were also estimated from the time–frequency diagrams. These widths were assumed to represent the uncertainty in the data. Acoustic signals from SUS explosions D2, D4, D6, C6, and C9 were analyzed for time–frequency behavior. All these explosions are in the shelf region and the bathymetry is gently varying. The modes are well defined and there is not much indication of mode coupling.

Forward propagation was modeled using adiabatic theory. In this inversion scheme, the propagation path is divided into five sections. Sediment properties in each section are assumed range independent. Up to nine layers of sediments with different compressional speeds and unequal thickness were considered. Close to the sediment–water interface the layers were very thin compared to deeper depths where penetration of acoustic energy is very low. Empirical orthogonal functions (EOFs) were generated using ocean sound-speed profiles at various locations in the propagation path calculated using SEASOAR temperatures. Four EOF coefficients were used to represent the variations at each section. The water depth at each section and the source–receiver

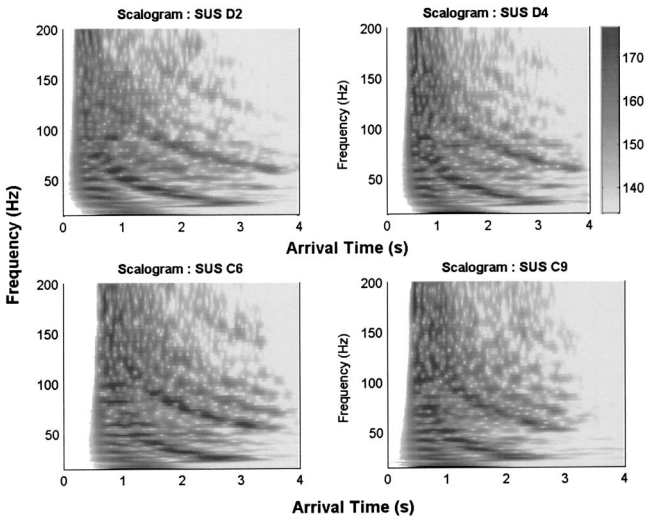


FIG. 8. Dispersion diagrams for four SUS signals received at a depth of 45.42 m. The arrival times are arbitrary. Locations of these SUS drops are shown in Fig. 7.

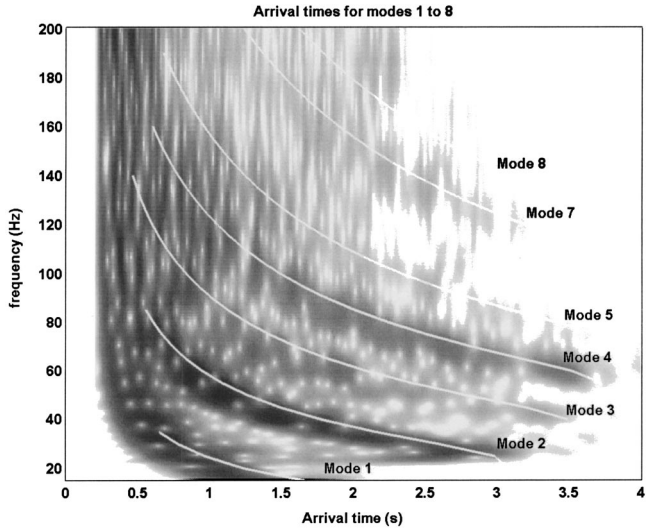


FIG. 10. Comparison of arrival times for SUS signal from D2. The continuous lines are the theoretical group speed curves calculated using the inversion.

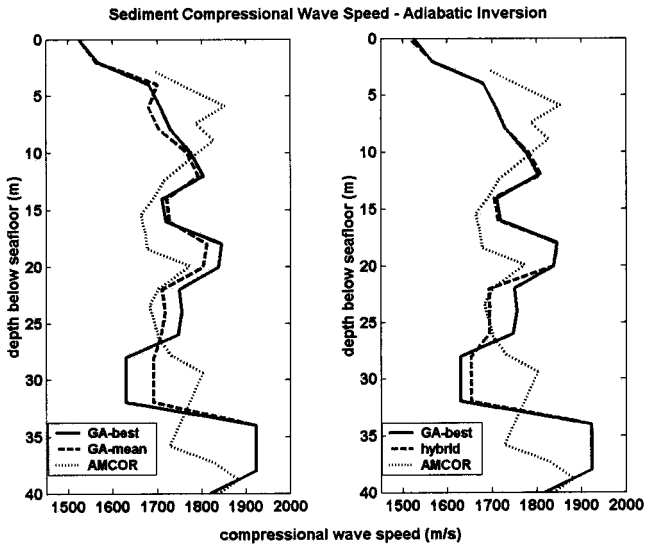


FIG. 11. Sediment compressional speeds obtained by GA inversion (left panel) and hybrid inversion (right panel). Compressional speeds calculated using the AMCOR data are also shown.

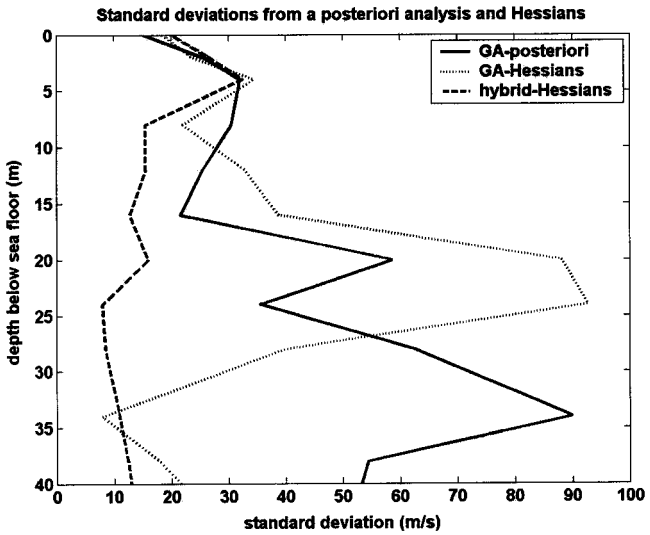


FIG. 12. Standard deviation of the sound-speed estimates computed by *a posteriori* error analysis. Local estimates calculated using Hessians are also shown.

range were also included as unknowns in the inversion scheme for compressional wave speeds. Hence, at each section four EOF coefficients, compressional wave speeds at nine layers, water depth, and range were used to model the environmental and geometric properties. Shear effects in the sediments were neglected. A normal-mode program (Porter and Reiss, 1984) was used to calculate the eigenvalues and mode shapes at each section. The group speed values for various frequencies and modes were calculated and matched with the experimentally observed group speeds. A GA was used to minimize the difference between the modeled and observed group speed values in a least-square sense via the objective function [Eq. (1)]. The GA was used with a stochastic universal sampling selection algorithm, real mutation, and discrete recombination. Three separate runs were made to verify the robustness of the model. The sampled model space is stored and used to calculate the *a posteriori* error estimates. The procedure for the calculation of the error estimates is described in detail by Potty *et al.* (2000).

Figure 11 (left panel) shows the path-averaged compressional wave speeds obtained by the inversion using signals from SUS D2. Compressional wave-speed profiles corresponding to the mean and best parameters obtained by the GA inversion are plotted along with the AMCOR profile for comparison. The inversion agrees with the AMCOR profile reasonably well considering the fact that the AMCOR location is approximately 40 km down slope. The mean and best profiles agree closely for the top 25 m, indicating good convergence at these depths. Figure 11 (right panel) shows the improvements achieved by the application of our hybrid method. It can be noted that the application of Levenberg–Marquardt methods did not produce appreciable improvement in this case, especially in the top 25 m. Figure 12 shows the standard deviations computed using *a posteriori* analysis and Hessians. The standard deviation is of the order of 15–20 m/s in the top 25 m. Down to a depth of 20 m, the Hessians are very high, which indicates very good convergence and explains the lack of success of the hybrid method. It can also be noted that the standard deviation calculated using these two different methods matches very well. Both the methods show very large errors in the 20–25-m depths. The hybrid method was effective at these depths, as it reduces the errors in this region. It should be noted that uncertainty generally increases with depth due to reduced modal penetration at greater depths. Uncertainty calculated by different methods showed this trend at depths greater than 30 m. We also see fluctuations in uncertainty at depths lower than 30 m. This may be due to the fact that different modes penetrate to different depths and hence sample different layers of sediment. This will be more pronounced as we go deep, as the inversion there is based on relatively smaller number of modes.

Figure 13 shows the compressional speed profile in the top 1.4 m of the sediment. Core data obtained from the grav-

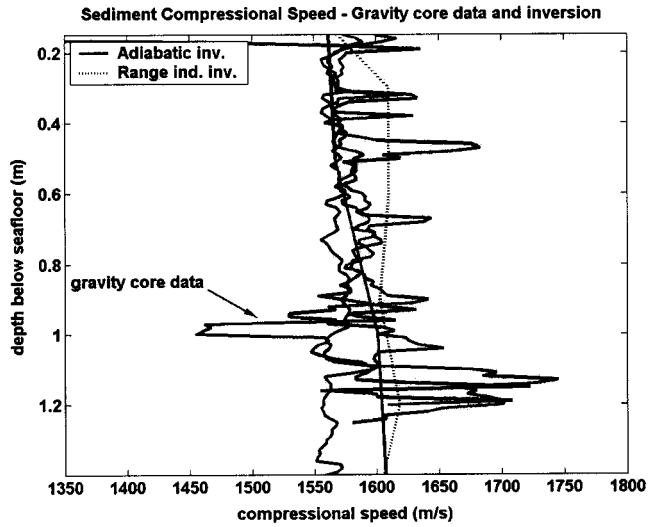


FIG. 13. Sediment compressional speeds obtained by hybrid inversion. Compressional speeds calculated using the gravity core data are also shown. The range-independent inversion shown in the figure is from Potty *et al.* (2000).

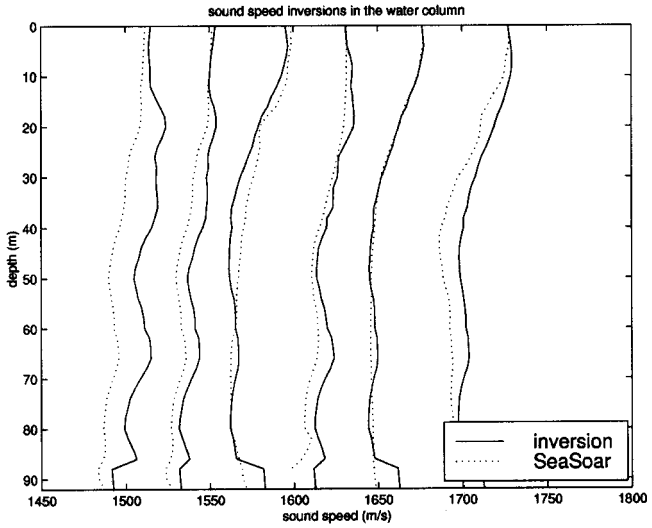


FIG. 14. Sound speeds in the water column obtained by GA inversion. The profiles at various sections are shown with sound-speed axis shifted by a small amount. Sound-speed profiles computed using SEASOAR data are also shown.

ity cores at locations 1, 2, and 3 (Fig. 7) and the range-independent inversion (Potty *et al.*, 2000) are also shown in that figure. Improvement in the inversion due to adiabatic modeling is evident, especially in the top 1 m of the sediment. Figure 14 shows the comparison of the sound speeds in the water column at six range points along the propagation path. Each profile is offset by 50 m/s for clarity of presentation, and hence only the difference between the two profiles is relevant. There is some disagreement between the two profiles, which is reflected in the group speed comparisons (Fig. 15). It should be noted that only four EOF coefficients were included to model the sound-speed fluctuations in the water column. Figure 15 shows the comparison of group speeds computed using the inversion and experimental data. Agreement is good for the late modal arrivals except for mode 1. There is appreciable difference in early arrivals, which may be due to the inferior ocean sound-speed values. It should also be noted that early arrivals are difficult to

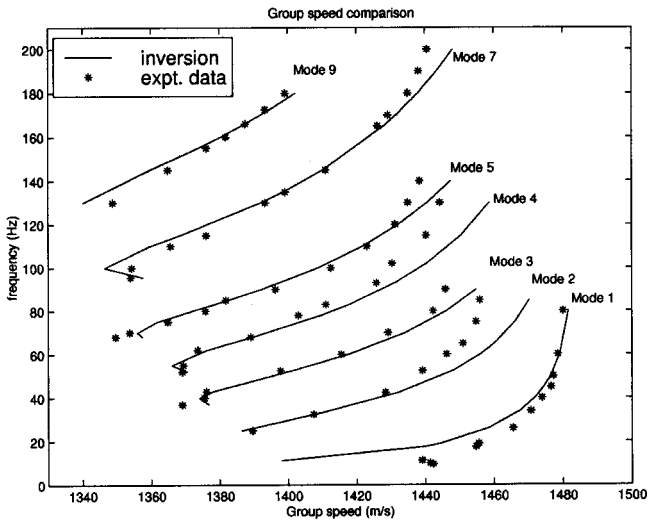


FIG. 15. Comparison of group speeds calculated using the inversion and experimental values of arrival times.

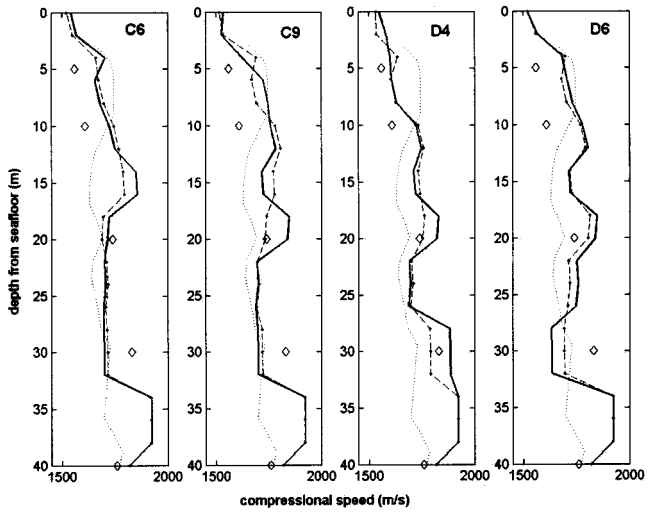


FIG. 16. Compressional speed profiles obtained by the inversion using SUS signals from D4, D6, C6, and C9. The continuous and dashed lines represent the sound-speed profile corresponding to the best and the mean inversions, respectively. The dotted line and the diamonds represent the AMCOR 6012 and 6010 data, respectively.

identify and separate and are prone to data errors compared to late arrivals. Later arrivals are more important to sediment inversions as they interact more with the bottom.

Figure 16 shows the compressional sound-speed profiles obtained by this inversion scheme using SUS signals corresponding to D4, D6, C6, and C9. The locations of these SUS charges are shown in Fig. 7. D4 and D6 belong to the downslope run, whereas the other two (C6 and C9) belong to the cross-slope run. The compressional wave speeds corresponding to the two AMCOR cores (6010 and 6012) are also shown in these figures. The variations in compressional wave speeds are not much, and in most cases they are within the standard errors. This was expected as the propagation paths for these four SUS signals are close to each other. Figure 17 shows the mean compressional speed profile on the shelf.

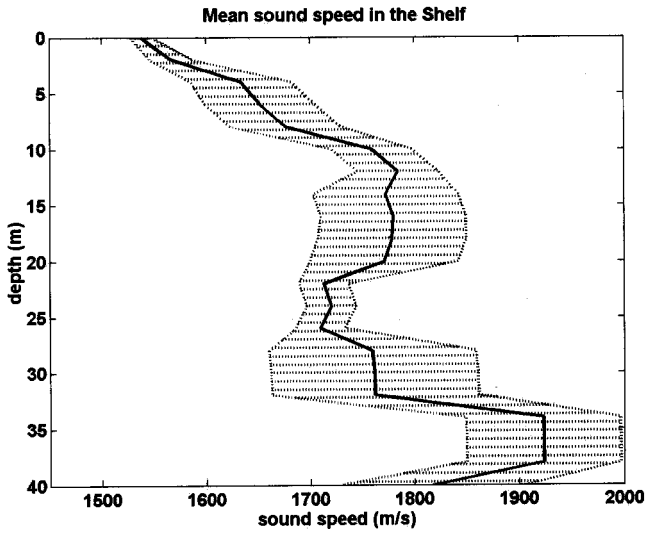


FIG. 17. Mean compressional speed profile in the shelf region. The shaded region corresponds to one standard deviation on either side of the mean. This mean profile is obtained using the inversions of SUS signals from various shots.

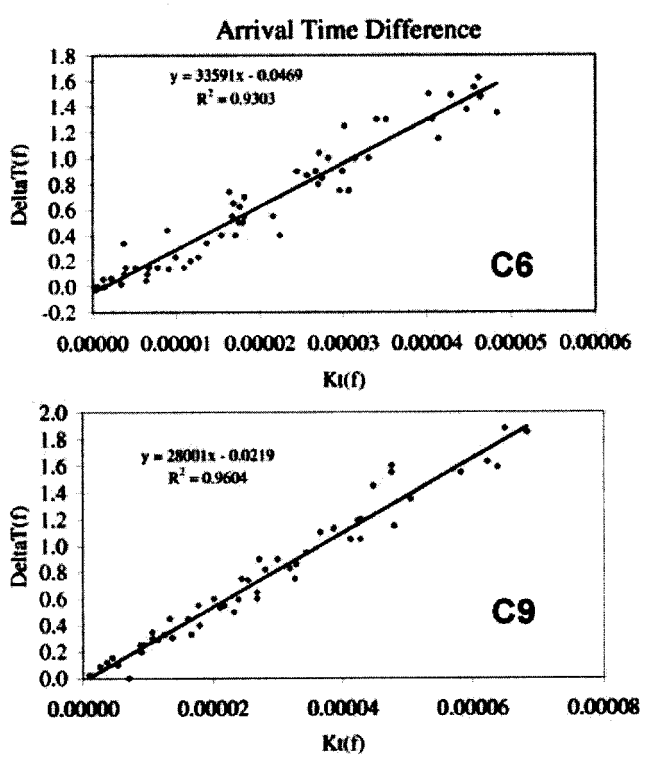


FIG. 18. Source–receiver range estimation for SUS charge C6 (top panel) and C9 (bottom panel). The estimated range is very close to the deployment range of 34.7 and 29.4 km, respectively. DeltaT(f)[$\Delta T(f)$] and $K_t(f)$ are defined in Sec. II C.

This mean profile is obtained using the inversion results corresponding to various SUS shots deployed in the shelf region. The mean compressional speed in the top 2 m of the sediment is of the order of 1550 m/s, which agrees well with the gravity core data. At greater depths the compressional speeds are of the order of 1700 m/s, which matches the average AMCOR-6012 data. The standard error in all the cases was of the order of 20–25 m/s in the top 15 to 20 m.

B. Source–receiver range estimation results

The differences in arrival times for various frequencies for a given mode and for various modes at a given frequency are calculated from the time–frequency distribution of the acoustic signal using Eqs. (5) and (6). The group speed values for various modes and frequencies are calculated using the compressional speed profile obtained from the adiabatic inversion. For two SUS charge explosions (C6 and C9), the experimental arrival time differences are plotted against the group slowness differences [Eqs. (5) and (6)]. It should be noted that these equations are strictly applicable only for range-independent propagation. The slopes of the lines obtained from these equations give the range for these SUS explosions (Fig. 18). The ranges were obtained as 33.6 and 28 km. These values are very close to the experimentally measured distance between the deployment locations of these SUS charges and the VLA (34.7 and 29.4 km, respectively). It should be noted that this estimation of range is based on range independence. The agreement between the estimated and actual ranges supports our assumption of range independence.

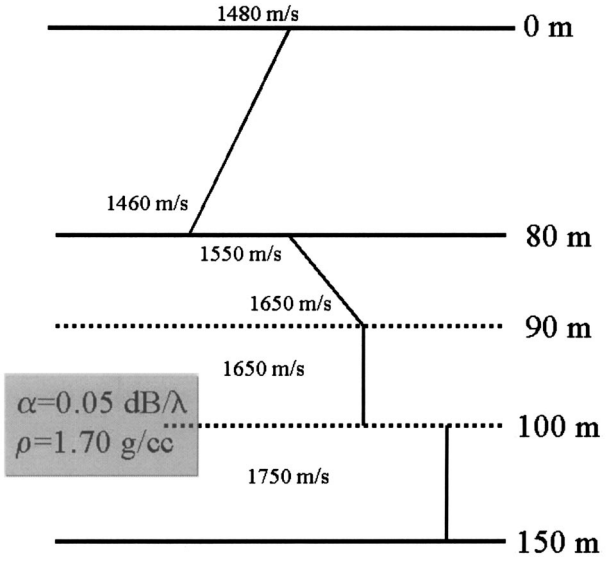


FIG. 19. Sound-speed profile used to generate synthetic time series for attenuation inversion.

C. Results of attenuation inversions: Inversion scheme II

1. Synthetic data

The inversion scheme II (Fig. 2) for obtaining the compressional wave attenuation was tested using synthetic data. Synthetic time series is generated for the sound-speed profile shown in Fig. 19. The attenuation coefficient for the sediment is assumed constant and equal to 0.05 dB/λ. Density also was assumed constant at 1.7 g/cc. The source was placed at a depth of 30 m and the receiver was at 40 km at a depth of 50 m. Acoustic pressure was generated for frequencies 10 to 200 Hz using the parabolic equation (PE)-based propagation code RAM (Collins, 1997). The time series thus obtained was analyzed using wavelets to separate individual mode arrivals. The arrivals obtained at 50 Hz are shown in Fig. 20. Modes 1, 2, and 3 can easily be identified in this figure. The ratios of pressure amplitudes are then calculated for use in the inversion scheme. Figure 21 shows the attenuation profile obtained using synthetic data. It can be noted that the inversion was successful in estimating the attenuation coefficient as the inversion closely matches the true attenuation value. It should also be noted that the synthetic data were noise-free and hence the spectral peaks were well separated and easily identifiable.

2. SUS data

This inversion scheme was next used to obtain the modal attenuation coefficients using the SUS data. The data from two explosions (D1 and D2) were chosen for this inversion. Signals received at the NE VLA at depths of 45.42, 53.12, 66.32, and 79.52 m were used. The propagation paths from these SUS explosions to the VLA are in a uniform depth of water and thus are assumed to be range indepen-

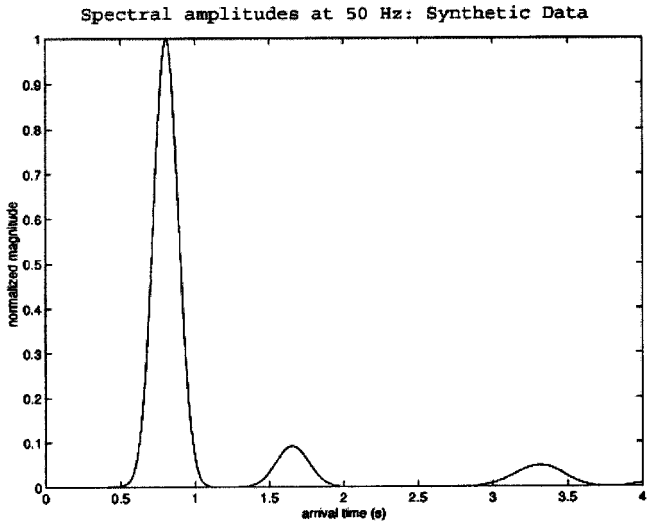


FIG. 20. Normalized magnitudes of the acoustic pressure for modes 1, 2, and 3 obtained using time–frequency analysis.

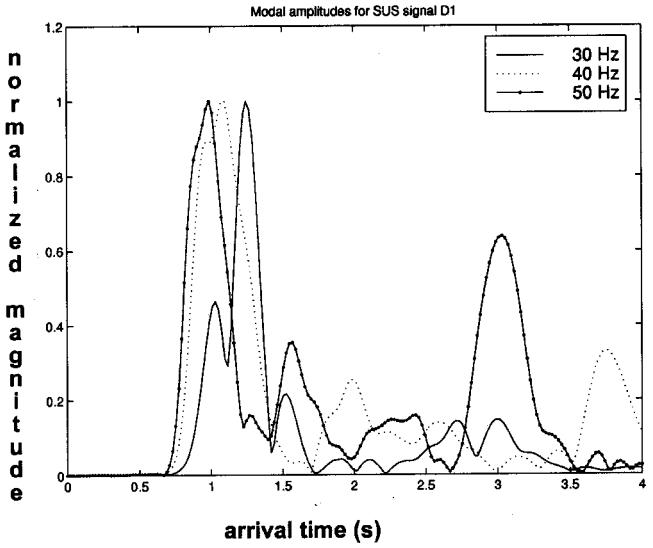


FIG. 22. Magnitudes of the acoustic pressure for SUS signal D1. Ratios between amplitudes of modes 1, 2, and 3 are used in the inversion scheme.

dent. This ignores any intrinsic medium range variability, which could be a source of some error. The time series from SUS D1 and D2 received at the VLA was analyzed using a Morlet wavelet. Figure 9 shows one such acoustic signal from SUS D2 received at the middle hydrophone at 66.32-m depth. Figure 10 shows the scalogram of this signal. It can be seen from this figure that in the 20- to 80-Hz region most of the acoustic energy is contained in the first three modes. Also, the first and second modes are well separated in this frequency band so that spectral ratios can be easily found. Figure 22 shows the normalized acoustic pressure amplitudes at frequencies 30, 40, and 50 Hz from the explosion D1. Spectral ratios R_{21}, R_{31}, and R_{32} are calculated using these mode amplitudes and are then used in the inversion to find the modal attenuation coefficients. It should be noted that at 70 Hz and higher, it is very difficult to identify the individual modal peaks, which makes the attenuation estimates less reliable at these frequencies. After obtaining the modal attenuation coefficients the attenuation profile $\alpha(z)$ was calculated by solving the integral equation (4). Inversion was carried out for frequencies in the range of 30–70 Hz and the results for frequency 30, 40, and 50 Hz are shown in Fig. 23. The attenuation profile calculated using gravity core data is also shown for comparison. Gravity core data are available only for a short depth of 1.5 to 2 m. The results agree reasonably well for the frequency shown in this figure and for other frequencies in the range 30–70 Hz. At higher frequencies the quality of the estimates becomes poor, which results in a comparatively larger disagreement. It should also be noted that the inversions corresponding to explosions D1 and D2 are in close agreement, as seen in the figure.

The attenuation coefficients are on the order of 0.04 to 0.045 dB/λ. In Fig. 24 the attenuation estimates (in dB/m) obtained by the present study are plotted along with previously available data reported by other investigators (Stoll, 1985). The shaded strip in this figure corresponds to the at-

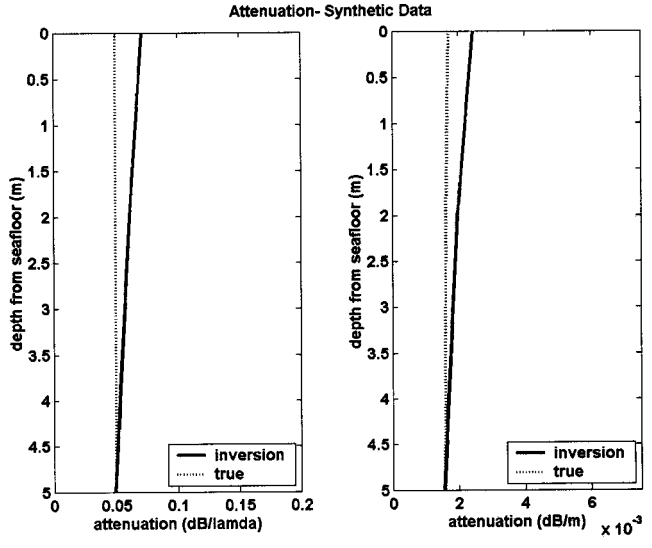


FIG. 21. Attenuation estimates obtained using the inversion scheme for the synthetic data. The true attenuation in this case was 0.05 dB/λ. Model attenuation coefficients were obtained from the inversion scheme and attenuation profile was estimated by linear inversion.

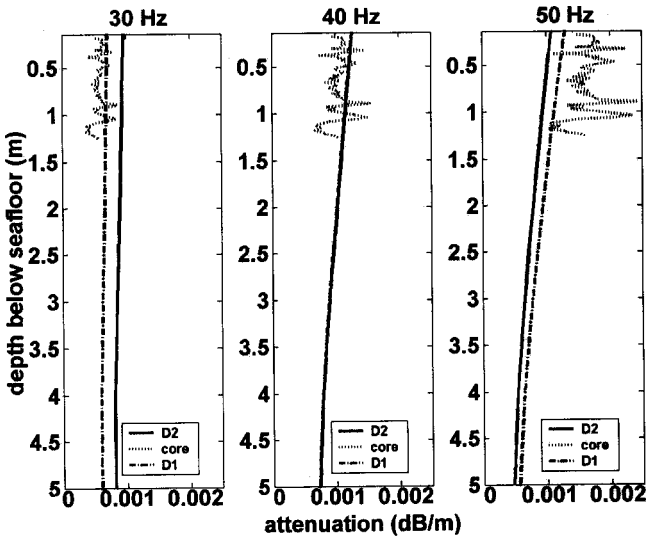


FIG. 23. Attenuation coefficient profile at 30, 40, and 50 Hz.

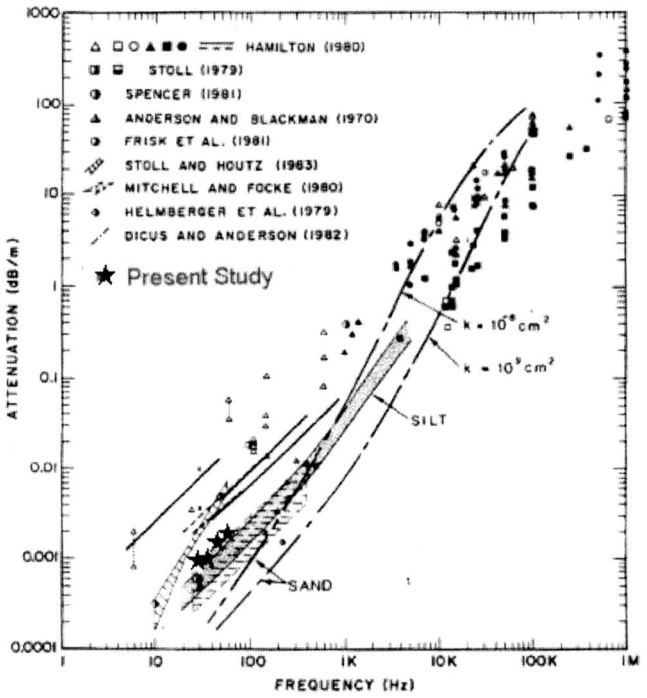


FIG. 24. Attenuation estimates (dB/m) obtained from present inversion. The historical data in the figure is taken from Stoll (1985).

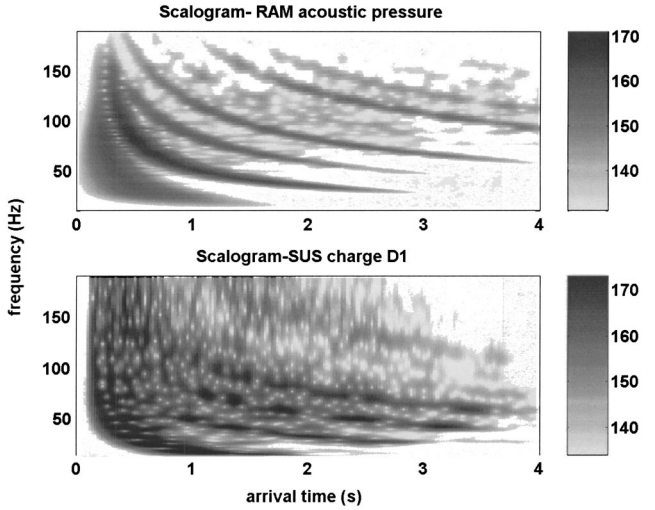


FIG. 25. Time–frequency diagrams of the synthetic and field data are shown in the top and bottom panels, respectively. The compressional wave speed and attenuation values from the inversion were used to generate the synthetic data.

tenuation values calculated using the Biot model (Badiey *et al.*, 1998) for sediments of the silt type. It should be noted that the amount of data at frequencies lower than 1 kHz is very small compared to higher frequencies. Values corresponding to our inversion range from 0.0009 to 0.0015 dB/m in the frequency range 30 to 70 Hz. These values are well below the values found by Hamilton (1972) for these frequencies. It should be noted that the Hamilton values were found by extrapolating the higher frequency findings based on linear first-power frequency dependence. If we calculate the value of attenuation coefficient with a value of $k=0.5$ (corresponding to very fine sand) from Hamilton (1972), and use the relationship

$$\alpha = k f^n \quad (n=1), \tag{21}$$

we get a value of 0.0025 dB/m for attenuation coefficient at a frequency of 50 Hz. The actual field values may be less than this since, in both sands and silts, the attenuation versus frequency relationship becomes nonlinear at low frequencies (Kibblewhite, 1989). This behavior is intimately related to the permeability of the sediment. Moreover, errors based on such extrapolation will be further compounded by the fact that, at low frequencies, propagation will also be controlled by deeper underlying structures rather than by the near-surface sediments alone. The vertical resolution length calculated was of the order of 4 m in the present inversion. The attenuation values then correspond to an average over a depth of 4 m, and surface sediments and deeper sandy sediments also. Another mechanism which can cause propagation at lower frequencies to be strongly attenuated is the presence of shear waves. However, this effect will be important only if the shear waves are on the order of 600 m/s. In the present study shear wave effects are probably not an important factor, as the shear speeds calculated using core data are on the order of 150 m/s.

Rogers *et al.* (1993), Mitchel and Focke (1983), and Zhao (1985) have reported attenuation values lower than Hamilton's predictions and closer to the present inversion. AMCOR-6010 site, which is southwest of the experimental location but in much shallower waters, consists of mostly silty sand silty clay overlying sand and a sandy clay layer. Rogers *et al.* (1993) have calculated the values of attenuation in the frequency range of 50 to 60 Hz from *in situ* measurements here. These values are 0.000 717 5 dB/m at 50 Hz and 0.001 09 dB/m at 75 Hz. Reflection data from the water–sediment interface measured by Mitchel and Focke (1983) yielded attenuation values of similar magnitude in deep ocean sediments in the 20–400-Hz frequency range. These values are also shown in Fig. 24. Zhao (1985) has reported a value of 0.0022 dB/m for attenuation coefficient at a site in the Yellow Sea off China's east coast at a water depth of 28.5 m. This value corresponds to a sand–silt–clay sediment at a frequency of 80 Hz. Even though these values correspond to different geographical locations, they all fall in the range of values closer to our inversion at these low frequencies.

Figure 25 shows the dispersion diagrams for synthetic data using the compressional speeds and attenuation obtained by the inversion. These are compared with the dispersion diagram for the experimental data. The synthetic data are noise-free and hence the individual modes are well separated. The first four modes are strong in both the real and synthetic data. Modes 5 and 6 are clearly identifiable in the synthetic data as these are noise-free as opposed to the real data. The pressure levels compare reasonably well. The source level for the explosive charge was evaluated using the method detailed by Urick (1983). The depths of receivers, depth of sources, and range obtained by the inversion are shown in Table I. They seem to agree well with the values noted at deployment.

TABLE I. Results of inversion scheme II using amplitude ratios.

Hydrophone number	Depth at deployment (m)	Depth-inversion (m)
1	45.42	44.60
4	53.12	52.00
8	66.32	65.30
12	79.52	79.59
Source depth (m)	18.29	18.62
Range (km)	40.96	40.82

VI. CONCLUSIONS

Sediment compressional speeds were evaluated using hybrid optimization schemes using broadband SUS data. Propagation was modeled using adiabatic mode theory. The inversions compare well with AMCOR and gravity core data. Compressional attenuation was obtained using another inversion scheme based on spectral ratios. This inversion scheme was tested using synthetically generated data. The attenuation values obtained fall within the reported compressional attenuation values in the frequency range 30–70 Hz. Source/receiver ranges were also estimated assuming approximate range independence, and agree well with measured values.

The possibilities of range variations in the bottom compressional wave speeds will be investigated further in another study. The up-slope propagation from the shots from the slope region will be more complicated because of the severe range variations in bathymetry, the range-dependent shelf break front, and probable variations in the bottom compressional wave speeds.

ACKNOWLEDGMENTS

The authors wish to thank the scientific team of the PRIMER experiment including Robert Beardsley, Ken Brink, Ching-Sang Chiu, Glen Gawarkiewicz, Robert Pickart, Alan Robinson, and Brian Sperry, Captain and Crew of R/V ENDEAVOR, and the pilots and crew from the Naval Air Warfare Center at Patuxent River NAS for their significant contributions to data collection and analysis. In particular, the authors recognize David Volak and Thomas Gabrielson for their contribution to the SUS component of the experimental design. This work was supported by Office of Naval Research, Dr. Jeffrey Simmen, Code 321 OA Program Manager and Dr. Steven Ramp, Code 322 PO Program Manager.

Badiey, M., Cheng, A. H.-D., and Mu, Y. (**1998**). ''From geology to geoacoustics—Evaluation of Biot–Stoll sound speed and attenuation for shallow water acoustics,'' J. Acoust. Soc. Am. **103**(1), 309–320.

Badiey, M., Jaya, I., and Cheng, A. (**1994**). ''Shallow water acoustic/geoacoustic experiments at the New Jersey Atlantic Generating Station site,'' J. Acoust. Soc. Am. **96**(6), 3593–3604.

Caiti, A., Akal, T., and Stoll, R. D. (**1994**). ''Estimation of shear wave velocity in shallow marine sediments,'' IEEE J. Ocean. Eng. **19**, 58–72.

Collins, M. D., Kuperman, W. A., and Schmidt, H. (**1992**). ''Nonlinear inversion for ocean-bottom properties,'' J. Acoust. Soc. Am. **92**, 2770–2783.

Collins, M. D. (**1997**). ''A split-step Pade solution for the parabolic equation method,'' J. Acoust. Soc. Am. **93**, 1736–1742.

Fletcher. (**1980**). *Practical Methods of Optimization* (Wiley, New York).

Gawarkiewicz, G., Bahr, F., Beardsley, R., and Brink, K. (**2001**). ''Interaction of a slope eddy with the shelfbreak front in the Middle Atlantic Bight,'' J. Phys. Oceanogr. **31**, 2783–2796.

Gerstoft, P. (**1994**). ''Inversion of seismoacoustic data using genetic algorithms and *a posteriori* probability distributions,'' J. Acoust. Soc. Am. **95**(2), 770–781.

Gerstoft, P., and Gingras, D. F. (**1996**). ''Parameter estimation using multifrequency range-dependent acoustic data in shallow water,'' J. Acoust. Soc. Am. **99**(5), 2839–2850.

Goldberg, D. (**1988**). *Genetic Algorithms* (Addison Wesley, Reading, MA).

Hamilton, E. L. (**1972**). ''Compressional wave attenuation in marine sediments,'' Geophysics **37**, 620–646.

Hamilton, E. L. (**1980**). ''Geoacoustic modeling of sea floor,'' J. Acoust. Soc. Am. **68**(5), 1313–1340.

Jensen, F. B., and Schmidt, H. (**1985**). ''Shear properties of ocean sediments determined from numerical modeling of Scholte wave data,'' in *Ocean Seismo-Acoustics*, edited by T. Akal and J. M. Berkson (Plenum, New York).

Kessel, R. T. (**1999**). ''A mode-based measure of field sensitivity to geoacoustic parameters in weakly range dependent environments,'' J. Acoust. Soc. Am. **105**(1), 122–129.

Kibblewhite, A. C. (**1989**). ''Attenuation of sound in marine sediments: A review with emphasis on new low frequency data,'' J. Acoust. Soc. Am. **86**, 716–738.

Lynch, J. F., Rajan, S. D., and Frisk, G. V. (**1991**). ''A comparison of broad band and narrow band modal inversions for bottom geoacoustic properties at a site near Corpus Christi, Texas,'' J. Acoust. Soc. Am. **89**(2), 648–665.

McGinnis, L. D., and Ottis, R. M. (**1979**). ''Compressional velocities from multichannel refraction arrivals on George's Bank-northwest Atlantic Ocean,'' Geophysics **44**, 1022–1032.

Mitchel, S., and Focke, K. (**1983**). ''The role of the sea bottom attenuation profile in the shallow water acoustic propagation,'' J. Acoust. Soc. Am. **73**, 465–473.

Porter, M. B., and Reiss, E. L. (**1984**). ''A numerical method for ocean acoustic normal modes,'' J. Acoust. Soc. Am. **76**(1), 244–252.

Potty, G., Miller, J. H., Lynch, J. F., and Smith, K. B. (**2000**). ''Tomographic mapping of sediments in shallow water,'' J. Acoust. Soc. Am. **108**(3), 973–986.

Rajan, S. D., Frisk, G. V., and Lynch, J. F. (**1992**). ''On the determination of modal attenuation coefficients and compressional wave attenuation profiles in a range dependent environment in Nantucket sound,'' IEEE J. Ocean. Eng. **17**(1), 118–127.

Rajan, S. D., Lynch, J. F., and Frisk, G. V. (**1987**). ''Perturbative inversion methods for obtaining bottom geoacoustic parameters in shallow water,'' J. Acoust. Soc. Am. **82**(3), 998–1017.

Rogers, A. K., Yamamoto, T., and Carey, W. (**1993**). ''Experimental investigation of sediment effect on acoustic wave propagation in the shallow ocean,'' J. Acoust. Soc. Am. **93**, 1747–1761.

Sen, M., and Stoffa, P. L. (**1995**). *Global Optimization Methods in Geophysical Inversion* (Elsevier, Amsterdam).

Siderius, M., Gerstoft, P., and Nielsen, P. (**1998**). ''Broadband geo-acoustic inversion from sparse data by Genetic Algorithms,'' J. Comput. Acoust. **6**, 117–134.

Stoll, R. D. (**1985**). ''Marine sediment acoustics,'' J. Acoust. Soc. Am. **77**(5), 1789–1799.

Stoll, R. D., Bryan, G. M., and Bautista, E. O. (**1994**). ''Measuring lateral variability of sediment geoacoustic properties,'' J. Acoust. Soc. Am. **96**(1), 427–438.

Taroudakis, M. I., and Markaki, M. G. (**1997**). ''On the use of matched-field processing and hybrid algorithms for vertical slice tomography,'' J. Acoust. Soc. Am. **102**(2), 885–895.

Tindle, C. T. (**1982**). ''Attenuation parameters from normal mode measurements,'' J. Acoust. Soc. Am. **71**(5), 1145–1148.

Tolstoy, A., Diachok, O., and Frazer, L. N. (**1991**). ''Acoustic tomography via matched field processing,'' J. Acoust. Soc. Am. **89**(3), 1119–1127.

Trevorrow, M. V., and Yamamoto, T. (**1991**). ''Summary of marine sedimentary shear modulus and acoustic speed profile results using gravity wave inversion technique,'' J. Acoust. Soc. Am. **90**(1), 441–456.

Urick, R. J. (**1983**). *Principles of Underwater Sound for Engineers* (McGraw-Hill, New York).

Zhao, J. (**1985**). ''Normal mode measurements and remote sensing of sea bottom sound velocity and attenuation in shallow water,'' J. Acoust. Soc. Am. **78**(3), 1003–1009.

Zhou, J., Zhang, X., Rogers, P. H., and Jarzynski, J. (**1987**). ''Geoacoustic parameters in a stratified sea bottom from shallow water acoustic propagation,'' J. Acoust. Soc. Am. **82**(6), 2068–2074.

Echo strength and density structure of Hawaiian mesopelagic boundary community patches

Kelly J. Benoit-Bird[a)]
Hawaii Institute of Marine Biology and Department of Zoology, University of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734

Whitlow W. L. Au[b)]
Hawaii Institute of Marine Biology, P.O. Box 1106, Kailua, Hawaii 96734

(Received 18 September 2002; accepted for publication 4 August 2003)

A broadband sonar system and digital camera with strobe lights were mounted on a vertically profiling frame with a depth sensor. The echo strengths and densities of animals within individual mesopelagic boundary community patches were investigated as a function of depth. Time and distance from shore were also investigated. Simultaneous surface echosounder surveys permitted comparison of density estimates from two techniques. Echo strength values suggest nearshore boundary community animals are primarily myctophid fishes, which was confirmed by preliminary photographic evidence. Echo strength varied significantly as a function of distance from the shoreline and time. These measures of echo strength are important for estimating density from a surface echosounder. Density estimates from these revised echo strengths compare well with those made with echo highlight counting, which is independent of echo strength. These density measures suggest that previous density estimates were too low but do not change the conclusions of these studies. Vertical microstructure in density was apparent but animal size and compositional structure was not evident within a patch. Patch edges were abrupt, with no differences in the density or echo strength from patch interiors. These edges were generally straight, with a sharp drop in density to the background density of zero. Estimates of animal size as a function of time provide information about the diel migration patterns of these mesopelagic animals. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1612484]

PACS numbers: 43.30.Sf, 43.30.Ft, 43.30.Vh [WMC]

I. INTRODUCTION

The mesopelagic boundary community in Hawaii is a land-associated sound-scattering layer of small fishes, shrimps, and squids (Reid *et al.*, 1991). This community undergoes diel vertical and horizontal migrations (Benoit-Bird *et al.*, 2001). Near its inshore boundary, the boundary community is heterogeneously distributed with discrete patches forming in areas with lower animal densities (Benoit-Bird and Au, 2003c). The boundary community serves as an important food resource to many predators whose foraging behavior is strongly affected by both the layer's migration patterns and its spatial distribution (Benoit-Bird and Au, 2003b). While patch microstructure, including density and composition, is known to affect the animals within the patch and its predators, previous work has been unable to observe microstructure because of methodological limitations.

Taxonomic composition is probably the most obvious form of microstructure. Because of the large differences in swimming ability of equivalent size fishes, shrimps, and squids (Cowles, 2001; Reid, 1994; Sambilay, 1990; Yatsu *et al.*, 1999), this could have important consequences for predators. The degree of compositional heterogeneity could also have significant effects on mating opportunities and competition for food. Previous work has implied that vertical stratification in species composition and/or size classes is absent (Benoit-Bird and Au, 2003c). However, the composition of the nearshore component (within 3 km of shore) of the mesopelagic boundary community has not been directly observed.

The spatial distribution of mesopelagic animals by size is obviously important to its predators as it is correlated with the energy content (Benoit-Bird and Au, 2002) and swimming ability (Sambilay, 1990) of these prey items. The differences in swimming ability of mesopelagic animals also affect their potential migration range and consequently, their temporal distribution. In nearshore areas, migration rates of up to 1.6 km/h have been measured (Benoit-Bird and Au, 2003a). These speeds are not likely achievable by the entire size range (1–10 cm) of mesopelagic animals within the boundary community, or by all taxonomic groups. The swimming ability of animals also affects their foraging ability. In particular, the size of animals affects the size (Elner and Hughes, 1978) and type of prey they can consume (Robinson and Motta, 2002), and how much prey they require to support themselves. Consequently, larger animals probably also produce more waste, which may have important ecosystem level effects (Benoit-Bird and Au, 2003a). Body size also affects the reproductive ability; as in many nektonic marine species, fecundity increases with body size (Gross and Sargent, 1985), and body size is correlated with male mating success (Howard, 1980). The distribution of mesopelagic animals of different sizes has implications for population

a)Electronic mail: benoit@hawaii.edu
b)Electronic mail: wau@hawaii.edu

 0001-4966/2003/114(4)/1888/10/$19.00

level processes and for the role of the layer in the nearshore and pelagic ecosystems surrounding the Hawaiian Archipelago.

The density and distribution of mesopelagic animals also has consequences for individual animals, their predators, and their prey. The behavior of predators of mesopelagic boundary community animals can be directly affected by their prey's density and distribution (Benoit-Bird and Au, 2003b). Longer term, these density patterns determine the amount of food the predator can obtain, affecting its growth and survival (Beyer, 1995). The density of mesopelagic animals affects their predation risk (Walter *et al.*, 2000), particularly because some of their predators specifically forage on high densities of prey (Benoit-Bird and Au, 2003b). The impact of mesopelagic micronekton on the zooplankton on which it feeds and the competition experienced by mesopelagic animals are also likely density dependent.

Variability in animal density, size, and composition within a patch is often most noticeable near patch edges (Nakaoka and Iizumi, 2000). Animals at patch edges are often different sizes from animals in the interior (Crawford and Jorgenson, 1996), and even sexes can be different within and at edges of patches (Bertiller *et al.*, 2000). Differences in behavior are also often observed in individuals near the edge of patches (Meadows, 2001). These differences in characteristics measured from a snapshot of a patch are probably caused by the influence of position within a group on fitness measures. Animals at the edges often have greater feeding opportunities (Black *et al.*, 1992; Romey, 1995), but also have greater predation risk (Cocroft, 2002; Hamilton, 1971). Grazers have been observed to feed most heavily on patch edges (Clarke *et al.*, 1995), which is also probably a factor with some predators of the boundary community (Benoit-Bird and Au, 2003b). Costs and benefits associated with living ''on the edge'' can be seen in territorial fish that grow slower and are smaller at group edges, but also have more matings and offspring (Meadows, 2001). In mobile prey, predation risk changes can affect spatial behavior and patch characteristics. For example, some nocturnally active animals avoided patch edges more at full moon when predation risk was higher (Bowers and Dooley, 1993). The need to balance hunger and predation risk affects the distribution of animals within a patch. Changes in characters such as patch shape and size can serve as indicators of changes in this tradeoff as they change the percentage of animals experiencing edge conditions.

The primary technique that has been used to study the distribution of the mesopelagic boundary community and other sound-scattering layers has been acoustics. Downward looking echosounders have been the principal tool employed because of their high spatial coverage with relatively high spatial and temporal resolution. Interpretation of this acoustic data requires information on the acoustic size, or target strength, of the animals. In order to estimate density from volume scattering, the mean target strength of individual animals within the study area must be known (MacLennan and Simmonds, 1992). Within the boundary community, target strength is most strongly affected by the taxonomic identity of the scatterer, i.e., whether the animal is a shrimp, a fish, or a squid (Benoit-Bird and Au, 2001). A second important factor is the size of the animal; smaller animals have lower target strengths and larger animals have higher target strengths. To calculate animal density using a technique called echo-energy integration, the total volume scattering is divided by the mean target strength for an individual (MacLennan and Simmonds, 1992). Previous estimates of mesopelagic animal density have been conservative, using the highest scattering possible by an individual animal from the mesopelagic layer as the divisor. Differences in animal density might be masked by concurrent differences in animal size and composition, and the limited resolution of downward looking sonars makes observation of microstructure difficult.

The goals of this work were to (1) obtain estimates of the lateral aspect echo strength of mesopelagic boundary community animals; (2) determine how echo strength varies with sampling variables including depth, time, and location; (3) identify taxonomic groups of animals present in the boundary layer; (4) estimate the density of mesopelagic boundary community animals using echo highlight counting; and (5) observe the distribution of mesopelagic animals near patch edges.

II. METHODS

A. Surveys

Acoustic surveys were conducted using a 200-kHz calibrated echosounder system as in Benoit-Bird *et al.* (2001). On each night between 1 and 5 June 2001, one 18-km long transect off the Kona coast of Hawaii was surveyed for 1 h at 2100, 0000, and 0300, local time, from the National Oceanic and Atmospheric Administration vessel, R/V TOWNSEND CROMWELL. Sampling began well after twilight each night and ended before morning twilight. During the sampling, the moon was between 85% and 100% illuminated. Ongoing work has shown a minimal effect of lunar cycle on the overall migration patterns of the nearshore scattering layer (Benoit-Bird and Au, unpublished data). Two transects were located 1.5 km (inshore) from the shoreline and two were located 3.0 km from the shore (offshore). Transects were selected in this way because previous studies have indicated that the effects of distance from the shoreline were significant on the distribution of the scattering layer but bottom topography was not (Benoit-Bird *et al.*, 2001). The order of transects and the direction in which they were surveyed was randomized and no transect was surveyed more than once each night.

Twice along each transect, an acoustically identified patch was further investigated with a vertical transect using both a broadband sonar system and a digital still camera with strobe lights. The sonar transducer, the pressure-housed camera, two strobe lights, and a conductivity, temperature, depth (CTD) sensor were mounted to a profiling frame and lowered at a rate of 0.1 m/min to 70 m. The CTD sent real time depth information through a conductive cable to the surface. This information was simultaneously stored with digitized echo data collected every 0.33 m and still images every 1 m during the entire descent. Temperature and conductivity data in

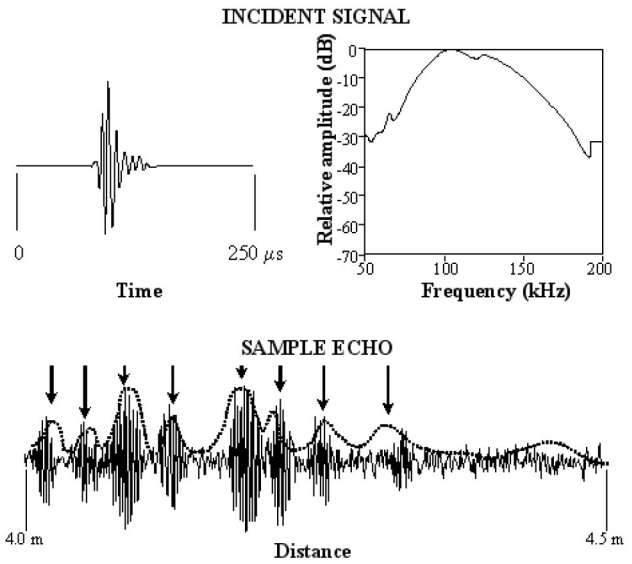


FIG. 1. Waveform and spectrogram of the incident sonar signal. The bottom shows a sample of a resulting echo waveform, solid line, and cross-correlation envelope, broken line. Arrows indicate echo highlights.

this well-mixed area were not found to significantly correlate with acoustic scattering and are not presented here.

B. Broadband sonar

The sonar used a 50-μs, broadband signal (Fig. 1. bandwidth 35 kHz, center frequency 120 kHz), modeled after a dolphin echolocation click. The outgoing signal was produced using a function generator computer plug-in board. A Rapid System R1200 analog-to-digital (A/D) converter was used to digitize and store the data to hard disk. Sampling rates of 500 kHz were used for the function generator and the A/D converter. This system has a temporal resolution of 3 cm (Au and Benoit-Bird, 2003) and at the peak frequency of the signal, a 3-dB beamwidth of 8°.

C. Camera

A 3.3 mega pixel digital camera in a pressure housing was modified so that it could be controlled from the surface. The flash of the camera triggered two, digital strobe lights. Calibration of the camera system showed that it effectively sampled 0.3–1.3 m from the camera, with an angle of 32°, under these lighting conditions. The total effective sampling volume of the camera system was 0.18 m^3.

D. Data analysis

1. Survey data

Survey data were utilized to select focal patches within the boundary layer and place them into a larger context. In particular, this information was utilized to compare the characteristics of the boundary layer to those observed in previous surveys. Echo energy integration (MacLennan and Simmonds, 1992) was utilized as in Benoit-Bird *et al.* (2001) to estimate animal density, and patch geometry was determined as in Benoit-Bird and Au (2003c).

2. Broadband sonar data

The total echo energy in each 0.5-m interval from 2.0 to 6.0 m from the transducer was calculated by taking the ratio of the frequency spectrum of the echo signals referenced to 1 m and the frequency spectrum of the incident signal. The transmitted signal was measured by directing the transducer toward an acoustic mirror made of a 0.64×45×45 cm sheet of aluminum covered with closed-cell neoprene.

The envelope of the cross-correlation between each echo and the incident signal was examined to determine basic time-domain characteristics. The cross correlation function was determined by the Fourier transform technique using the equation

$$c(t)=\Im^{-1}[E(f)U(f)] \tag{1}$$

(Brigham, 1988), where $E(f)$ and $U(f)$ are the Fourier transform of the echo and incident signals, respectively, and $\Im^{-1}$ denotes the inverse Fourier transform of the terms in the brackets. The envelope of the cross-correlation function was calculated by converting $c(t)$ into an analytic signal using the Hilbert transform method where the absolute value of the analytic signal represents the envelope of the signal (Barr, 2001; Burdic, 1968). Individual highlights, or glints, in the echo waveforms represent distinct surfaces from which sound is reflected. The broadband signal had temporal resolution of 3 cm (Au and Benoit-Bird, 2003), larger than the structures within a single boundary community animal (Benoit-Bird and Au, 2001). Individual highlights probably represent reflections from different animals. At the highest animal densities estimated in previous studies (Benoit-Bird *et al.*, 2001), mesopelagic animals within the boundary community would have an average spacing of 12 cm, much larger than the resolution of the sonar signal. Highlights most likely represent reflections from distinct individuals from within the boundary community permitting echo highlight counting estimates of mesopelagic animal density (MacLennan and Simmonds, 1992). Animal density was calculated in each 0.5-m interval from the transducer.

The average backscatter cross-section in each 0.5-m interval from the transducer was calculated by dividing the total energy within the distance class by the number of individual echo highlights. These values were converted to echo strength for easier presentation. The effects of time, distance from the shoreline, depth, and range from the transducer on echo strength and animal density estimates were analyzed using two univariate analyses of variance (ANOVAs).

3. Camera data

The digital camera produced limited results because of focusing difficulties. This limited the identification of animals in most of the images. However, the number of animals in each digital photo could be analyzed. The rough calibration of the sampling volume of the camera permits approximate estimates of density to be made.

III. RESULTS

Patches observed with the echosounder surveys were not significantly different from those observed during previous

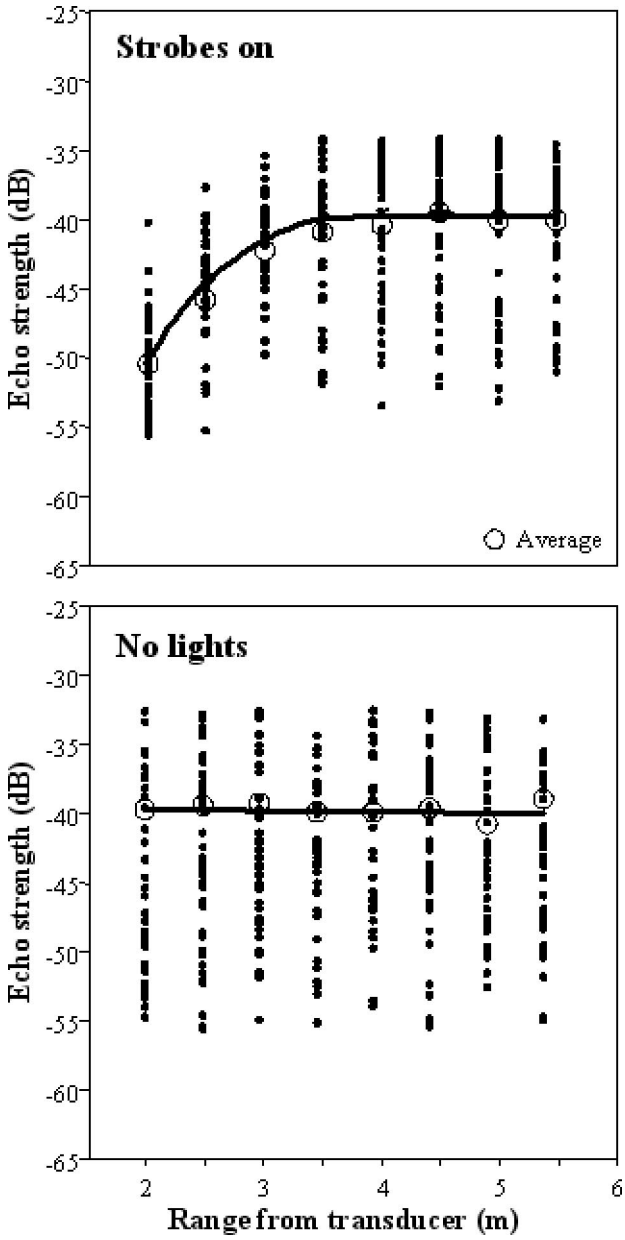


FIG. 2. Echo strength of individual highlights within each 0.5-m interval from the transducer from transects during which the strobe were lights on, top, and those during which strobe lights were not used, bottom.

surveys of the same transects. Both in their geometry and density characteristics, the patches surveyed with the vertical transects were typical of the patches in the remainder of each transect. Patches ranged in horizontal extent from 20 to 110 m. The vertical extent of patches matched the distribution of layer reported in this location by Benoit-Bird *et al.* (2001), ranging from 70 to 100 m. No differences were observed in surface echosounder results when the broadband, dolphin-like signal was used and at other times suggesting there is no strong effect of the signal on the animals in the layer.

Strong light-avoidance effects were observed on the echo strength of mesopelagic animals (Fig. 2). During several transects, no strobe lights were utilized because of equipment failure providing a no-light condition for comparison. The strongest avoidance effects were observed nearest to the transducer with the effect apparently lost at around 4 m. When the first 4 m from the transducer were ignored, no significant effects of range were observed on mean echo strength (Table I). There was also no significant effect of depth on echo strength ($p>0.05$). Both time and distance from the shoreline had significant effects on echo strength (Fig. 3). There was also a significant interaction of time and distance from the shoreline on echo strength.

TABLE I. Analysis of variance (ANOVA) of between-subject effects on mean echo strength. SS=sum of the squared deviations of scores from their mean (a measure of variability or dispersion), df=degrees of freedom (the number of independent pieces of information available in the estimation of population parameters), MS=mean square (a calculated variance estimate), $F=F$ ratio (an index permitting assessment of the statistical significance), and p=probability where ns=non significant at the 0.05 level.

Source	SS	df	MS	F	p
Time	2576	2	1288	7625	<0.05
Depth	73	69	1	1	ns
Range	15	4	4	34	ns
Distance	1368	1	1368	20484	<0.05
Time*Depth	145	138	1	14	ns
Time*Range	16	8	2	6	ns
Depth*Range	294	276	1	3	ns
Time*Distance	563	2	282	3804	<0.05
Depth*Distance	355	69	5	20	ns
Range*Distance	10	4	3	5	ns
Error	6	1	6		
Total	298 815	1680			

The mean echo strength values were relatively high for micronektonic animals. The lateral aspect length-target strength relationships established for these mesopelagic animals by Benoit-Bird and Au (2001) were used to estimate the average size of animals with these echo strengths oriented randomly in the lateral plane relative to the transducer. The variance in target strength in this orientation measured for other fish by Benoit-Bird and Au (Benoit-Bird *et al.*, 2003)

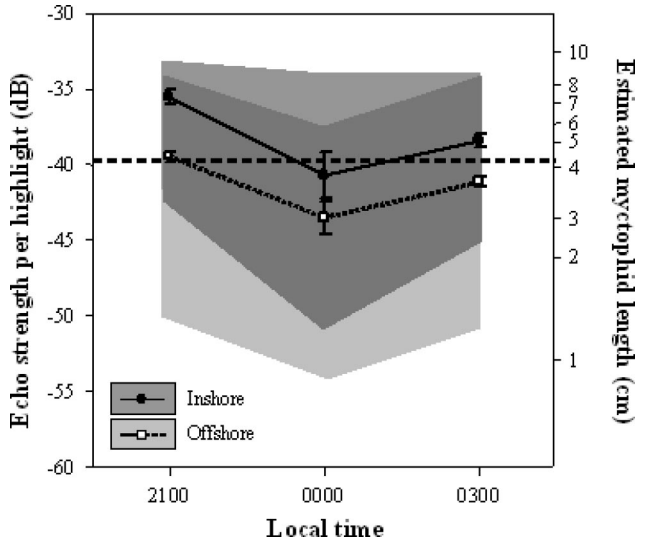


FIG. 3. Mean echo strength from 4 to 6 m from the transducer at each sampling time for inshore and offshore sampling times. Error bars represent the standard error of the mean of the entire vertical range for each of four replicates. Gray areas show the entire range of echo strengths observed with the overlap zone between inshore and offshore in the darkest shade. The dashed line represents the average of all the samples. On the right axis is the estimated length of a myctophid randomly oriented in the lateral plane.

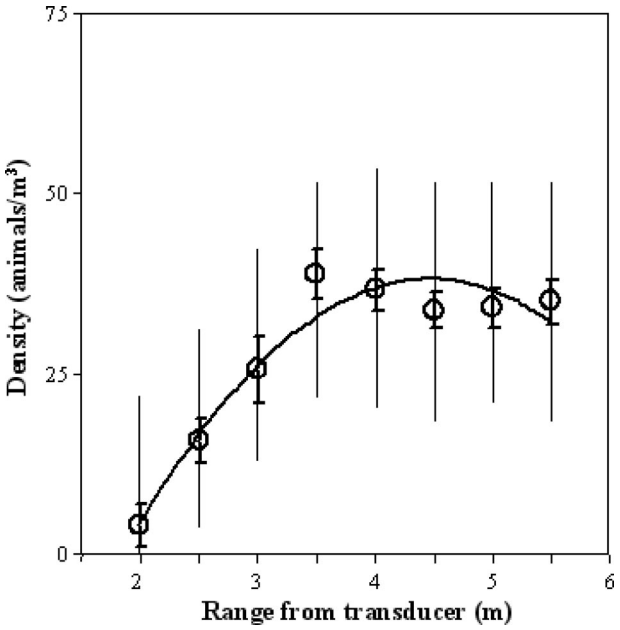


FIG. 4. Mesopelagic animal density at midnight, inshore as a function of distance from the transducer when strobe lights were on. Like echo strength, no range effect was observed when the lights were off. Similar patterns were observed both inshore and offshore at other sampling times. Error bars represent the standard error of the mean while vertical lines show the entire range of each sample.

was used as an estimate of the expected change for myctophids. The measured variation for squid (Arnaya *et al.*, 1989) and shrimp (Imazeki *et al.*, 1989) was utilized. All of the mean lateral aspect echo strengths were too high to have come primarily from mesopelagic shrimps. Only the largest squids in the boundary layer could have produced the echo strengths measured at midnight inshore. All other echo strengths measured were too strong for backscattering primarily from squid. The mean echo strengths measured are estimated to be equivalent to myctophids between 2.9 and 7.3 cm long. While limited, the photographic results support the conclusion that the scattering layer was primarily comprised of fish. Of the nearly 1700 photographs taken, 880 of which were within the acoustically identified layer, only 27 had animals identifiable to taxonomic class. All of the identified animals were myctophid fishes.

Light avoidance was also observed in the density of mesopelagic animals estimated from echo highlight counting (Fig. 4). When the first 4 m were ignored, no significant effect of range from the transducer was observed (Table II). Depth, distance from the shoreline, and time all significantly affected density. Significant interactions between time and depth, time and distance, and distance and depth on echo strength were also observed. The effects of depth, time, and distance from the shoreline on density are shown in Figs. 5 and 6. The mean echo strength at each depth is also shown.

Figure 7 shows the mean density of mesopelagic animals estimated through echo highlight counting as a function of time. Also shown are the mean densities of the same patches measured from the surface using echo energy integration. The target strength value used as the average for the population was −37 dB, the same as that used in previous studies (Benoit-Bird and Au, 2003a, c; Benoit-Bird *et al.*,

TABLE II. Analysis of variance (ANOVA) of between-subject effects on density (all abbreviations are the same as those in Table I).

Source	SS	df	MS	*F*	*p*
Time	8903	2	4452	13425	<0.05
Depth	7671	69	111	456	<0.05
Range	68	4	17	29	ns
Distance	874	1	874	1981	<0.05
Time*Depth	6712	138	49	303	<0.05
Time*Range	107	8	13	21	ns
Depth*Range	981	276	4	13	ns
Time*Distance	1873	2	937	1593	<0.05
Depth*Distance	4098	69	59	87	<0.05
Range*Distance	67	4	17	27	ns
Error	8	1	8		
Total	423 931	1680			

2001). Differences in both the density values and the patterns in density were observed between the two estimates. The mean lateral aspect echo strengths measured here were converted to the equivalent dorsal aspect target strength of myctophid fishes (measured by Benoit-Bird and Au, 2001) and used to reestimate the density by echo energy integration. Both the values of the revised density estimates and their patterns fit the echo highlight counting estimates more closely than those based on the fixed target strength from previous studies. The density calculated by echo energy in-

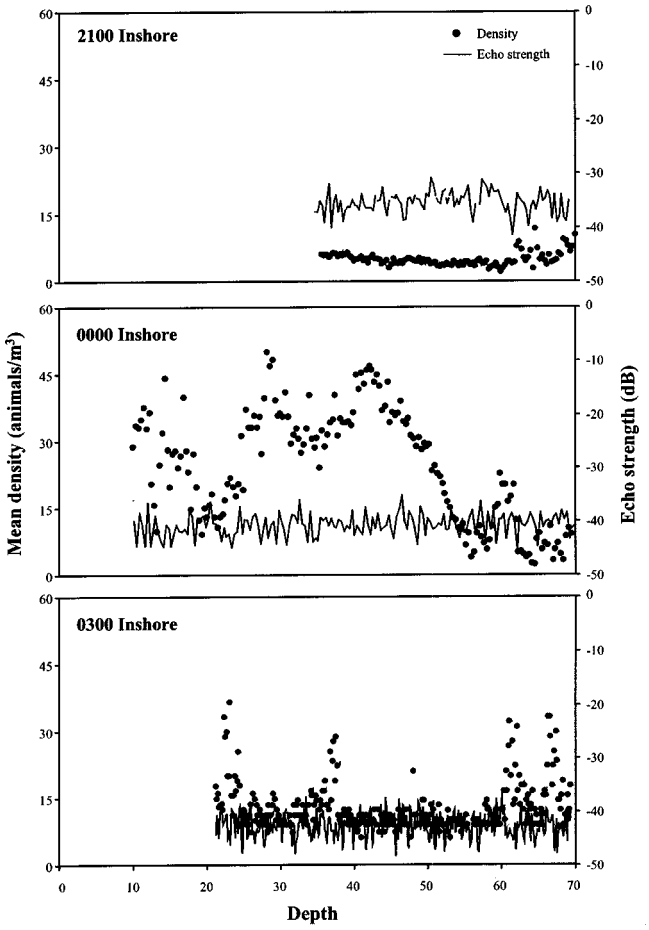


FIG. 5. Mean density from 4 to 6 m from the transducer of mesopelagic animals as a function of depth at inshore sites measured with echo highlight counting. The mean echo strength at each depth is also shown.

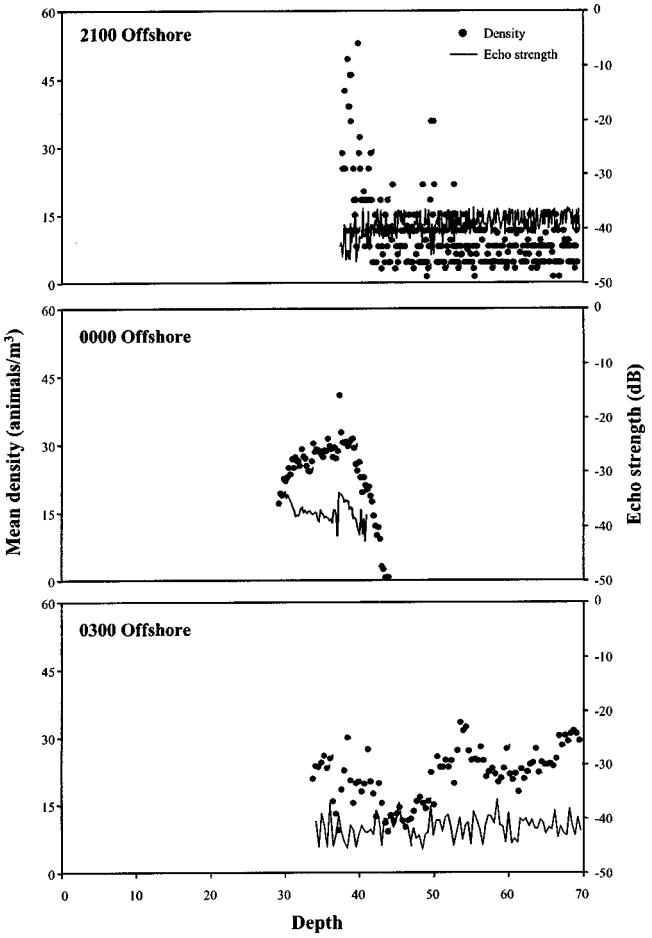


FIG. 6. Mean density from 4 to 6 m from the transducer of mesopelagic animals as a function of depth at offshore sites measured with echo highlight counting. The mean echo strength at each depth is also shown.

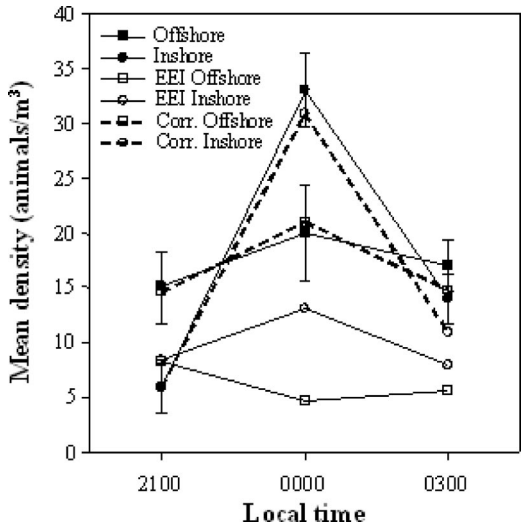


FIG. 7. Mean density of animals as a function of sampling time, measured from 4 to 6 m from the transducer using echo highlight counting from the broadband sonar is shown by the solid lines and filled symbols. Error bars show the standard error of the mean. Open symbols and solid lines show the mean density for the upper 70 m of the same patches estimated with echo energy integration (EEI), using the target strength estimates from previous studies (Benoit-Bird *et al.*, 2001). Utilizing the mean lateral aspect echo strengths measured in this study converted to equivalent dorsal aspect echo strengths to estimate density with echo energy integration greatly improved the density estimates made from the surface (dashed lines).

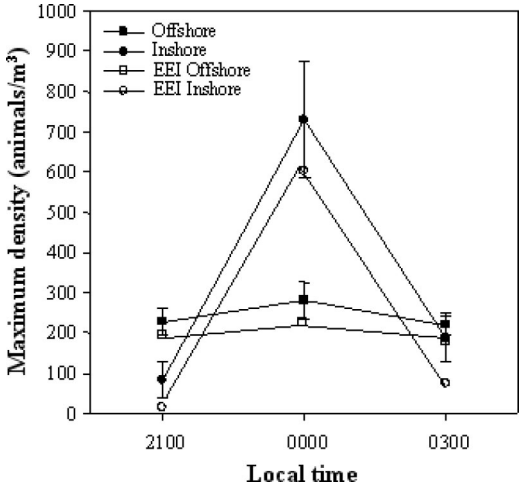


FIG. 8. Maximum density of mesopelagic animals as a function of sampling time and location estimated through echo highlight counting (filled symbols). Error bars represent the standard error of the mean. Open symbols show the maximum density estimated by echo energy integration (EEI) using the fixed target strength from previous studies.

tegration (using the measured echo strengths converted to dorsal aspect) and that calculated from echo highlight counting in each 1-m depth interval for all the vertical transects were compared. The correlation was significant ($p<0.01$), with a coefficient of 0.81. An *F*-test revealed that the slope of the relationship was not significantly different from 1 ($p>0.05$, observed power=0.85). Figure 8 shows the effect of time on the maximum density of mesopelagic animals. Echo energy integration estimates of maximum density using the fixed target strength value from previous work were more similar to the echo highlight counting estimates than the fixed estimate-based estimates of mean density. Information from underwater photographs was limited. However, preliminary confirmation of the acoustic estimates of density was made by comparing the number of animals predicted from each photograph with the observed number of animals (Fig. 9). The number of expected animals was calculated from the density of animals observed between 2.0 and 2.5 m from the transducer (closest to the range photographed) at the same depth as each photograph. The observed and predicted number of animals were significantly correlated (Pearson correlation coefficient=0.52, $p<0.01$).

In seven patches, the edges of patches were observed with the sonar. An example from the patch edge observed for the greatest distance is shown in Fig. 10. Patch edges were very distinct, that is the change in density from within the patch to outside of it dropped dramatically, even over half a meter. Figure 11 shows the density of mesopelagic animals as a function of distance from the edge of the patch both within and outside its boundary. One-tailed *t*-tests were utilized to compare the density of animals in the 0.5 m nearest the patch edge to their density in the 0.5 m furthest from the edge. Density of mesopelagic animals was not significantly less near the edge of the patch than in the interior of the patch for any of the seven edges observed ($p>0.05$ for all comparisons).

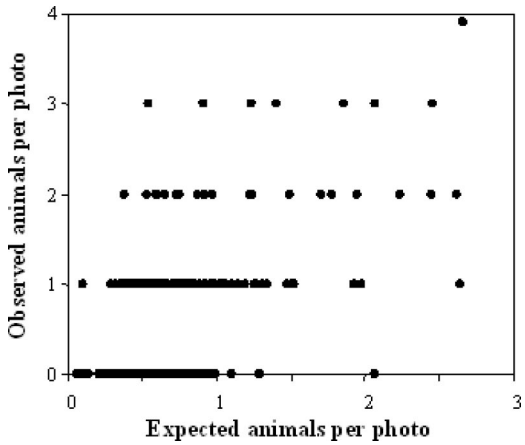


FIG. 9. Number of mesopelagic animals expected based on echo highlight counting density estimates from 2 to 3 m from the transducer, the range closest to that photographed, is shown on the *x*-axis. The *y*-axis shows the number of animals counted from a photograph at the same depth. The two variables were significantly correlated (Pearson correlation coefficient $=0.52$, $p<0.01$).

IV. DISCUSSION

Strong light avoidance was observed in mesopelagic animals. Future sampling devices will need to reduce this effect by eliminating light or otherwise reducing its effects. Reduced echo strength for the first 4 m from the transducer was evident. This could be accounted for by a change in animal orientation as animals started to move away from the light source. The end-on echo strength of animals, where their cross-section is smallest, is decreased relative to all other orientations (Benoit-Bird *et al.*, 2003; Martin Traykovski *et al.*, 1998). The decreases in animal density near the transducer suggest that animals are able to respond rapidly enough to move away from the transducer and the lights. However, the continuous increase in density with range from the lights suggests that some animals are able to move further away from the source in the same amount of time. The corresponding increase in echo strength with range from the lights indicates that larger animals are able to make it further. While questions about the orientation of the animals make it impossible to measure the size of the animals within the range of the light effect, this information suggests that animals from a range of sizes are present within the scattering layer, even at one depth. This supports previous conclusions that the mesopelagic boundary community off the Kona coast is not vertically stratified by animal size (Benoit-Bird and Au, 2003c). Further evidence to support this conclusion is provided from the echo strength values from 4 to 6 m from the transducer, where a range effect was no longer detected. No significant effects of depth were observed on echo strength in echoes from this range.

The consistency of echo strength as a function of depth also supports the conclusion that there is no vertical stratification in the species assemblage in the upper 70 m of the nearshore component of the boundary layer. While the low animal densities near the transducer because of light avoidance and photographic difficulties limited the number of images containing taxonomically identifiable animals, all discernable images were of myctophid fishes. In addition, all

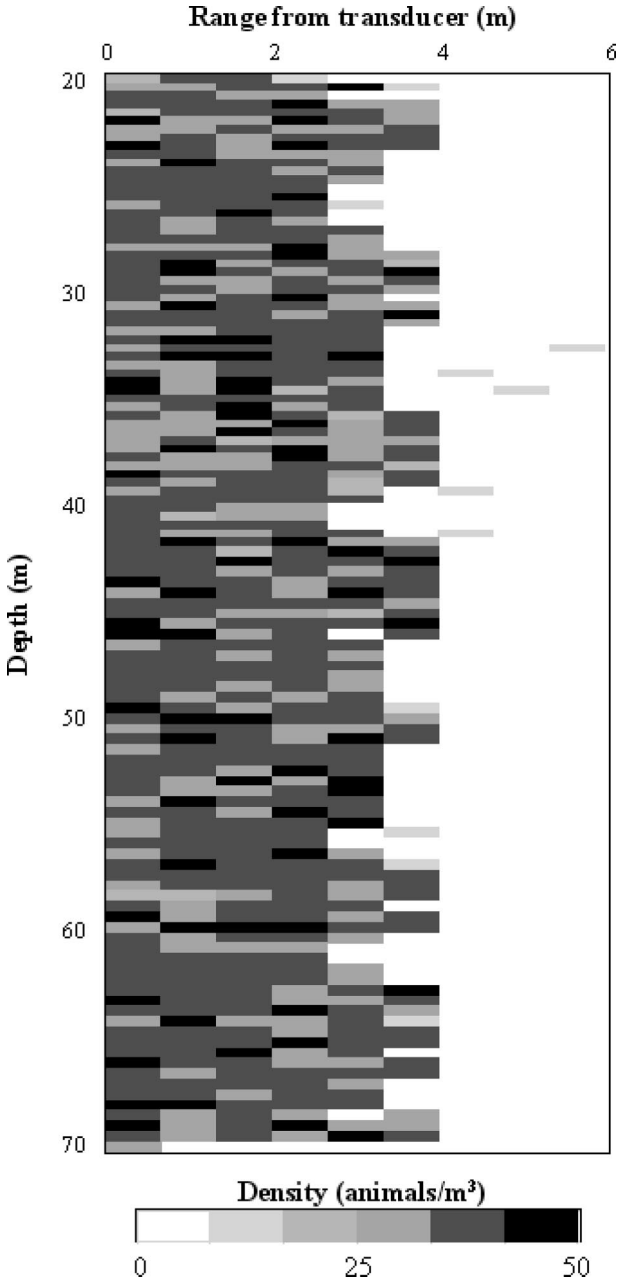


FIG. 10. Sample of animal density estimates from the edge of an inshore patch observed at midnight. The edge of the patch is relatively straight and very distinct in the 50-m depth range over which it was observed.

mean echo strengths were relatively high for micronektonic animals. These echo strengths could only have come from an assemblage of animals primarily composed of fish. While direct observation of composition of the layer was not possible, this is the first evidence of composition of this nearshore boundary layer component. Further offshore, net samples of the boundary layer's upper depths were also primarily comprised of myctophids (Reid, 1994).

While all mean echo strengths were relatively high, echo strength was significantly affected by time and distance from the shoreline. Animals with higher echo strengths, presumably larger animals, migrated inshore and upwards early in the night, before 2100 hours, and were followed by midnight by smaller animals with lower echo strengths. Smaller animals migrated down and away from shore before 0300

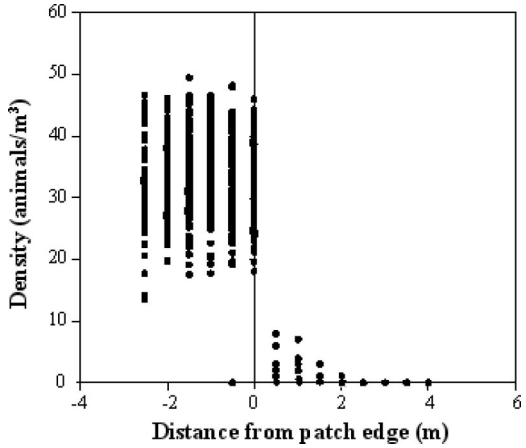


FIG. 11. Density of animal density in each 0.5-m interval near the edge of the midnight inshore patch shown in Fig. 10. Density drops dramatically at the patch's edge. A one-tailed *t*-statistic revealed no significant differences in density between the 0.5 m nearest the patch edge and the 0.5 m furthest from the edge.

hours. The larger, higher echo strength animals migrated down and away from shore later in the morning. Mean echo strength, and presumably animal size, decreased at midnight both inshore and offshore. The variance in echo strength also increased at midnight. This suggests that the mean size of animals near the beginning and end of the migration were larger than those near the midpoint of the diel migration near midnight. Further, animal sizes were larger inshore than offshore at all sampling times. The migration rates of the leading and trailing edges of the boundary layer have been measured at greater than 1.6 km/h. Based on the fin dimensions of myctophids, the smaller animals observed inshore at midnight are not likely capable of sustaining this swimming speed (Sambilay, 1990) while neither the largest mesopelagic shrimps or squids are likely capable of these speeds (Cowles, 2001; Yatsu *et al.*, 1999). The animals of the mean size predicted from echo strengths inshore at 2100 and 0300 hours could easily maintain this rate. The range of estimated animal sizes and the variance of estimated sizes also increased at midnight while the mean size of animals decreased. This suggests that the larger animals that migrated inshore first, before 2100 hours, were still present but were later joined by more, smaller animals. The increase in the density of animals from 0900 to 0000 hours supports this.

Analysis of echo strength provides information about the biological processes at work within the boundary layer. It also provides significant practical information for estimating density. To estimate density with downward looking echosounders requires information on the average acoustic size of targets within the study area. Accounting for all sources of target strength variance such as time, depth, and location is difficult, particularly when animal densities are high and target strength cannot be determined using a split-beam echosounder. While dorsal aspect target strengths cannot be directly measured from the techniques employed here, the effects of sampling variables on echo strength can be assessed. Depth was not found to significantly affect echo strength. However, time and distance from the shoreline did significantly affect echo strength. This means that a single estimate of target strength can be applied to the upper 70 m of the water column at a specific time and location. However, the static estimate of target strength utilized to estimate density in previous studies introduced error. The lateral aspect echo strengths measured can be converted into estimates of equivalent dorsal aspect echo strengths to estimate the effect of previous methods on the results. Utilizing the revised echo strength measurements confirms that previous density estimates were conservatively low both inshore and offshore, at all sampling times. The changes in echo strength as a function of time also suggests that previous estimates of density differences between midnight and the other sampling times were not large enough. However, these new analyses of density do not affect the major conclusions of these studies, instead they provide further support of the high abundance of animals in this layer.

Animal density was also measured utilizing echo highlight counting, which is independent of echo strength. Preliminary photographic measures confirm these estimates. Echo highlight counting density estimates were remarkably close to those obtained with data from the echosounder at the surface analyzed using dorsal-aspect equivalent echo strengths applied to echo-energy integrations. This method of obtaining *in situ* estimates of patterns in mean echo strength over space and time can be useful for interpreting volume scattering data from an echosounder. The echo highlight density estimates are significantly higher than those obtained previously using fixed-target strength echo-energy integration, however, the conclusions from the estimates are the same. The density patterns as a function of time and distance from the shoreline show the diel, horizontal migration of the mesopelagic boundary community (Benoit-Bird *et al.*, 2001).

Animal density, measured with echo highlight counting, changed not only as a function of time and distance from the shoreline as in previous studies (Benoit-Bird *et al.*, 2001), but also with depth. These animal density patterns were consistent among the four replicates of each treatment. Abrupt spikes in animal density, covering narrow depth zones, were sometimes observed, primarily when the overall animal density was low. When overall animal density was higher, animal density peaks tended to cover a greater range of depths and the changes in animal density were much more gradual. As the entire layer migrated upward from 2100 to 0000 hours, the depths of animal density peaks also moved upward. While vertical microstructure in echo strength was not observed, fine-scale vertical microstructure in animal density was apparent.

Edge effects were not observed in the mesopelagic boundary community. Neither echo strength, related to animal size, nor animal density varied significantly between the patch's edge and the patch's interior for any of the seven patch edges observed. Variance in these characteristics also remained the same. In all cases the density dropped dramatically at the patch edge to an average of zero outside the patch. For all seven patches, the vertical patch edge was nearly straight as observed in previous work (Benoit-Bird and Au, 2003c). This is unusual when compared with studies of other scattering communities (see, for example, Greene *et al.*, 1994). The reasons for this sharp change between the

patch and the background are not clear; however, the result is that the perimeter of each patch is as low as possible for its size. This might reduce the number of animals exposed to edge risks but also reduces the number of animals experiencing its potential foraging benefits. Off the coast of Oahu, a sharp delineation between a high animal density mesopelagic layer and lower density patches was observed (Benoit-Bird and Au, 2003c). Because distinct patches are only found in sites with low animal density while high animal density sites have continuous layers, the distinct edges also suggest a density threshold for mesopelagic animal distribution. These animals could disperse into a continuous layer, reducing the overall density and eliminating edge effects except at the top and bottom of the layer. Maintenance of a threshold density appears to be more important than the avoidance of edge effects.

The microstructure observed within the mesopelagic boundary layer off the coast of Hawaii was similar to that predicted from echosounder surveys (Benoit-Bird and Au, 2003c). Most animals in this nearshore boundary layer appear to be myctophid fishes. Strong effects of time, distance from shore, and depth were observed on density. However, no vertical effect on echo strength was evident, suggesting a lack of stratification in animal size and composition. Depth-dependent animal density patterns provide important information for assessing the impact of the boundary layer and the behavior of its predators. Finally, revision of echo strength values reveals the underestimates of animal density estimates from previous studies, but supports the conclusions of these studies. High animal density sites with continuous layers show strong vertical stratification in echosounder surveys and may produce different results with respect to echo strength and density microstructure. The technique of utilizing a high-resolution, relatively short-range system could provide important information to explain the differences between these sites and assist in the interpretation of echosounder data from these, and other locations.

ACKNOWLEDGMENTS

We thank the National Marine Fisheries Service, Honolulu Laboratory, especially Robert Humphreys for collaborating and providing ship time. The crew of the TOWNSEND CROMWELL, especially Phil White, was instrumental in conducting this research. Christopher Bird assisted in the construction of the underwater camera housing and with field work. K. Benoit-Bird was partially supported by a Leonida Family Fisheries Research Scholarship and an ARCS Foundation award. This work was supported in part by Office of Naval Research Grant No. N00014-98-1-068, Dr. Robert Gisiner, Program Manager. This work was also funded by a grant from the National Oceanic and Atmospheric Administration, Project No. R/FM-7, which is sponsored by the University of Hawaii Sea Grant College Program, SOEST, under Institutional Grant No. NA86RG0041 from NOAA Office of Sea Grant, Department of Commerce. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subagencies. UNIHI-SEAGRANT-JC-02-17. This is HIMB contribution 1154.

Arnaya, I. N., Sano, N., and Iida, K. (**1989**). ''Studies on acoustic target strength of squid III. Measurement of the mean target strength of small live squid,'' Bull. Fac. Fish. Hokkaido Univ. **40**, 110.

Au, W. W. L., and Benoit-Bird, K. J. (**2003**). ''Acoustic backscattering by Hawaiian lutjanid snappers II: Broadband temporal and spectral structure,'' J. Acoust. Soc. Am. (in press).

Barr, R. (**2001**). ''A design study of an acoustic system suitable for differentiating between orange roughy and other New Zealand deep-water species,'' J. Acoust. Soc. Am. **109**, 164–178.

Benoit-Bird, K. J., and Au, W. W. L. (**2001**). ''Target strength measurements of animals from the Hawaiian mesopelagic boundary community,'' J. Acoust. Soc. Am. **110**, 812–819.

Benoit-Bird, K. J., and Au, W. W. L. (**2002**). ''Energy: Converting from acoustic to biological resource units,'' J. Acoust. Soc. Am. **111**, 2070–2075.

Benoit-Bird, K. J., and Au, W. W. L. (**2003a**). ''Fine-scale diel migration dynamics of an island-associated sound-scattering layer,'' Deep-Sea Res. submitted for review.

Benoit-Bird, K. J., and Au, W. W. L. (**2003b**). ''Prey dynamics affect foraging by a pelagic predator (*Stenella longirostris*) over a range of spatial and temporal scales,'' Behav. Ecol. Sociobiol. **53**, 364–373.

Benoit-Bird, K. J., and Au, W. W. L. (**2003c**). ''Spatial dynamics of a nearshore micronekton sound-scattering layer,'' ICES J. Mar. Sci. **60**, 899–913.

Benoit-Bird, K. J., Au, W. W. L., and Kelley, C. D. (**2003**). ''Acoustic backscattering by Hawaiian lutjanid snappers I: Target strength and swimbladder characteristics,'' J. Acoust. Soc. Am. (in press).

Benoit-Bird, K. J., Au, W. W. L., Brainard, R. E., and Lammers, M. O. (**2001**). ''Diel horizontal migration of the Hawaiian mesopelagic boundary community observed acoustically,'' Mar. Ecol.: Prog. Ser. **217**, 1–14.

Bertiller, M. B., Ares, J. O., Graff, P., and Baldi, R. (**2000**). ''Sex-related spatial patterns of Poa ligularis in relation to shrub patch occurrence in northern Patagonia,'' J. Veg. Sci. **11**, 9–14.

Beyer, J. E. (**1995**). ''Functional heterogeneity: Using the interrupted Poisson process (IPP) model unit in addressing how food aggregation may affect fish ration,'' ICES Council Meeting Papers, 14.

Black, J. M., Carbone, C., Wells, R. L., and Owen, M. (**1992**). ''Foraging dynamics in goose flocks: The cost of living on the edge,'' Anim. Behav. **4**, 41–50.

Bowers, M. A., and Dooley, J. L. J. (**1993**). ''Predation hazard and seed removal by small mammals: Microhabitat versus patch scale effects,'' Oecologia **94**, 247–254.

Brigham, E. O. (**1988**). *The Fast Fourier Transform and its Applications* (Prentice–Hall, Englewood Cliffs, NJ).

Burdic, K. J. (**1968**). *Radar Signal Analysis* (Prentice–Hall, Englewood Cliffs, NJ).

Clarke, J. L., Welch, D., and Gordon, I. J. (**1995**). ''The influence of vegetation pattern on the grazing of heather moorland by red deer and sheep: II. The impact on heather,'' J. Appl. Ecol. **32**, 177–186.

Cocroft, R. B. (**2002**). ''Antipredator defense as a limited resource: Unequal predation risk in broods of an insect with maternal care,'' Behav. Ecol. **13**, 125–133.

Cowles, D. L. (**2001**). ''Swimming speed and metabolic rate during routine swimming and simulated diel vertical migration of *Sergestes similis* in the laboratory,'' Pac. Sci. **55**, 215–226.

Crawford, R. E., and Jorgenson, J. K. (**1996**). ''Quantitative studies of Arctic cod (*Boreogadus saida*) schools: Important energy stores in the Arctic food web,'' Arctic **42**, 181–193.

Elner, R. W., and Hughes, R. N. (**1978**). ''Energy maximization in the diet of the shore crab, *Carcinus maenas,*'' J. Anim. Ecol. **47**, 103–116.

Greene, C. H., Wiebe, P. H., and Zamon, J. E. (**1994**). ''Acoustic visualization of patch dynamics in oceanic ecosystems,'' Oceanogr. **7**, 4–12.

Gross, M. R., and Sargent, R. C. (**1985**). ''The evolution of male and female parental care in fishes,'' Am. Zool. **25**, 807–822.

Hamilton, W. D. (**1971**). ''Geometry for the selfish herd,'' J. Theor. Biol. **31**, 295–311.

Howard, R. D. (**1980**). ''The evolution of mating strategies in bullfrogs, Rana catesbiana,'' Evol. **32**, 850–871.

Imazeki, A., Yonemoto, H., Saotome, Y., Isouchi, T., and Hamada, Y. (**1989**). ''Multifrequency measurements of target strength of the sergestid shrimp *Sergia lucens,*'' J. Tokyo Univ. Fish. **76**, 37–44.

MacLennan, D. N., and Simmonds, E. J. (**1992**). *Fisheries Acoustics* (Chapman and Hall, New York).

Martin Traykovski, L. V., O'Driscoll, R. L., and McGehee, D. E. (**1998**). ''Effect of orientation on broadband acoustic scattering of Antarctic krill *Euphasia superba*: Implications for inverting zooplankton spectral acoustic signatures for angle orientation,'' J. Acoust. Soc. Am. **104**, 2121–2135.
Meadows, D. W. (**2001**). ''Center-edge differences in behavior, territory size and fitness in clusters of territorial damselfish: Patterns, causes, and consequences,'' Behaviour **138**, 1085–1116.
Nakaoka, M., and Iizumi, H. (**2000**). ''Magnitude of within-patch variation in seagrass *Halophila ovalis* growth affected by adjacent *Thalassia hemprichii* vegetation,'' Ecol. Res. **15**, 415–424.
Reid, S. B. (**1994**). ''Spatial structure of the mesopelagic fish community in the Hawaiian boundary region,'' Ph.D. dissertation, University of Hawaii.
Reid, S. B., Hirota, J., Young, R. E., and Hallacher, L. E. (**1991**). ''Mesopelagic-boundary community in Hawaii: Micronekton at the interface between neritic and oceanic ecosystems,'' Mar. Biol. **109**, 427–440.
Robinson, M. P., and Motta, P. J. (**2002**). ''Patterns of growth and the effects of scale on the feeding kinematics of the nurse shark (*Ginglymostoma cirratum*),'' J. Zool. **256**, 449–462.
Romey, W. L. (**1995**). ''Position preferences within groups: Do whirligigs select positions which balance feeding opportunities with predator avoidance?'' Behav. Ecol. Sociobiol. **37**, 195–200.
Sambilay, Jr., V. C. (**1990**). ''Interrelationships between swimming speed, caudal fin aspect ratio and body length of fishes,'' Fishbyte **83**, 16–20.
Walter, C., Pauly, D., Christensen, V., and Kitchell, J. F. (**2000**). ''Representing density dependent consequences of life history strategies in aquatic ecosystems: EcoSim II,'' Ecosystems **3**, 70–83.
Yatsu, A., Yamanaka, K., and Yamashiro, C. (**1999**). ''Tracking experiments of the jumbo flying squid, *Dosidicus gigas*, with an ultrasonic telemetry system in the eastern Pacific Ocean,'' Bull. Nat. Res. Inst. Far Seas Fish. **36**, 55–60.

Bubble oscillations driven by aspherical ultrasound in liquid

Wenjie Wang, Weizhong Chen,[a] Meijun Lu, and Rongjue Wei
Institute of Acoustics and State Key Laboratory of Modern Acoustics, Nanjing University, Nanjing 210093, China

(Received 8 November 2002; revised 28 April 2003; accepted 4 August 2003)

In this paper we study the bubble dynamics driven by an aspherical acoustic field, based on the theory of hydrodynamics. Evolution equations for an aspherical bubble are derived under the aspherical acoustic driving. The numerical calculations show that the aspherical bubble can oscillate stably and periodically under suitable conditions, which is out of the explanation of spherical perturbation theory. Furthermore, under some controlling parameters, the aspherical distortion can either grow rapidly, leading to the bubble's breakdown, or decay gradually making the bubble spherical, which is similar to the result of the perturbation theory driven by a spherical ultrasound. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1612486]

PACS numbers: 43.35.Hl, 78.60.Mq, 43.25.+y [AJS]

I. INTRODUCTION

Single-bubble sonoluminescence (SBSL) refers to the emission of light from an acoustically trapped bubble undergoing the highly nonlinear, presumably radial oscillation.[1] Experiments have been performed on emission spectra to estimate the bubble's size and its shape.[2,3] Recently, the shape departure of the bubble has been accessed by dipole radiation angular correlations.[4] The analysis of data yields that at the emission instant, the bubble's surface is actually an ellipsoid of eccentricity very close to 0.2 and major radius between 1.5 and 2 μm.[5] What interests us is how cavitation bubbles develop into the aspherical shape and perform stable, periodic oscillations. As is well known, the surface distortion of an originally spherical bubble must arise from some asymmetrical factors in its environment. Typical asymmetries are gravity, proximity to the boundary walls of cells or other oscillating bubbles, and perhaps just the pressure gradient across the bubble when its diameter is no longer very small compared with the acoustic wavelength.[6] In the authors' opinion, the aspherical acoustic pressure is one of the factors leading to the deformation of the bubble in SBSL. Figure 1 shows a typical system for SBSL. As is well known, the acoustic field is driven by a pair of transducers (PZT) stuck on the outside of the cell. It is considered that the distribution of the acoustic pressure is no better than rotational symmetrical in 3-D space unless the cell's wall is thick enough to perform purely radial vibrations. Theoretically, the dynamics for a spherical bubble driven by a spherical sound field can be well described by the Rayleigh–Plesset (R–P) equation:

$$R\frac{d^2R}{dt^2}+\frac{3}{2}\left(\frac{dR}{dt}\right)^2 = \frac{1}{\rho_l}[P_g(R,t)-P_a(t)-P_0] + \frac{R}{\rho_l c_l}\frac{d}{dt}[P_g(R,t)-P_a(t)]-\frac{2\xi}{\rho_l R}-\frac{4\nu}{R}\frac{dR}{dt}, \quad (1)$$

where R is the radius of the spherical bubble, $P_g(R,t)$ the pressure inside the bubble, P_0 ambient pressure, ρ_l liquid density, c_l sound speed in the liquid, ξ the interfacial tensor, and ν kinematic viscosity, respectively. According to Rayleigh's theory, the driving pressure is only related to the time t, and independent on the spatial variables. The simplest typical case is

$$P_a(t)=-P_a\sin\omega t, \quad (2)$$

with P_a and $\omega/2\pi$ being the amplitude and frequency of driving pressure, respectively. The more complicated forms of $P_a(t)$ have been used to drive the SBSL.[7] In our point of view, the R–P equation only describes the spherical bubble's periodic radial oscillations driven by the isotropic sound field. This means that the precondition of the R–P equation, Eq. (1), is that the system possesses a perfect spherical symmetry. Since any aspherical factor will lead to complex, we have to give up the pretty and simple form of Eq. (1). Fortunately, if the bubble suffers a small initial disturbance leading to a tiny distortion from the sphere, we can use the perturbation method to extend the R–P equation to the aspherical case.[9] Traditionally, the perturbation theory of the R–P equation shows that the initial aspherical bubble will either break up rapidly or recover the sphere gradually.[11] Therefore, the perturbation theory also serves as an instability analysis of the bubble dynamics. In this paper, we are interested in the bubble dynamics driven by the aspherical ultrasound, which is different from Eq. (2). In Sec. II a theoretical model describing the bubble's oscillations under the aspherical driving pressure is derived from hydrodynamics. In Sec. III, we give the numerical calculations based on this aspherical model. Some typical results show the role of the aspherical acoustic field. Finally, there is a brief summary.

II. THEORY

If the sound field in a cell is driven by a pair of transducers (see Fig. 1), it can be considered as an axial symmetry, $p(r,\theta,t)$, instead of spherical symmetry. In the compressible liquid, the pressure distribution, solution of the linear wave equation, can be written as follows:

[a]Electronic mail: wzchen@nju.edu.cn

0001-4966/2003/114(4)/1898/7/$19.00

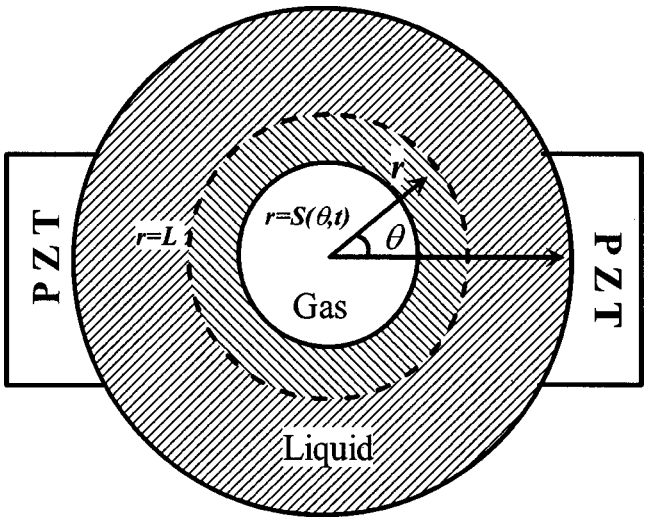


FIG. 1. A typical experiment cell for SBSL. A pair of transducers (PZT) are stuck symmterically on the outside of the spherical cell. The gas bubble's surface is at $r=S(\theta,t)$, the incompressible shell in $S(\theta,t)\leqslant r\leqslant L$, and the compressible shell in $r\geqslant L$, respectively.

$$p(r,\theta,t)=\sum_{n=0}^{+\infty} A_n(t)P_n(\cos\theta)j_n(kr), \tag{3}$$

where $A_n(t)$ is the nth time-dependence coefficient, $P_n(\cos\theta)$ and $j_n(kr)$ are Legendre polynomial and spherical Bessel function of degree n, and k is the wave number of the acoustic field, respectively.

In typical experiments of SBSL (Fig. 1), the driving sound frequency is approximately 25 kHz, which translates into a wavelength of $\lambda\approx 5.9$ cm given the sound velocity in water of $c_l=1481$ m/s. Relevant bubble radii scaled by the wavelength are small, that is, $R/\lambda\leqslant 10^{-3}$. Therefore, we can consider the external field in the long-wavelength limit. The leading term of Eq. (3) is the spherical component of the external field,

$$\begin{aligned} p_{\mathrm{sym}}(r,t)&=A_0(t)P_0(\cos\theta)j_0(kr)\\ &\approx A_0(t)[1+O((kr)^2)]. \end{aligned} \tag{4}$$

We ignore the term with $n=1$ in Eq. (3) because the translational effects are always disregarded in shape stability analysis. Thus, the leading aspherical pressure comes from the term with $n=2$, which is the dominant driving force for the aspherical components of the bubble oscillations,

$$\begin{aligned} p_{\mathrm{asym}}(r,\theta,t)&=A_2(t)P_2(\cos\theta)j_2(kr)\\ &\approx A_2(t)P_2(\cos\theta)[\tfrac{1}{15}(kr)^2+O((kr)^4)]. \end{aligned} \tag{5}$$

In this paper, we limit our scope with an acoustic field consisting of Eqs. (4) and (5) as follows:

$$p(r,\theta,t)=p_{\mathrm{sym}}(r,t)+p_{\mathrm{asym}}(r,\theta,t). \tag{6}$$

On the other hand, when the bubble is trapped in the pressure antinode, the compressibility of the liquid can be neglected, comparing to that of the gas inside the bubble. So the vicinity shell with thickness L ($R\ll L\ll\lambda$) outside the bubble can be regarded as incompressible. And the driving sound pressure is determined by the acoustic field on the spherical surface $r=L$,

$$P_a(\theta,t)=p_{\mathrm{sym}}(L,t)+p_{\mathrm{asym}}(L,\theta,t). \tag{7}$$

As

$$\frac{p_{\mathrm{asym}}(L,\theta,t)}{p_{\mathrm{sym}}(L,t)}\propto\frac{1}{15}(kL)^2\equiv\epsilon, \tag{8}$$

we rewrite the driving pressure Eq. (7) as follows:

$$P_a(\theta,t)=-(P_{\mathrm{sym}}+\epsilon P_{\mathrm{asym}}P_2(\cos\theta))\sin\omega t, \tag{9}$$

where the time-dependence factors in Eq. (7) have been assumed to be the harmonic form, $\sin\omega t$, and P_{sym} and $\epsilon P_{\mathrm{asym}}$ are amplitudes of spherical and aspherical driving pressures, respectively, with ϵ being defined by Eq. (8).

On this condition, it is natural to express the deformation in terms of spherical harmonics and we may take the equation for the bubble wall in spherical coordinates to be

$$F(r,\theta,t)=r-S(\theta,t)=0, \tag{10}$$

where $S(\theta,t)\equiv a_0(t)+\Sigma_{n=1}^{+\infty}\epsilon a_n(t)\cdot P_n(\cos\theta)$. We take $n=2$ to satisfy the pressure distribution. We ignore the $n=1$ mode, as described previously, consistent with the driving pressure distribution. We only retain one aspherical coefficient a_2, namely, the first and dominant correction of the bubble's deformation. As we mentioned above, the deformations of the bubble observed are small in SBSL. So the existence of the ϵ is reasonable. Though Eq. (10) is unable to describe the extreme condition such as the breakdown of the bubble or violent deformation, it does accord with such fact that the deformation is relatively small in SBSL. On the other hand, it has been proven that it is impossible to contain the stable deformation oscillations in viscous liquid if the distribution of the acoustic pressure is a purely radial and temporal function.[10,11] Thus, the introduction of spatial variable θ seems necessary.

In the following developments only terms of first order in ϵ will be retained. To this approximation the outward unit normal vector $\mathbf{n}$ at every point of the bubble's surface is

$$\mathbf{n}=\mathbf{e_r}-\epsilon\frac{a_2(t)}{r}\frac{\partial P_2(\cos\theta)}{\partial\theta}\mathbf{e}_\theta, \tag{11}$$

where $\mathbf{e_r}$ and $\mathbf{e}_\theta$ are orthogonal unit vectors tangent to the coordinate lines at every point.

The liquid velocity must obey

$$\nabla\cdot\mathbf{U}=0, \tag{12}$$

$$\frac{\partial\mathbf{U}}{\partial t}+(\mathbf{U}\cdot\nabla)\mathbf{U}=\frac{1}{\rho_l}\nabla\cdot\sigma, \tag{13}$$

where $\mathbf{U}$ is the velocity of the liquid and σ denotes the stress tensor, respectively. Now the velocity $\mathbf{U}$ and pressure field p are divided into three parts as follows:[9]

$$\mathbf{U}=\mathbf{U}_0+\epsilon\mathbf{U}_p+\epsilon\mathbf{U}_v \quad\text{and}\quad p=p_0+\epsilon p_p+\epsilon p_v. \tag{14}$$

The first terms $(\mathbf{U}_0,p_0)$ in these expressions describe the purely radial motion that would take place for $\epsilon=0$; the second terms $(\mathbf{U}_p,p_p)$ are the corrections of the potential flow induced by the aspherical ultrasound driving and the distor-

tion of the spherical bubble; and the third terms $(\mathbf{U}_v, p_v)$ are the corrections to the potential problems introduced by the presence of viscosity.[9]

The purely radial motion, the evolution of the radius $a_0(t)$, is described in terms of a potential source by writing

$$\mathbf{U}_0 = \nabla \phi_0, \tag{15}$$

with

$$\phi_0 = C_0(t) + \frac{D_0(t)}{r}, \tag{16}$$

where $C_0(t)$ and $D_0(t)$ are purely temporal functions.

The potential correction introduced by the deviation from sphericity of the sound field can be written as follows:

$$\mathbf{U}_p = \nabla \phi_2, \tag{17}$$

with

$$\phi_2 = (C_2(t) r^2 + D_2(t)/r^3) P_2(\cos\theta). \tag{18}$$

We first study the potential problem. In potential flow, Eq. (14) is reduced to

$$\mathbf{U} = \mathbf{U}_0 + \epsilon \mathbf{U}_p \quad \text{and} \quad p = p_0 + \epsilon p_p. \tag{19}$$

By integration, Eq. (13) is transformed to the Bernoulli equation,

$$\frac{\partial \phi}{\partial t} + \frac{1}{2}(\nabla \phi)^2 + \frac{p}{\rho_l} = f(t), \quad [S(\theta,t) \leqslant r \leqslant L], \tag{20}$$

where $\phi = \phi_0 + \epsilon \phi_2$, $p = p_0 + \epsilon p_p$, and $f(t)$ is an arbitrary temporal function. We apply Eq. (20) to the boundary surface of the incompressible shell, $r = L$ and $r = S(\theta,t)$, respectively. Here, we use $p_{l0} + \epsilon p_{lp} P_2(\cos\theta)$ as the substitute for $p|_{r=S(\theta,t)}$. Thus we get

$$\begin{aligned}
&\epsilon^0\text{:} \quad \left(\frac{1}{L}\frac{dD_0}{dt}\right) + \frac{1}{2}\left(\frac{D_0^2}{L^4}\right) + \frac{P_0 - P_{\text{sym}} \sin\omega t}{\rho_l} = f(t), \\
&\epsilon^1\text{:} \quad \left(L^2 \frac{dC_2}{dt} + \frac{1}{L^3}\frac{dD_2}{dt}\right) + \left(-\frac{D_0}{L^2}\right)\left(2C_2 L - \frac{3D_2}{L^4}\right) \\
&\qquad - \frac{P_{\text{asym}} \sin\omega t}{\rho_l} = 0,
\end{aligned} \tag{21}$$

and

$$\begin{aligned}
&\epsilon^0\text{:} \quad \frac{p_{l0}}{\rho_l} + \frac{\partial \phi_0}{\partial t} + \frac{1}{2}(\nabla \phi_0)^2 = f(t) \quad (r = a_0), \\
&\epsilon^1\text{:} \quad \frac{p_{lp}}{\rho_l} P_2(\cos\theta) + \left(\frac{\partial}{\partial r}\left(\frac{\partial \phi_0}{\partial t}\right)\right) a_2 P_2(\cos\theta) + \frac{\partial \phi_2}{\partial t} \\
&\qquad + \frac{1}{2}\left(\frac{\partial}{\partial r}\left(\frac{\partial \phi_0}{\partial r}\right)^2\right) a_2 P_2(\cos\theta) + (\nabla \phi_0 \cdot \nabla \phi_2) = 0 \\
&\qquad\qquad (r = a_0).
\end{aligned} \tag{22}$$

To introduce a viscous correction into the solution, the kinematical boundary conditions as well as Eq. (14) are taken into account. There should be no discontinuity in the normal velocities on the free surface,[9] so

$$\frac{\partial F}{\partial t} + (\mathbf{U}_0 + \epsilon \mathbf{U}_p) \cdot \mathbf{e_r} \frac{\partial F}{\partial r} = 0, \quad \text{on } F(r,\theta,t) = 0, \tag{23}$$

$$\mathbf{U}_v \cdot \mathbf{e_r} \frac{\partial F}{\partial r} = 0, \quad \text{on } F(r,\theta,t) = 0. \tag{24}$$

From Eq. (23), we get

$$\begin{aligned}
&\epsilon^0\text{:} \quad \frac{da_0}{dt} = \left.\frac{\partial \phi_0}{\partial r}\right|_{r=a_0}, \\
&\epsilon^1\text{:} \quad \frac{da_2}{dt} P_2(\cos\theta) = \left.\left(\frac{\partial}{\partial r}\left(\frac{\partial \phi_0}{\partial r}\right)\right)\right|_{r=a_0} a_2 P_2(\cos\theta) \\
&\qquad + \left.\frac{\partial \phi_2}{\partial r}\right|_{r=a_0}.
\end{aligned} \tag{25}$$

According to Ref. 9 we have

$$\mathbf{U}_v = T(r,t) Y_2^0(\theta,\varphi) \mathbf{e_r} - \nabla \Phi, \tag{26}$$

which satisfies Eq. (24) with

$$\begin{aligned}
\Phi = P_2(\cos\theta) \Bigg[&\left(\alpha(t) + \frac{3}{5}\int_{a_0}^{r} s^{-2} T(s,t)\, ds\right) r^2 \\
&+ \left(\frac{2}{3} a_0^5 \alpha(t) + \frac{2}{5}\int_{a_0}^{r} s^3 T(s,t)\, ds\right) r^{-3} \Bigg],
\end{aligned} \tag{27}$$

and

$$\alpha(t) = -\frac{3}{5}\int_{a_0}^{+\infty} s^{-2} T(s,t)\, ds. \tag{28}$$

As the discontinuity in the normal stress should equal the surface tension on the bubble surface, we get

$$p_l = p_g - \xi H + 2\eta \frac{\partial v_r}{\partial r} \quad [r = S(\theta,t)], \tag{29}$$

where p_l is the pressure exerted on the bubble surface in the liquid, p_g is the gas pressure inside the bubble, the mean curvature of the surface H and the radial component of the liquid velocity v_r are defined as follows:

$$H = \frac{2\xi}{a_0} + \epsilon \frac{4\xi a_2}{a_0^2} P_2(\cos\theta), \tag{30}$$

and

$$v_r = (\mathbf{U}_0 + \epsilon \mathbf{U}_p + \epsilon \mathbf{U}_v) \cdot \mathbf{e_r}. \tag{31}$$

The pressure inside the bubble p_g is related to the volume of the bubble V by an adiabatic process of the van der Waals gas,

$$p_g = p_{g0}\left(\frac{V_0 - b}{V - b}\right)^{\gamma}, \tag{32}$$

where V_0 and p_{g0} are the ambient bubble's volume and interior pressure, γ is the ratio of specific heats, and b is the collective hard core van der Waals volume, respectively. Furthermore, the bubble volume can be approximately expressed by $V = 4/3\pi a_0^3$. On the bubble wall, the pressure can be written as

$$p_l=(p_0+\epsilon p_p+\epsilon p_v)|_{r=S(\theta,t)}$$
$$=p_{l0}+\epsilon(p_{lp}+p_{lv})P_2(\cos\theta), \tag{33}$$

with

$$p_{lv}=2\rho_l\left\{\frac{\nu T(a_0,t)}{a_0}+\frac{1}{a_0}\frac{da_0}{dt}\int_{a_0}^{+\infty}\left[\left(\frac{a_0}{s}\right)^3-1\right]\times\left(\frac{a_0}{s}\right)^2 T(s,t)ds\right\}, \tag{34}$$

and

$$T(a_0,t)=a_0\left(2C_2-\frac{8D_2}{a_0^5}+\frac{10}{3}\alpha(t)+\frac{6a_2D_0}{a_0^4}\right). \tag{35}$$

As $\alpha(t)$ in Eq. (28) is a complicated integral, we simplify it with a boundary-layer-type model proposed by Prosperetii[9] and examined by Brener, etc.[10,11] According to this approach, considerable vorticity is localized within a small boundary layer of thickness δ around the bubble; and $T(r,t)\equiv 0$ outside the boundary layer, which means there is only potential flow outside the layer. We take $\delta=a_0/4$[8] and get

$$T(a_0,t)=\frac{2}{3}\left(2a_0C_2-\frac{8D_2}{a_0^4}+\frac{6a_2D_0}{a_0^3}\right). \tag{36}$$

So Eq. (34) is simplified to

$$p_{lv}=2\eta\frac{T(a_0,t)}{a_0}. \tag{37}$$

Thus we can get

$$\epsilon^0:\ p_{l0}=p_g-\frac{2\xi}{a_0}+4\eta\frac{D_0}{a_0^3},$$
$$\epsilon^1:\ p_{lp}=-\frac{4\xi a_2}{a_0^2}+2\eta\left(2C_2+\frac{12D_2}{a_0^5}\right)-12\eta\frac{D_0a_2}{a_0^4}+\eta\frac{T(a_0,t)}{a_0}. \tag{38}$$

Let us define the following dimensionless variables and constants:

$$\tilde{r}=\frac{r}{R_0},\quad \tilde{t}=\omega t,\quad \tilde{a}_0=\frac{a_0}{R_0},\quad \tilde{a}_2=\frac{a_2}{R_0},$$
$$\tilde{p}=\frac{p_l}{\rho_lR_0^2\omega^2},\quad \tilde{\eta}=\frac{\eta}{R_0^2\omega\rho},\quad \tilde{\xi}=\frac{\xi}{\rho_lR_0^3\omega^2},\ldots. \tag{39}$$

Here, R_0 is the initial value of a_0, namely, the initial radius of the bubble. From Eqs. (21), (22), (25), (38), we obtain

$$\left(\frac{1}{\tilde{a}_0}-\chi\right)\left(\tilde{a}_0^2\frac{d^2\tilde{a}_0}{d\tilde{t}^2}+2\tilde{a}_0\left(\frac{d\tilde{a}_0}{d\tilde{t}}\right)^2\right)-\frac{1}{2}\left(\frac{d\tilde{a}_0}{d\tilde{t}}\right)^2$$
$$=(\tilde{P}_g(\tilde{R},\tilde{t})-\tilde{P}_{\rm sym}\sin\tilde{t}-\tilde{P}_0)-\frac{2\tilde{\xi}}{\tilde{a}_0}-\frac{4\tilde{\nu}}{\tilde{a}_0}\frac{d\tilde{a}_0}{d\tilde{t}}, \tag{40}$$

and

$$\frac{d\tilde{C}_2}{d\tilde{t}}+\chi^5\frac{d\tilde{D}_2}{d\tilde{t}}+\chi^3\tilde{a}_0^2\frac{d\tilde{a}_0}{d\tilde{t}}(2\tilde{C}_2-3\chi^5\tilde{D}_2)+\chi^2\tilde{P}_{\rm asym}\sin\tilde{t}=0,$$
$$\frac{d\tilde{a}_2}{d\tilde{t}}+\frac{2\tilde{a}_2}{\tilde{a}_0}\frac{d\tilde{a}_0}{d\tilde{t}}-\left(2\tilde{C}_2\tilde{a}_0-\frac{3\tilde{D}_2}{\tilde{a}_0^4}\right)=0,$$
$$\frac{\tilde{a}_2}{\tilde{a}_0^2}\frac{d\tilde{D}_0}{d\tilde{t}}-\left(\tilde{a}_0^2\frac{d\tilde{C}_2}{d\tilde{t}}+\frac{1}{\tilde{a}_0^3}\frac{d\tilde{D}_2}{d\tilde{t}}\right)+2\tilde{D}_0^2\frac{\tilde{a}_2}{\tilde{a}_0^5}+\frac{\tilde{D}_0}{\tilde{a}_0^2}\left(2\tilde{C}_2\tilde{a}_0-\frac{3\tilde{D}_2}{\tilde{a}_0^4}\right)-\tilde{p}_{lp}=0,$$
$$\tilde{p}_{lp}+\frac{4\tilde{\xi}\tilde{a}_2}{\tilde{a}_0^2}-2\tilde{\eta}\left(2\tilde{C}_2+\frac{12\tilde{D}_2}{\tilde{a}_0^5}\right)+\frac{12\tilde{\eta}\tilde{D}_0\tilde{a}_2}{\tilde{a}_0^4}-\tilde{\eta}\frac{\tilde{T}(\tilde{a}_0,\tilde{t})}{\tilde{a}_0}=0, \tag{41}$$

where $\chi=R_0/L$. Equation (40) describes the evolution of the radial component of the bubble surface, and if we take acoustic radiation into account and eliminate χ for $1/\tilde{a}_0\gg\chi$, we may get

$$\tilde{a}_0\frac{d^2\tilde{a}_0}{d\tilde{t}^2}+\frac{3}{2}\left(\frac{d\tilde{a}_0}{d\tilde{t}}\right)^2=(\tilde{P}_g(\tilde{R},\tilde{t})-\tilde{P}_{\rm sym}\sin\tilde{t}-\tilde{P}_0)+\frac{\tilde{a}_0}{\tilde{c}_l}\frac{d}{d\tilde{t}}(\tilde{P}_g(\tilde{a}_0,\tilde{t})+\tilde{P}_{\rm sym}\sin\tilde{t})-\frac{2\tilde{\xi}}{\tilde{a}_0}-\frac{4\tilde{\nu}}{\tilde{a}_0}\frac{d\tilde{a}_0}{d\tilde{t}}, \tag{42}$$

which is exactly the same as the R–P equation (1).

As mentioned above, under the spherical pressure there is a parameter regime in which the disturbance can die out, if the bubble shape has an initial disturbance,[9] which is well known as the Rayleigh–Taylor instability analysis for the bubble dynamics driven by a spherical pressure. However, the governing equations in our paper for $\tilde{a}_2$, Eqs. (41), are new and differ from those in Ref. 9 due to the aspherical driving pressure. So now we have two sets of equations to describe $\tilde{a}_2$: one is for spherical driving, Eq. (2), which has been detailed in previous works, and the other for the aspherical driving, Eq. (9), which is derived in this paper. How will the bubble oscillate under the drive of the aspherical pressure, Eq. (9)? We now take a numerical study on the bubble dynamics driven by an aspherical pressure on the basis of Eqs. (41).

III. CALCULATION RESULTS

In numerical calculations we choose a set of typical values[12] for the physical parameters in Eqs. (40) and (41). These are $\sigma=0.076$ kg/s^2, $b=V_0/(8.5)^3$, $\gamma=1.33$, $\eta=1\times10^{-3}$ kg/m s, $\rho_l=10^3$ kg/m^3, $c_l=1480$ m/s, and $P_0=1$ bar, respectively. Furthermore, we set $P_{\rm sym}=1.3$ bar, $\omega/2\pi=26.5$ kHz, $R_0=3.5\,\mu$m, which are experimental data for SBSL. SBSL will occur in such parameters and a small disturbance with an $n=2$ mode oscillation will die away under purely spherical driving pressure.[10–12] In our paper, the

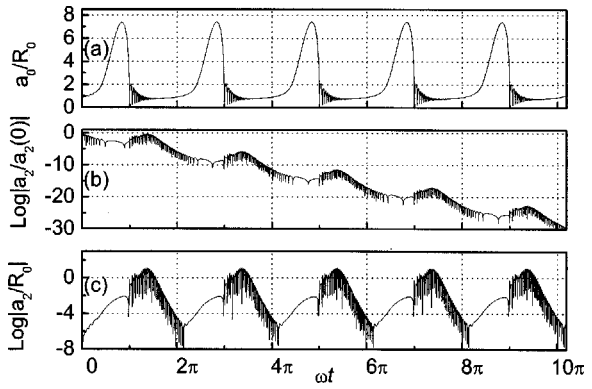


FIG. 2. The evolution of the aspherical bubble in SBSL for $a_0(0)=3.5\ \mu$m and $P_{\mathrm{sym}}=1.3$ bar. (a) The spherical component of oscillation, $\tilde{a}_0(\tilde{t})$, as a function of $\tilde{t}$; (b) $\log|a_2/a_2(0)|$ with an initial small disturbance and spherical driving pressure, $P_{\mathrm{asym}}/P_{\mathrm{sym}}=0$; (c) $\log|\tilde{a}_2|$ with aspherical driving pressure, $\tilde{a}_2(0)=0$, $P_{\mathrm{asym}}/P_{\mathrm{sym}}=1$.

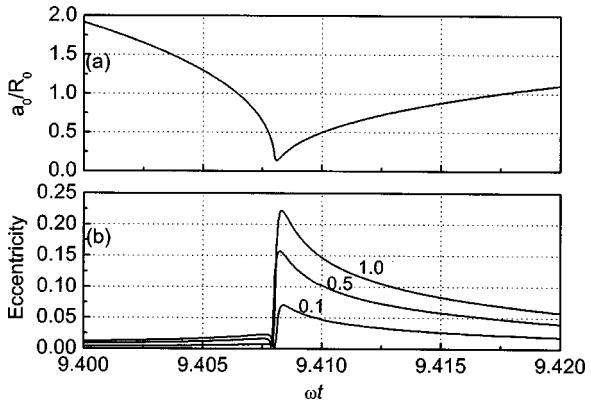


FIG. 3. Eccentricity of the bubble at emission instance under a different aspherical driving amplitude. (a) The bubble's radial component $\tilde{a}_0$ approaching the minimum. (b) Eccentricity of the bubble for different aspherical driving amplitudes, $P_{\mathrm{asym}}/P_{\mathrm{sym}}=0.1-1$. The initial conditions are $\tilde{a}_2(0)=0$ and $a_0(0)=3.5\ \mu$m, and the spherical driving amplitude is $P_{\mathrm{sym}}=1.3$ bar.

acoustic field deviating from spherical symmetry is discussed, as is described by Eq. (9) with $\epsilon=0.01$, which is related to the asymmetry of acoustic driving decided by the experimental system. Thus, from Eq. (8), we have $L=\sqrt{15\epsilon}/k=\sqrt{15\epsilon}c_l/\omega$ and $\chi=kR_0/\sqrt{15\epsilon}=\omega R_0/(c_l\sqrt{15\epsilon})$. Figure 2 compares the three results under different circumstances. Figure 2(a) shows the purely radial oscillation of the bubble under the purely spherical driving pressure. An aspherical oscillation of mode $n=2$ will die out even if the bubble has suffered an initial small disturbance under spherical pressure [see Fig. 2(b)], which agrees with the Rayleigh–Taylor instability theory. However, if the driving pressure becomes an aspherical one, the bubble will contain an aspherical oscillation of mode $n=2$ stably [see Fig. 2(c)]. In addition, it must be pointed out that we take the initial value of the $\tilde{a}_2$ as zero in our calculations while taking the aspherical driving into account, as shown in Fig. 2(c). So the aspherical oscillations are derived from the aspherical driving, namely, $P_{\mathrm{asym}}\sin\omega t$, instead of an initial small disturbance. We can surely add the initial value to $\tilde{a}_2$ just as we do in the spherical condition. The effect caused by the initial condition will die out after several periods. Furthermore, though we choose $\epsilon=0.01$ in our numerical simulation, the results are insensitive to the choice of the value of ϵ. Table I shows the results in different ϵ.

In order to understand the role of the aspherical driving pressure, we turn to the analysis of the details of $|\tilde{a}_2|$. In Fig. 2 we have plotted $\tilde{a}_0$ and $\tilde{a}_2$ for five periods. Comparing Fig. 2(b) with (c), one can say that there are no main differences between them in a single period. Both Figs. 2(b) and (c) show an initial growth from $\tilde{t}\approx\pi+2n\pi$ to $4\pi/3+2n\pi$ ($n=1,2,3,...$). Then, they decay from $\tilde{t}\approx4\pi/3+2n\pi$ to $\pi+2(n+1)\pi$. The remarkable difference arises when we compare them for several periods. In Fig. 2(b), at $\tilde{t}=0$, $\tilde{a}_2(0)$ is normalized to unity. At $\tilde{t}=2\pi$, $\log|a_2/a_2(0)|$ is negative, $|\tilde{a}_2|$ decays by several orders, which means the energy for shape oscillation decays rapidly under the purely spherical driving pressure. On the contrary the $|\tilde{a}_2|$ in Fig. 2(c) almost remains unchanged from a cycle to its next under an aspherical driving pressure. The energy for shape oscillation is made up from the aspherical driving of each period to overcome the attenuation due to damping such as viscosity and sound radiation. Therefore, the behavior of $\tilde{a}_2$ is related to two factors, both spherical and aspherical driving pressure. The aspherical driving $P_{\mathrm{asym}}\sin\omega t$ acts as an energy source of the aspherical oscillation, which determines the long time behaviors of $|\tilde{a}_2|$.

TABLE I. Insensitiveness of results to ϵ.

ϵ	χ	Peak value of $\log\|\tilde{a}_2\|$	Peak value of $\|\epsilon\tilde{a}_2\|$
0.1	3.22×10^{-4}	~0	~0.1
0.01	1.02×10^{-3}	~1	~0.1
0.001	3.22×10^{-3}	~2	~0.1

In retrospect to the beginning part of this paper, we have mentioned the bubble of the ellipsoid in SBSL. Numerical simulation shows that the eccentricity of the bubble is proportional to the amplitude of the aspherical sound pressure (see Fig. 3). The peak of the eccentricity of the bubble is comparable to the observed eccentricity during light emission, which occurs at the minimum of $\tilde{a}_0$. As a matter of fact, we are unable to quantitatively know the extension to which the sound field deviates from spherical symmetry in the experimental system, so our results are significant in qualitative description of the bubble. Of course, we can fit the experiments by choosing some suitable parameters. Our result confirms us that aspherical driving is a key factor leading to stable deformed oscillations.

Here we compare our result with the perturbation theory under spherical driving. One is based on perturbation theory, which emphasizes the evolution of the initial small disturbance. As shown in Fig. 2(b), the initial small disturbance lacks the persistence under the spherical driving, i.e., after several cycles, the bubble either recovers to the sphere, or does not exist at all. So it is not appropriate to account for the durative SBSL with the ellipsoidal bubble. Furthermore, the initial small distortion may take on a variety of forms besides the ellipsoid form. Our result based on our model introduces the aspherical driving pressure, as mentioned above. Such an introduction engenders a durative, stable, pe-

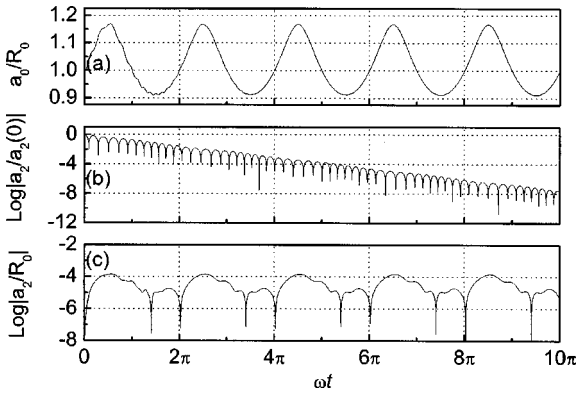


FIG. 4. The evolution of an aspherical bubble of the nonsonoluminescence bubble for $a_0(0)=10\ \mu$m and $P_{\mathrm{sym}}=0.5$ bar. (a) The spherical component of oscillation, $\tilde{a}_0(\tilde{t})$, as a function of $\tilde{t}$; (b) $\log|a_2/a_2(0)|$ with an initial small disturbance and spherical driving pressure, $P_{\mathrm{asym}}/P_{\mathrm{sym}}=0$; (c) $\log|\tilde{a}_2|$ with aspherical driving pressure, $\tilde{a}_2(0)=0$, $P_{\mathrm{asym}}/P_{\mathrm{sym}}=1$.

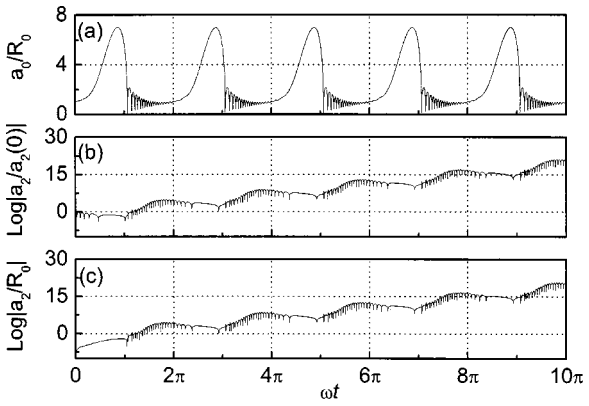


FIG. 5. Evolution of the aspherical bubble tending to break up, where $a_0(0)=5\ \mu$m and $P_{\mathrm{sym}}=1.3$ bar. (a) The spherical component of oscillation, $\tilde{a}_0(\tilde{t})$, as a function of $\tilde{t}$; (b) $\log|a_2/a_2(0)|$ with an initial small disturbance and spherical driving pressure, $P_{\mathrm{asym}}/P_{\mathrm{sym}}=0$; (c) $\log|\tilde{a}_2|$ with aspherical driving pressure, $\tilde{a}_2(0)=0$, $P_{\mathrm{asym}}/P_{\mathrm{sym}}=1$.

riodical driving force to make the bubble's deformation for every acoustic cycle.

We apply our theory to the nonsonoluminescencing bubble with initial radius $R_0=10\ \mu$m. As is well known, a large bubble can only be alive at lower driving.[10] We adjust the driving amplitude $P_{\mathrm{sym}}=0.5$ bar and keep other parameters unchanged. The numerical calculation shows that the bubble oscillates in a nearly linear way and with a small amplitude as well [see Fig. 4(a)]. Figure 4(b) shows that the bubble with initial shape disturbance will recover to the sphere soon in a purely spherical sound field. Figure 4(c) shows that stable shape deformation is able to exist due to the aspherical driving. It is easy to find out that the amplitude of $\tilde{a}_2$ in Fig. 4(c) is tiny. A small amplitude is due to low driving power. If the aspherical driving were allowed to be amplified, the amplitude of $\tilde{a}_2$ would be enhanced.

Two cases mentioned above do not include the bubble tending to break up. In perturbation theory, there exists the case in which an initially small disturbance will evolve into infinity, namely, the bubble will break up ultimately. Such a case also exists under the aspherical driving pressure, because the aspherical driving is an energy source of the aspherical oscillation. Figure 5 shows the bubble's breakdown, where the bubble's initial radius $R_0=5\ \mu$m, and the amplitude of spherical component of driving pressure $P_{\mathrm{sym}}=1.3$ bar We may say that if a bubble cannot exist stably under a spherical driving pressure, neither can it under an aspherical driving pressure.

Finally, the authors would like to point out that if we eliminate the aspherical driving, namely, $P_{\mathrm{asym}}\sin\omega t$, Eqs. (41) will act just as the perturbation theory, aiming at the evolution of disturbance in mode $n=2$ sphere distortion.

IV. SUMMARY

Our aim of present work is to consider the actual experiment system without isotropy. The traditional Rayleigh theory of bubble based on space isotropy may encounter some conflicts with actual experiments' results, because the purely spherical sound fields are hard to obtain. In our point of view, we should pay attention to the environments of the bubble such as sound fields. Therefore, based on the hydrodynamics, we derive our equations, which include the deviation from purely radial spherical driving. Our calculations reveal some interesting facts: some of them are similar to those in traditional spherical theory; a stable aspherical oscillation is out of the previous result. In spherical theory, the bubble cannot perform a stable aspherical oscillation. The surface distortion caused by an initial small disturbance will evolve into two futures in liquid with damping. One is to recover to the sphere; the other is to break up. By the introduction of the aspherical factor such as the deviation from the spherical driving pressure, the spherical bubble without an initial small surface disturbance will be forced to deviate from its sphere and to perform a stable aspherical oscillation. In other words, the aspherical oscillation will get energy from the corresponding aspherical driving continuously. On the other hand, as the dominant factor that destroys the spherical symmetry, the aspherical driving pressure can also force the bubble to break up. In the actual experiments, it is nearly impossible to get sound fields perfectly symmetrical. So our work shows its importance. Our results reveal that if the acoustic field deviates from spherical symmetry, there will be part of acoustic energy obtained by a high mode of oscillation. And the purely radial mode, which determines the volume of the bubble, will get less energy. So at least in SBSL, if the environment's symmetry get improved, the light emission intensity will be enhanced. An other factor leading to the bubble's deformation such as gravity has been studied in Ref. 13. We would like to point out that our work aims at the sonoluminescencing bubble levitated at the antinode of the acoustic field. If the bubble is translating, a_n in Eqs. (40) and (41) will be coupled together, which makes the problem more complicated, as proposed in Ref. 14.

ACKNOWLEDGMENTS

This work was partly supported by the National Natural Science Foundation of China (Grant No. 10174036), the Trans-Century Training Program Foundation for the Talents by the State Education Commission of China, and the Spe-

cial Funds for Major State Based Research Projects of China. We thank Professor Andrew Szeri and the referees for their valuable comments.

[1] D. F. Gaitan, L. A. Crum, C. C. Church, and R. A. Roy, ''Sonoluminescence and bubble dynamics for a single, stable, cavitation bubble,'' J. Acoust. Soc. Am. **91**, 3166–3188 (1992).
[2] R. Hiller, S. J. Putterman, and B. P. Barber, ''Spectrum of synchronous picosecond sonoluminescence,'' Phys. Rev. Lett. **69**, 1182–1184 (1992).
[3] B. P. Barber and S. J. Putterman, ''Light scattering measurements of the repetitive supersonic implosion of a sonoluminescencing bubble,'' Phys. Rev. Lett. **69**, 3839–3842 (1992).
[4] K. Weninger, S. J. Putterman, and B. P. Barber, ''Angular correlations in sonoluminescence: Diagnostic for the sphericity of a collapsing bubble,'' Phys. Rev. E **54**, 2205–2208 (1996).
[5] A. Madrazo, N. Garcia, and M. Nieto-Vesperinas, ''Determination of the size and shape of a sonoluminescent sing bubble: theory on angular correlations of the emitted light,'' Phys. Rev. Lett. **80**, 4590–4593 (1998).
[6] F. R. Young, *Cavitation* (McGraw-Hill, New York, 1989), Chap. 3.
[7] W. Z. Chen, X. Chen, M. J. Lu, and R. J. Wei, ''Single bubble sonoluminescence driven by non-simple-harmonic ultrasounds,'' J. Acoust. Soc. Am. **111**, 2632–2637 (2002).
[8] V. A. Bogoyavlenskiy, ''Single-bubble sonoluminescence: Shape stability analysis of collapse dynamics in a semianalytical approach,'' Phys. Rev. E **62**, 2158–2167 (2000).
[9] A. Prosperetti, ''Viscous effects on perturbed spherical flows,'' Q. Appl. Math. **34**, 339 (1977).
[10] M. P. Brenner, D. Lohse, and T. F. Dupont, ''Bubble shape oscillations and the onset of sonoluminescence,'' Phys. Rev. Lett. **75**, 954–957 (1995).
[11] S. Hilgenfeldt, D. Lohse, and M. P. Brenner, ''Phase diagrams for sonoluminescing bubbles,'' Phys. Fluids **8**, 2808–2857 (1996).
[12] C. C. Wu and P. H. Roberts, ''Bubble shape instability and sonoluminescence,'' Phys. Lett. A **250**, 131–136 (1998).
[13] T. J. Matula, ''Single-bubble sonoluminescence in microgravity,'' Ultrasonics **38**, 559–565 (2000).
[14] A. J. Reddy and A. J. Seriz, ''Shape stability of unsteadily translating bubbles,'' Phys. Fluids **14**, 2216–2224 (2002).

A cascade thermoacoustic engine

D. L. Gardner and G. W. Swift[a)]
Condensed Matter and Thermal Physics Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 11 April 2003; revised 29 July 2003; accepted 4 August 2003)

A cascade thermoacoustic engine is described, consisting of one standing-wave stage plus two traveling-wave stages in series. Most of the acoustic power is produced in the efficient traveling-wave stages. The straight-line series configuration is easy to build and allows no Gedeon streaming. The engine delivers up to 2 kW of acoustic power, with an efficiency (the ratio of acoustic power to heater power) of up to 20%. An understanding of the pressure and volume-velocity waves is very good. The agreement between measured and calculated powers and temperatures is reasonable. Some of the measured thermal power that cannot be accounted for by calculation can be attributed to Rayleigh streaming in the two thermal buffer tubes with the largest aspect ratios. A straightforward extension of this work should yield cascade thermoacoustic engines with efficiencies of around 35–40% of the Carnot efficiency. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1612483]

PACS numbers: 43.35.Ud, 43.25.Vt [RR]

I. INTRODUCTION

Rayleigh's criterion[1] for self-sustained heat-driven acoustic oscillations, ''If heat be given to the air at the moment of greatest condensation or be taken from it at the moment of greatest rarefaction, the vibration is encouraged,'' is the foundation of the many thermoacoustic schemes for producing acoustic power from heat without moving parts. Early concepts include the heat-driven electric generators described by Marrison,[2] Carter and Feldman,[3,4] and Ceperley.[5,6] Marrison's engine used standing-wave phasing and deliberately imperfect thermal contact to achieve Rayleigh's criterion, and had mature heat exchangers but without anything like today's stack between them.[7] The engines of Carter, Feldman *et al.* also used standing-wave phasing and deliberately imperfect thermal contact, and had stacks but no mature heat exchangers. Taking a completely different approach, Ceperley realized that Stirling engines[8] had always been meeting Rayleigh's criterion, using traveling-wave phasing and excellent thermal contact in a regenerator surrounded by mature heat exchangers, so he suggested acoustic networks to eliminate the Stirling engines' pistons.

Thermoacoustic research has progressed steadily over succeeding decades, and profited greatly from the publication of Rott's mathematical treatment.[9–11] A wide variety of standing-wave engines,[12–15] as illustrated in Fig. 1(a), has been built combining practical aspects of Marrison's and Feldman's work and based on Rott's theory. More recently, traveling-wave engines with the toroidal topology suggested by Ceperley, shown in Fig. 1(b), have also been demonstrated.[16–18] However, efforts to commercialize such engines have not yet succeeded.

The straight-line, closed–open topology shown in Fig. 1(a) works because the imperfect thermal contact in a stack can meet Rayleigh's criterion for acoustic-power production using pressure–velocity phase differences anywhere near that of a standing wave.[19] Acoustic power can flow out of both ends of the stack, mostly at the ambient end as in Fig. 1(a) or mostly at the hot end.[20,21] Although this straight-line topology is the simplest to build, the deliberately imperfect thermal contact on which it relies limits its efficiency.[19] The best traveling-wave engines[17,22] (and traditional Stirling engines) have 50% higher efficiency than the best standing-wave engines because the thermal contact in regenerators is excellent. However, they require phasing near that of a traveling wave to create acoustic power: Traveling-wave engines only amplify, so power can only flow out of the hot end of a regenerator if a smaller power is fed into the ambient end.[19,23] The toroidal topology shown in Fig. 1(b) achieves this by feeding some of the engine's own power, from the regenerator's hot end, back to its ambient end. Unfortunately, the toroidal topology is more difficult to build than the straight-line topology, and also suffers from a circulating second-order mass flow, Gedeon streaming,[24] that can reduce or eliminate the efficiency advantage by convecting heat from the hot heat exchanger to one of the ambient heat exchangers. Gedeon streaming has been successfully suppressed by exploiting the time-averaged pressure gradient developed in oscillating flow through an asymmetric channel,[17] but this consumes acoustic power. In addition, fabricating asymmetric channels that are adjustable, to stop Gedeon streaming under a variety of operating conditions, adds complexity that is undesirable in commercial devices.

In this paper we describe our first attempt to enjoy the best features of standing-wave engines and traveling-wave engines simultaneously, by arranging a standing-wave engine and one or more traveling-wave engines in a series, as shown in Fig. 1(c), a topology we call a ''cascade.'' The straight-line topology is easy to build, automatically prevents Gedeon streaming, and has high velocities only where streamlining is easy, while the efficiency is reasonably high because most of the acoustic power can be created in the traveling-wave stage or stages.

a)Electronic mail: swift@lanl.gov; URL: www.lanl.gov/thermoacoustics/

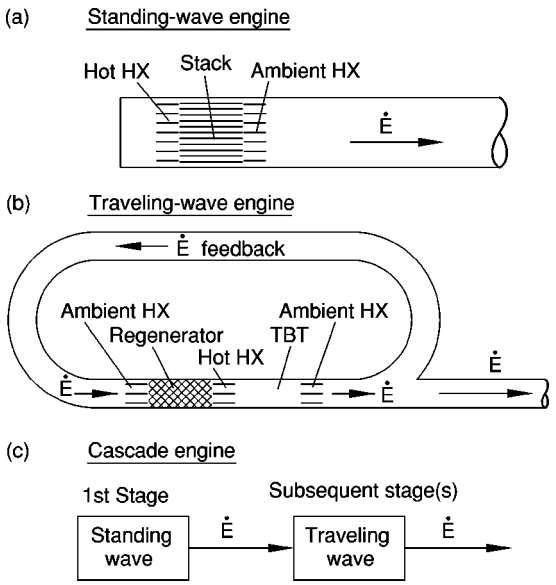


FIG. 1. Some thermoacoustic engine topologies. HX=heat exchanger, TBT=thermal buffer tube, $\dot{E}$=acoustic power. (a) In a standing-wave engine, the temperature difference between the hot heat exchanger and the ambient heat exchanger falls across the stack, whose pore dimensions are of the order of a few thermal penetration depths. Here, the standing-wave engine is in a simple cylindrical resonator, closed at its hot end and delivering acoustic power through its ambient end. (b) In a traveling-wave engine, the temperature difference between the hot heat exchanger and the ambient heat exchanger falls across the regenerator, whose pore dimensions are much smaller than a thermal penetration depth. Acoustic power can only be produced if some is fed into the ambient end of the regenerator, such as through the acoustic feedback path shown here. (c) The cascade engine combines one standing-wave engine with one or more traveling-wave engines. The standing-wave engine supplies the acoustic power needed at the ambient end of the adjacent traveling-wave engine.

Much of our fundamental thermoacoustics work at Los Alamos is motivated by our collaboration with Praxair, Inc., directed toward the development of powerful thermoacoustic natural-gas liquefiers.[25] Hence, we decided to build a cascade engine that might provide guidance to some future aspect of that collaboration, while still at low enough power for convenient laboratory experimentation. The experimental engine is ''similar'' to a hypothetical 35-kW engine. Similitude[26] in thermoacoustics shows that reducing all dimensions of such a hypothetical engine by a factor of two and changing from helium to argon (at a slightly lower pressure) preserves all dimensionless thermoacoustic variables such as Mach numbers, Reynolds numbers, and ratios of penetration depths to pore sizes. All thermoacoustic behavior, both linear and nonlinear, is identical when expressed with dimensionless variables. This scaling lets the experimental engine fit vertically in our lab and reduces the total heater power required so that electric heaters easily suffice.

In Sec. II we summarize the qualitative considerations that led us to adopt this particular cascade configuration: one standing-wave stage and two traveling-wave stages located within one pressure maximum of the wave. In Sec. III we describe the construction of this apparatus. The principal complexity was the need to design and build about three times more parts than for previous thermoacoustic engines, due to the presence of three stages. This complexity is reflected throughout the paper, as there are three sets of parts to describe, often three sets of data to explain, and many references to cite. In Sec. IV we describe the initial debugging, measurements of heat leak, and the evolution of the thermoacoustic measurements, as some problems were solved while others grew. In Secs. V and VI we describe the thermoacoustic measurements in detail, with pressure and volume–velocity waves in Sec. V and temperatures and powers in Sec. VI. Our understanding of the waves is very good, but an understanding of the powers leaves room for future work, including the study of heat transport by streaming in thermal buffer tubes. Finally, in Sec. VII we discuss the efficiency of the cascade. Although this engine converted heater power to acoustic power with an efficiency of only 20%, many of the sources of inefficiency are due to our decision to build a similitude scale model, and others are probably associated with one design error discussed below. We are confident that cascade engines can be built with efficiencies in the range of 35%–40% of Carnot's efficiency, the upper end of this range equaling the highest measured efficiency in a thermoacoustic engine to date.[17]

II. BROAD DESIGN CONSIDERATIONS

To design an experimental test of the cascade idea, we began with decisions about the number of stages and the resonator type, guided by the qualitative and approximate considerations (principally of efficiency and size) described in this section. After these decisions were made, we relied on a numerical analysis to select specific dimensions and other details, compromising as necessary to satisfy practical fabrication constraints.

The number of stages is an important issue. A traveling-wave stage can create acoustic power from thermal power with high efficiency, but, in the linear topology of the cascade shown in Fig. 1(c), the traveling-wave stage must be supplied with acoustic power from another stage, such as a standing-wave stage. If τ is the ratio of hot and ambient (absolute) temperatures, the acoustic-power gain[5] of a traveling-wave stage is approximately τ and the acoustic power created in it is approximately proportional to $\tau-1$. Then a series combination of a standing-wave stage with efficiency of $\sim 1/5$ and $N-1$ traveling-wave stages each with efficiency $\sim 1/3$ has an overall efficiency,

$$\eta \sim \frac{\tau^N}{2+3\tau^N}, \tag{1}$$

showing how the efficiency grows with the number of stages, as more and more acoustic power is created in the efficient traveling-wave stages: $\eta=0.20$ for one stage (standing-wave only), $\eta=0.26$ for two stages and $\tau=2.5$, $\eta=0.30$ for three stages and $\tau=2.5$, $\eta=0.32$ for four stages and $\tau=2.5$, and $\eta=0.33$ for an infinite number of stages. Hence, three or four stages seems to be an acceptable compromise between the optimum efficiency of a large number of stages and the simplicity of a small number of stages.

Ceperley[5] realized that the best efficiency in traveling-wave engines is obtained when the time phase ϕ by which oscillating pressure p_1 leads oscillating velocity u_1 is near zero. Figures 2(a)–(d) show four ways to satisfy this crite-

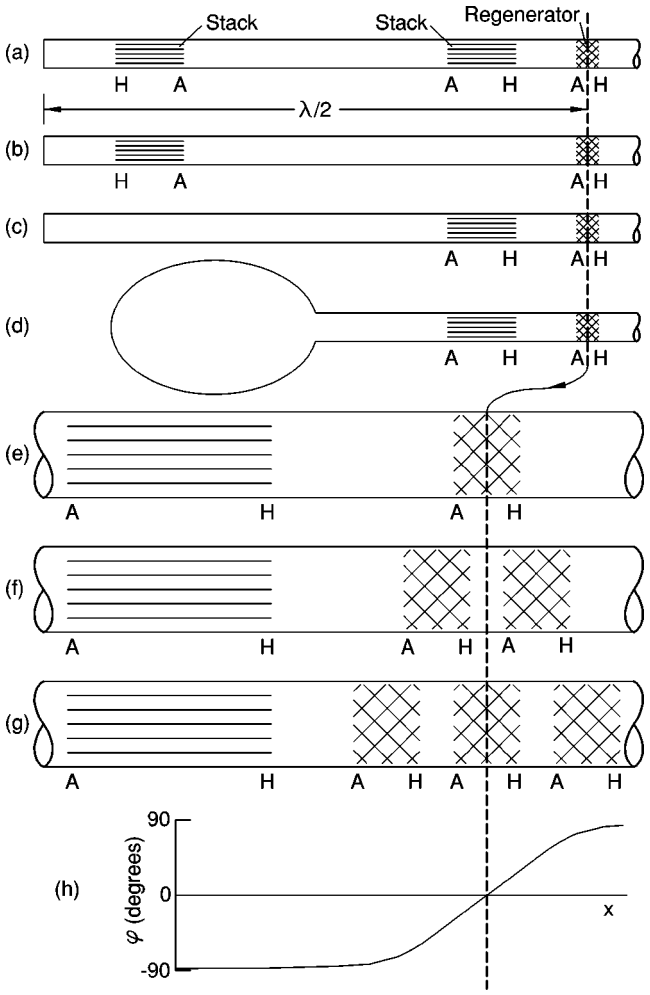


FIG. 2. An illustration of some alternatives in the design of cascade engines. The dashed line running through all parts of the figure marks the location of the sweet spot where the phase ϕ by which pressure leads velocity is zero. Acoustic power flows out of the right end of the apparatus in all cases. H and A indicate the hot and ambient ends, respectively, of each stack or regenerator. For clarity, heat exchangers and thermal buffer tubes are not shown. Parts (a) through (d) illustrate options for the resonator and the standing-wave stage(s); we chose (d). Parts (e) through (g) illustrate a closer view of different numbers of traveling-wave stages; we chose (f). Part (h) illustrates ϕ as a function of position in (e) through (g).

rion: with a traveling-wave stage at the ''sweet spot'' of zero phase, fed by one or two standing-wave stages. From these, we chose (d) to keep the total height of the apparatus as short as possible; other criteria (such as minimum weight) might encourage a different choice.

Ceperley[6] also realized that the magnitude of the specific acoustic impedance z of a pure traveling wave is too low for efficient traveling-wave engine performance, because the relatively high velocity present in a pure traveling wave causes high viscous dissipation of acoustic power in the regenerator and high shuttle transport of heat through the regenerator. The best performance requires $|z| \sim 10\rho_m a$, where ρ_m is mean density and a is sound speed. Initially, we simply assumed that we would be able to reach such a condition in the traveling-wave stages by changing area from stage to stage (not shown in Fig. 2); later detailed design calculations confirmed this assumption.

Another factor affecting efficiency is the number of traveling-wave stages that can be collocated at the sweet spot of zero phase, as illustrated in Figs. 2(e)–(h). When $|z| \gg \rho_m a$, the sweet spot is a significant local minimum in velocity. Viscous dissipation of acoustic power in a regenerator and heat transport through a regenerator due to imperfect thermal contact are both proportional to $|u_1|^2$, so a doubling of $|u_1|^2$ from its minimum value at the sweet spot can be taken as a rough measure of how far from the sweet spot a regenerator might acceptably be. Moving away from the sweet spot, the volume velocity U_1 grows according to the continuity equation,[27]

$$\frac{dU_1}{dx} \sim -i\omega c p_1, \tag{2}$$

where ω is the angular frequency, c is the compliance per unit length, $i = \sqrt{-1}$, and the subscript 1 denotes a complex variable. With changes in U_1 90° out of phase from U_1 itself, a doubling of $|U_1|^2$ occurs when $\phi \sim 45°$. Hence, multiple traveling-wave stages should be crowded together within $\phi \sim \pm 45°$ of the sweet spot. However, a thermal buffer tube is needed between stages to insulate the ambient end of one stage from the hot end of its neighbor. For effective insulation, the length L_{tbt} of a thermal buffer tube must be of the order of 10 times the gas displacement amplitude $|u_1|/\omega$. Combining this requirement with $|z| \sim 10\rho_m a$ yields

$$L_{\mathrm{tbt}} \sim 10|u_1|/\omega \sim |p_1|/\omega\rho_m a \sim \lambda/100, \tag{3}$$

for $|p_1|/p_m \sim 0.1$, where λ is the acoustic wavelength. Taking $L_{\mathrm{tbt}}/\gamma p_m$ as the minimum compliance per unit area between stages, Eqs. (2) and (3) show that $\Delta|U_1| \sim |U_1|$. Hence, for $|p_1|/p_m \sim 0.1$, the practical limit is two or possibly three traveling-wave stages in one pressure maximum. The heat exchangers and the regenerator itself have compliance in addition to that of the thermal buffer tube, so it seems unlikely that more than two traveling-wave stages can effectively share one sweet spot.

Hence, we chose a combination of Figs. 2(d) and (f) for our apparatus.

III. APPARATUS

The cascade apparatus, shown in Fig. 3, has its three engine stages in the center of a tall, thin resonator with a single pressure antinode in the center. It was built of welded steel, mostly 304 and 316 stainless steel for strength at high temperatures and rust resistance in water-cooled heat exchangers, but with carbon steel for some parts of the pressure vessel. The wall thicknesses of all pressure barriers were determined using Sec. VIII of the Pressure Vessel Code.[28] Two flanged rubber-O-ring seals, shown as collars just above and below the pressure housing in Fig. 3, enable handling the apparatus in three parts of roughly equal height. In the following top-to-bottom description of the apparatus, lengths reported are along the acoustic axis (generally vertical) and diameters reported are inside diameters.

Above the 1st stage, the upper portion of the resonator, which is of the type shown in Fig. 2(d), was designed for the minimum dissipation of acoustic power without regard for size. The 1.0-m long compliance at the top has a volume of

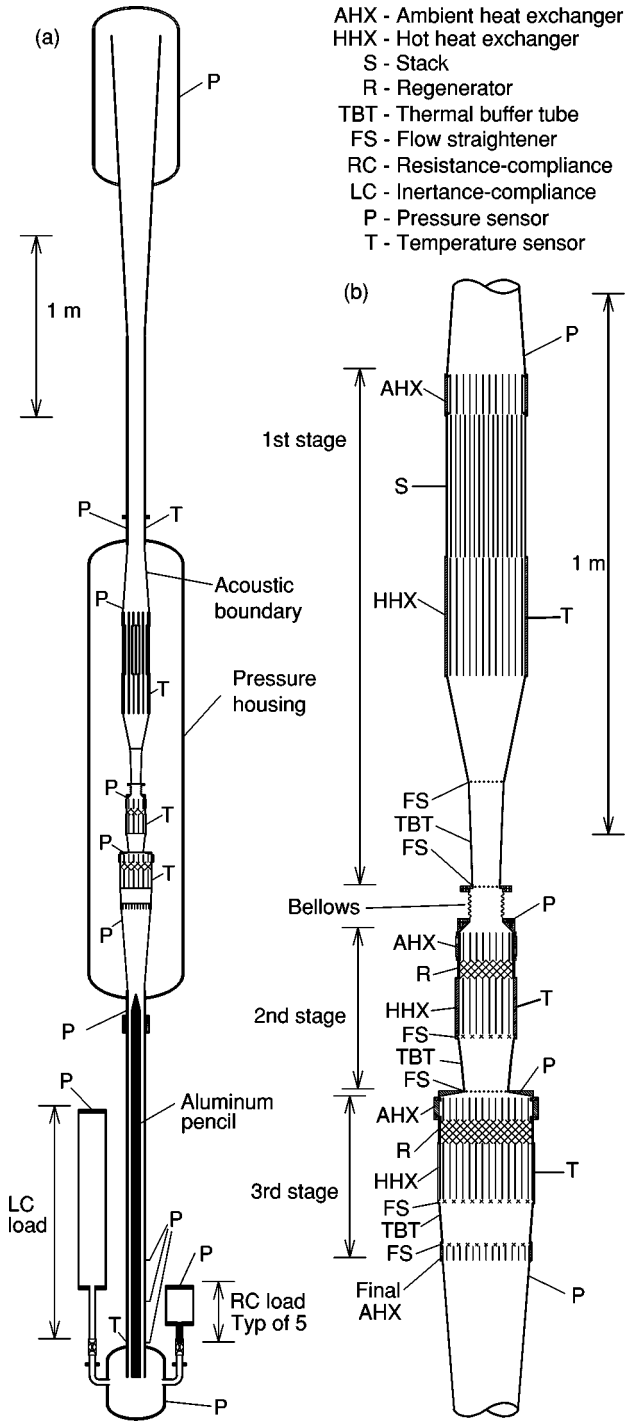


FIG. 3. The cascade engine built for this work. Everything is drawn to scale except the pores inside the heat exchangers, stack, regenerators, and flow straighteners. (a) The entire system, including (from top to bottom) the upper resonator, the cascade engine inside the pressure housing, and the lower resonator and load components. (b) A more detailed view of the three stages in the cascade engine. The 1st stage is a standing-wave engine; the 2nd and 3rd stages are traveling-wave engines. The major components and sensors are identified by the legend in the figure.

$0.11\ m^3$ outside the conical penetration in it. This cone is 1.66-m long and penetrates 0.83 m into the compliance, ending there with an inside diameter of 0.26 m. The cone lowers the acoustic velocity of the gas entering the compliance, reducing minor-loss dissipation of acoustic power there, and its penetration far into the compliance keeps the overall height of the apparatus (7.8 m) as short as possible. The small diameter of the cone smoothly joins the 9.0-cm-diam, 1.14-m long upper resonator pipe that penetrates the pressure housing. A flowing-water jacket (not shown in Fig. 3) surrounds most of the exposed portions of this pipe and half of the cone above it, but there was so little dissipation of acoustic power in the upper portion of the resonator that shutting off its water made no noticeable difference to the engine's performance. A 0.38-m long cone connects the lower end of the resonator pipe smoothly to the 1st-stage ambient heat exchanger.

The pressure housing surrounding the three stages contains the 2.4-MPa mean argon pressure, so that the inner shells labeled an ''acoustic boundary'' in Fig. 3(a) need to support only the acoustic pressure amplitude. Thus, the walls of this acoustic boundary around the stack, regenerators, and thermal buffer tubes are only 1–2 mm thick (depending on diameter and temperature), much thinner than would be required if they supported the mean pressure, with a concomitant reduction of heat leaks. The space between the acoustic boundary and the pressure housing was packed with refractory ceramic fiber thermal insulation,[29] to a density of $10\ lb/ft^3$, to reduce heat leak from the hot heat exchangers. Despite this dense packing of the insulation, the argon permeating it experienced significant gravity-driven convection, which is discussed in the next section.

Most of the quantitative details of the construction of the three stages of the cascade engine are presented in Table I. The uppermost component, the 1st-stage ambient heat exchanger, is of traditional, welded tube-and-shell design, with the argon oscillating vertically through the tubes and ambient-temperature water flowing crosswise in the spaces between the tubes. The water is delivered through three 10-mm-diam ports on one side, and removed similarly from the other side. Below this heat exchanger is a honeycomb stack[30] with 1 mm between opposite flats in each hexagonal channel in the honeycomb. Next comes the electrically heated 1st-stage hot heat exchanger, consisting of a cross-drilled stainless-steel cylinder. The oscillating argon passes through the axial holes, and the 6.4-mm-diam transverse holes hold electric cartridge heaters of various lengths corresponding to the hole lengths. The heaters are wired in three parallel groups in a Delta configuration and supplied with three-phase 60-Hz power through three mechanically linked variable autotransformers connected to the 208-Volt mains.

The taper angle of the 1st-stage thermal buffer tube was designed to suppress Rayleigh streaming,[31] and the stainless-steel-screen flow straighteners at the ends of the thermal buffer tube were designed to prevent jet-driven streaming.[19] Together, these features were intended to encourage thermally stratified oscillating flow in the thermal buffer tube, without gross convection of heat from hot to ambient. A 0.20-m long ''hot cone'' connects the bottom of the 1st-stage hot heat exchanger to the top of the 1st-stage thermal buffer tube. Without the hot cone, the velocity at the top of the 1st-stage thermal buffer tube would have been too high to preserve laminar flow in the boundary layer, and the taper angle calculated to suppress Rayleigh streaming in this ther-

TABLE I. Dimensions and other details of the three stages of the cascade. HX = heat exchanger, TBT = thermal buffer tube. Ranges of diameters for thermal buffer tubes indicate a conical taper. Gaps of 2–3 mm at interfaces between components are not tabulated.

	Length (cm)	Diam (cm)	Porosity	Number of bores on axis	Bore diam (mm)	Number of heaters	Mesh (wires/inch)	Wire diam (μm)	Number of screens
1st stage									
Ambient HX	7.62	14.3	43%	283	5.54				
Stack	26.0	14.3	92%	~20,000	0.97				
Hot HX	21.5	14.3	17%	88	6.35	108			
Flow str.		6.39					16	410	2
TBT	19.4	6.39–4.94							
Flow str.		4.94					16	410	1
2nd stage									
Ambient HX	5.40	10.1	20%	361	2.36				
Regenerator	3.10	10.1	75%				145	56	278
Hot HX	10.2	10.1	12%	124	3.18	54			
Flow str.		10.1					50	190	13
TBT	9.90	10.1–7.92							
Flow str.		7.92					50	140	1
3rd stage									
Ambient HX	4.30	17.1	20%	1139	2.36				
Regen. (1st half)	2.05	17.1	75%				180	46	224
Regen. (2nd half)	2.15	17.1	76%				120	66	155
Hot HX	10.2	17.1	12%	362	3.18	78			
Flow str.		17.1					50	190	13
TBT	7.50	17.1–16.0							
Flow str.		16.0					50	190	13
Final amb. HX	2.54		20%	913	2.36				

mal buffer tube would have been too steep to prevent flow separation.

The small end of the 1st-stage thermal buffer tube is bolted and sealed to a ten-convolution, 8-cm long flanged bellows having diameter 4.9 cm, which accommodates thermal expansion of the three heated engine stages relative to the pressure housing. Below, the bellows connects to a 2-cm long transition cone, which, in turn, connects to the 2nd stage.

The 2nd and 3rd stages are similar in character to the 1st, with tube-and-shell ambient heat exchangers, electrically heated hot exchangers, and tapered thermal buffer tubes between stainless-steel-screen flow straighteners. In place of the honeycomb stack of the 1st stage, the 2nd and 3rd stages have regenerators made of stacked circular pieces of stainless-steel screen. The velocities at the transitions from the hot heat exchangers to the thermal buffer tubes are low enough in these stages that nothing like the hot cone of the 1st stage is needed. A 1.2-cm long transition cone accommodates the diameter mismatch between the bottom of the 2nd stage and the top of the 3rd stage.

A 0.45-m long cone reaches from the final ambient heat exchanger to the 2.13-m long, 9.0-cm-diam lower resonator pipe exiting the pressure housing and penetrating 16 cm into the lower compliance. A flowing-water jacket (not shown in Fig. 3) surrounds most of the exposed portion of this pipe above the pressure transducers, and dissipation of acoustic power in this part of the apparatus was high enough that shutting off the flowing water endangered the transducers. An error in the initial design of the apparatus, described in the next section, required a large increase of inertance in the lower resonator pipe after the apparatus was built. This was accomplished by retrofitting a 2.11-m long, 5.1-cm-diam aluminum ''pencil'' [shown in Fig. 3(a)], streamlined at its upper end, into the lower resonator pipe, to reduce the pipe's area.

The lower compliance, which has a volume of 0.024 m^3, has six penetrations leading through six ball valves to five RC loads and one LC load, to consume power from the engine and to change the phase of the impedance of the lower resonator, respectively. The resistive element ''R'' in each RC load is a 10-cm long brass cylinder having diameter 3.3 cm and drilled axially with nineteen 2.4-mm holes, providing a resistance of 6 MPa s/m^3 at a typical operating amplitude. The 3.0-liter compliance volume ''C'' of each RC load was chosen to cancel the inertial impedance of its penetration, connecting piping, and ball valve so that the impedance of each RC load was real. The 16-liter compliance volume of the LC load was chosen to make its compliant impedance much smaller than the inertial impedance of its penetration, connecting piping, and ball valve, and no deliberate resistive component was present, so that this load was as inertial as possible. The inertial parts of the LC load have a total length of 41 cm and a diameter of 2.5 cm, resulting in an inertial impedance of 5 MPa s/m^3. (However, the resistive impedance due to turbulence in these parts was not insignificant.)

Piezoresistive pressure transducers[32] at the locations shown in Fig. 3, sensed with a lock-in amplifier, were used to measure oscillating pressure amplitudes and phases. The four pressure transducers inside the pressure housing measured the pressure differences between the cascade and the pres-

sure housing, while the others measured the differences between the apparatus and atmospheric pressure. Sheathed, ungrounded, type-K thermocouples were used to measure temperature. Temperatures of metal parts in the cascade, where shown in Fig. 3, were measured with thermocouples inserted into short, close-fitting tubes that had been tack welded to the metal parts, and these were recorded with a personal computer. In hindsight, we should have ensured better thermal contact between these thermocouples and their metal parts, because heat carried to them through the fiber-insulation space, either by conduction or convection, was often noticeable. The temperatures of water streams flowing into and out of each heat exchanger were measured with thermocouples inserted into the flowing streams, not shown in Fig. 3, and these were sensed with a hand-held digital readout. Rotameters on each water stream sensed the flow rates, and together with the density and specific heat of water these yielded the measured heat-rejection powers removed from the engine at the four ambient heat exchangers. Electric powers to the three hot heat exchangers were measured with commercial electronic power meters.[33]

IV. DESCRIPTION AND HISTORY OF THE MEASUREMENTS

Measurements unfolded in the course of several episodes, with unforeseen problems occurring and evolving from episode to episode.

Initially, heating the three hot heat exchangers to 500 °C produced no oscillations at all, even though we expected the oscillations to appear below 400 °C. An analysis of this disheartening situation revealed a serious design error: The volume of the bottom compliance was half of what it should have been, distorting the wave so badly that the phase of U_1 in the middle of the apparatus would differ from what was intended by as much as 70°. The constraints of the floor-to-ceiling height of the room and the width of the system's support frame discouraged making a new, larger compliance, so the inertance of the lower resonator pipe was increased by insertion of the 2.11-m long pencil described above. An analysis showed that this retrofit would re-shape the expected wave well enough to make the engine work, though not optimally: It increased the dissipation of acoustic power by increasing both the overall surface area of the lower resonator tube and the velocity in it at a given pressure amplitude in the engine. Accordingly, this would not be a good design for efficiently delivering acoustic power to an external load. The retrofit was adequate to bring the acoustic impedances in the 2nd- and 3rd-stage regenerators near their design values. Streaming in the thermal buffer tubes is especially sensitive to impedance, and the retrofit left the 3rd-stage thermal-buffer-tube impedance significantly different than its optimum, as is discussed in Sec. VI.

Following insertion of the pencil, heating the hot heat exchangers to 600 °C still produced no oscillations. One thermometer at the axial center of the 1st-stage stack's shell indicated a temperature near 550 °C, suggesting that strong gravity-driven convection might be occurring in the stack. The dimensionless number governing such convection, the Grashof number $\rho_m^2 G\beta(T_{\text{hot}}-T_{\text{ambient}})hr_h^2/\mu^2$, where β and μ are the thermal-expansion coefficient and viscosity of the gas, G is the acceleration of gravity, T is temperature, h is the height of the stack, and r_h is the hydraulic radius of its pores, is not one of the similitude-conserved thermoacoustic numbers; G was not included in the similitude analysis of Ref. 26. The Grashof number in the argon in this stack was 6 times larger than it would be in the stack of the hypothetical 35-kW helium engine, to which this engine is otherwise similar. Hence, this gravity-driven convection was an unfortunate consequence of our decision to build an argon "scale-model" engine. Fortunately, lowering the mean pressure by only 30% reduced ρ_m and the strength of the convection enough to let the acoustic oscillations begin, near 25 Hz as expected. The strong tendency of the acoustic oscillations to maintain an energy-conserving, nearly linear temperature profile in the stack (expressed by the presence of the axial mean temperature gradient dT_m/dx in the thermoacoustic energy equation[10,19]) then stabilized the 1st stage against gravity-driven convection as the mean pressure was returned to its design value. We followed this awkward but reliable startup procedure at the beginning of each day of measurements.

TABLE II. Values of the constants A and B appearing in Eq. (4).

	A (W/K)	B (W)
1st ambient HX	0.15	0
1st hot HX	0.24	10
2nd ambient HX	0.40	10
2nd hot HX	0.33	30
3rd ambient HX	0.52	30
3rd hot HX	0.44	100
Final ambient HX	0	100

Measurements of heat leaks with no acoustic oscillations showed that the 2nd- and 3rd-stage heat leaks increased only about 30% when the argon mean pressure was raised from 0.1 to 2.3 MPa, while the 1st-stage heat leak increased tenfold over the same pressure range, confirming the presence of strong convection in the 1st stage. We did not study this carefully, because we saw no way to distinguish experimentally between convection in the stack and regenerators and convection in the pressurized fiber-packed insulation space around the outsides of the stack and regenerators. Using the density and viscosity of argon and the porosity and average fiber size of the insulation, we crudely estimated that convection in the insulation space might carry three times more heat leak than conduction in it. This supported the observation that almost all of the power applied to each hot heat exchanger during heat-leak measurements appeared at the ambient heat exchanger above it, and that the outside of the pressure vessel felt cool, except near the 1st stage. From these measurements, we arrived at a rough estimate of the heat leaks $\dot{Q}_{\text{leak}}$ that we added to the thermoacoustic calculations of thermal power at each heat exchanger:

$$\dot{Q}_{\text{leak}}=A(T_{\text{hot}}-T_{\text{ambient}})+B\frac{T_{\text{hot}}^4-T_{\text{ambient}}^4}{(900\ \text{K})^4}, \tag{4}$$

where T is absolute temperature and the values of A and B are given in Table II. At the highest operating temperatures,

these estimated heat leaks range from 100 W to 400 W, and probably have uncertainties of 100 to 200 W.

We first ran the engine with no RC loads to postpone the risky higher hot temperatures associated with higher powers as long as possible. During the measurements, we eventually noticed that the lower compliance (with its pressure transducer) became uncomfortably hot to the touch, so we installed a fan to blow room air on it.

Shortly after beginning to add loads to the engine, we noticed a time-dependent, weak resistance to ground in one pressure transducer, which is discussed in more detail in the next section. At about the same time, a small leak developed between the acoustic space and the fiber-insulation space. The first manifestation of this leak was a diode-like characteristic: a dc pressure difference of about 50 kPa between the two spaces whenever the acoustic oscillations were present. Most of the loaded data were obtained while this small leak was present. We did not notice the presence of this leak immediately, so many of the data sets were obtained at a mean pressure of 2.33 MPa, instead of 2.38 MPa used for most data before the leak appeared. Shortly after our first use of the LC load took the hot temperatures to new highs (660 °C on the hot cone in the 1st stage, 605 °C on the three hot heat exchangers), the leak increased quickly, and, fearing the intrusion of ceramic fiber insulation into the acoustic space, we ended the measurements. Earlier, repeated measurements at 10% amplitude, both with and without load, had been reproducible, suggesting that the appearance and initial evolution of this leak had not been affecting the thermoacoustic performance.

After all measurements were complete, disassembly revealed probable causes of these problems. First, we found about half a liter of water in the fiber-insulation space. There were no leaks from the ambient heat exchangers or their water pipes, so this water must have been present from the beginning. We surmise that the fiber insulation became damp when we stored it outdoors for a few months before assembling the engine. (The insulation packaging had been damaged; we would not have noticed the dampness during the assembly of the engine because we always wore rubber gloves when handling the insulation.) Engine operation at high temperatures and for long times would have gradually redistributed the water, driving it toward cooler and lower locations. The pressure transducer that developed the time-dependent resistance to ground was the lowest and presumably coolest of the four pressure transducers in the insulation space, so we ascribe its failure to this water. In the future, we will gently warm and patiently evacuate all such fiber-insulation spaces. Second, we found that the leak between the acoustic space and the fiber-insulation space was due to two fatigue cracks in welds on the uppermost and lowermost cones of the acoustic boundary inside the fiber-insulation space. In hindsight we realized that these ambient-temperature cones required slightly greater thickness to reliably survive the oscillating pressure in them.

To obtain a set of data with the engine oscillating, we opened selected load valves as desired, adjusted the powers to the three hot heat exchangers by hand until the pressure amplitude was at a desired value and the three hot temperatures were within 10 °C of each other, and then waited to ensure that a steady state was actually achieved. A minor adjustment of the total argon content of the apparatus was usually necessary to keep the mean pressure at the desired value as temperatures evolved. After at least an hour of adjustment and waiting, complex pressure amplitudes, temperatures, electric heater powers, and water flow rates were recorded, forming one steady-state data set. Twenty such data sets were obtained at mean pressures between 2.32 and 2.45 MPa, and are reported in this paper.

To calculate values with which to compare the measurements, we modeled the apparatus, from the upper compliance through the three stages to the lower compliance, with DeltaE.[34] DeltaE numerically integrates the one-dimensional acoustic continuity, momentum, and (when appropriate) energy equations. It does so in a geometry defined by the user as a sequence of segments, such as ducts, cones, stacks, regenerators, and heat exchangers. Solutions to the appropriate 1-d equations are found for each segment, with acoustic pressures, acoustic volume velocities, and mean temperatures matched at the junctions between segments. In stacks, regenerators, and thermal buffer tubes, the energy equation is solved simultaneously with the momentum and continuity equations to find the temperature profile as well as the acoustic variables. The energy flow through such segments is determined by temperatures and/or heat flows at adjacent heat exchangers. The user can define the geometry of each segment and the global variables such as mean pressure. In general, a single pass of DeltaE's integration requires values for some variables that the user does not know, and yields values for some other variables that the user does know. Hence, a shooting method is used to accommodate these irregularities. The DeltaE model of this apparatus included rough estimates for minor losses at the locations of a dozen abrupt area changes, based on tables published by Idelchik[35] for steady flow. With one exception discussed in Sec. VI, the model is very insensitive to the exact values of these minor-loss coefficients. When load was applied at the bottom of the engine, each RC or LC load was modeled with a separate DeltaE file, and the resulting net load impedance was imposed on the DeltaE model for the whole engine. (For the RC loads, the individual DeltaE load models yielded powers differing by only 1% from the simplest lumped-RC expression,[36]

$$\dot{E} = \frac{\omega V_C}{2\gamma p_m} \mathrm{Im}[p_{1,\mathrm{in}} \tilde{p}_{1,C}], \tag{5}$$

where V is volume, γ is the ratio of specific heats, the tilde denotes complex conjugation, and the subscripts C and ''in'' refer to the compliance and the inlet to the flow resistance, respectively. However, the more extended LC load required a DeltaE model for acceptable accuracy.) For each data set, we forced the whole-engine DeltaE model to have the measured mean pressure, the measured complex pressure at the transducer between the 2nd and 3rd stages, the measured common water inlet temperature as DeltaE's ''metal'' temperature at all four ambient heat exchangers, and the measured temperatures of the upper and lower resonator tubes. We also forced the model to have a common ''metal'' temperature for all three hot temperatures, but we did not choose its value. Del-

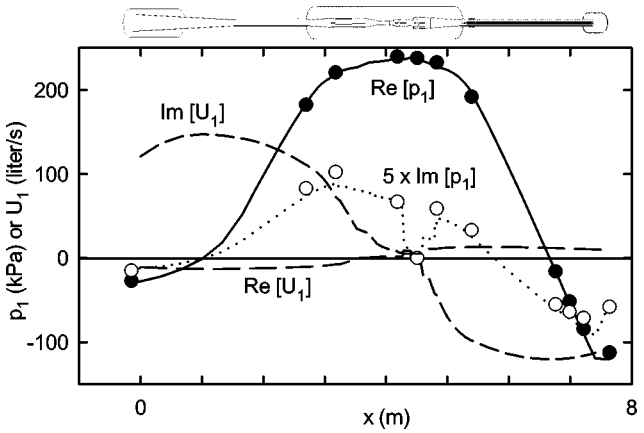


FIG. 4. Components of pressure and volume velocity for a typical operating point. Circles are measured values of the pressure, and lines are calculations. Re[ ] and Im[ ] denote real and imaginary parts. The calculated pressure is forced to agree with the measured pressure at $x=4.5$ m, where the phase is taken to be zero without loss of generality.

taE then computed as results the operating frequency, the wave profile throughout the apparatus, the common temperature of the three hot heat exchangers, and the thermal powers at the three hot heat exchangers and the four ambient heat exchangers.

For one of the data sets, we compared the calculated results for ''as-built'' dimensions and ''as-running'' dimensions, to learn whether thermal expansion of the hot parts and consequent compression of the bellows were significant. Differences were only of the order of 0.1% and 0.1°, so we performed all other calculations with as-built dimensions.

V. THE WAVE

We begin presentation of the data with the wave in the engine. Establishing agreement between the measured and calculated pressure phasors yields confidence in the calculated volume-velocity phasors, and together these two components of the wave form the basis for trying to understand the powers in the engine.

Figure 4 shows the components of the wave at 10% amplitude (i.e., $|p_1|/p_m=0.10$ at the pressure maximum between the 2nd and 3rd stages) with no RC or LC load applied to the bottom of the apparatus. Overall, the wave qualitatively resembles a half-wavelength standing wave, with its pressure maximum in the middle and volume-velocity maxima at the two ends. However, the compliances at the two ends have less than infinite volume, and the area of the intervening piping varies dramatically with x, so the actual wave is more intricately structured than the simple trigonometric functions associated with idealized geometries like those of Fig. 2. Note that the imaginary part of pressure is magnified $\times 5$ in the figure—like Re$[U_1]$, it is small enough to be unimportant for a qualitative overview of the standing wave.

The disagreements between measured and calculated complex pressures in Fig. 4 are typically 2%, and at most 6%, in amplitude, and typically 1°, and at most $3\frac{1}{2}$°, in phase. This accurate agreement between measurement and calculation is typical of all twenty data sets. The calculated frequency exceeded the measured frequency by 1.0% to 1.5% after the addition of the fan to cool the lower compliance (without the fan, the excess ranged from 0.6% to 1.6%), more evidence that the computer model describes the apparatus very well and that a more detailed discussion of the wave, presented next, is justified.

The thermoacoustic continuity equation,[27]

$$\frac{dU_1}{dx}=-i\omega c p_1-\frac{1}{r_\kappa}p_1+gU_1, \tag{6}$$

where $1/r_\kappa$ is the thermal-relaxation conductance per unit length and g is the complex gain constant, describes the changes in U_1 from place to place in the apparatus. Figure 5(a) shows the spatial distribution of the U_1 phasors throughout the three stages. Compared with this broad range of U_1 phasors, the p_1 phasors in this region are almost independent of position: 220 kPa at 5° above the 1st stage, 240 kPa at 3° between the 1st and 2nd stages, 280 kPa at 0° between the 2nd and 3rd stages, and 230 kPa at 3° below the 3rd stage. Hence, for a qualitative interpretation of Fig. 5(a) in terms of the continuity equation, the pressure phasors throughout all three stages can be taken to have essentially constant magnitude and zero phase.

The overall appearance of Fig. 5(a) shows Im$[U_1]$ varying smoothly from large and positive above the 1st stage to large and negative below the 3rd stage. This overall variation simply reflects the compressibility of the gas throughout the stages, appearing in the first term of Eq. (6). Meanwhile, Re$[U_1]$ varies from negative above the 1st stage to positive below the 3rd stage, as the acoustic power $\dot{E}=\frac{1}{2}\mathrm{Re}[p_1\tilde{U}_1]$ increases from a negative value to a positive value.

The 1st stage, with the largest volume of the three stages, has the largest compliance, and so it spans the largest range of Im$[U_1]$. In this stage, the increase in Re$[U_1]$ occurs in the stack, where the g term in Eq. (6), which contributes positively to Re$[U_1]$, dominates the r_κ term, which contributes negatively. The phase difference between U_1 and p_1 is approximately 90° in the stack, as it should be for a standing-wave engine.

The 2nd stage, with the smallest volume of the three stages, spans the smallest range in U_1. Across the regenerator, where $g\simeq(1/T_m)dT_m/dx$ is real, the third term in Eq. (6) by itself would cause a multiplication[5] of the magnitude of U_1 by $T_{\mathrm{hot}}/T_{\mathrm{ambient}}$, without a change in the phase of U_1. Since U_1 has a positive real component and p_1 is essentially constant, the acoustic power $\dot{E}$ is multiplied by the same factor, $T_{\mathrm{hot}}/T_{\mathrm{ambient}}$. The change in phase of U_1 in the regenerator is due to the first term in Eq. (6) and the compressibility of the gas in the regenerator. The 2nd-stage thermal buffer tube displays a significant change in Im$[U_1]$, because it has a significant volume.

Between the 2nd and 3rd stages, U_1 passes through the ''sweet spot'' where it has the same phase as p_1. This is the most efficient location for a regenerator; optimally, the two ends of the regenerator should straddle the sweet spot, as discussed in Sec. II. Large deviations from this condition cause the volume velocity through one or both ends the regenerator to be larger than it need be for the desired acoustic

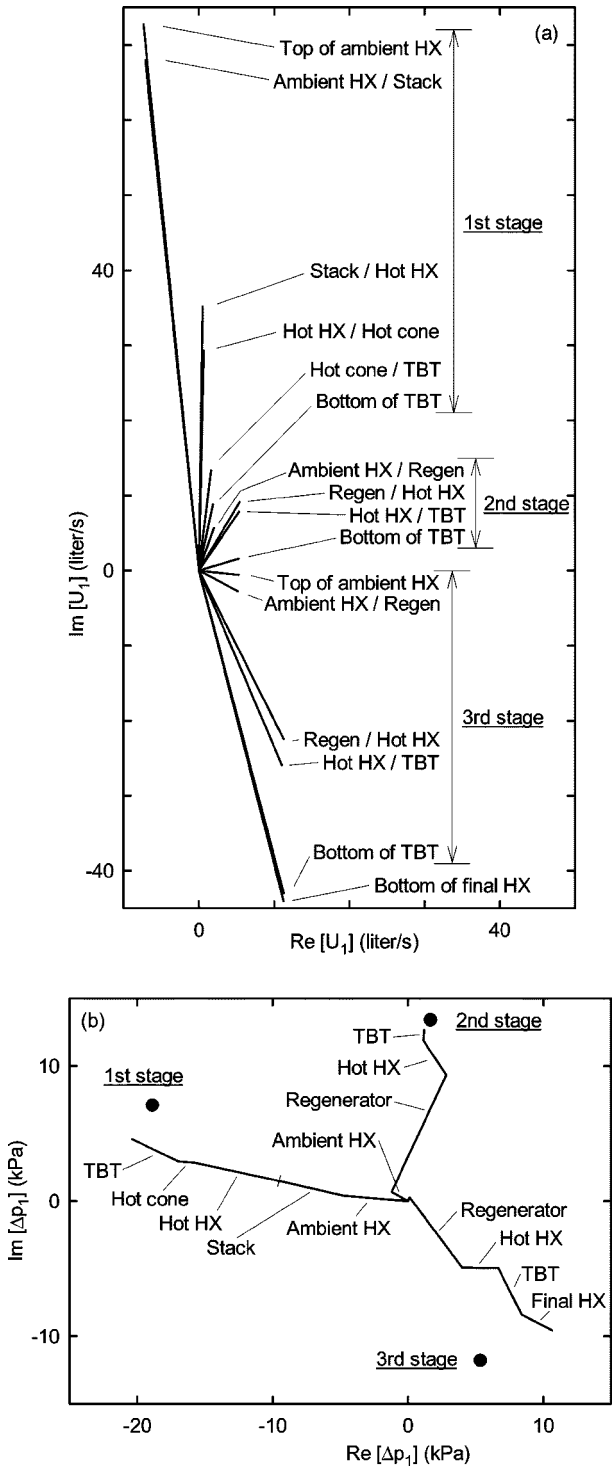


FIG. 5. Phasor diagrams for the operating point of Fig. 4. HX=heat exchanger; TBT=thermal buffer tube. As in Fig. 4, the zero of phase is chosen to be that of the pressure between the 2nd and 3rd stages. To avoid clutter, the conventional arrowheads have been omitted from the ends of the phasors. (a) Calculated volume-velocity phasors at several locations in all three stages. (b) Pressure-difference phasors for the same operating point. Circles are measured values, and lines are calculations showing how the total pressure-difference phasor across each stage is made up of smaller pressure differences across the components within that stage.

power, with attendant larger losses due to viscosity and imperfect thermal contact. The 2nd and 3rd stages in this apparatus are as close to the sweet spot as possible, given the requirement that the thermal buffer tube between them must have a volume significantly larger than the volumetric displacement within it. However, neither regenerator is actually close to traveling-wave phasing. (A calculation showed that the sweet spot was between the 1st and 2nd stages before retrofitting with the aluminum pencil discussed in previous sections.)

The U_1 phasors in the 3rd stage are similar to those of the 2nd, but with a different initial phase atop the ambient heat exchanger, and with larger changes from location to location, which reflect larger volumes in the 3rd stage. Again, across the regenerator, where $g \simeq (1/T_m)dT_m/dx$ is real, the third term in Eq. (6) by itself would cause a multiplication of the magnitude of U_1 by $T_{\text{hot}}/T_{\text{ambient}}$, without a change in the phase of U_1. Since U_1 has a positive real component, the acoustic power $\dot{E}$ is multiplied by the same factor. The accompanying change in phase of U_1 is due to the first term in Eq. (6) and the compressibility of the gas in the regenerator.

The thermoacoustic momentum equation,[37]

$$\frac{dp_1}{dx} = -i\omega l U_1 - r_\nu U_1, \tag{7}$$

where l is the inertance per unit length and r_ν is the viscous resistance per unit length, describes the changes in p_1 from place to place in the apparatus. The complex pressure throughout the three stages is so nearly spatially uniform that a pressure-phasor plot displays little detail. Hence, in Fig. 5(b) we display pressure *differences* across the stages and components. In the figure, the sign convention is chosen so that a resistive impedance creates a pressure-difference phasor in Fig. 5(b) that is parallel (not anti-parallel) with the corresponding volume-velocity phasor in Fig. 5(a). [This is opposite the sign convention of ordinary calculus, used in Eq. (7).] The three experimental points in Fig. 5(b) are differences between the complex pressures measured at the ends of each of the three stages, i.e., differences between adjacent pressure measurements at $x=3.2$, 4.2, 4.5, 4.8 m in Fig. 4. Each of the three corresponding calculated lines extending from the origin is a sequence of straight line segments, each of which shows the calculated pressure difference across a component within the stage. (Here, the thermal buffer tubes include the flow straighteners at their ends.) Hence, the three experimental points should lie close to the ends of the three lines.

This method of displaying the data accentuates experimental uncertainties, because the experimental points are small differences between large numbers. The uncertainty in the measured pressure phasors is about $1\frac{1}{2}\%$ in amplitude, i.e., 3 kPa, due to a combination of transducer calibration uncertainty, transducer-calibration temperature sensitivity, and lock-in gain uncertainty, so we expect the uncertainties in the differences $\text{Re}[\Delta p_1]$ shown in Fig. 5(b) to be a few kPa. The uncertainties in $\text{Im}[\Delta p_1]$ are smaller, about 1 kPa.

In the 1st stage, the net calculated pressure-difference phasor in Fig. 5(b) and all of its individual-component phasors lead the U_1 phasors of Fig. 5(a) by approximately 90°, indicating the mostly inertial nature of the impedance of this stage—each component within the 1st stage has much more

inertial impedance than resistive impedance. Additional resistance in the calculation of any one of the components within the 1st stage would rotate the corresponding straight line segment, and hence also the net Δp_1 phasor for the entire stage, clockwise toward the experimental point. For example, an additional resistance corresponding to 50 W of acoustic-power dissipation in the stack—only 10% of the power generated in the stack—would bring the calculated phasor into agreement with experiment. The validity of the ''tubular'' heat-exchanger algorithm used in DeltaE is doubtful at this level of accuracy, because it assumes laminar oscillatory flow while the Reynolds-number amplitude in the hot heat exchanger is 25,000 and in the ambient heat exchanger is 100,000. Furthermore, the applicability of the ''circular pore'' algorithm used in the DeltaE model for this hexagonal-honeycomb stack has never been checked to this level of accuracy. Hence, we believe that the small disagreement between the calculation and experiment in the 1st stage is probably due to the inaccuracy of one or both of these calculation algorithms.

The calculated 2nd-stage regenerator phasor in Fig. 5(b) is the largest segment of the net 2nd-stage Δp_1 phasor, and it points in a direction between those of the two U_1 phasors at the ends of the 2nd-stage regenerator in Fig. 5(a), indicating the resistive nature of the regenerator. The Δp_1 phasor segments for the 2nd-stage ambient and hot heat exchangers point somewhat counterclockwise from this direction, indicating the inertial contributions in these components. In the 2nd stage, the disagreement between the net calculated Δp_1 phasor and the experimental Δp_1 phasor is within the experimental uncertainty.

As in the 2nd stage, the regenerator in the 3rd stage makes the largest contribution to the 3rd-stage net Δp_1 phasor in Fig. 5(b). The impedances of the regenerator and of the thermal buffer tube are resistive, the latter because of its flow straighteners, its location where $|U_1|$ is high, and the low inertance of its large aspect ratio. The impedances of the 3rd-stage hot heat exchanger and final ambient heat exchanger have substantial inertial components, and the tiny pressure drop in the 3rd-stage ambient heat exchanger is barely visible in the figure.

The difference between the experimental and calculated 3rd-stage Δp_1 in Fig. 5(b) is difficult to explain. It is slightly larger than the expected experimental uncertainty of a few kPa, described above. To rotate the calculated net phasor to match the experimental phasor would require either elimination of all of the inertial impedance in the hot heat exchanger and final ambient heat exchanger, or a large clockwise rotation of most or all of the 3rd-stage U_1 phasors in Fig. 5(a); neither of these possibilities seems plausible. One possible explanation is that the pressure transducer at $x=4.8$ m, just below the final ambient heat exchanger, was already developing an electrical problem when this data set was taken. It is the one mentioned above that developed a time-dependent, weak resistance to ground, which we noticed later when experimentation with loads commenced. Thereafter it deteriorated erratically with time and experienced a change in ac sensitivity of several percent even when it was powered by a floating power supply and detected with a differential amplifier. Hence, this sensor cannot be trusted for the data with the engine loaded, and is also suspect for the earlier, unloaded data, although there is no explicit evidence of time dependence in its sensitivity for the earlier data. A 2.5% change in the calibration constant of this transducer would move the 3rd-stage experimental point in Fig. 5(b) into perfect agreement with the calculation, without changing the other two experimental points in Fig. 5(b).

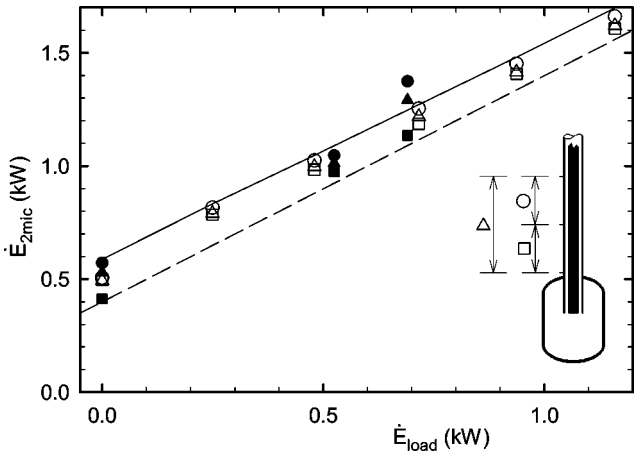


FIG. 6. Acoustic power $\dot{E}_{2\mathrm{mic}}$ flowing down the lower resonator pipe as a function of acoustic power $\dot{E}_{\mathrm{load}}$ dissipated in the loads, with pressure amplitudes of 242–252 kPa between the 1st and 2nd stages. The open symbols represent measurements with the RC loads and with the fan cooling the lower compliance. The filled symbols are measurements under other circumstances, as described in the text. The inset shows which symbols are associated with which pair of pressure transducers for $\dot{E}_{2\mathrm{mic}}$. The dashed line, a guide to the eye, has a slope of unity and an arbitrary vertical offset. The solid line links DeltaE results corresponding to each of the open triangles, with a minor-loss coefficient of 0.7 as described in the text.

Despite this uncertainty about one pressure transducer, the overall evidence for our understanding of the pressure and volume-velocity phasors in this apparatus is very good.

VI. POWERS

The series of three pressure transducers in the lower resonator pipe give three ''pairs'' for three measurements[36] of acoustic power $\dot{E}_{2\mathrm{mic}}$ flowing past this location. Figure 6 shows a comparison of these measurements with the power dissipated below this location. The circles, squares, and triangles indicate which pair of the three transducers is used to obtain $\dot{E}_{2\mathrm{mic}}$, as illustrated in the inset. The horizontal axis is the power $\dot{E}_{\mathrm{load}}$ dissipated in the RC or LC loads, obtained as described in Sec. IV from measured complex pressures in the bottom compliance and the load compliances, and the geometry of the loads (most importantly the volumes of the load compliances). The open symbols represent measurements with 0, 1, 2, 3, 4, and 5 RC loads, and with the fan cooling the lower compliance. The filled symbols near $\dot{E}_{\mathrm{load}}=0.7$ kW represent measurements with the LC load and with the fan cooling the lower compliance. The filled symbols near 0.0 and 0.5 kW represent measurements with 0 and 2 RC loads, respectively, without the fan cooling.

The overall slope near unity displayed by the measurements inspires confidence in both $\dot{E}_{\mathrm{load}}$ and $\dot{E}_{2\mathrm{mic}}$. The slight

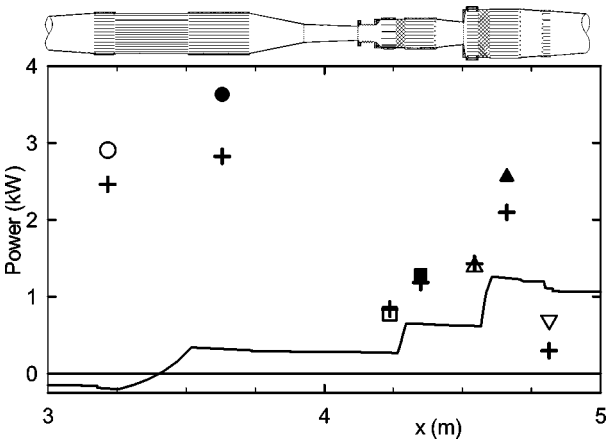


FIG. 7. Powers as a function of position through the three stages, at the same operating point as for Figs. 4 and 5. The solid curve is the calculated acoustic power $\dot{E}$. The symbols are thermal powers at the heat exchangers; on the plot, each symbol is placed horizontally at the center of its heat exchanger. The three filled symbols are experimental hot powers, and the four open symbols are experimental ambient powers. Plus signs are the corresponding calculated thermal powers.

deviation from unity slope shared by the open-symbol measurements and the solid calculated line is due to slightly reduced $|p_1|$ and $|U_1|$ throughout the lower resonator as more power is drawn through the 3rd stage. Of the vertical offset of 0.5 kW in the open-symbol measurements and almost 0.6 kW in the calculated line, 25% is due to dissipation between the sensors and the bottom of the tube and 75% is due to the minor loss at the abrupt transition between the bottom of the tube and the bottom compliance. (We chose to retain the *a priori* estimated value of the minor-loss coefficient, $K=0.7$, in all calculations reported in this paper, instead of changing to $K=0.65$, which would be more compatible with the data in Fig. 6, because we did not want to introduce *any* fitting parameters into the calculations.) The self consistency in Fig. 6 adds to our confidence in the DeltaE modeling of the apparatus.

The three sets of filled symbols in Fig. 6 have a greater vertical spread than do the open symbols. A greater vertical spread seems to indicate a greater dissipation of acoustic power in the resonator pipe between the transducers, yet there is no reason to expect such additional dissipation. Hence, we take this spread as strong evidence that the theory of the measurement[36] of $\dot{E}_{2\mathrm{mic}}$ is not perfectly applicable here. That theory is based on a laminar analysis, but the flow here is turbulent, with a Reynolds-number amplitude of 2×10^6. The increased vertical spread indicates that, at least for turbulent flow, effects not included in the present theory of the measurement should be added. Candidates include an axial temperature gradient in the measurement zone, which was larger when the fan was absent, and an unforeseen sensitivity to the phase of the acoustic impedance, which changes dramatically when substituting one LC load for three RC loads.

Figure 7 displays powers for the same typical operating point as has been described with Figs. 4 and 5. The features shown in Fig. 7 are typical of all data sets. The measured hot powers, indicated by three filled symbols, are simply the electrical powers delivered to the heaters. The measured ambient heat-removal powers, indicated by four open symbols, are the product of the measured water flow rates, the measured temperature differences between inflowing and outflowing water streams, and the density and specific heat of water. The calculated $\dot{E}$ represented by the solid curve and the calculated thermal powers represented by $+$'s are results of the same DeltaE calculation that produced the calculated results in Figs. 4 and 5. The acoustic power is obtained straightforwardly from the calculated phasors, as $\dot{E}=\frac{1}{2}\mathrm{Re}[p_1\tilde{U}_1]$. The calculated thermal powers include the effects of thermoacoustic heat transport through the regenerators and stack, boundary-layer heat transport at the surfaces of the thermal buffer tubes, and ordinary axial conduction in the argon gas and stainless steel through these components. In addition to these standard thermoacoustic heat effects, the DeltaE calculation includes the conduction of heat through the metal shells around the stack, regenerators, and thermal buffer tubes. Estimates of black-body radiation heat transport in the thermal buffer tubes and of heat transport via conduction and convection in the pressurized fiber-packed insulation space around the stages, based on measurements as described in Sec. IV, were added to the DeltaE results. No estimates of Rayleigh streaming and jet-driven streaming in the thermal buffer tubes or internal streaming in the stack and regenerators were included in the calculations.

In Fig. 7, $\dot{E}$ is negative above the 1st stage, showing the delivery of acoustic power from the 1st stage to the unavoidable dissipation in the upper part of the resonator. Here, about a third of the acoustic power produced in the 1st stage is spent in the upper resonator, but this fraction is smaller for other data sets having RC loads in use at the bottom of the resonator. Acoustic power from the 1st stage is more than doubled in the 2nd stage, and then redoubled in the 3rd stage. Dissipation in the hot heat exchanger and flow straighteners below the 3rd stage is larger than for the 2nd stage, because in the 3rd stage these components are farther from the sweet spot and so they experience higher velocities.

The thermal powers in Fig. 7 display interesting features, in common with all other data sets. The 3rd-stage ambient power and both 2nd-stage powers are in excellent agreement with calculations. Similar measurements and calculations in another traveling-wave engine[17] and in the late-20th-century experience with Stirling engines[38] are also in good agreement, so this is to be expected. However, the 3rd-stage hot power exceeds the corresponding calculation significantly, by about the same amount that the final ambient heat exchanger's power exceeds its calculation. This strongly suggests a common cause for both disagreements: streaming in the 3rd-stage thermal buffer tube, carrying heat from the 3rd-stage hot heat exchanger to the final ambient heat exchanger. This thermal buffer tube has an aspect ratio (diameter/length) of 2.2, while that of the 2nd stage is only 0.9 and those of the 1st stage and nearly all previous thermal buffer tubes and pulse tubes[39] are much smaller. The 1st-stage thermal powers are also significantly higher than the corresponding calculations. The disagreement is comparable to that seen in other standing-wave engines near 10%

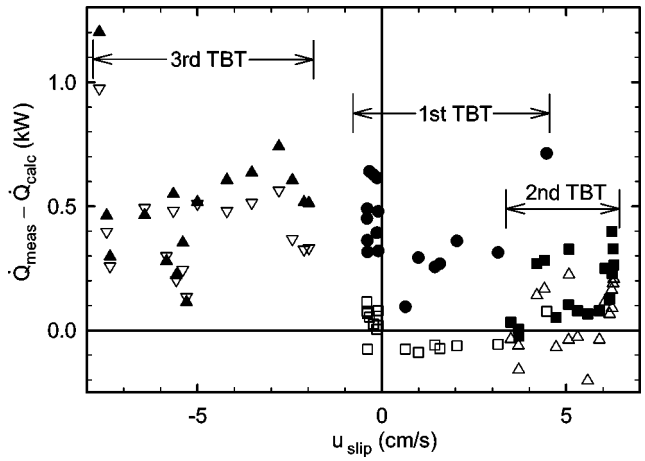


FIG. 8. The difference between the measured and calculated thermal powers associated with each thermal buffer tube, as a function of the calculated effective Rayleigh-streaming slip velocity near the wall of the thermal buffer tube midway along its length. Filled symbols represent power flowing from hot heat exchangers into thermal buffer tubes, and open symbols represent power flowing from thermal buffer tubes into ambient heat exchangers.

amplitude[13,26] and, hence, is not surprising. The observation that the extra 1st-stage hot power exceeds the extra 1st-stage ambient power by a few hundred Watts is initially disturbing—no extra power appears at the neighboring 2nd-stage ambient heat exchanger, so it seems that energy conservation is not obeyed. We suspect that extra heat leak from the 1st-stage hot heat exchanger to the room accounts for these few hundred Watts. An estimation of the correct heat leak to use for the 1st stage based on heat-leak measurements was difficult, because the enormous convective heat transport through the stack at high mean pressure masked heat leak carried from the hot heat exchanger to the pressure-vessel shell by the slow convection in the insulation space. Additionally, during heat leak measurements the hot cone below the hot heat exchanger shared the hot heat exchanger's temperature only at its upper edge, the rest developing a temperature distribution determined by conduction, while during operation the hot cone was kept uniformly hot by vigorous gas motion and, in fact, was hotter than the hot heat exchanger by typically 50 °C, the adiabatic oscillatory temperature amplitude, an effect that has been reported before.[13] Hence, with more hot metal surface area against the insulation space, the 1st-stage hot heat leak to the room should indeed be higher during engine operation than during the heat leak measurements whose results are included in the calculation shown in Fig. 7.

To try to better understand the large heat transport down the 3rd-stage thermal buffer tube, in Fig. 8 we display the excess thermal power, i.e., the difference between measured and calculated thermal powers, for the six heat exchangers abutting the three thermal buffer tubes. For the 1st-stage hot heat exchanger, we have subtracted the excess power in the 1st-stage ambient heat exchanger, because this excess power flows through the stack, not through the thermal buffer tube, as is explained in the previous paragraph. In the 2nd- and 3rd-stage thermal buffer tubes, there is a good correlation between the hot and ambient powers, while in the 1st-stage thermal buffer tube the hot power usually exceeds the ambient power by a few hundred Watts, which we ascribe to increased heat leak, as discussed in the previous paragraph. Overall, except for a few outlying points, around 500 W flows down the 3rd-stage thermal buffer tube, while the 1st- and 2nd-stage thermal buffer tubes carry between zero and a few hundred Watts (and the 1st-stage hot heat exchanger shows a few hundred additional Watts of heat leak).

The horizontal axis in Fig. 8 is the effective slip velocity $u_{\mathrm{slip}}=\dot{m}_{2,w}/\rho_m$ just outside the boundary-layer against the wall of the thermal buffer tube, at the axial center of the thermal buffer tube, calculated using the modified Rayleigh-streaming expression[31] for the streaming mass flux density $\dot{m}_{2,w}$ near the wall. Originally, the tapers of the three thermal buffer tubes were designed to make $u_{\mathrm{slip}}=0$ at 10% amplitude and full load, with the recognition that u_{slip} would be nonzero for other amplitudes or loads. Then we might have hoped to see data for each thermal buffer tube fall on a big "V" with its tip at the origin in Fig. 8, because streaming-driven heat transport down a thermal buffer tube might be roughly proportional to $\rho_m c_p|u_{\mathrm{slip}}|(T_{\mathrm{hot}}-T_{\mathrm{ambient}})$, where c_p is isobaric heat capacity per unit mass, over some range of u_{slip}. A little experimental evidence for such a minimum has been previously reported.[31,40] The 1st stage, for which data fall on both sides of $u_{\mathrm{slip}}=0$, shows only a slight suggestion of such a "V" shape, asymmetrical at best. Unfortunately, the use of the lower-resonator pencil to partly "fix" the design error in the bottom-compliance volume made the acoustic conditions in the other two stages so far from the original design conditions that they do not operate near $u_{\mathrm{slip}}=0$ for any of the data sets. It is unclear how far from $u_{\mathrm{slip}}=0$ the analysis of Ref. 31 should be applicable; perhaps the 2nd- and 3rd-stage thermal buffer tubes' data fall outside this range.

To summarize the thermal aspects of the present experiment, Fig. 9 shows the seven thermal powers and the three hot temperatures as functions of acoustic power flow past the $\dot{E}_{2\mathrm{mic}}$ measurement location near the bottom of the resonator, at full amplitude.

The thermal powers in Fig. 9(a) display the features already discussed above: The 1st-stage measured powers exceed the calculations, as for previous standing-wave engines, with some of the disagreement in the hot power probably due to heat leak to the room. The measured 3rd-stage hot power and the final ambient power exceed their calculated values by equal amounts, presumably due to heat transport down the 3rd-stage thermal buffer tube. The other measured powers are in better agreement with their calculated values.

The measured temperatures in Fig. 9(b) show the same increase with $\dot{E}$ as the calculated T_{metal} shows, but fall about 60 °C below the calculation, in the vicinity of the calculated values of gas temperature. An error in the temperature calculation of this magnitude is not surprising, given the laminar nature of DeltaE's heat-exchanger algorithms and the highly turbulent flow in the apparatus. For these data sets, the DeltaE algorithms yield typically 50 °C gas-to-metal temperature differences in both ambient and hot heat exchangers, so if these calculated differences were instead about 30 °C then the calculations would agree with the measurements. In addition, the drilled stainless-steel hot heat exchanger bodies

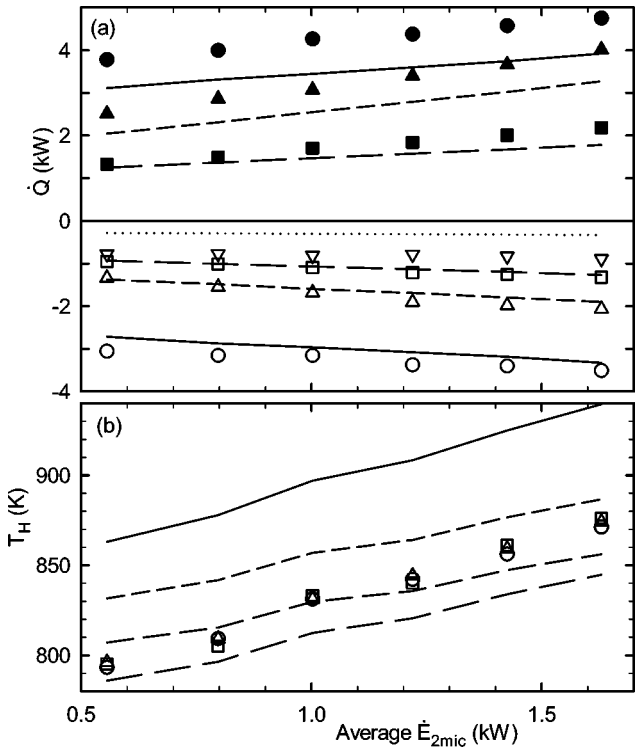


FIG. 9. (a) Thermal powers as a function of load, at full oscillation amplitude. Filled symbols are measured hot powers, and open symbols are measured ambient powers. The lines represent associated calculated values. Circles and solid curves, 1st stage; squares and long-dashed curves, 2nd stage; erect triangles and short-dashed curves, 3rd stage; inverted triangles and dotted curve, final ambient heat exchanger. The sign convention for the ambient powers here is opposite to that used in Figs. 7 and 8, to avoid clutter. (b) Hot temperatures as a function of load, at full oscillation amplitude. Symbols are measured metal temperatures. The solid curve is the calculated metal temperature, forced to be the same for all three stages. The dashed curves are calculated gas temperatures. Circles and long dashes are 1st stage, squares and middle dashes are 2nd stage, and triangles and short dashes are 3rd stage.

are not spatially isothermal. In each heat exchanger, the heaters of various lengths have various powers per unit length, and variations in the number of argon channels close to a unit length of each heater also exist. Because stainless steel has a poor thermal conductivity, spatial nonuniformity of as little as 10% in hot power density can cause spatial nonuniformity of 50 °C in the temperature in the larger (1st- and 3rd-stage) hot heat exchangers.

VII. EFFICIENCY

To discuss the efficiency of this engine, we rely mostly on measurements but also slightly on calculations. We define the efficiency to be the ratio of the acoustic power delivered below the final ambient heat exchanger to the sum of the measured electric powers supplied to the three hot heat exchangers. We choose that location to account for the acoustic power because it is where an alternator could be in a heat-driven electricity-generating system and it is equivalent to where acoustic power was benchmarked in a previous, exemplary thermoacoustic-Stirling engine.[17] Anticipating this choice, we designed the present apparatus for minimum height—not maximum efficiency—below this location. The two-microphone measurements of acoustic power were made 2 m below this location, but we are confident in the ability of the computer model of the apparatus to account for the difference in $\dot{E}$ between these two locations. To define a Carnot efficiency $1-T_{\text{ambient}}/T_{\text{hot}}$ with which to compare the engine's efficiency, we use the incoming water-stream temperature as T_{ambient} and the average of the three measured hot temperatures as T_{hot}.

The efficiencies of the operating points shown in Fig. 9 were, from left to right, 0.15, 0.16, 0.17, 0.18, 0.19, and 0.20. At this fixed amplitude of 10%, efficiency rises as delivered power increases because the deliberate load applied at the bottom of the engine becomes a larger fraction of the total acoustic power that the three stages create, and also the thermodynamically useful heat consumed by the three stages dominates their heat leaks more strongly. Another interesting operating point, not shown in Fig. 9, had all five RC loads open and an amplitude of only 8.5%, yielding an efficiency of 0.18. At this lower-power operating point, the heat leaks were a larger fraction of the delivered hot powers, but this was mostly compensated by lower dissipation of acoustic power in the upper resonator and the regenerators, where the dissipation is more nearly proportional to the cube of the amplitude than to its square.

A more detailed analysis of the sources of dissipation in the engine is conveniently made in terms of exergy, the thermodynamic energy (or, here, power) that accounts for the ability to do useful work.[41] The exergy power associated with acoustic power in open space equals the acoustic power itself, but the exergy power of acoustic power in a porous medium is more complicated.[42] The exergy power associated with a heat source $\dot{Q}$ at temperature T is

$$\dot{Q}(1-T_{\text{ambient}}/T), \tag{8}$$

where T_{ambient} is the temperature at which heat is freely available in the environment. An analysis of a thermodynamic system in terms of either exergy or irreversible entropy generation allows a quantitative discussion of the irreversibilities in the system by subdividing it into logical sets of processes, locations, etc.[41,43]

At its most efficient operating point, the total exergy input power to the cascade engine via the three hot-heat-exchanger powers, given by Eq. (8), was 7322 W. The acoustic power (and, hence, exergy power) delivered below the final ambient heat exchanger was 2179 W. Hence, 5143 W of exergy power was lost in the conversion of thermal to acoustic power by the engine. Four per cent of that was lost in the upper resonator, 49% in the 1st stage, 18% in the 2nd stage, and 29% in the 3rd stage. That the 3rd stage is responsible for more irreversibility than the 2nd is to be expected, since it is also responsible for more acoustic-power production. That almost half of the irreversibility occurs in the 1st stage, even though it creates less than a third of the acoustic power, is due to the intrinsically irreversible nature of standing-wave thermoacoustics.

The ''Experiment'' column of Table III shows a further subdivision of lost exergy, by location and process, for this most efficient operating point. The largest irreversibility is the sum of the well-understood viscous, thermal-relaxation, and axial conduction phenomena in the stack, effects that are

TABLE III. Rate of exergy loss in the engine at 10% amplitude and full load, subdivided according to location and process, in order of importance. The first column of numbers gives the exergy accounting for the most efficient experimental operating point, the second column shows the improvement possible with circumstances that we are very confident could be achieved with a scaled-up helium engine, and the third column shows the further improvement possible with reasonable probability. HX = heat exchanger; TBT = thermal buffer tube.

	Experiment	Confident future	Probable future
—Lost exergy power:—			
1st stage, stack, known processes	1177 W	14.1 kW	14.1 kW
1st stage, unknown heat	554 W	7.1 kW	7.1 kW
1st stage, dT, ambient HX	532 W	8.5 kW	4.3 kW
3rd stage, unknown heat	496 W	7.9 kW	4.0 kW
3rd stage regen., known processes	309 W	2.3 kW	2.3 kW
2nd stage regen., known processes	294 W	3.3 kW	3.3 kW
2nd stage, unknown heat	266 W	4.3 kW	2.2 kW
3rd stage dT, ambient HX	261 W	4.2 kW	2.1 kW
3rd stage, expt heat leak	230 W	1.8 kW	1.8 kW
3rd stage, TBT, known processes	213 W	2.3 kW	2.3 kW
Upper resonator	192 W	3.1 kW	3.1 kW
2nd stage, expt heat leak	145 W	1.2 kW	1.2 kW
2nd stage, dT, ambient HX	144 W	2.3 kW	1.2 kW
1st stage, expt heat leak, fiber insul.	98 W	7.8 kW	7.8 kW
Sum of smaller effects	234 W	3.2 kW	3.2 kW
Total lost exergy power	5143 W	66.4 kW	52.8 kW
Delivered exergy power	2179 W	34.9 kW	34.9 kW
Percent of Carnot	30%	34%	40%

inherent in standing-wave engines and are included in the DeltaE calculations describing the apparatus. This is a price one chooses to pay in using the cascade arrangement for a thermoacoustic engine. Next largest is the ''extra'' hot power consumed by the 1st stage that can be accounted for neither in the calculations nor by the heat leak measurements. Standing-wave engines near 10% amplitude have always suffered from such extra heat consumption.[13,26] Next in importance in Table III is irreversible heat transfer at the 1st-stage ambient heat exchanger, where a calculation shows that the ambient waste heat of the 1st stage must flow across a difference of 50 °C between the average temperature of the oscillating gas and the metal temperature.

The fourth entry in the table is the thermal power that we suspect is carried by streaming in the 3rd-stage thermal buffer tube, as discussed in the previous section. The fifth and sixth entries are the well-understood lossy processes in the two regenerators, caused by viscosity and imperfect thermal contact. These are large in this cascade arrangement, relative to traditional Stirling engines, because both regenerators are forced significantly away from the sweet spot of small velocity amplitude where the phase difference between pressure and velocity is zero.

The exact values of some of the numbers in this table should not be taken too seriously. The excellent agreement between calculated and measured complex pressures gives us confidence in numbers associated with the wave and the acoustic power, but some heat and temperature effects are less certain. The worst example is the gas-to-metal temperature difference in the 1st-stage ambient heat exchanger, which is calculated with a laminar algorithm in DeltaE and using the calculated heat flow, even though the oscillations in the heat exchanger are turbulent and the experimental heat flow is larger. The corresponding entry in the table might be in error by as much as 50%. Differences between measured and calculated hot temperatures shown in Fig. 9(a) further highlight the uncertainties in our knowledge of heats and temperatures. Nevertheless, the total lost exergy in the table is a number based on measured thermal and acoustic powers, and on the confidently calculated difference between acoustic powers at the 2-microphone location and the final ambient heat exchanger; most of the individual entries in the table are based on measurements or on dependable calculations.

Two decisions made during the design of this engine had a major impact on its efficiency. First, we decided to use a honeycomb stack instead of a parallel-plate stack, because a stainless-steel honeycombs are commercially available. Calculations predict that the honeycomb performs about 20% less efficiently than parallel plates. In one experimental engine in which we used both types of stack (unpublished), the hot power consumed in excess of the calculated amount was also reduced, by at least 20%, by parallel plates instead of honeycomb. Second, we decided to use argon instead of helium in this engine, for reasons described in the Introduction. Thermoacoustically, the present engine is strictly ''similar'' to[26] an engine in which every spatial dimension would be double that of the present engine and which would be filled with 3.1-MPa helium instead of 2.4-MPa argon. However, perfect similitude would also require that the present engine be built of metals with a thermal conductivity 8.5× lower than that of the metals (presumably stainless steel) of the hypothetical helium engine, and that the acceleration of gravity be 5× lower and the insulation fiber sizes in the pressure housing be 2× smaller for the present engine, in order for metallic heat conduction and convection-borne heat leaks to scale correctly. Accounting for these easily calculable thermal effects while otherwise scaling up powers by a factor of 16, and simultaneously incorporating the difference between honeycomb stacks and parallel-plate stacks, yields the numbers in the ''Confident'' column of Table III, showing that a cascade engine of double the present size, filled with helium and of stainless-steel construction, (or, equivalently, one of the present size and argon, but with an imaginary low-conductivity construction material and in reduced gravity) can certainly achieve 34% of Carnot's efficiency.

Other straightforward improvements should raise the efficiency further. Smaller passages in the heat exchangers would reduce the exergy wasted in the temperature defects in them. Reducing these passage sizes by a factor of two might reduce these losses by a similar factor. Avoiding the mistake in the design of the resonator, which put the 2nd- and especially 3rd-stage thermal buffer tubes far from zero Rayleigh streaming in the present engine, would presumably reduce the heat transported by streaming in these two thermal buffer tubes significantly, perhaps by about a factor of 2. Accounting for these two improvements yields the ''Probable'' column in Table III, showing that the efficiency of a cascade engine can probably reach 40% of Carnot's efficiency, the same as that of the most efficient thermoacoustic-Stirling engine built to date.[17]

Whether the cascade configuration is superior for any actual applications will depend on practical issues, such as the cost of construction, that are beyond the scope of the present paper.

ACKNOWLEDGMENTS

We are grateful to Chris Espinoza, Mike Torrez, Brad Sims, and Bob Miller for attention to detail paid during the construction of this apparatus, to Scott Backhaus for helpful discussions during the analysis of the data, and to John Wollan for a careful reading of the manuscript. This work was supported by OBES/DMS in the U.S. Department of Energy's Office of Science, and by Praxair, Inc.

[1] Lord Rayleigh, ''The explanation of certain acoustical phenomena,'' Nature (London) **18**, 319–321 (1878).
[2] W. A. Marrison, ''Heat-controlled acoustic wave system,'' U.S. Patent No. 2,836,033, 1958.
[3] R. L. Carter, M. White, and A. M. Steele (private communication of Atomics International Division of North American Aviation Inc., 1962). See also K. T. Feldman, ''A study of heat generated pressure oscillations in a closed end pipe,'' Ph.D. dissertation, Mechanical Engineering Department, University of Missouri, 1966.
[4] K. T. Feldman, ''Review of the literature on Sondhauss thermoacoustic phenomena,'' J. Sound Vib. **7**, 71–82 (1968), and hard-to-find references therein.
[5] P. H. Ceperley, ''A pistonless Stirling engine—The traveling wave heat engine,'' J. Acoust. Soc. Am. **66**, 1508–1513 (1979).
[6] P. H. Ceperley, ''Gain and efficiency of a short traveling wave heat engine,'' J. Acoust. Soc. Am. **77**, 1239–1244 (1985).
[7] R. S. Wakeland and R. M. Keolian, ''Thermoacoustics with idealized heat exchangers and no stack,'' J. Acoust. Soc. Am. **111**, 2654–2664 (2002).
[8] G. Walker, *Stirling Engines* (Clarendon, Oxford, 1960).
[9] N. Rott, ''Damped and thermally driven acoustic oscillations in wide and narrow tubes,'' Z. Angew. Math. Phys. **20**, 230–243 (1969).
[10] N. Rott, ''Thermally driven acoustic oscillations, part III: Second-order heat flux,'' Z. Angew. Math. Phys. **26**, 43–49 (1975).
[11] N. Rott, ''Thermoacoustics,'' Adv. Appl. Mech. **20**, 135–175 (1980).
[12] A. Migliori and G. W. Swift, ''Liquid sodium thermoacoustic engine,'' Appl. Phys. Lett. **53**, 355–357 (1988).
[13] G. W. Swift, ''Analysis and performance of a large thermoacoustic engine,'' J. Acoust. Soc. Am. **92**, 1551–1563 (1992).
[14] T. J. Hofler and M. S. Reed, ''Measurements with wire mesh stacks in thermoacoustic prime movers,'' J. Acoust. Soc. Am. **99**, 2559–2574 (1996).
[15] J. A. Lightfoot, W. P. Arnott, H. E. Bass, and R. Raspet, ''Experimental study of a radial mode thermoacoustic prime mover,'' J. Acoust. Soc. Am. **105**, 2652–2662 (1999).
[16] T. Yazaki, A. Iwata, T. Maekawa, and A. Tominaga, ''Traveling wave thermoacoustic engine in a looped tube,'' Phys. Rev. Lett. **81**, 3128–3131 (1998).
[17] S. Backhaus and G. W. Swift, ''A thermoacoustic-Stirling heat engine: Detailed study,'' J. Acoust. Soc. Am. **107**, 3148–3166 (2000).
[18] C. M. de Blok, ''Thermoacoustic system, 1998,'' Dutch Patent: International Application Number PCT/NL98/00515, U.S. Patent 6,314,740, November 13, 2001.
[19] G. W. Swift, *Thermoacoustics: A Unifying Perspective For Some Engines and Refrigerators* (Acoustical Society of America Publications, Sewickley, PA, 2002).
[20] T. J. Hofler, ''High-efficiency heat-driven acoustic cooling engine with no moving parts, '' U.S. Patent No. 5,901,556, 1999.
[21] A. J. Lesperance, ''Hardware modifications and instrumentation of the thermoacoustically driven thermoacoustic refrigerator,'' Master's thesis, U.S. Naval Postgraduate School, Monterey, CA, 1997.
[22] S. Backhaus and G. W. Swift, ''Fabrication and use of parallel-plate regenerators in thermoacoustic engines,'' in *Proceedings of the 36th Intersociety Energy Conversion Engineering Conference* (American Society of Mechanical Engineers, New York, NY, 2001), pp. 453–458.
[23] G. Petculescu and L. A. Wilen, ''Traveling-wave amplification in a variable standing wave ratio device,'' Acoustic Research Letters Online **3**, 71–76 (2002).
[24] D. Gedeon, ''DC gas flows in Stirling and pulse-tube cryocoolers,'' in *Cryocoolers 9*, edited by R. G. Ross (Plenum, New York, 1997), pp. 385–392.
[25] G. W. Swift and J. J. Wollan, ''Thermoacoustics for liquefaction of natural gas,'' GasTIPS **8**(4), 21–26 (2002). Also available at www.lanl.gov/thermoacoustics/Pubs/GasTIPS.pdf.
[26] J. R. Olson and G. W. Swift, ''Similitude in thermoacoustics,'' J. Acoust. Soc. Am. **95**, 1405–1412 (1994).
[27] Ref. 19, Eq. (4.76).
[28] Boiler and Pressure Vessel Committee of the American Society of Mechanical Engineers, *ASME Boiler and Pressure Vessel Code* (American Society of Mechanical Engineers, New York, 2001).
[29] Fiberfrax®, Unifrax, Niagara Falls, New York.
[30] Kentucky Metals, New Albany, Indiana.
[31] J. R. Olson and G. W. Swift, ''Acoustic streaming in pulse tube refrigerators: Tapered pulse tubes,'' Cryogenics **37**, 769–776 (1997).
[32] Endevco, San Juan Capistrano, CA.
[33] Ohio Semitronics, Hilliard, OH.
[34] W. C. Ward and G. W. Swift, ''Design environment for low amplitude thermoacoustic engines (DeltaE),'' J. Acoust. Soc. Am. **95**, 3671–3672 (1994). Software and user's guide available either from the Los Alamos thermoacoustics web site at www.lanl.gov/thermoacoustics/ or from the Energy Science and Technology Software Center, U.S. Department of Energy, Oak Ridge, Tennessee.
[35] I. E. Idelchik, *Handbook of Hydraulic Resistance*, 3rd ed. (Begell House, New York, 1994).
[36] A. M. Fusco, W. C. Ward, and G. W. Swift, ''Two-sensor power measurements in lossy ducts,'' J. Acoust. Soc. Am. **91**, 2229–2235 (1992).
[37] Ref. 19, Eq. (4.72).
[38] A good overview of this large body of work can be found in proceedings of the annual Intersociety Energy Conversion Engineering Conferences and the less frequent International Stirling Engine Conferences.
[39] R. Radebaugh, ''A review of pulse tube refrigeration,'' Adv. Cryog. Eng. **35**, 1191–1205 (1990).
[40] G. W. Swift, M. S. Allen, and J. J. Wollan, ''Performance of a tapered pulse tube,'' in *Cryocoolers 10*, edited by R. G. Ross, Jr. (Plenum, New York, 1999).
[41] A. Bejan, *Advanced Engineering Thermodynamics*, 2nd ed. (Wiley, New York, 1997).
[42] Ref. 19, Eq. (6.21).
[43] R. C. Tolman and P. C. Fine, ''On the irreversible production of entropy,'' Rev. Mod. Phys. **20**, 51–77 (1948).

Beampattern control of a microphone array to minimize secondary source contamination

Peter Jordan, John A. Fitzpatrick,[a] and Craig Meskell
Department of Mechanical Engineering, Trinity College, Dublin 2, Ireland

(Received 28 November 2001; revised 17 January 2003; accepted 24 February 2003)

A null-steering technique is adapted and applied to a linear delay-and-sum beamformer in order to measure the noise generated by one of the propellers of a $\frac{1}{8}$ scale twin propeller aircraft model. The technique involves shading the linear array using a set of weights, which are calculated according to the locations onto which the nulls need to be steered (in this case onto the second propeller). The technique is based on an established microwave antenna theory, and uses a plane-wave, or far field formulation in order to represent the response of the array by an nth-order polynomial, where n is the number of array elements. The roots of this polynomial correspond to the minima of the array response, and so by an appropriate choice of roots, a polynomial can be generated, the coefficients of which are the weights needed to achieve the prespecified set of null positions. It is shown that, for the technique to work with actual data, the cross-spectral matrix must be conditioned before array shading is implemented. This ensures that the shading function is not distorted by the intrinsic element weighting which can occur as a result of the directional nature of aeroacoustic systems. A difference of 6 dB between measurements before and after null steering shows the technique to have been effective in eliminating the contribution from one of the propellers, thus providing a quantitative measure of the acoustic energy from the other.
[DOI: 10.1121/1.1568755]

PACS numbers: 43.38.Hz, 43.28.Tc, 43.50.Lj [SLE]

I. INTRODUCTION

Microphone array processing consists of combining the signals from a collection of sensors such that the device becomes maximally sensitive to sound originating from a chosen location, while its sensitivity to sound emanating from all other regions is reduced. Such devices have been used in a wide range of applications to identify and measure noise sources. For aeroacoustics, a planar array was designed by Marcolini and Brooks[1] for the measurement of helicopter rotor noise, allowing the beam-width of the array to be maintained constant, independent of frequency and look direction. Billingsley and Kinns[2] used a linear array for the location of noise sources on a full scale jet engine, where, by statistical averaging, the time-varying source distributions could be calculated with respect to position and frequency. More recent examples are the spiral array used by Humphreys *et al.*[3] for the measurement of flap side-edge noise and the linear phased array used by Siller *et al.*[4] to measure engine core noise.

One aspect of array technology, identified as problematic at an early stage, is the susceptibility of the device to contamination from sources other than those at the focus position. This has led to a great deal of research focused on control of the spatial sensitivity (or beampattern) of the array. Beamformer design approaches for the control of sidelobe levels and mainlobe width were proposed by both Dolph[5] and Riblet[6] for microwave antenna applications. This work was extended to steered and compensated arrays for acoustic systems by Pritchard[7] who established the relationship between lobe width and array length. More recent developments for beamformer optimization are typified by the work of Humphreys *et al.*[3] whose spiral array layout results in an improvement of the dynamic gain of the device, or that of Dougherty *et al.*[8] which also focuses on design optimization for sidelobe suppression. These kinds of application seek to improve the general performance characteristics of the array, such that application to systems with unknown source distributions are optimal. However, in systems where the source distributions (or target directions where emitting arrays are concerned) are known, it can be advantageous to use a more application-specific approach whereby the response of the array is modified to be optimum for a look direction of interest. This kind of approach has been used in speech acquisition applications, where, for example, to eliminate the contaminating effect of secondary sources, Kootsookos *et al.*[9] have derived the conditions necessary to impose a broadband null on an array's response, while Ryan *et al.*[10] considered a beampattern subject to such a constraint, where the array was used to measure a nearfield source located in the same direction as farfield interference which the null was to reject.

In this work such an application-specific optimization is considered for aeroacoustic measurements. The nulls of a linear delay-and-sum beamformer are steered so as to enhance its capacity for the quantitative measurement of a real aeroacoustic source, in the form of installed propeller noise. Tests were performed on a $\frac{1}{8}$ scale twin-propeller aircraft model using an out-of-flow linear microphone array in a broadside configuration. A procedure was developed to enable a beampattern null to be steered onto one of the propellers, thus virtually eliminating its contribution to the array

[a]Author to whom correspondence should be addressed. Electronic mail: john.fitzpatrick@tcd.ie

0001-4966/2003/114(4)/1920/6/$19.00

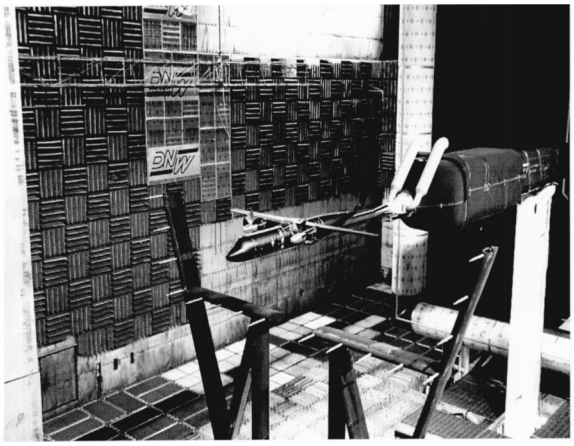


FIG. 1. APIAN model in DNW open jet.

measurement when the focus position is directed on the other and hence to improve the reliability of the measurement from the propeller of interest.

II. TEST SETUP

The model wingspan was 3 m with propeller diameters of 0.5 m. Figure 1 shows the model installed in the open jet anechoic facility of DNW (The German-Dutch Windtunnels) in the Netherlands. The measurements were performed as part of the European project APIAN (Advanced Propulsion Integration Aerodynamics and Noise). As the primary objective of the project was the validation of numerical codes for the prediction of installed propeller aeroacoustics, the scope of the array experiments was limited to a small number of test cases. Tests were performed at a flow Mach number of 0.2, with angles of attack of 0° and 10°. The blade pass frequency for this velocity was 800 Hz, corresponding to a wavelength of 0.43 m. Measurements were made using an out-of-flow linear array, comprising 29 microphones at 0.01-m intervals, giving an overall length of 2.8 m. The array was located 5.7 m beneath the model and 6.5 m upstream of the propeller plane as shown schematically in Fig. 2. When the shear layer refraction effects are accounted for as detailed

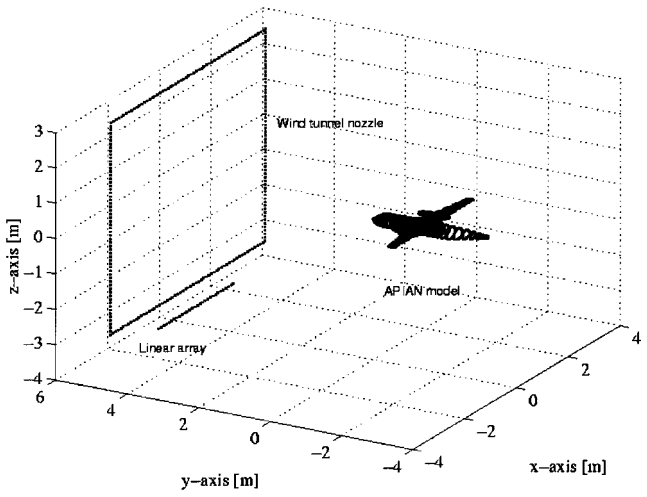


FIG. 2. Schematic of model setup and array.

by Amiet,[11] the distance from source to receiver is of the order of 22 acoustic wavelengths so that the array can be considered to be in the far field.

The low-frequency, narrow-band source characteristics together with the distance of the array from the source region result in poor spatial resolution. As a consequence, it is extremely difficult to separate the noise from each of the propellers using conventional delay-and-sum beamforming techniques. In order to enable the noise from each propeller to be estimated, the null positions need to be controlled, so as to eliminate the contaminating effect of one propeller when the array is focused on the other. Section IV considers how this can be achieved.

III. DATA ACQUISITION AND ARRAY POSTPROCESSING

The microphone array consisted of 29 KE4 Senheiser electret microphones with a 20–20 000-Hz pass band and integrated preamplifiers. Data acquisition was achieved using two 16-channel Kinetic systems V200 acquisition cards, mounted on a National Instruments VXI chassis. A controller provided local control of the system and was connected to a PCI card on a PC via an MXI interface. Each V200 card was capable of acquiring 16 channels of data simultaneously up to a sampling rate of 200 kHz.

A single acquisition consisted of 150 000 data points acquired at 25 kHz and band-pass filtered between 100 Hz and 10 kHz. The signals were then divided into blocks of 8192 points, Fourier transformed and averaged, and the upper triangular elements of the cross-spectral matrix calculated as follows:

$$R_{ij}(f)=\frac{1}{B}\sum_{b=1}^{B}[X_{ib}^{*}(f)X_{jb}(f)], \tag{1}$$

where R_{ij} is the cross-spectral matrix entry for row i and column j, B is the number of averages, X_{ib} and X_{jb} are the Fourier transforms of the temporal signals for channels i and j and data block b, and the superscript * denotes the complex conjugate. The cross-spectral matrix is of Hermitian form and so the lower triangular elements were calculated via complex conjugation of the upper triangular elements.

Focusing was performed to the broadside in the conventional way as detailed, for example, by Johnson and Dudgeon,[12] using a steering vector of the form

$$w=[e^{-j(kR_{fm})}], \tag{2}$$

where R_{fm} is a vector of length M containing the acoustic ray distances from a given focus position to each of the microphones and includes a correction to account for refraction at the shear layer,[9] and k is the wavenumber at which the beamformer is applied.

The beamformer output for a given focus position is then calculated as

$$P=w^{*}Rw^{T}, \tag{3}$$

where R is the cross spectral matrix, the superscript * again denotes complex conjugate, and the superscript T denotes a nonconjugate transpose.

IV. BEAMFORMER IMPLEMENTATION

A. Theory

The principle of microphone array measurement is based on the fact that sound waves arriving at elements of the array do so at times dependent on the source-microphone distance and the speed of sound. The relative phase of the signals is therefore a function of source-microphone distance, the speed of sound, and the frequency of the source. If each of the signals is manipulated in order to eliminate this phase difference and they are then summed, they will do so constructively. In this way an array can be made more sensitive to sound radiating from a given direction and a set of phase weights can be chosen so that the directional sensitivity of the array can be controlled. When the array is focused in a given direction, it becomes less sensitive to sound radiating from other directions, as the set of weights used to focus the array will cause sound waves coming from those other directions to interfere destructively when the signals are summed. This spatial variation of sensitivity, P, is given by

$$P(k,R_{fm},R_{sm})=\sum_{m=1}^{N} A_m \frac{e^{-jk(R_{jm}-R_{sm})}}{R_{sm}}, \tag{4}$$

which expresses the sensitivity of an N microphone array, focused on a location given by the microphone-focus distances R_{fm}, to a point monopole source whose position relative to the array is given by the set of source–microphone distances R_{sm}. As the array is linear, this sensitivity is axisymmetric when considered in three-dimensional space, and so the locus of the set of points corresponding to a single set of microphone-focus or microphone-source distances describes a circle centered on the array axis. For this particular application, however, given the anechoic environment in which the measurements take place, regions away from the aircraft model are considered free of any acoustic sources. The propellers are thus assumed to be the only major sources of sound and can be considered to be two concentrated sources located on a horizontal axis in the flow field, hereafter referred to as the source axis. The array is orientated parallel to the source axis as shown in Fig. 2, in a broadside configuration, and the vectors R_{sm} and R_{fm} refer to locations on the source axis.

The spatial sensitivity of an array is usually determined by the source frequency and the position of the source relative to the array. However, the sensitivity of the array can be controlled independently of the experimental parameters through use of the weights A_m, which can be used to specify the sensitivity pattern. Thus, Eq. (4) can be used to generate the system of M simultaneous equations given by

$$[P]=[A_m]^*[S],$$

corresponding to M equispaced locations on the source axis, where $[P]$ is an $M\times 1$ matrix defining the desired array sensitivities, $[A_m]$ is a $1\times N$ matrix of unknown microphone weights, and $[S]$ is an $N\times M$ matrix, where the columns c_i contain N element signals, $[e^{-jk(R_{fm}-R_{sm})}/R_{sm}]_i$, corresponding to a point monopole at position i on the source axis, for a fixed focus position defined by the phase delays $e^{-jkR_{fm}}$.

This equation system can be easily overdetermined by specifying a larger number of source positions than microphones and solved using a pseudo-inverse method for the set of weights A_m which would give the desired sensitivity pattern, P. However, the equation system is badly conditioned and useful weighting functions are impossible to obtain. This ill conditioning is due in part to the fact that Eq. (4) assumes spherical wavefronts propagating from a series of discrete points in space, in short a near-field assumption.

An alternative approach is to formulate the problem using a far-field or plane wave model as is often the case in telecommunications research. In this way, Schelkunoff[13] modeled an array system such that the response of an n-element linear array could be represented by the nth-order polynomial

$$P(\theta,k,d)=|A_0+A_1z+A_2z^2+A_3z^3+\cdots+A_mz^m|, \tag{5}$$

where

$$\begin{aligned} &z(\phi)=e^{j\phi},\\ &\phi(\theta,k,d)=kd\ \mathrm{Cos}\ \theta, \end{aligned} \tag{6}$$

with d the element spacing, k the wavenumber, and θ the look direction of the beamformer for broadside application. This polynomial can be factorized to give

$$P(\theta,k,d)=|(z-t_1)(z-t_2)\cdots(z-t_{m-1})|, \tag{7}$$

where t_m are the roots and the $(m-1)$ binomials correspond to the nulls of the directivity pattern. By then defining the roots t_m as

$$t_m(\theta_m,k,d)=e^{jkd\ \mathrm{Cos}\ \theta_m},$$

where θ_m are the desired null locations, and substituting back into Eq. (7), a polynomial is generated whose coefficients give the shading necessary to steer the nulls of the array to the locations defined by θ_m. Thus it can be seen that, while the spatial sensitivity of the array cannot be prespecified arbitrarily, a certain level of control in the form of null steering is possible.

B. Simulation

A simulation was performed to demonstrate the feasibility of the null steering operation for the measurement of sound generated at the blade-pass frequency by one of the model propellers. The setup was the same as that used in the tests, consisting of a 2.8-m-long, 29-element linear array in a broadside configuration, 9.5 m from a source region containing two sources with frequencies of 800 Hz. The sources are located at the port (left) and starboard (right) propellers, respectively, and the array is focused on the latter. The theoretical spatial sensitivity of the array, defined over a linear region containing the model, is shown in Fig. 3 before and after null steering was implemented. It can be seen that for a conventional delay-and-sum beamformer focused on the starboard propeller, the sensitivity to sound emanating from the port propeller is -10 dB relative to the maximum, so it attenuates by this amount. After the array has been shaded, and a null steered onto the port propeller, the attenuation has been increased to -45 dB, so that any contribution from the

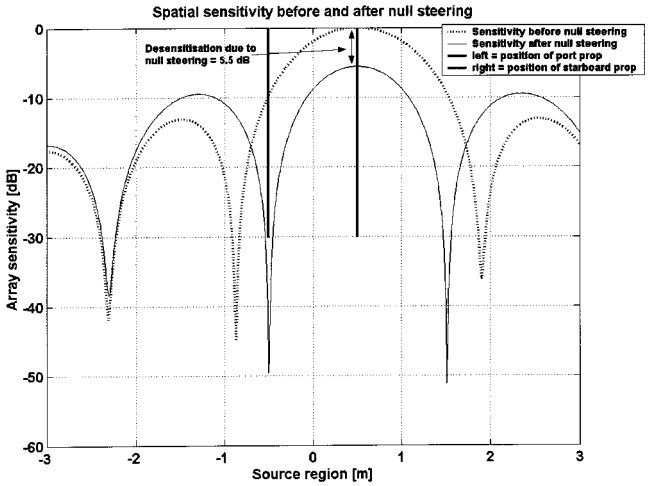


FIG. 3. Beampattern spatial sensitivity before and after null steering.

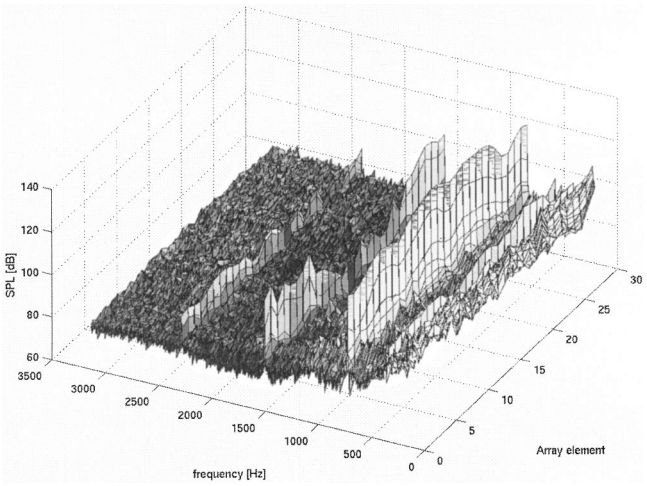


FIG. 5. Microphone auto-spectra for test case 1.

port propeller is now eliminated when the array is focused on the starboard propeller. An additional effect of the null steering operation is a reduction in mainlobe level and width and an increase in the sidelobe levels. The array is now less sensitive at its focus position and, so, to correct for this shading-induced attenuation, measured levels must be increased by 5.5 dB. This level can be estimated for any null steering operation in advance of the implementation of the method.

To assess the array's capacity to perform quantitative measurement of a source in the presence of a secondary contaminating source, a simulation for a linear source system containing two uncorrelated monochromatic point monopoles of equal strength was performed. The source frequency was 800 Hz with source separation of 1 m and source axis parallel to the array at a distance of 9.5 m. Figure 4 shows results from this simulation, with and without null steering, together with those from a single source system which serve as a reference. When both sources are present and conventional beamforming is applied, the estimated noise level at the starboard propeller is -2 dB relative to the maximum which is located midway between the sources. The actual level of noise from this source is -4.5 dB as indicated by the

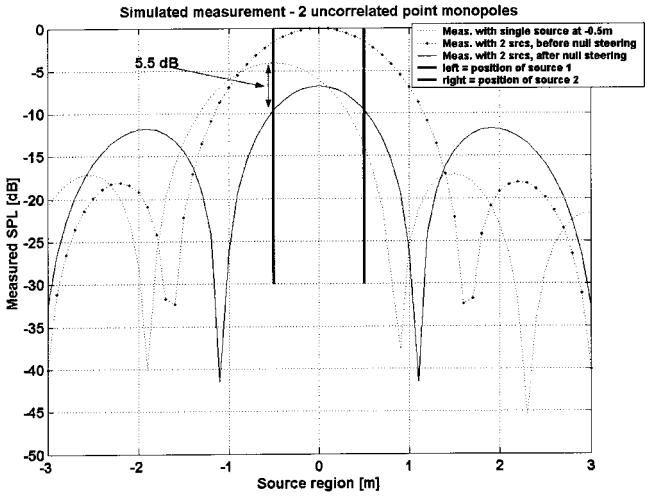


FIG. 4. Simulated beamformer measurements.

results for the single source system where the source location has been correctly identified. When the beampattern is controlled to place a null on the port propeller and the array focused on the starboard propeller, the measured level is now -10 dB, which is to be corrected for the predicted 5.5-dB attenuation resulting in a measurement of -4.5 dB, which is the expected value. Thus, the use of the null steering to eliminate the second source has been effective.

C. Conditioning of the cross-spectral matrix

A necessary condition for accuracy in the steering of array pattern nulls is that, prior to shading, the individual microphone amplitudes be nominally equal. For receiving arrays, and particularly in aeroacoustic applications, the element responses are subject to the directional nature of the incident waveform. Shading functions calculated to effect a desired null-steering operation will be distorted by this intrinsic array shading, resulting in a loss of accuracy in the operation. The directional nature of the aeroacoustic system under investigation is demonstrated in Fig. 5 in which the autospectrum from each microphone in the array is plotted for a particular test. The directivity of the sound field is clearly evident, both at the blade-pass frequency and the first two harmonics. The beamforming operation is defined in Eq. (3) in terms of P, the beamformer output, R, the cross-spectral matrix, and w, the steering vector. To account for the directional effect of the sound field, the cross-spectral matrix has to be normalized prior to application of the null-steering vectors. This normalization consists in averaging the cross-spectral matrix along each of the columns prior to premultiplication by the steering vector, and along the resulting row prior to postmultiplication by the steering vector. The absolute values of the cross-spectral matrix prior to premultiplication by the steering vector are thus given by

$$A_j^{pr} = \frac{\Sigma_{i=1}^{N} |R_{ij}^{pr}|}{N}, \tag{8}$$

where the superscript pr indicates values prior to premultiplication, i and j refer to row and column numbers, respectively, as before, and the absolute values of the resulting 1

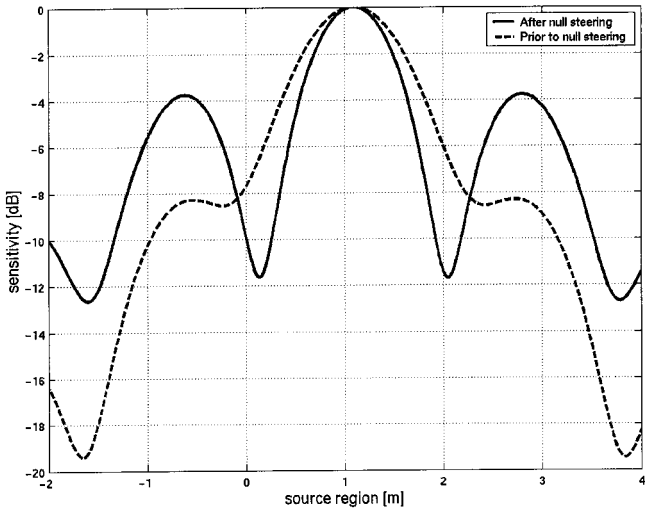


FIG. 6. Effect of intrinsic weighting on response.

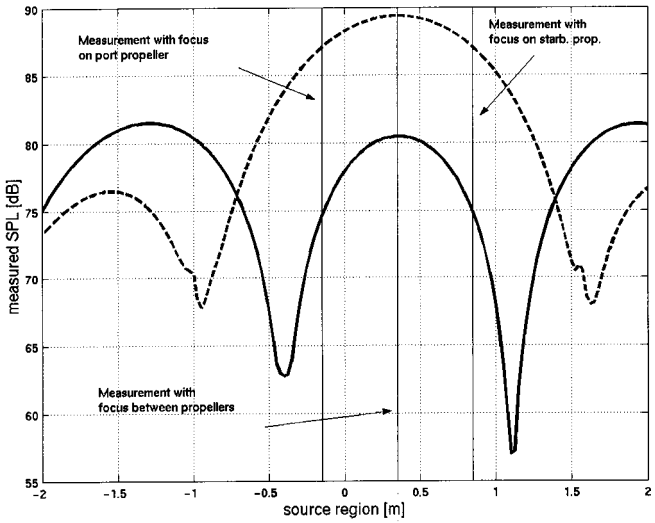


FIG. 7. Array measurement results for test case 1 (solid line—modified: dashed line—original).

$\times N$ cross-spectral matrix prior to postmultiplication are given by

$$A^{po}=\frac{\Sigma_{j=1}^{N}|R_j^{po}|}{N}, \tag{9}$$

where the superscript *po* refers to values prior to postmultiplication.

To demonstrate the influence on the array response, Eq. (4) was implemented numerically with an arbitrary set of weights, A_m, and null steering then implemented using the original cross spectra and then the conditioned cross spectra. The results are shown in Fig. 6. It can be seen that prior to null steering the characteristic shape of the beampattern has changed considerably. The sidelobe troughs, which previously showed very large levels of attenuation, now show an attenuation of the order of -8 dB, and while the null-steering operation has been effective in moving the nulls, the maximum attenuation achieved has been considerably reduced. This clearly indicates that if the cross-spectral matrix is not conditioned (i.e., normalized), the array will underperform when a conventional delay-and-sum approach is used, and the null-steering operation will no longer be effective for the elimination of secondary source contamination.

V. EXPERIMENTAL RESULTS AND DISCUSSION

Three tests at a flow Mach number of 0.2 were performed, with the propellers at zero incidence for two of tests and at 10° for the third. The results are summarized in Table I, where the corrected and uncorrected levels are shown for the three test cases, together with the levels measured by a single microphone. As can be seen, the levels of the original measurements are some 6 dB above the levels estimated once null steering or beamformer optimization has been implemented. The results show that the noise level generated by each propeller is ~80 dB at zero incidence and ~85 dB at 10° incidence. The result for the array measurement for the first test is shown in Fig. 7. The array focus has been swept incrementally across a linear region containing the two propellers using delay-and-sum beamforming [Eq. (7)]. The dashed and solid curves represent, respectively, the results before and after null steering has been applied, and the cross-spectral matrix treatment is applied in both cases. The asymmetry of the results represents a positional error of 0.35 m (2°). This is most probably a consequence of phase discrepancies which may arise as the propellers are considered to be point monopole sources when the shear layer correction and phase weighting operations are implemented. In practice, the noise source is distributed over each of the propellers.

The result of the conventional beamformer in Fig. 7 appears to identify a single source at a level of 89 dB, located at approximately 0.35 m on the source axis whereas the propellers are located at -0.15 and 0.85 m, respectively. This is because the spatial resolution of the array is poor, the distance between the two sources being small compared with their acoustic wavelength and the distance of the array from the source region, so that the two sources appear as a single source. However, as the spatial structure of the source system is known, the result can be interpreted using the simulation described in Sec. IV B. The array measurement can thus be used to effect a quantitative measure of the true contribution of each source to the far-field sound. With the beamformer optimized for measurement of the propellers (the solid curve

TABLE I. Original and corrected measurements

Mach no.	Angle of attack	Single mic level (dB)	Starboard SPL (dB)	Port SPL (dB)	Starboard $SPL_{corrected}$ (dB)	Port $SPL_{corrected}$ (dB)
0.2	0°	93.57	87.03	87.03	80.29	80
0.2	0°	93.78	86.39	86.48	80.2	80.31
0.2	10°	102.22	91.33	91.45	85.12	85.31

in Fig. 7), levels of ~75 dB are measured, which when corrected for the 5.5-dB attenuation of the array gives a value of ~80 dB for the individual propeller levels. Thus, 6 dB of the original beamformer measurement for either of the propellers was due to contamination from the other propeller.

It should be noted, however, that this optimization is only effective for focus on the propellers. For foci in other directions the beampattern is not optimal and so such measurements are subject to significant and varying levels of contamination, depending on the relative positions of the mainlobe, the sidelobes, and the sources. Thus, the levels measured by the array at locations other than those of the port and starboard propellers are superfluous, and represent only in a very complex way information pertaining to the source distribution of this particular aeroacoustic system. For example, the high levels measured when the array is focused at the wingtips of the model arise because the sidelobes now lie in the direction of the propellers. As a result of this increase in sensitivity, the array output is contaminated for these regions.

VI. CONCLUSIONS

A procedure has been developed to control the beampattern of a linear delay-and-sum beamformer. The resultant shading of the array pattern can steer the nulls onto a prespecified location, resulting in an improvement in the capacity of the array to measure a source of interest in the presence of a second ''contaminating'' source. The technique has been validated using numerical simulations and then used to measure the noise emitted from one propeller of a twin propeller model in wind tunnel tests. The results show that about 6 dB of the original beamformer measurement was due to contamination from the second propeller. The procedures give reliable control of an array beampattern and it has been demonstrated that treatment of the cross-spectral matrix is required to remove the intrinsic element weighting caused by the directivity of the aeroacoustic sound field.

ACKNOWLEDGMENTS

The work described in this paper was performed as part of the APIAN project financially supported by the European Union under Contract No. BRPR-CT96-0201. The authors would like to acknowledge the staff of DNW for their support during the test campaign.

[1] M. A. Marcolini and T. F. Brooks, ''Rotor noise measurement using a directional microphone array,'' J. Am. Helicopter Soc. **37**, 11–22 (1992).
[2] J. Billingsley and R. Kinns, ''The acoustic telescope,'' J. Sound Vib. **48**, 485–510 (1976).
[3] W. M. Humphreys, Jr., T. F. Brooks, W. W. Hunter, Jr., and K. R. Meadows. ''Design and use of a microphone directional array for aeroacoustic measurements,'' AIAA Paper No 98-0471 (1998).
[4] H. A. Siller, F. Arnold, and U. Michel, ''Investigation of Aero-Engine Core Noise using a Phased Microphone Array,'' AIAA Paper No. 2001-2269, 7th AIAA/CEAS Aeroacoustics Conference, Maastricht, Netherlands (2001).
[5] C. L. Dolph, ''A current distribution for broadside arrays which optimises the relationship between beamwidth and sidelobe level,'' Proc. Inst. Radio Eng. **34**, 335–348 (1946).
[6] H. J. Riblet, ''Microwave omnidirectional antennae,'' Proc. Inst. Radio Eng. **35**, 474–478 (1947).
[7] R. L. Pritchard, ''Optimum directivity patterns for linear point arrays,'' J. Acoust. Soc. Am. **25**, 879–891 (1953).
[8] R. P. Dougherty, ''Source location with sparse acoustic arrays: interference cancellation,'' Proc. CEAS/ASC Workshop on Wind Tunnel Testing in Aeroacoustics, DNW, Holland, November (1997).
[9] P. J. Kootsookos, D. B. Ward, and R. C. Williamson, ''Imposing pattern nulls on broadband array responses,'' J. Acoust. Soc. Am. **105**, 3390–3398 (1999).
[10] J. G. Ryan and R. A. Goubran, ''Optimum near-field performance of microphone arrays subject to a far-field beampattern constraint,'' J. Acoust. Soc. Am. **108**, 2248–2255 (2000).
[11] R. K. Amiet, ''Refraction of sound by a shear layer,'' J. Sound Vib. **58**, 467–482 (1978).
[12] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1993).
[13] A. Schelkunoff, ''A mathematical theory of arrays,'' Bell Syst. Tech. J. **22**, 80–107 (1943).

High-sensitivity photoacoustic leak testing

Eric Huang and David R. Dowling[a)]
Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109-2133

Timothy Whelan
Honeywell Federal Manufacturing & Technologies, LCC, P.O. Box 419159, Kansas City, Missouri 64141-6159

John L. Spiesberger
Department of Earth and Environmental Science, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6316

(Received 27 January 2003; revised 25 June 2003; accepted 14 July 2003)

The photoacoustic effect may be exploited for the detection and localization of gas leaks from otherwise sealed components. The technique involves filling the test component with a tracer gas, and radiating the component to produce photoacoustic sound from any leak site where tracer gas is present. This paper describes demonstration experiments utilizing 10.6-μ radiation from a carbon-dioxide laser and sulfur hexafluoride as a tracer gas for photoacoustic leak testing at leak rates between 6×10^{-5} cm^3/s (1 cm^3 in 4.6 h) and 5×10^{-9} cm^3/s (1 cm^3 in 6.3 years). The technique may reach or exceed the capabilities of the most sensitive commercial leak test systems using helium mass-spectrometers. In addition, comparison of the measured results to a simple scaling law suggests that tracer gas cloud geometry influences the photoacoustic signal amplitude.
 [DOI: 10.1121/1.1605386]

PACS numbers: 43.38.Zp, 43.35.Ud, 43.60.Gk [AJZ]

I. INTRODUCTION

Photoacoustics is the generation of sound due to unsteady heating of a photoactive material—gas, liquid, or solid—by an unsteady source of light or invisible electromagnetic radiation. Over the last decade or so, gas-phase photoacoustics has been developed into a noncontacting means of detecting and locating small gas leaks on the exterior surfaces of components that are intended to be sealed. Although implementations may be different, photoacoustic leak test systems all rely on the same basic steps. First, the part under test is charged with a photoactive tracer gas (typically sulfur hexafluoride, SF_6) to a pressure greater than the local ambient pressure so that a cloud or plume of tracer gas forms adjacent to any unintended leak. Next, the component under test is scanned with a radiation source (typically a carbon dioxide, CO_2, laser) having a wavelength tuned to an absorption line of the photoactive gas (10.6 μ for SF_6). When a tracer gas plume is illuminated, photoacoustic sound is generated. Recordings of this sound can be analyzed to detect and locate the tracer gas plume when the test geometry and environment are known. Here, the tracer gas plumes are detected and localized, not the actual physical defect in the part that leads to an unintended gas pathway. However, for the volumetric leak rates of interest here ($\sim10^{-4}$ cm^3/s and lower), gas-phase diffusive transport ensures that the highest tracer gas concentrations will only occur at the location where tracer gas molecules emerge from the test component's surface.

Photoacoustic leak testing has advantages in sensitivity, speed, ease of implementation, and robustness compared to other leak test technologies for leak rates of $\sim10^{-4}$ cm^3/s and lower. It can be implemented at room temperature and pressure. Photoacoustic leak testing requires neither the complex high-vacuum system typically essential for helium-mass spectrometry (HeMS), nor the use of potentially hazardous chemicals like the ammonia-phenolphthalein method. Unlike dunk tanks and soap film systems, it is remote and noncontacting. It is more sensitive than pressure decay measurements or backscatter absorption systems, and is more robust (less likely to be fouled and much less operator dependent) than sniffer-based systems. In addition to the other listed advantages, photoacoustic leak tests can be completed in a matter of seconds for phone-book size components. Larger objects can be tested with scaled up versions of the test setup described here. The main liabilities of photoacoustic leak testing are its reliance on potentially dangerous laser radiation and its inability to accurately detect or locate leaks lying on interior or shadowed surfaces that cannot be illuminated.

The experimental results in this paper show that photoacoustic leak testing can be conducted at sensitivities that rival or possibly exceed that available from leak test systems based on HeMS, the method most utilized for sensitive testing in commercial technology. Helium mass spectrometry has been used for the detection of small leaks since World War II (Dushman and Lafferty, 1962) and is now used pervasively in manufacturing due to its high sensitivity, 10^{-10} cm^3/s or better (see Rasmussen and Jeppesen, 1998), relatively low cost, and good reliability. As of this writing, some commercial manufacturers (for example, Alcatel and BOC Edwards) are advertising HeMS systems capable of detecting leaks into the 10^{-12}-cm^3/s range and many others advertise systems capable of detecting leaks in the

[a)]Author to whom correspondence should be addressed. Electronic mail: drd@engin.umich.edu

10^{-11}-cm^3/s range. A good technical description of HeMS is given in Hablainian (1997). In addition to addressing sensitivity limits, the present results appear to defy a simple diffusive-transport scaling law for photoacoustic signal amplitudes as the leak rate decreases.

The main advantage of photoacoustic leak testing is that it allows small leaks, 10^{-5} cm^3/s or smaller, to be detected and located quickly. By comparison, locating leaks of such low flow rates with helium mass spectrometry is a time consuming, heavily operator-dependent process. There are two methods used to locate the leak with helium mass spectrometry. In the first method, the item of interest is filled with helium (or a mixture of helium and another gas) and a wand or sniffer is passed over the surface of the item. The wand is usually a tube with a small diameter hole that draws in gas which then passes through a tube into the mass spectrometer. In the second method, the item is connected to the mass spectrometer and a vacuum is created inside the unit. The operator then sprays the outside of the item with a slow flow of helium and watches the mass spectrometer for an indication of its presence. Due to the length of time necessary to perform these tests and the skill level required of the operators, these methods are frequently employed only after other tests have indicated a leak exists somewhere on the item.

Photoacoustic leak testing was first reported in McRae and Dewey (1992) and in McRae (1994), and is similar to work done on remote detection and ranging of gas clouds (Brassington, 1982). Since then commercial developments of the technique have seen it extended to include signal-to-noise ratio improvements via proper laser-scanning frequency selection (Olender *et al.*, 1998a), multiple laser beams (Olender *et al.*, 1998b), enclosures to improve signal characteristics (Schroff and Stetter, 1999), multiple microphones for acoustic localization (Yonak and Dowling, 2001), and multiple laser wavelengths for background noise assessment (Veronesi *et al.*, 2001). A general audience description of the approach followed here is available in Sharke (2000) with more detailed accounts provided in Yönak and Dowling (1999, 2002). The current results differ from these prior efforts in their emphasis on increased sensitivity.

II. EXPERIMENTAL SETUP

The experimental setup was nearly the same as for the cylindrical geometry leak tests described in Yönak and Dowling (2002) except that the optical and acoustic path lengths were shorter, 12 microphones were used instead of four, and a CO_2 laser capable of delivering a nominal 145 W, instead of the previously utilized 12 W, provided the input radiation. All experiments were conducted on an optical tabletop in a laboratory environment with painted cinder block walls and tiled floor.

The radiation source used in these experiments was a DeMaria ElectroOptics Systems, Inc. (DEOS) LC-100NV carbon dioxide (CO_2) laser. Its output power was controlled via pulse width modulation from a DEOS PC-1 controller and was measured using a SYNRAD PowerWizard™ PW-250 hand held laser power meter. A NESLAB M75 STD 208/60 PD2 chiller kept the laser head and power supply at 20 °C. A rotating 12-sided polygonal mirror assembly (Lincoln Laser Model DT-12-200-028) was used to repetitively sweep the incoming CO_2 laser through a plane lying parallel to, and 125 mm above, the optical tabletop. The rotational speed of the polygonal mirror assembly (18750 rpm) was set by a Lincoln Laser MC-5 motor controller, and the electrical power to the rotating mirror motor was supplied by an Acopian Model A48HT600 48 V dc power supply. For safety purposes, a coated ZnSe flat from II-VI Optics was used to combine and co-align the invisible and potentially dangerous CO_2 laser beam with a visible and relatively harmless 5 mW red He–Ne laser beam.

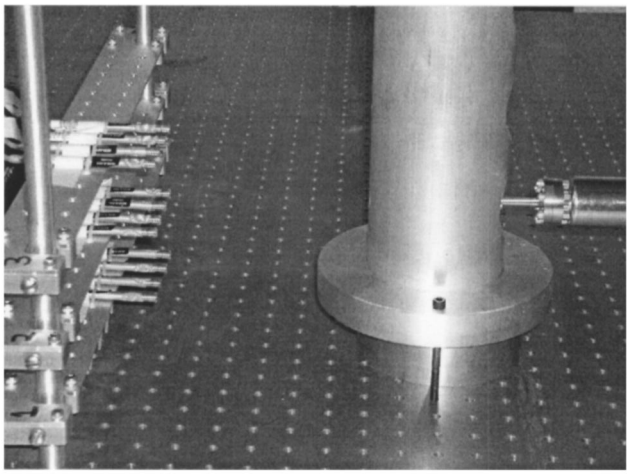


FIG. 1. Picture of the experimental test geometry. The 12-microphone array appears on the left and the vertical test cylinder on the right. The leak is located at the left edge of the test cylinder at the height of the SF_6 canister shown at the right edge of the picture. The spacing between holes in the optical tabletop is 25.4 mm in each direction.

A brushed aluminum cylinder having a 10-cm diameter provided the test surface on which the calibrated SF_6 leaks were mounted. The axis of this test cylinder was vertical and perpendicular to the tabletop (see Fig. 1). The test surface of the cylinder was the semi-circle determined by the intersection of the CO_2-laser scan plane with the cylinder. The six calibrated leaks in this study were manufactured by Vacuum Technology, Inc. and ranged in rate from 6.0×10^{-5} cm^3/s down to 5.0×10^{-9} cm^3/s. The front 25% of a SF_6-reservoir canister for one of these leaks is shown at the right edge of Fig. 1. The leaking gas was brought to the test surface with special high-vacuum fittings and a leak cap having a diagonal hole. This hole forced the leaking SF_6 to escape from a known location without letting the CO_2 laser beam shine into the gas volume trapped between the leak orifice and the outside of the leak cap (see Fig. 2).

Twelve 7.6-mm-diam condenser microphones (Brüel and Kjaer type 4939-A011) were used to record the photoacoustic sounds created by the swept laser beam as it passed through the SF_6 gas cloud formed at the opening of the leak cap. The microphones were placed 152 mm from the leak in a 3 by 4 array as shown in Fig. 1. The microphone rows were 57, 108, and 159 mm above the tabletop, and the horizontal spacing between microphones was 25.4 mm. In Fig. 1, the invisible CO_2-laser beam emerges between the upper two

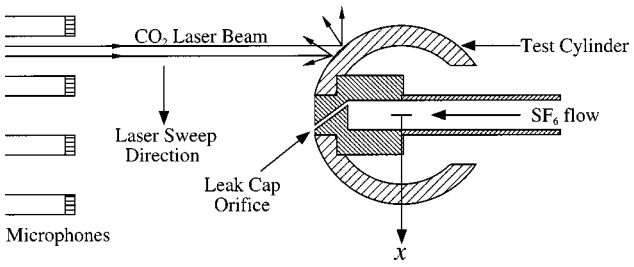


FIG. 2. Schematic top view of the experimental test geometry.

rows of microphones and traverses the picture from left to right, hitting the test cylinder at the height of the leak. The microphone signals were amplified using three four-channel Brüel and Kjaer NEXUS Type 2690 microphone conditioning amplifiers to an output sensitivity of 316 mV/Pa, high-pass filtered above 3 kHz with three four-channel Krohn-Hite Model 3384 active filters with an output gain of 10 dB, digitized with 12-bit resolution at 400 kHz using National Instruments data acquisition hardware, and stored in a laboratory PC computer running Labview® software. The nominal experimental bandwidth was 3 to ~80 kHz, the upper limit being determined by the roll-off of the microphones with their protection grids in place.

The digitized microphone signals were Fourier analyzed and comb-filtered to select the signal frequencies. Bartlett matched-field processing (MFP), computed along the intersection of the laser scan plane and the front of the test cylinder (the semi-circular test surface), was conducted at each signal frequency using the tabletop as a hard reflecting surface to detect and locate the various leaks. These single-frequency Bartlett ambiguity surfaces were incoherently averaged across 16 signal frequencies from 15 to 75 kHz to obtain the final broadband Bartlett ambiguity-surface results presented in the next section. For the present purposes, these final Bartlett MFP results can be interpreted as indicating the likelihood of a leak lying at a particular location. When a peak in the Bartlett MFP output rises well above the noise level, a leak is detected and its acoustically determined location is the position of the ambiguity-surface peak. Bartlett MFP is described in Jensen *et al.* (1994) and has been used in prior leak test studies (see Yönak and Dowling, 1999, 2002).

III. RESULTS

For simplicity, all of the results reported here are from leaks centered on the swept laser beam's path across the cylinder ($x=0$). Figure 3 shows Bartlett MFP output on the vertical axis vs x—the horizontal Cartesian coordinate perpendicular to the test cylinder's axis and the line connecting the axis of the spinning mirror to the test cylinder axis—for the smallest leak in this study at CO_2 laser power levels between 0 and 140 W. Here $x=\pm 5$ cm corresponds to the edges of the test cylinder when it is viewed from the microphone array. The main feature of this figure is the prominence of the Bartlett MFP peak at $x=0$ for all nonzero laser powers. Furthermore, the peaks are monotonically increasing in height with increasing laser power. These results show that 5×10^{-9} cm^3/s of SF_6, which corresponds to approximately 1 cm^3 every 6 years, does indeed produce measurable photoacoustic sound. It should be mentioned that the 5×10^{-9} cm^3/s of SF_6 was the smallest leak for which a supplier was found. This leak was also at the NIST traceable detectable limit of the supplier's capabilities for SF_6.

Figures 4 and 5 show how the Bartlett MFP results vary with leak rate for laser powers of 10 and 140 W. Here, the plug and cylinder test cases involve a blank leak cap inserted into the test cylinder in place of a calibrated leak. Thus, these test cases represent the background noise signature of this experimental geometry at nominal laser powers of 10 and 140 W. In all cases shown in Figs. 4 and 5, the leaks are indicated by a substantial peak in the Bartlett output near $x=0$, which exceeds the background peak obtained without a leak at the same laser power. The minor misalignment of the signal peaks from $x=0$, about ± 1 mm, and the limitation on the height (~0.7) of the Bartlett peaks for the larger leaks were traced to a minor synchronization problem in the

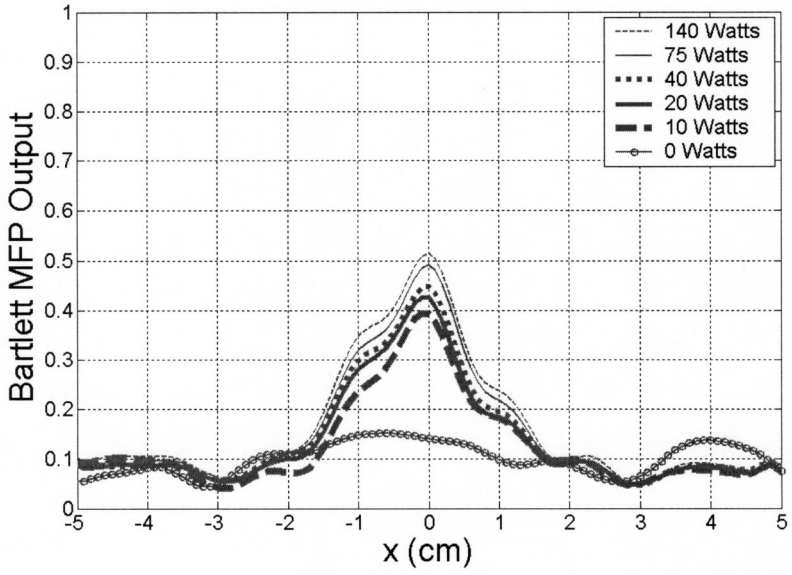


FIG. 3. Bartlett MFP output along the surface of the test cylinder versus horizontal distance for the 5.0×10^{-9} cm^3/s SF_6 leak at laser powers from zero to 140 W.

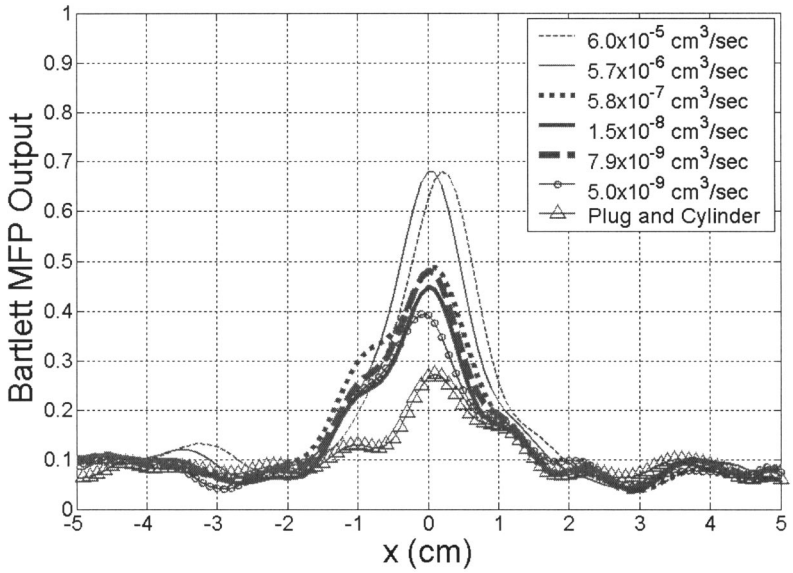


FIG. 4. Bartlett MFP output along the surface of the test cylinder versus horizontal distance for the various SF_6 leak rates at laser a laser power of 10 W. The "plug and cylinder" case represents background noise because it involves the entire experiment except no leak is present.

microphone-signal digitization system. Thus, these imperfections are artifacts of the experimental setup. Overall, Figs. 4 and 5 demonstrate that better results are achieved at higher laser powers, and that even the smallest leak of this study exceeds the background noise (the plug and cylinder case). However, Figs. 4 and 5 also show that background noise for this experiment can produce Bartlett MFP peaks of 0.28 at 10 W and 0.35 at 140 W.

Background noise peaks are caused by at least two phenomena. First, the surface of the test cylinder is mildly excited by the sweep of the hot CO_2 laser beam causing the surface to radiate coherent noise. The test cylinder is centered with respect to the microphone array so this surface noise appears as a false central peak in the Bartlett MFP output in Figs. 4 and 5 for the plug and cylinder case. Second, the rotating mirror assembly, which is centered behind the planar array, hisses at the frequencies of interest while it is in operation. So, just like the cylinder surface noise, the symmetry of this rotating-mirror sound leads to the broad but weak central peak seen in Fig. 3 for 0 W. These noise sources are also discussed in Yönak and Dowling (2002).

IV. CHARACTERISTICS AND SCALING OF PHOTOACOUSTIC SIGNALS

One of the objectives of this research was to determine or at least estimate the sensitivity limit of photoacoustic leak testing. Three features of the photoacoustic signals were examined for their dependence on volumetric leak rate: signal amplitude, signal waveform, and Bartlett MFP peak height. Although the first two signal characteristics are readily measured and amenable to theoretical analyses, the final signal characteristic proved to be the most useful for estimating the sensitivity limit of the present experiments.

A simple scaling law for the photoacoustic signal amplitude can be developed from a little perfect gas thermodynamics and the sound radiation characteristics of compact monopole sound sources. The leaks in this study are presumed to form localized gas clouds whose size is somehow determined by the leak rate and diffusive transport. The

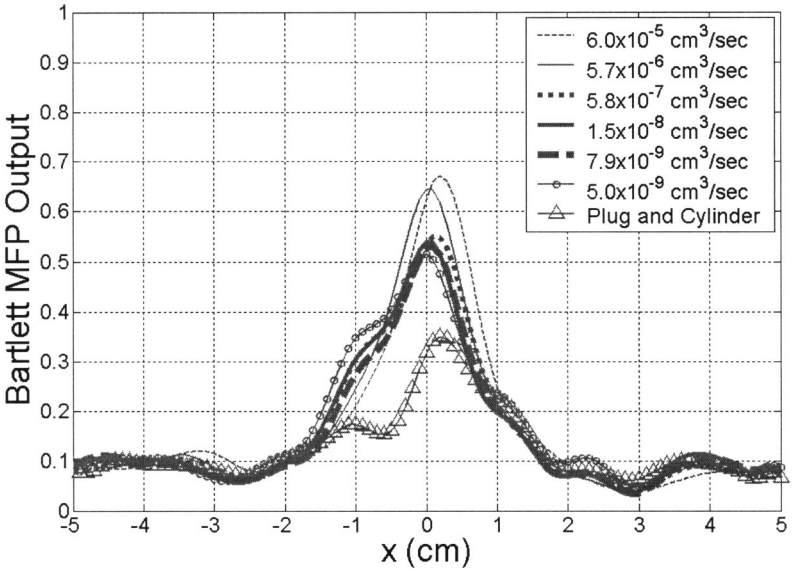


FIG. 5. Same as Fig. 4 except the laser power is 140 W.

amount of photoacoustic sound produced by a gas cloud during a leak test should depend on its size and the characteristics of the swept laser beam. Here, we merely seek to set the parametric dependencies of the sound amplitude. A more comprehensive model of photoacoustic sound production is available in Yönak and Dowling (2003).

In the experiments of this study, photoacoustic sound is generated as the laser beam passes through the tracer gas cloud with volume V, and heats the gas, causing it to expand to size $V+\Delta V$. Assuming that this volumetric expansion is small enough such that the static pressure P is nearly constant, the volume change, ΔV, of the gas cloud is represented by

$$\Delta V = \frac{mR\Delta T}{P} = \frac{mR\Delta E}{Pc_p} = \frac{\gamma - 1}{\rho c^2}\Delta E, \tag{1}$$

and thus will be proportional to the energy, $\Delta E = mc_p\Delta T$, deposited in the tracer gas by the laser beam. In (1), m is the mass of the gas in the cloud, R is the gas constant for the gas in the cloud, c_p is the specific heat at constant pressure for the gas, and γ is the ratio of c_p divided by c_v, the specific heat at constant volume. In addition, the pressure-density (ρ) form of the speed of sound, c, in a perfect gas, $c^2 = \gamma P/\rho$ where ρ is the gas density, has been used to achieve the final equality in (1).

The small volume change specified in (1) is the mechanism that launches the photoacoustic sound. In fact, the volume flux, Q, from an acoustic source is commonly considered to be a measure of its strength (see Kinsler *et al.*, 2000), and the sound amplitude, p, from a monopole is proportional to $\rho f Q/r$, where f is the sound frequency and r is the distance between the source and the listening location. For the present purposes, Q can be approximated by $\Delta V/\Delta t$, where Δt is the time it takes for the laser beam to be swept through the gas cloud. Combining these facts and assuming $f \sim 1/\Delta t$ produces

$$p \propto \frac{\rho f Q}{r} \sim \frac{\gamma - 1}{rc^2(\Delta t)^2}\Delta E. \tag{2}$$

In the present experiments, the gas clouds are much smaller than the laser beam diameter so $\Delta t = d/a$, where d is the laser beam diameter and a is the sweep rate, or speed (in m/s), of the point where the laser beam touches the test surface.

The laser energy deposited in the gas cloud, ΔE, will depend on the laser beam intensity; the length of time that the cloud is illuminated, Δt; the cross section of the gas cloud; the product of the tracer gas absorption constant, β; the laser path length in the gas cloud; and the mole fraction, X, of the tracer gas in the cloud. Here, the laser beam intensity is proportional to W/d^2, where W is the laser power, and the gas cloud cross section and laser path length are taken to be proportional to l^2 and l, respectively, where l is the characteristic size of the gas cloud. These scalings for the cloud cross section and path length presume the cloud to possess nominal spherical symmetry. Thus, the laser energy deposited in the tracer gas cloud should have the following parametric dependence:

$$\Delta E \propto \frac{W\Delta t}{d^2} \cdot l^2 \cdot \beta l X. \tag{3}$$

When (3) and $\Delta t = d/a$ are substituted into (2), a parametric scaling for the amplitude of leak-test photoacoustic sound is obtained:

$$p \propto \frac{(\gamma - 1)Wa}{rc^2 d^3} l^3 \beta X. \tag{4}$$

This scaling law still lacks one important feature: a relationship between the size of the gas cloud, l, and the volumetric leak rate, q. Such a relationship must depend on the mechanisms of tracer-gas mass transport gas away from the leak location. Clearly this might involve the local geometry of the test surface and the type of surface defect from which the tracer gas emerges—a small hole, a crack, a region of porosity, etc. If the leaking tracer gas emerges from a point, the analytic mole-fraction profile derived in Yönak and Dowling (1999) allows the gas-cloud length scale to be set by representative diffusion length. If $R_{1/2}$ is the distance from the point source, where $X = 1$, to the radius where $X = \frac{1}{2}$, then scaling $l \sim 2R_{1/2}$ means

$$l \sim 2R_{1/2} = \frac{q}{\pi D \ln(2)}, \tag{5}$$

where D is the diffusivity of the tracer gas into air ($D \sim 0.1\,\mathrm{cm}^2/\mathrm{s}$ for SF_6 diffusing into air). For the range of leak rates used in this study, (5) produces $l \sim 3\,\mu\mathrm{m}$ at $q = 6 \times 10^{-5}\,\mathrm{cm}^3/\mathrm{s}$, and $l \sim 0.2\,\mathrm{nm}$ for $q = 5 \times 10^{-9}\,\mathrm{cm}^3/\mathrm{s}$. These estimates for l are very small compared to the size of the experimental leak-cap orifice (0.8 mm) which suggests that the geometrical characteristics of the leak may not be well represented by (5). However, use of (5) in (4) leads to the prediction that $p \propto q^3$.

Unfortunately, the experimental results for photoacoustic sound amplitude do not support a cubic dependence on volumetric leak rate. The experiments covered slightly more than four orders of magnitude in q, but the measured signal amplitudes span less than half of the 12 orders of magnitude predicted by a cubic power law. This lack of agreement suggests that the scaling suggested by the combination of (4) and (5) cannot be used to estimate the sensitivity limit of photoacoustic leak testing. The predicted cubic dependence on leak rate comes from assuming that the tracer gas cloud can be characterized by a single length scale, i.e., the cloud is assumed to be nearly spherical. Thus, the lack of scaling-law–experiment agreement suggests that the leak cloud geometry is not nearly spherical but may instead be better modeled as a thin disk lying on the test surface having a planform area set by the leak orifice size with the disk thickness being set by diffusion. Additional understanding of leak cloud geometry requires further investigation and lies beyond the scope of the current effort.

By comparison with the signal amplitude, which did vary with leak rate, the photoacoustic signal waveform was nearly constant for increasing leak rate. Extracting the experimental photoacoustic waveform involved a series of steps in which much of the noise was removed from the

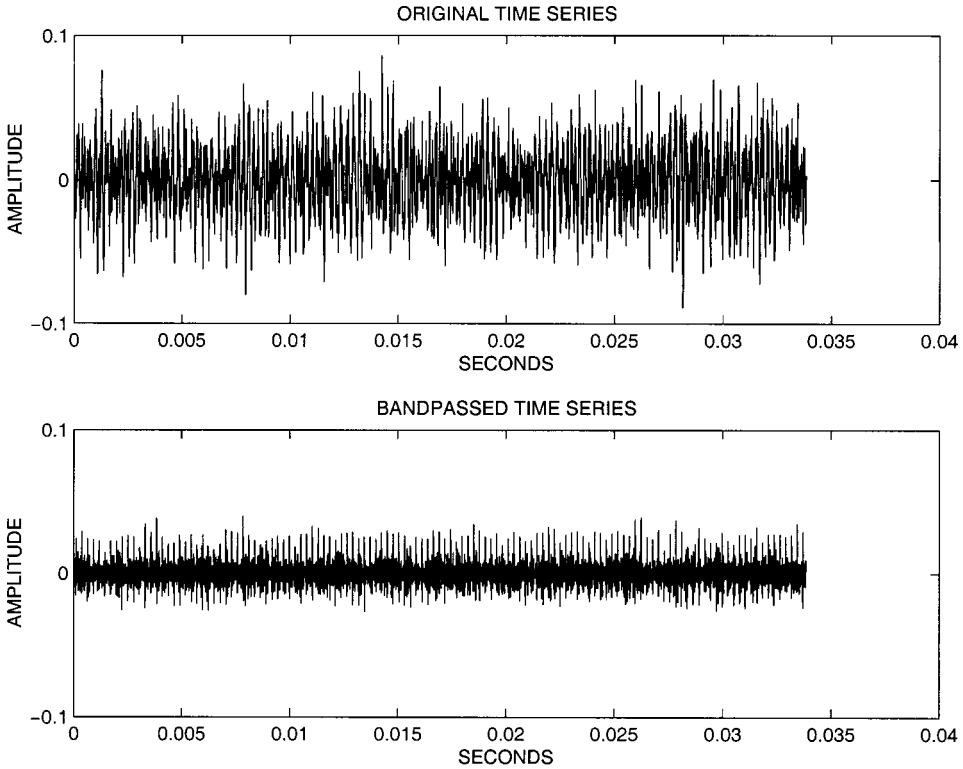


FIG. 6. Time series of measured photoacoustic data with 140 W of laser power scanning the 7.9×10^{-9} cm^3/s leak (top), and the same time series bandpass filtered between 15 and 71.25 kHz. Note the periodicity in the bandpassed time series.

measurements. These steps are illustrated in Figs. 6–8. All of these results were collected by a single microphone with the laser running at 140 W and the $q=7.9\times10^{-9}$ cm^3/s leak mounted on the test cylinder. Figure 6 shows the measured signal versus time for a 34-ms segment of data, and the same segment after bandpass filtering between 15 and 71.25 kHz. The regular periodic spikes seen in the lower panel of Fig. 6 are the photoacoustic signal pulses and four of these are shown with expanded scales in Fig. 7. The letters A, B, C, and D mark direct-path-arrival photoacoustic signal pulse peaks. These peaks actually correspond to negative acoustic pressures because the analog signal electronics of the microphones invert the signal. The temporal spacing between these peaks matches the repetition rate of the laser scan: $\frac{1}{3750}\approx0.267$ ms. The sound measured between the peaks is assumed to be a combination of noise and reflected photoa-

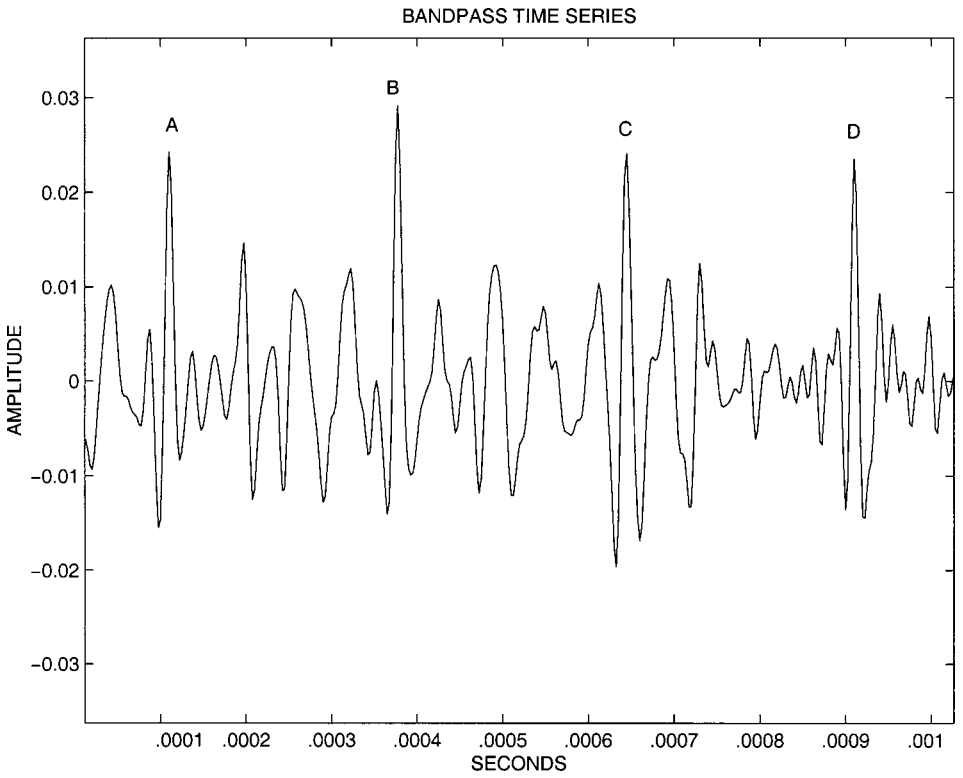


FIG. 7. Bandpass-filtered time series data from the bottom panel of Fig. 6 shown on expanded scales. The letters A–D mark four photoacoustic waveform peaks that occur in the first millisecond of the time series. The positive peaks shown above actually correspond to negative acoustic pressures because the measurements are inverted in the microphone's analog electronics.

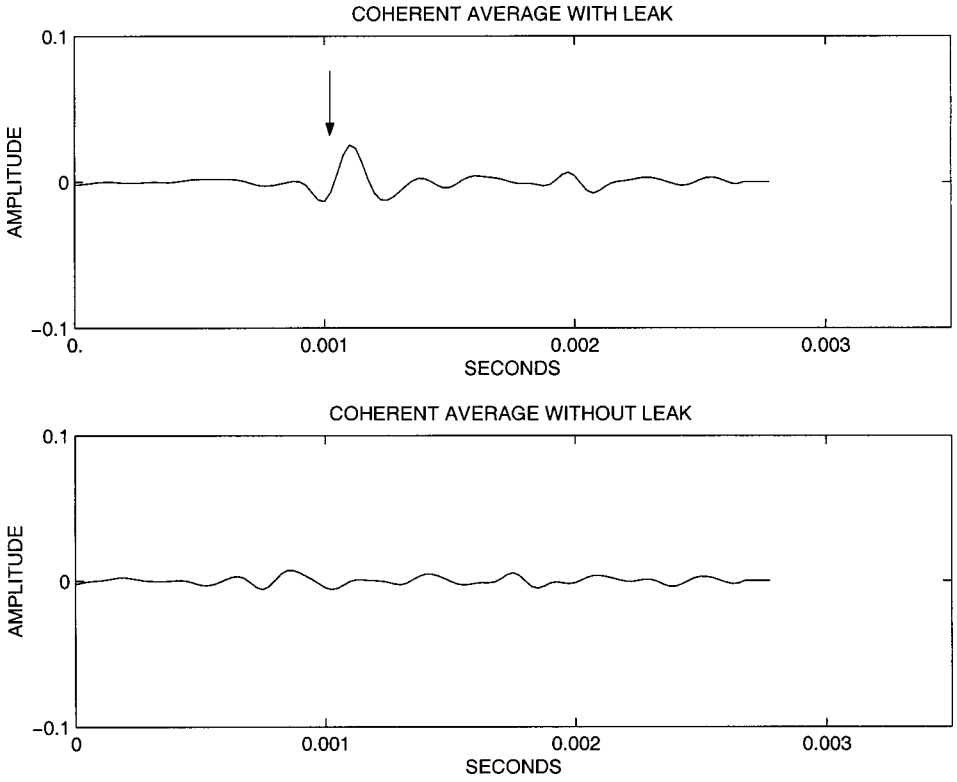


FIG. 8. (Top) Bandpass-filtered time series data from the bottom panel of Fig. 6 coherently averaged over a time window of $\frac{1}{3750}$ of a second. The above results correspond to 126 time windows. The arrow points to the direct-path photoacoustic waveform. It has the same amplitude as the four peaks in Fig. 7, thus it is coherent. (Bottom) Coherently averaged time series results as in the top panel except in this case there is no leak present. An arbitrary and unknown time shift exists between the top and bottom panel.

coustic sounds. In order to assess the extent to which the measured peak and the between-peak sounds were coherent, the measured time series was divided into windows $\frac{1}{3750}$ of a second long and averaged. A sample result of this process is shown in Fig. 8, which was constructed from the 34-ms data segment sample shown in Fig. 6 (upper panel) and an equivalent duration measurement without the leak present (lower panel). An absolute reference to the rotating prism assembly was not available so there is an unknown time shift between the two averaged waveforms in Fig. 8. A comparison of the amplitude levels in Figs. 7 and 8 shows that while the photoacoustic signal level is largely unchanged by averaging, the between-peak sound amplitude is largely suppressed by averaging. This indicates that the photoacoustic signal is coherent while most of the remaining sound (noise) is not. The between-peak sound that is not fully suppressed by averaging, and is thus coherent, is believed to be photoacoustic sound reflected from the optical tabletop and other parts of the experimental environment.

The final means attempted to estimate the sensitivity limit for photoacoustic leak detection was based on the measured MFP peak heights shown in Figs. 3–5. Here, the Bartlett output peak values for each leak at laser powers of 10 and 140 W were plotted versus leak rate and a simple line fit was passed through each set of data as shown in Fig. 9. The leak rates at which these fitted lines cross the measured background-noise-peak level for the same laser power (horizontal lines in Fig. 9) provide an estimate of the sensitivity limit of these photoacoustic leak test experiments. Straight line fits were chosen for this extrapolation task for their simplicity and presumed robustness in the presence of the data scatter shown in Fig. 9. The projected sensitivity limits, $\sim 10^{-11}$ and $\sim 10^{-13}\,\mathrm{cm}^3/\mathrm{s}\,(\mathrm{SF}_6)$ for laser powers of 10 and 140 W, respectively, are not precise and have estimated uncertainties of approximately half an order of magnitude. Even so, while these projected sensitivity limits are certainly impressive, they may yet be quite conservative because they do not account for any noise rejection capability that might be implemented in real applications of photoacoustic technology at these sensitivity levels.

V. SUMMARY AND CONCLUSIONS

A series of demonstration experiments, involving variable laser power and leak rate, have been conducted to estimate the sensitivity limit for photoacoustic leak testing. Three main conclusions can be drawn from this effort. First, the measured results spanning four orders of magnitude in leak rate suggest that the geometry of the tracer gas cloud formed near the leak is nonspherical in nature. Investigating gas cloud shapes and how they influence photoacoustic signal levels appears to be an important future research step. Second, the extrapolated results from the MFP peaks suggest that photoacoustic techniques can provide leak location information while rivaling or even possibly surpassing the sensitivity of conventional helium-mass spectrometry systems for which leak location determination is arduous. However, the implementation details for a photoacoustic leak test system intended to operate in this projected sensitivity range ($q \sim 10^{-11}$ to $10^{-13}\,\mathrm{cm}^3/\mathrm{s}$) will need careful consideration. Past experience has shown that clean test pieces and accurate acoustic models of the test geometry are needed to reject noise and ensure leak localization accuracy. However, novel means of acoustic noise rejection and good metrology may address these concerns. And for the final conclusion, the pho-

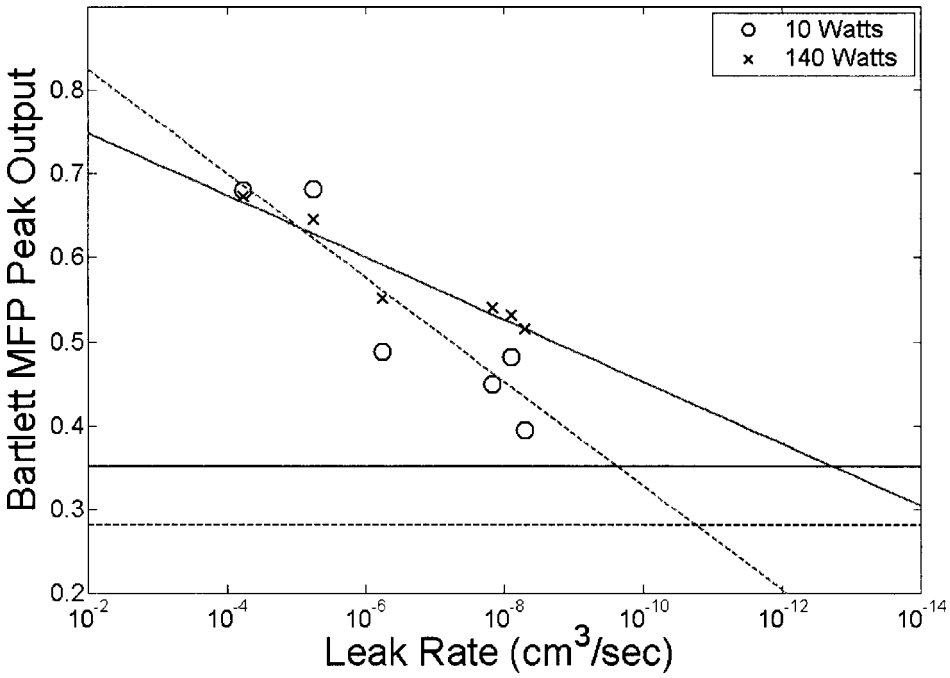


FIG. 9. Extrapolation of the results in Figs. 4 and 5 to smaller leak rates. The sloped lines are fitted to the measured Bartlett MFP peak values. The horizontal lines represent the noise-peak limits. Dashed lines are for a laser power of 10 W. Solid lines are for a laser power of 140 W. The intersection of like lines represents an estimate of the sensitivity limit of the current setup at that laser power. Noise control efforts may extend the sensitivity of the technique to even smaller leak rates.

toacoustic signals from small leaks are coherent so that coherent signal processing techniques can be used to address noise concerns.

ACKNOWLEDGMENTS

This research project was supported by the U.S. Department of Energy through the Kansas City Plant which is operated by Honeywell Federal Manufacturing and Technologies in Kansas City, Missouri. For U.S. Department of Energy under Contract No. DE-ACO4-01AL66850.

Brassington, D. J. (**1982**). ''Photo-acoustic Detection and Ranging—A New Technique for the Remote Detection of Gases,'' J. Phys. D **15**, 219–228.

Dushman, S., and Lafferty, J. M. (**1962**). *Scientific Foundations of Vacuum Technique* (Wiley, New York).

Hablanian, M. H. (**1997**). *High Vacuum Technology: A Practical Guide* (Marcel Dekker, New York).

Jensen, F. B., Kuperman, W. A., Porter, M. B., and Schmidt, H. (**1994**). *Computational Ocean Acoustics* (American Institute of Physics, New York), Chap. 10.

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**2000**). *Fundamentals of Acoustics*, 4th ed. (Wiley, New York).

McRae, T. G. (**1994**). ''Photo Acoustic Leak Location and Alarm on the Assembly Line,'' Mater. Eval. **52**, 1186–1190.

McRae, T. G., and Dewey, A. H. (**1992**). ''Photo-acoustic leak detection system and method,'' US Patent No. 5,161,408.

Olender, F. T., Woody, B. A., and Newman, L. A. (**1998a**). ''Photoacoustic leak detector with improved signal-to-noise response,'' US Patent No. 5,780,724.

Olender, F. T., Woody, B. A., and Newman, L. A. (**1998b**). ''Photoacoustic leak detector with multiple beams,'' US Patent No. 5,834,632.

Rasmussen, H. H., and Jeppesen., L. (**1998**). ''Industrial Applications of Helium Leak Test,'' NDT.net, Vol. 3, no. 12. (http://www.ndt.net/article/ecndt98/offshore/268/268.htm)

Schroff, G., and Stetter, M. (**1999**). ''Method and apparatus for detecting leaks in a container,'' US Patent No. 5,917,193.

Sharke, P. (**2000**). ''Looking for leaks in all the small spaces,'' Mech. Eng. **122**, 66–68.

Vernonesi, W. A., Olender, F. T., and Hart, R. A. (**2001**). ''Photo-acoustic leak detection system,'' U.S. Patent No. 6,327,896.

Yönak, S. H., and Dowling, D. R. (**1999**). ''Photoacoustic leak detection and localization,'' J. Acoust. Soc. Am. **105**, 2685–2694.

Yönak, S. H., and Dowling, D. R. (**2001**). ''Multiple microphone photoacoustic leak detection and localization system and method,'' US Patent No. 6,227,036.

Yönak, S. H., and Dowling, D. R. (**2002**). ''Parametric dependencies for photoacoustic leak localization,'' J. Acoust. Soc. Am. **112**, 145–155.

Yönak, S. H., and Dowling, D. R. (**2003**). ''Gas-phase generation of photoacoustic sound in an open environment,'' J. Acoust. Soc. Am. (to be published).

Passively minimizing structural sound radiation using shunted piezoelectric materials

M. Bulent Ozer and Thomas J. Royston[a)]
University of Illinois at Chicago, Chicago, Illinois 60607

(Received 25 October 2002; revised 11 July 2002; accepted 14 July 2003)

Two methods are presented to determine optimal inductance and resistance values of the shunt circuit across a piezoceramic material, which is bonded to a simply supported plate in order to minimize sound radiation from the plate. The first method (DH) makes use of den Hartog's damped vibration absorber principle. The second method (SM) uses the Sherman Morrison matrix inversion theorem. The effectiveness of each method is compared with regard to minimizing total acoustic sound-power radiation and acoustic pressure at a point. Optimization algorithms and case studies are presented using a linearized model for the piezoceramic and using a nonlinear model for the piezoceramic that accounts for the inherent dielectric hysteresis. Case studies demonstrate that the second method (SM) results in superior performance, under both linear and nonlinear system assumptions. Studies also illustrate that, if the nonlinearity in the system is significant, it must be incorporated in the optimization process. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1605390]

PACS numbers: 43.40.Dx, 43.40.Ga, 43.38.Fx [JHG]

I. INTRODUCTION

Electrically shunted piezoelectric materials have been considered for passive vibration and acoustic control of structures. If the shunted circuit has an inductance element as well as a resistance element, the electrical resonance can be utilized and an effect similar to a mechanical dynamic vibration absorber is obtained.[1–8] Although the equations of motion are different for the system with the electro-mechanical absorber versus the mechanical absorber, the optimal shunt parameters can be estimated[2,3] using den Hartog's[9] invariant points concept, which assumes a linear single-degree-of-freedom main system (without the absorber) that is undamped. However, if there is considerable damping in the main system, and/or if the mode that is to be suppressed is close to other modes of the structure (particularly with respect to resonant frequency), and/or if the inherent dielectric hysteresis in the piezoceramic is significant the shunting circuit parameters calculated using den Hartog's (the DH) method will not be optimal. In this paper a new analytical method is proposed for calculation of shunting parameters. This new method makes use of the Sherman Morrison (SM) matrix inversion. It will be shown that this method can be applied for minimization of vibration, acoustic pressure, and acoustic sound-power radiation. It is also possible to incorporate dielectric hysteresis into this optimization method, albeit resulting in an iterative optimization instead of the closed-form solution obtained for the linear problem.

In the section to follow the constitutive piezoelectric equations are reviewed. The equation of motion of a structure consisting of a simply supported plate with a bonded, electrically shunted PZT wafer are presented, as well as expressions for the radiated acoustic pressure and sound power. Application of den Hartog's method to the electro-mechanical system is discussed briefly. Then, the proposed new method is introduced and its applications to acoustic pressure and acoustic sound-power radiation are considered. Results of linear case studies are then given showing the effectiveness of the proposed method. Finally, a model for dielectric hysteresis is reviewed and integrated into the proposed optimization method.

[a)]Electronic mail: troyston@uic.edu

II. CONSTITUTIVE RELATIONSHIPS AND EQUATIONS OF MOTION

A. Constitutive piezoelectric equations

In this study a monolithic piezoceramic (PZT) will be considered. The two-dimensional piezoceramic equations will be given under the following assumptions: $T_3=D_1=D_2=E_1=E_2=0$. Here, the variables are mechanical stress T, electric displacement D, mechanical strain S, and electric field E. Numerical subscripts refer to geometric direction (1 being $+x$ direction, 2 being $+y$, and 3 being $+z$ of a right-handed coordinate system). Also, for simplicity in presentation, the case will be taken where the ''1'' and ''2'' directions have identical properties

$$T_1=\frac{E_{pz}}{1-\nu_{pz}^2}(S_1+\nu_{pz}S_2)-h_{31}D_3, \tag{1a}$$

$$T_2=\frac{E_{pz}}{1-\nu_{pz}^2}(S_2+\nu_{pz}S_1)-h_{32}D_3, \tag{1b}$$

$$E_3=-g_{31}T_1-g_{32}T_2+\beta_{33}^T D_3. \tag{1c}$$

Superscript T refer to ''at constant'' mechanical stress. Coefficients E_{pz}, ν_{pz}, h, g, and β refer to the elastic modulus, Poisson's ratio, two piezoelectric constants, and the dielectric impermeability, respectively. Note that ''one-dimensional'' formulation of the constitutive relations can be extracted

from Eqs. (1a)–(1c) with $T_2=0$ if interest is confined to the relation between variables in the 1 and 3 direction.

B. Equations of motion

The system under investigation is shown in Fig. 1. The PZT is perfectly bonded to a simply supported plate. There is a shunt *LR* (resistance *R* and inductance *L*) circuit across the PZT. This system can be excited with a mechanical point forcing $F(t)$ and/or with an external voltage $V(t)$ through the shunt circuit. For passive applications $V(t)$ equals zero.

In order to derive the equations of motion, Hamilton's energy principle is utilized

$$\int_{t_1}^{t_2}\delta(L)dt=\int_{t_1}^{t_2}(\delta K_{pl}+\delta K_{pz}-\delta U_{pl}-\delta U_{pztot}+\delta W)dt =0. \tag{2}$$

Here, K_{pl} and K_{pz} are the kinetic energy of the plate and PZT wafer, respectively. And, U_{pl} is the mechanical strain energy of the plate; U_{pztot} is the electro-mechanical energy of the PZT. Also, *W* is the total work done on the system by external mechanical and electrical forces. In the work of Ozer and Royston,[10] one can find the detailed derivation of the equations of motion, which are

$$[\mathbf{M}]\ddot{\mathbf{q}}+[\mathbf{C}]\dot{\mathbf{q}}+[\mathbf{K}]\mathbf{q}+\frac{1}{2}J_2\left(h_{31}+g_{31}\frac{E_{pz}}{1-\nu_{pz}}\right)\times\left(\int_{y_1}^{y_2}\mathbf{\Phi}_x\Big|_{x_1}^{x_2}dy+\int_{x_1}^{x_2}\mathbf{\Phi}_y\Big|_{y_1}^{y_2}\right)\frac{Q}{(x_2-x_1)(y_2-y_1)} =F(t)\mathbf{\Phi}(x_f,y_f), \tag{3}$$

$$L\ddot{Q}+R\dot{Q}+h_{pz}(2g_{31}h_{31}+\beta_{33}^T)\frac{Q}{(x_2-x_1)(y_2-y_1)}+\frac{1}{2}J_2\left(h_{31}+g_{31}\frac{E_{pz}}{1-\nu_{pz}}\right)\times\left(\int_{y_1}^{y_2}\mathbf{\Phi}_x^T\Big|_{x_1}^{x_2}dy+\int_{x_1}^{x_2}\mathbf{\Phi}_y^T\Big|_{y_1}^{y_2}dx\right)\frac{\mathbf{q}}{(x_2-x_1)(y_2-y_1)} =V(t). \tag{4}$$

Here, [**M**], [**C**], and [**K**] are the mass, damping, and stiffness matrices, respectively. The "$|_{x_1}^{x_2}$" symbol in the integration terms represents the evaluation of the function at points x_1 and x_2. Also, $\mathbf{\Phi}$ and J_2 are the mode shape vector and moment of inertia term, respectively. Superscript "*T*" on the $\mathbf{\Phi}$ term denotes the transpose of the vector. Subscripts "*x*" and "*y*" in Eqs. (3) and (4) show the differentiation with respect to *x* and *y*, respectively. Also, **q** is the modal displacement vector, *Q* is the charge on the electrical circuit, h_{pz} is the thickness of the PZT patch, *F* is the amplitude of the point harmonic forcing, and x_f and y_f are the *x* and *y* coordinate locations of the point force, respectively. The terms *L* and *R* are the inductance and the resistance of the shunt circuit, respectively. Additional parameters are defined in Fig. 1 and Sec. II A, and mathematical expressions for some of the terms are given in Eqs. (5a)–(5d). The first set of equations of motion (3) is actually composed of *N* coupled equations; *N* is the number of modes considered in the analysis. Equation (3) is dominantly a mechanical equation and the last term on the left-hand side is the electro-mechanical coupling term. The other equation of motion (4) is dominantly an electrical equation and the last term on the left-hand side is the electro-mechanical coupling term.

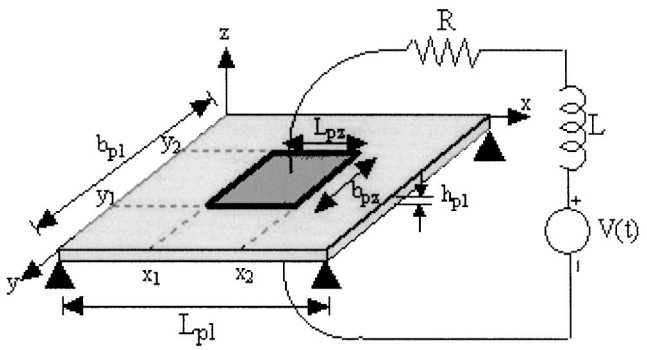


FIG. 1. Simply supported plate–PZT-shunt circuit structure.

$$\mathbf{\Phi}_{mn}=\sin\left(\frac{m\pi}{L_{pl}}x\right)\sin\left(\frac{n\pi}{b_{pl}}y\right), \tag{5a}$$

$$J_2=\frac{1}{2}\left(\left(\frac{h_{pl}}{2}+h_{pz}\right)^2-\left(\frac{h_{pl}}{2}\right)^2\right), \tag{5b}$$

$$K_{mn,kl}=\left[\left(\frac{m\pi}{L_{pl}}\right)^2+\left(\frac{n\pi}{b_{pl}}\right)^2\right]\frac{E_{pl}h_{pl}^3}{12(1-\nu_{pl}^2)}\delta(ij,kl)+\left(\frac{E_{pz}}{1-\nu_{pz}^2}J_3\right)\left[\left(\frac{m\pi}{L_{pl}}\right)^2+\left(\frac{n\pi}{b_{pl}}\right)^2\right]\times\left[\left(\frac{k\pi}{L_{pl}}\right)^2+\left(\frac{1\pi}{b_{pl}}\right)^2\right]\int_{y_1}^{y_2}\int_{x_1}^{x_2}\left(\sin\left(\frac{m\pi}{L_{pl}}x\right)\times\sin\left(\frac{n\pi}{b_{pl}}y\right)\sin\left(\frac{k\pi}{L_{pl}}x\right)\sin\left(\frac{1\pi}{b_{pl}}y\right)\right)dx\,dy, \tag{5c}$$

$$M_{mn,kl}=\tfrac{1}{4}\rho_{pl}h_{pl}L_{pl}\delta(ij,kl)+\int_{y_1}^{y_2}\int_{x_1}^{x_2}\rho_{pz}h_{pz}\sin\left(\frac{m\pi}{L_{pl}}x\right)\times\sin\left(\frac{n\pi}{b_{pl}}y\right)\sin\left(\frac{k\pi}{L_{pl}}x\right)\sin\left(\frac{1\pi}{b_{pl}}y\right)dx\,dy. \tag{5d}$$

Here, ρ_{pl}, ν_{pl}, E_{pl} are the density, Poisson's ratio, and modulus of elasticity of the plate, respectively, and ρ_{pz} is the density of the PZT. The Kronecker delta function is $\delta(ij,kl)$.

C. Equations for acoustic pressure at a point

Acoustic pressure at a point over a simply supported thin plate is given by Wang and Fuller[11] as follows:

$$p(x,y,z,t)=\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}q_{mn}p_{mn}e^{j\omega t}, \tag{6}$$

where

$$p_{mn}=\frac{-\rho\omega^2}{2\pi}\int_{-L_{pl}/2}^{L_{pl}/2}\int_{-b_{pl}/2}^{b_{pl}/2}\frac{e^{-j[(\omega/c)\sqrt{(x-\xi)^2+(y-\eta)^2+z^2}]}}{\sqrt{(x-\xi)^2+(y-\eta)^2+z^2}}\times\sin(\theta_\xi)\sin(\theta_\eta)d\xi\,d\eta, \tag{7}$$

$$\theta_\xi=\frac{m\pi}{L_{pl}}\left(\xi+\frac{L_{pl}}{2}\right), \tag{8a}$$

$$\theta_\eta=\frac{n\pi}{b_{pl}}\left(\eta+\frac{b_{pl}}{2}\right). \tag{8b}$$

Here, ρ is the surrounding medium density, ω is the circular excitation frequency, and q_{mn} is the modal displacement amplitude of the system. Calculation of the modal displacement can be done using the equations given in the previous subsection. Using Eq. (7), one can find the modal and cross-modal pressure amplitudes by numerically integrating the double integral expression.

D. Sound-power radiation

Radiated acoustic power from a simple supported plate can be calculated using

$$W=\mathbf{v}^{\mathbf{H}}\mathbf{M}\mathbf{v}, \tag{9}$$

where $\mathbf{v}$ is the complex structural modal velocity amplitudes, superscript$^{\mathrm{H}}$ denotes the Hermitian, and $\mathbf{M}$ is the real, symmetric, and positive definite radiation resistance matrix with[12]

$$M_{mm'nn'}=\frac{\omega\rho_0 mm'nn'\pi^4}{8\pi^2a^2b^2}\mathrm{Re}\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{f_{mm'}(k_xa)f_{nn'}(k_ya)dk_xdk_y}{\left(k_x^2-\left(\frac{m\pi}{a}\right)^2\right)\left(k_x^2-\left(\frac{m'\pi}{a}\right)^2\right)\left(k_y^2-\left(\frac{n\pi}{b}\right)^2\right)\left(k_y^2-\left(\frac{n'\pi}{b}\right)^2\right)\sqrt{k^2-k_x^2-k_y^2}}\right) \tag{10}$$

$$f_{mm'}(k_xa)=\begin{cases}2(1-\cos(k_xa)) & m \text{ even, } m' \text{ even}\\ 2(1+\cos(k_xa)) & m \text{ odd, } m' \text{ odd}\\ 2j\sin(k_xa) & m \text{ odd, } m' \text{ even}\\ -2j\sin(k_xa) & m \text{ even, } m' \text{ odd}\end{cases}. \tag{11}$$

Here, a is the length of the panel, b is the width of the panel, m and n are the mode numbers, and k is the wave number ($k=\omega/c$). The numerical evaluation of the above equation can be difficult because of singularities in the integrand. In the work of Elliott and Johnson[13] a different approach is taken. The plate is divided into smaller elements and each element is treated as a piston. Using the specific acoustic impedance of an elementary radiator, it is possible to form the radiation resistance matrix for an array of elemental radiators as follows:

$$\mathbf{R}=\frac{\omega^2\rho S^2}{4\pi c}\begin{bmatrix}1 & \frac{\sin(kr_{12})}{kr_{12}} & \cdots & \frac{\sin(kr_{1I})}{kr_{1I}}\\ \frac{\sin(kr_{12})}{kr_{12}} & 1 & \cdots & \cdots\\ \cdots & \cdots & \cdots & \cdots\\ \frac{\sin(kr_{1I})}{kr_{1I}} & \cdots & \cdots & 1\end{bmatrix}_{I\times I}. \tag{12}$$

Here, c is the speed of sound in the medium, k is the wave number, ρ is the medium density, S is the area of the elemental radiator, ω is the excitation frequency, and r_{ij} is the distance from the ith element to jth element. One can easily find the modal radiation resistance matrix as follows:

$$\mathbf{M}=\mathbf{\Phi}^T\mathbf{R}\mathbf{\Phi}, \tag{13}$$

where $\mathbf{\Phi}$ is an $I\times N$ matrix, which is essentially a matrix of N mode shapes of the panel for I number of radiating elements. Once the modal radiation resistance matrix is obtained, the radiated power can be calculated from Eq. (9). Figure 2 shows the radiation efficiency calculations for different modes of a plate with an aspect ratio of 0.595. The results given in this plot are in agreement with results reported elsewhere.[11,13,14]

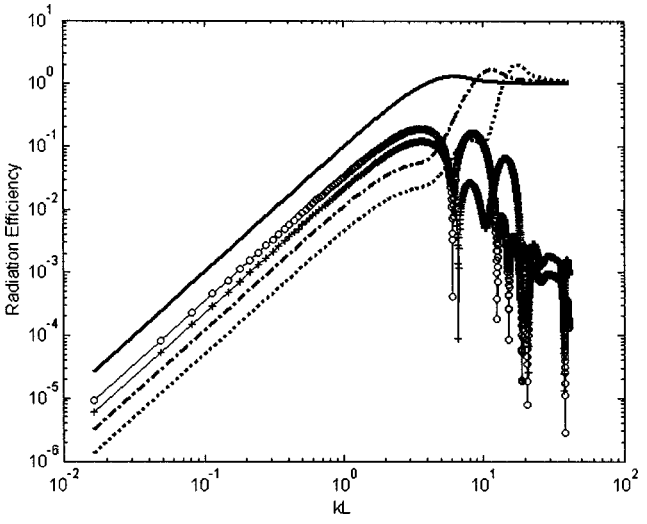


FIG. 2. Acoustic radiation efficiency plots for different modes of a simply supported plate with an aspect ratio of 0.595. Key: —— (1,1) radiation mode, —·—· (3,1) radiation mode, - - - (5,1) radiation mode, ○○○ [(1,1), (1,3)] radiation mode, +++ [(1,1), (1,5)] radiation mode.

III. OPTIMAL SHUNT PARAMETERS USING DEN HARTOG'S INVARIANT POINTS CONCEPT

den Hartog's damped vibration absorber optimization (DH) method makes use of the two points on a frequency

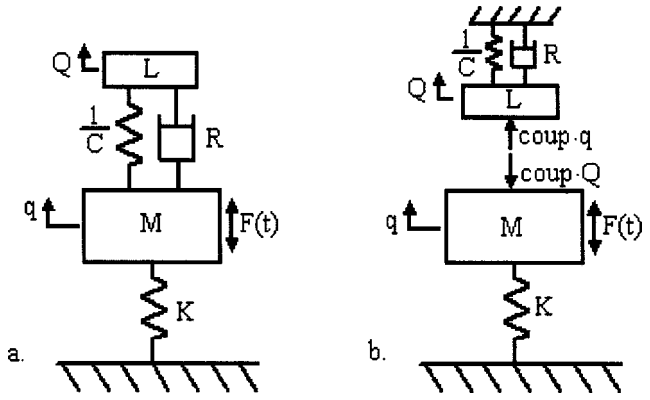


FIG. 3. Mechanical schematics of vibration absorber systems. (a) Classic two-degree-of-freedom system studied by den Hartog. (b) Two-degree-of-freedom electro-mechanical vibration absorber system.

response function (for a single-degree-of-freedom undamped main system and attached damped single degree-of-freedom absorber) whose locations are independent of the damping value of the vibration absorber. New variables will be defined and used in the equations of motion (3)–(4) to facilitate a comparison between the classic mechanical vibration absorber and the electro-mechanical vibration absorber system considered here. The equations of motion can be put in a short-hand form

$$[\mathbf{M}]\ddot{\mathbf{q}}+[\mathbf{K}]\mathbf{q}+\mathbf{coup}Q=\mathbf{F}(t), \tag{14}$$

$$L\ddot{Q}+R\dot{Q}+\frac{1}{C}Q+\mathbf{coup}^T\mathbf{q}=0, \tag{15}$$

where

$$\text{coup}=\frac{1}{2}J_2\frac{\left(h_{31}+g_{31}\frac{E_{pz}}{1-\nu_{pz}}\right)\left(\int_{y_1}^{y_2}\mathbf{\Phi}_x\Big|_{x_1}^{x_2}dy+\int_{x_1}^{x_2}\mathbf{\Phi}_y\Big|_{y_1}^{y_2}\right)}{(x_2-x_1)(y_2-y_1)}, \tag{16a}$$

$$\frac{1}{C}=\frac{h_{pz}(2g_{31}h_{31}+\beta_{33}^T)}{(x_2-x_1)(y_2-y_1)}, \tag{16b}$$

$$\mathbf{F}(t)=F\mathbf{\Phi}(x_f,y_f). \tag{16c}$$

An undamped linear two degree-of-freedom electromechanical vibration absorber system can be visualized as shown in Fig. 3(b). This system is different from the classic system shown in Fig. 3(a). Although the systems are different, there are two invariant points in the frequency response function whose locations are independent of the damping value. If the derivation procedure of den Hartog[9] is followed, one can find the conditions that would create two invariant points with equal amplitudes. This condition is given below and matches the condition derived by Tang and Wang[3]

$$\frac{\omega_{n_{\text{abs}}}}{\omega_{n_{\text{main}}}}=1, \tag{17}$$

or equivalently

$$L_{\text{opt}}=\frac{M}{K}\frac{h_{pz}(2g_{31}h_{31}+\beta_{33}^T)}{(y_2-y_1)(x_2-x_1)}, \tag{18}$$

where $\omega_{n\ \text{abs}}$ and $\omega_{n\ \text{main}}$ are the natural frequency of the absorber and main system, respectively. In order to find the optimal damping, den Hartog proposed to choose a damping value that creates a zero slope at one of the two invariant points. It is impossible for one damping value to produce zero slopes at both points. Therefore, the values that create zero slopes at each invariant point are calculated and the average of these two values is used as the optimal damping value. If that procedure is applied on the electro-mechanical system it is possible to come up with the following expression for a damping value:

$$R_{\text{opt}}=\sqrt{\frac{3}{2}\frac{T}{1-\frac{T}{2}}}\sqrt{\frac{M}{K}}\frac{h_{pz}(2g_{31}h_{31}+\beta_{33}^T)}{(x_2-x_1)(y_2-y_1)}, \tag{19}$$

where

$$T=\frac{\text{coup}^2M}{K^2L}. \tag{20}$$

This procedure for finding the optimal shunting parameters is valid not only for the simply supported-PZT-shunt structure but for any electro-mechanical system that has equations of motion in the form of Eqs. (14)–(15). However, there are a few shortcomings in this approach. It assumes linearity. It assumes an undamped main system; it is that assumption that creates invariant points. It assumes the main system has one degree of freedom, neglecting the contribution of the neighboring modes (other degrees of freedom). That is, it minimizes the response of a particular vibration mode. This approach will work for vibration minimization as long as the target mode is an isolated mode; but generally, in structural systems, particularly as one increases in frequency, the modal density increases. Additionally, if the objective is to minimize the sound power radiated or sound pressure at a point then this method may not give the best result since these are frequency-dependent weighted sums of multiple modes, with the weighting function being quite different from that which is used to construct the vibration response as a weighted sum of the modes.

IV. OPTIMAL SHUNT PARAMETERS USING SM MATRIX INVERSION METHOD

A. Setup

The Sherman Morrison[15] (SM) matrix inversion method may be utilized for the determination of passive shunt circuit parameters. Unlike the other method, this method takes into account the damping of the main structure and the effects of the neighboring modes. The SM method for matrix inversion states that

$$[\mathbf{A}+\mathbf{uv}^T]^{-1}=\mathbf{A}^{-1}-\frac{\mathbf{A}^{-1}\mathbf{uv}^T\mathbf{A}^{-1}}{1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}, \tag{21}$$

where A is $N\times N$ matrix and u and v are $N\times 1$ vectors. The above relationship will be applied to the equations of motion. However, the short-hand form of the equation of motion in Eq. (14) will be slightly modified because of the addition of the main system damping to the analysis

$$[\mathbf{M}]\ddot{\mathbf{q}}+[\mathbf{B}]\dot{\mathbf{q}}+[\mathbf{K}]\mathbf{q}+\mathbf{coup}Q=\mathbf{F}(t). \tag{22}$$

The second equation of motion will remain the same (15). Making use of Eq. (15) and considering harmonic forcing, the electrical charge is:

$$Q=\frac{-\mathbf{coup}^T\mathbf{q}}{\frac{1}{C}-\omega^2L+i\omega R}. \tag{23}$$

If one substitutes Eq. (23) into (22), the following is obtained:

$$[\mathbf{M}]\ddot{\mathbf{q}}+[\mathbf{B}]\dot{\mathbf{q}}+[\mathbf{K}]\mathbf{q}-\frac{\mathbf{coup}\ \ \mathbf{coup}^T}{\frac{1}{C}-\omega^2L+i\omega R}\mathbf{q}=\mathbf{F}. \tag{24}$$

For convenience, the following is defined:

$$E\equiv\frac{1}{C}-\omega^2L+iwR. \tag{25}$$

Using Eqs. (24)–(25), one can obtain the expression for modal displacement $\mathbf{q}$

$$\mathbf{q}=\left[[\mathbf{K}]-\omega^2[\mathbf{M}]+i\omega[\mathbf{B}]-\frac{\mathbf{coup}\ \ \mathbf{coup}^T}{E}\right]^{-1}\mathbf{F}. \tag{26}$$

The receptance matrix of the open circuit system can be written as

$$[\boldsymbol{\alpha}]=[[\mathbf{K}]-\omega^2[\mathbf{M}]+i\omega[\mathbf{B}]]^{-1}. \tag{27}$$

Now, the SM method may be applied while inverting the matrix in Eq. (26); and, if one examines Eqs. (21) and (26) the following relationships apply:

$$[\mathbf{A}]=[\boldsymbol{\alpha}]^{-1}, \tag{28a}$$

$$\mathbf{u}=\frac{\mathbf{coup}}{E}, \tag{28b}$$

$$\mathbf{v}^T=\mathbf{coup}^T. \tag{28c}$$

Looking at Eqs. (28a)–(c), it is possible to see that the expression for modal displacement $\mathbf{q}$ is actually in SM form and, if one applies the SM theorem, the following is obtained:

$$\mathbf{q}=\left[[\boldsymbol{\alpha}]+\frac{[\boldsymbol{\alpha}]\mathbf{coup}\ \ \mathbf{coup}^T[\boldsymbol{\alpha}]}{E-\mathbf{coup}^T\boldsymbol{\alpha}\ \ \mathbf{coup}}\right]\mathbf{F}. \tag{29}$$

Expanding Eq. (29), one can obtain the expression below for modal displacements:

$$\begin{bmatrix}q_1\\ \cdots\\ q_n\end{bmatrix}=\begin{bmatrix}\alpha_{11}+\frac{\sum_{i=1}^N\boldsymbol{\alpha}^{\langle 1\rangle^T}\mathbf{coupsq}^{\langle i\rangle}\alpha_{i1}}{\text{const}} & \cdots & \alpha_{1n}+\frac{\sum_{i=1}^N\boldsymbol{\alpha}^{\langle 1\rangle^T}\mathbf{coupsq}^{\langle i\rangle}\alpha_{in}}{\text{const}}\\ \cdot & \cdot & \\ \cdot & \cdot & \\ \cdot & \cdot & \\ \alpha_{n1}+\frac{\sum_{i=1}^N\boldsymbol{\alpha}^{\langle n\rangle^T}\mathbf{coupsq}^{\langle i\rangle}\alpha_{i1}}{\text{const}} & \cdots & \alpha_{nn}+\frac{\sum_{i=1}^N\boldsymbol{\alpha}^{\langle n\rangle^T}\mathbf{coupsq}^{\langle i\rangle}\alpha_{\text{in}}}{\text{const}}\end{bmatrix}\begin{bmatrix}F_1\\ \cdot\\ \cdot\\ \cdot\\ F_n\end{bmatrix}, \tag{30}$$

where superscript$^{\langle i\rangle}$ means ith column of the matrix and

$$[\mathbf{coupsq}]=\mathbf{coup}\ \ \mathbf{coup}^T, \tag{31}$$

$$\text{const}=E-\mathbf{coup}^T[\boldsymbol{\alpha}]\mathbf{coup}. \tag{32}$$

Using Eq. (30), one can write the modal displacement expression for the kth mode as follows:

$$q_k=\boldsymbol{\alpha}^{\langle k\rangle^T}\mathbf{F}+\frac{\sum_{i=1}^N\boldsymbol{\alpha}^{\langle k\rangle^T}\mathbf{coupsq}^{\langle i\rangle}\sum_{j=1}^N\alpha_{ij}F_j}{\text{const}}. \tag{33}$$

Renaming the following variables:

$$\alpha^{\langle k\rangle^T}F=q_{\text{open}_k}, \tag{34}$$

$$\sum_{j=1}^N\alpha_{ij}F_j=q_{\text{open}_i}, \tag{35}$$

where q_{open} is the open circuit modal response of the structure and inserting equations (34)–(35) into (33), the following simple expression for the kth modal response can be obtained:

$$q_k=q_{\text{open}_k}+\frac{\sum_{i=1}^N\boldsymbol{\alpha}^{\langle k\rangle^T}\mathbf{coupsq}^{\langle i\rangle}q_{\text{open}_i}}{\text{const}}. \tag{36}$$

Here, q_{open_k} is the open circuit response of the PZT–plate structure. When the PZT is open circuited, it acts like an elastic plate and there are no electrical effects of the PZT. The second term in Eq. (36) accounts for the effect of the shunt circuit and piezoelectric material on the kth modal response.

B. Shunt parameter values for minimization of acoustic pressure at a point using the SM method

The sound pressure at a point above a vibrating simply supported plate can be found using Eqs. (6)–(7). Along with Eq. (36), one can find the expression for acoustic pressure at a point as follows:

$$p(x,y,z)=p_{\text{open}}(x,y,z)+\sum_{k=1}^{N} p_k\left(\frac{\Sigma_{i=1}^{N}\boldsymbol{\alpha}^{\langle k\rangle T}\mathbf{coupsq}^{\langle i\rangle}q_{\text{open}_i}}{\text{const}}\right). \quad (37)$$

It can be seen from Eq. (37) that the sound pressure at a point is separated into the open circuit and shunted circuit contributions. If the left-hand side of Eq. (37) is set to zero and the equation is solved for the ''const'' term, one can obtain Eq. (38)

$$\text{const}=\frac{\Sigma_{k=1}^{N}p_k(\Sigma_{i=1}^{N}\boldsymbol{\alpha}^{\langle k\rangle T}\mathbf{coupsq}^{\langle i\rangle}q_{\text{open}_i})}{-p_{\text{open}}(x,y,z)}. \quad (38)$$

The above equation gives the electrical impedance expression, which will result in zero acoustical pressure at a point. Using Eqs. (25) and (32) and substituting them in (38), Eq. (39) can be obtained as below:

$$-\omega^2 L+i\omega R=\frac{\Sigma_{k=1}^{N}p_k(\Sigma_{i=1}^{N}\boldsymbol{\alpha}^{\langle k\rangle T}\mathbf{coupsq}^{\langle i\rangle}q_{\text{open}_i})}{-p_{\text{open}}(x,y,z)}+\mathbf{coup}^T\boldsymbol{\alpha}\mathbf{coup}-\frac{1}{C}. \quad (39)$$

Using Eq. (39), one can find the inductance and resistance values that result in zero pressure at a point as follows:

$$L=\text{Re}\left(\sum_{k=1}^{N} p_k\left(\frac{\Sigma_{i=1}^{N}\boldsymbol{\alpha}^{\langle k\rangle T}\mathbf{coupsq}^{\langle i\rangle}q_{\text{open}_i}}{-p_{\text{open}}}\right)-\frac{1}{C}+\mathbf{coup}^T\boldsymbol{\alpha}\mathbf{coup}\right)\cdot\frac{1}{-\omega^2}, \quad (40)$$

$$R=\text{Im}\left(\sum_{k=1}^{N} p_k\left(-\frac{\Sigma_{i=1}^{N}\boldsymbol{\alpha}^{\langle k\rangle T}\mathbf{coupsq}^{\langle i\rangle}q_{\text{open}_i}}{-p_{\text{open}}}\right)-\frac{1}{C}-\mathbf{coup}^T\boldsymbol{\alpha}\mathbf{coup}\right)\cdot\frac{1}{\omega}. \quad (41)$$

C. Shunt parameter values for minimization of acoustic sound power using SM method

In Sec. II D the procedure for calculation of the sound power radiated from a simply supported plate was discussed. In this section the expression for the inductance value of the shunt circuit, which will minimize the sound power radiated, is derived. The sound-power expression is given in Eq. (9). Since the radiation resistance matrix (**M**) is positive definite, it is not possible to set the power expression to zero by adjusting values of the shunt circuit. However, the minimum value of the energy functional [Eq. (9)] can be obtained by adjusting the shunt circuit inductance value.

The sound-power expression in Eq. (9) can be written as follows:

$$W=\sum_{j=1}^{N}\sum_{i=1}^{N} v_i^* v_j M_{ij}, \quad (42)$$

where superscript * denotes the complex conjugate and as mentioned before ''v'' is the modal velocity. Using Eq. (36), one can write the modal velocity expression as

$$v_k=v_{\text{open}_k}+\frac{(\Sigma_{i=1}^{N}\boldsymbol{\alpha}^{\langle k\rangle T}\mathbf{coupsq}^{\langle i\rangle}v_{\text{open}_i})}{\text{const}}. \quad (43)$$

Here, v_{open} denotes the modal velocity of the open circuited structure. Substituting Eq. (42) into (43) results in the following short-hand expression:

$$W=A+\frac{(\text{Re}[B])(\text{Re}[\text{const}])+(\text{Im}[B])(\text{Im}[\text{const}])+C}{(\text{Re}[\text{const}])^2+(\text{Im}[\text{const}])^2}, \quad (44)$$

where

$$A=\sum_{j=1}^{N}\sum_{i=1}^{N}(v_{\text{open}_i}^* v_{\text{open}_j})M_{ij}, \quad (45)$$

$$B=\left(\sum_{j=1}^{N}\sum_{i=1}^{N}\left(v_{\text{open}_i}^*\sum_{k=1}^{N}\boldsymbol{\alpha}^{\langle j\rangle}\mathbf{coupsq}^{\langle k\rangle}v_{\text{open}_k}\right)M_{ij}\right), \quad (46)$$

$$C=\sum_{j=1}^{N}\sum_{i=1}^{N}\left[\left(\sum_{k=1}^{N}(\boldsymbol{\alpha}^{\langle i\rangle})^*\mathbf{coupsq}^{\langle k\rangle}(v_{\text{open}_k})^*\times\sum_{m=1}^{N}\boldsymbol{\alpha}^{\langle j\rangle}\mathbf{coupsq}^{\langle m\rangle}v_{\text{open}_m}\right)\right]M_{ij}. \quad (47)$$

The aim is to find the ''L'' value, which will minimize the expression given in Eq. (44). Note that the inductance value is affecting the real part of ''const.'' In order to find the L value that makes the sound power radiated minimal, the following should be performed:

$$\frac{\partial W}{\partial(\text{Re}[\text{const}])}=0. \quad (48)$$

If the above expression is evaluated the optimal inductance is

$$L=\frac{\left[\frac{-(\text{Im}[B])(\text{Im}[\text{const}])-\sqrt{[(\text{Im}[B])(\text{Im}[\text{const}])]^2-[(\text{Re}[B])(\text{Im}[\text{const}])]^2}}{(\text{Re}[B])}+\text{Re}\left[-\frac{1}{C}+\text{coup}^T\boldsymbol{\alpha}\text{coup}\right]\right]}{-\omega^2}. \quad (49)$$

TABLE I. The geometric and forcing properties of the simply supported plate.

Plate properties		
L_{pl}	0.560	Meters
b_{pl}	0.532	Meters
h_{pl}	0.0015	Meters
E_{pl}	195	GPa
υ_{pl}	0.3	⋯
ρ_{pl}	7800	Kg/m^3
x_f	0.28	Meters
y_f	0.266	Meters
x coord of accel.	0.28	Meters
y coord of accel.	0.266	Meters
F	0.01	N
x coord of mic	0.28	Meters
y coord of mic	0.266	Meters
z coord of mic	1	Meters

V. LINEAR SYSTEM CASE STUDIES

The plate and the PZT system considered in the following numerical case studies is shown in Fig. 1, with geometric and material properties given in Tables I, II, and III. Note that the rectangular plate is approximately, but not quite a square, so that the resonant frequency of modes like (1,3) and (3,1) are very close. Also, the plate is driven at its center only exciting the acoustically efficient odd modes.

The first three case studies focus on the (1,1) mode of the plate. The inductance values that minimize the acoustic pressure at a given point and total radiated acoustic power are calculated using Eqs. (40)–(41) and (49). Acoustic pressure values in simulations are measured 1 meter above the center of the plate. Figure 4 shows the effect of using shunt parameters that are calculated with the Sherman Morrison (SM) method and the den Hartog (DH) method. It can be seen that the shunt value of resistance in the Sherman Morrison method is very close to zero. As noted in Sec. IV B, the aim of the Sherman Morrison method is to set the acoustic pressure to zero at the resonant frequency. The Sherman Morrison method results in very effective narrow-band absorption. However, broadband attenuation is desired in most applications. In the case studies to follow, the resistance value that is calculated using the DH method will be used for both methods. The other two simulation results around the first mode are shown in Figs. 5 and 6. It can be seen that

TABLE II. Geometric and electrical properties of PZT-5H.

PZT-5H properties		
L_{pz}	0.072 4	Meters
b_{pz}	0.072 4	Meters
h_{pz}	0.000 267	Meters
E_{pz}	100	GPa
υ_{pz}	0.4	⋯
ρ_{pz}	7550	Kg/m^3
x_1	0.243 8	Meters
y_1	0.098 8	Meters
x_2	0.316 2	Meters
y_2	0.171 2	Meters
h_{31}	$-1.35.10^9$	Volts/m
g_{31}	$-1.14.10^{-2}$	Volt-m/N
β_{33}	25 438 800	Volt-m/C

TABLE III. Modal properties of the PZT–simply supported plate structure.

Mode	Natural freq (Hz)	Modal damping ratio
(1,1)	24.05	0.003 7
(2,1)	58.08	0.001 5
(1,2)	62.17	0.000 8
(2,2)	95.89	0.006
(3,1)	115.29	0.000 86
(1,3)	125.19	0.001 5

compared to the response in Fig. 4 a broadband reduction is achieved. The inductance value that is calculated from the SM method gives a better performance. The first mode is isolated in terms of frequency, and the contribution of neighboring modes at its resonance is not significant. So, it is concluded that the improvement of the SM method over the DH method is due to its accounting for main system damping.

Next, minimization of acoustic pressure and sound-power radiation in the vicinity of the resonant frequency of the (3,1) mode is considered (115.29 Hz). Since (3,1) and (1,3) mode resonant frequencies are very close, both modes contribute significantly to the response at either mode's resonant frequency. It can be observed from Figs. 5–8 that the SM method further reduces the acoustic pressure and acoustic power than the DH method, primarily because the SM method takes into consideration the effect of neighboring modes in its optimization algorithm.

VI. OPTIMAL PASSIVE SHUNT PARAMETERS ACCOUNTING FOR HYSTERESIS

A. Ishlinskii hysteresis model

Especially at moderate to high forcing levels, piezoceramics exhibit hysteretic nonlinearity[16–24] and the shunting parameters that are calculated under the assumption of linearity will not be optimal. It has been shown that hysteretic behavior in piezoceramics can be accurately modeled using a

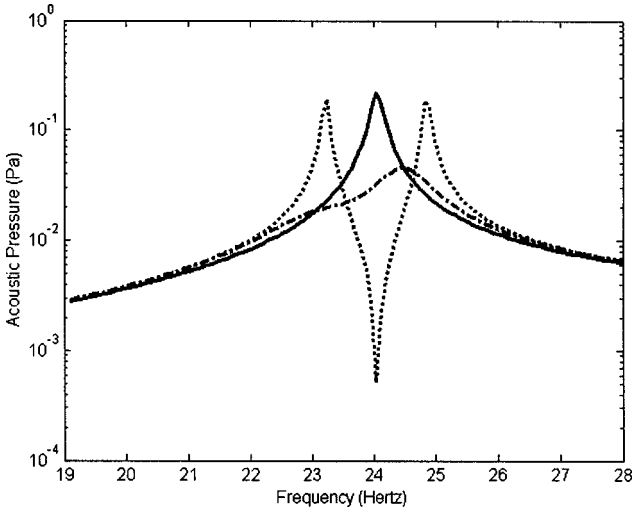


FIG. 4. Acoustic sound power radiated near (1,1) mode resonance. Key: —— open circuit configuration ($L=0,R=\infty$), —·—·— using DH principle ($L=125.23H,R=1738\Omega$), - - - using SM method ($L=119.582$, $R\approx 0$).

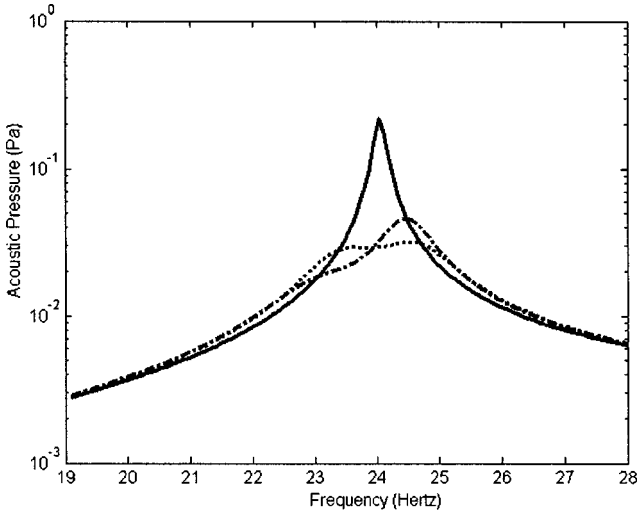


FIG. 5. Acoustic pressure near (1,1) mode resonance. Key: —— open circuit configuration ($L=0,R=\infty$), —·—· using DH principle ($L=125.23H,R=1738\Omega$), - - - using SM method ($L=119.582$, $R=1738\Omega$).

discretized Ishlinskii (elasto-slide) model.[19–22] Here, a brief review of the model is given; further details can be found in the references.

The Ishlinskii hysteresis model (IM) is schematically represented in Fig. 9 using mechanical analogies. The model is composed of $(n-1)$ massless elasto-slide elements on a surface with Coulomb friction. The input to the model is the electric displacement D and output from the model is the sum of the spring forces. For smaller amplitude inputs fewer elasto-slide elements move, resulting in less nonlinearity. Referring to Fig. 9, the model may be implemented into the otherwise linear constitutive equations as follows:

$$T_1=\frac{E_{pz}}{1-\nu_{pz}^2}(S_1+\nu_{pz}S_2)-h_{31}D_3, \tag{50a}$$

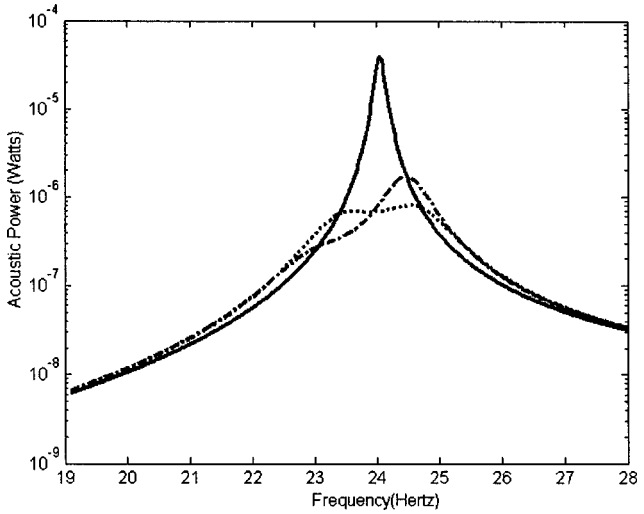


FIG. 6. Acoustic sound power radiated near (1,1) mode resonance. Key: —— open circuit configuration ($L=0,R=\infty$), —·—· using DH principle ($L=125.23H,R=1738\Omega$), - - - using SM method ($L=119.577$, $R=1738\Omega$).

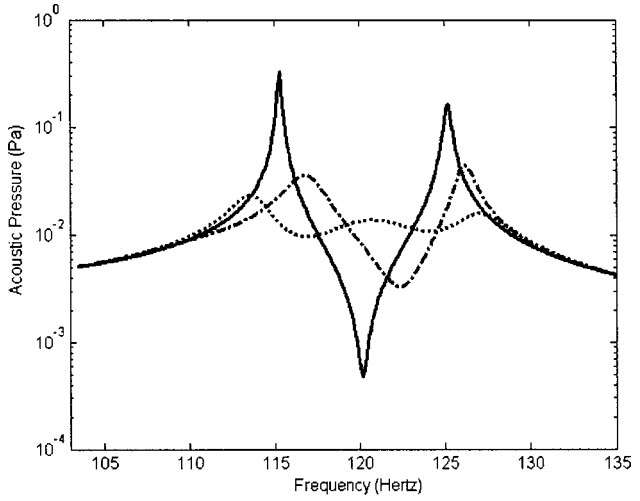


FIG. 7. Acoustic pressure near (3,1) mode resonance. Key: —— open circuit configuration ($L=0,R=\infty$), —·—· using DH principle ($L=5.45H,R=343\Omega$), - - - using SM method ($L=4.72,R=343\Omega$).

$$T_2=\frac{E_{pz}}{1-\nu_{pz}^2}(S_2+\nu_{pz}S_1)-h_{32}D_3, \tag{50b}$$

$$E_3=-g_{31}T_1-g_{32}T_2+\mathrm{IM}[\beta_{33}^T D_3], \tag{50c}$$

$$\mathrm{IM}[\beta_{33}^T D_3]=\sum_{i=1}^{n} E_{rc}^{(i)}, \tag{51}$$

where

$$e_{rc_i}=\mu e_{N_i} \tag{52a}$$

$$\text{if } |\beta_{33}^{T(i)}(D_3-D_b^{(i)})|<e_{rc}^{(i)} \text{ then } E_{rc}^{(i)}=\beta_{33}^{T(i)}(D_3-D_b^{(i)}); \tag{52b}$$

$$\text{otherwise } E_{rc}^{(i)}=e_{rc}^{(i)}\mathrm{sign}[\dot{D}_3] \tag{52c}$$

and

$$D_b^{(i)} \text{ is set such that } |\beta_{33}^{T(i)}(D_3-D_b^{(i)})|=e_{rc}^{(i)}. \tag{52d}$$

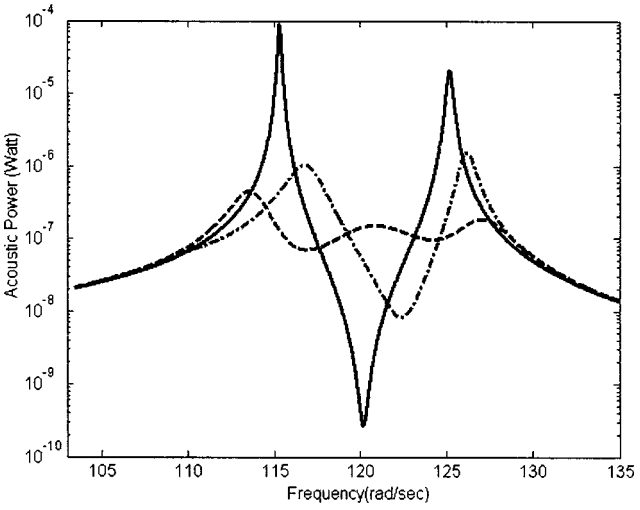


FIG. 8. Acoustic sound power radiated near (3,1) mode resonance. Key: —— open circuit configuration ($L=0,R=\infty$), —·—· using DH principle ($L=5.45H,R=343\Omega$), - - - using SM method ($L=4.70,R=343\Omega$).

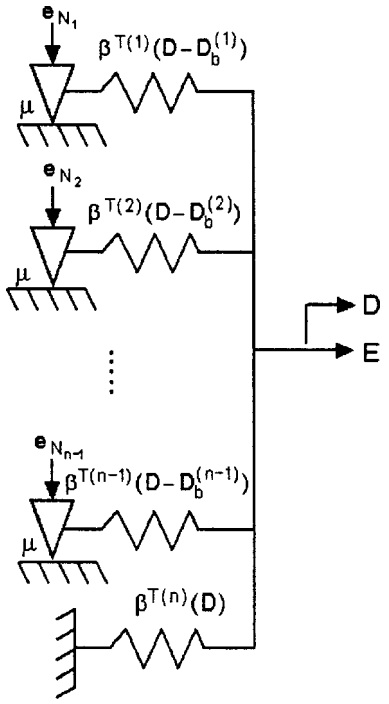


FIG. 9. The discretized Ishlinskii model (IM) for dielectric hysteresis.

Here, the terms β_{33}^{T}, e_N, μ, e_{rc}, and D_b may be viewed as electrical analogies to a mechanical spring stiffness, normal force, Coulomb friction coefficient, the force due to Coulomb friction, and the displacement from an equilibrium position of the massless box, respectively. $\beta_{33}^{T(i)}$ terms are supposed to be determined through experimental measurement of first-order reversal curves.

Figure 10 shows the effect of the increase in the number of slide elements on the hysteresis loop. The top plot shows the hysteresis curve generated using a sinusoidal input to the model with one element only. The middle plot shows the effect of addition of an elasto-slide element to the hysteresis curve. The addition of an elasto-slide element improves the hysteresis loop. However, the loop still looks discrete. At the bottom plot the hysteresis loop with eight elements is shown. The hysteresis loop is improved significantly with the additional elements. The values for β_{33}^{T}, e_n, and μ can be determined from the experimental hysteresis loop. The procedure for extracting these constants is discussed in detail by Lee and Royston.[23,24]

Using the hysteretic constitutive relationships in Eqs. (50a)–(50c), one can derive the equations of motion for the nonlinear case.[10,22] Equation (53) is a linear equation, but Eq. (54) carries the nonlinearity and these equations are coupled. The term in brackets in Eq. (54) is the nonlinear Ishlinskii hysteretic operator described above

$$[\mathbf{M}]\ddot{\mathbf{q}}+[\mathbf{C}]\dot{\mathbf{q}}+[\mathbf{K}]\mathbf{q}+\frac{1}{2}J_2\left(h_{31}+g_{31}\frac{E_{pz}}{1-\upsilon_{pz}}\right)\times\left(\int_{y_1}^{y_2}\mathbf{\Phi}_x\Big|_{x_1}^{x_2}dy+\int_{x_1}^{x_2}\mathbf{\Phi}_y\Big|_{y_1}^{y_2}dx\right)\frac{Q}{(x_2-x_1)(y_2-y_1)}=F(t)\mathbf{\Phi}(x_f,y_f), \tag{53}$$

$$L\ddot{Q}+R\dot{Q}+h_{pz}(2g_{31}h_{31})\frac{Q}{(x_2-x_1)(y_2-y_1)}+\left[\beta_{33}^{T}\frac{Q}{(x_2-x_1)(y_2-y_1)}\right]+\frac{1}{2}J_2\left(h_{31}+g_{31}\frac{E_{pz}}{1-\upsilon_{pz}}\right)\times\left(\int_{y_1}^{y_2}\mathbf{\Phi}_x^{T}\Big|_{x_1}^{x_2}dy+\int_{x_1}^{x_2}\mathbf{\Phi}_y^{T}\Big|_{y_1}^{y_2}dx\right)\frac{\mathbf{q}}{(x_2-x_1)(y_2-y_1)}=V_c(t). \tag{54}$$

B. A multiterm describing function of the IM model and nonlinear system equations

Application of the IM as a causal model in the time domain is straightforward. However, analysis in the fre-

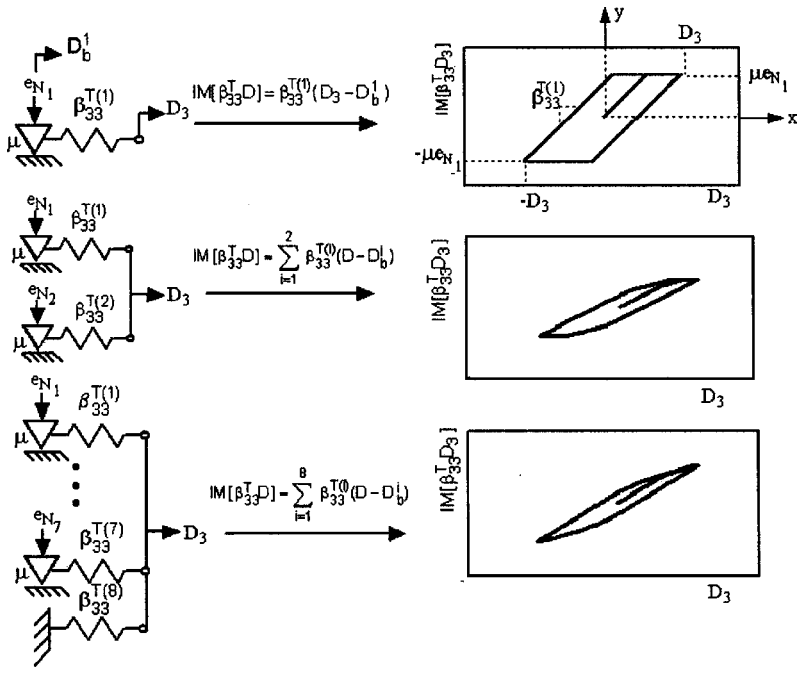


FIG. 10. Hysteresis curves with different numbers of elasto-slide elements.

quency domain can be more informative and computationally less intensive. It is common to use describing functions to approximate nonlinear functions in the frequency domain, particularly in controls,[25] as well as in structural dynamics.[26,27] Here, the sinusoidal describing functions will be used for approximating the nonlinear IM operator.

For simplicity, only one element will be considered for the derivation of the sinusoidal describing function of the nonlinearity, which is defined as in Fig. 9. The plot of force on the spring versus the sinusoidal input for a single element is given in Fig. 10. In order to derive the describing function representation of the IM model, one can make use of the describing function of the backlash type of nonlinearity, schematically depicted in Fig. 11. It can be seen that the bottom element (D_b) does not move until the backlash distance ($\mu e_N/\beta_{33}^T$) is traveled by the input (D_3). After the backlash distance is traveled, the input bar and the element move together until the input reverses its direction of motion and then motion repeats itself, as discussed before. Similar kinematic behavior is observed in the system at the top of Fig. 10. The element (D_b) does not move until the multiplication of the spring stiffness and the input (D_3) exceeds the frictional force. Once the frictional force is exceeded the input and the element move together. This continues until the direction of the input reverses. So, one can make use of this kinematic similarity and use the describing function representation of the backlash nonlinearity to find the describing function of the IM model.

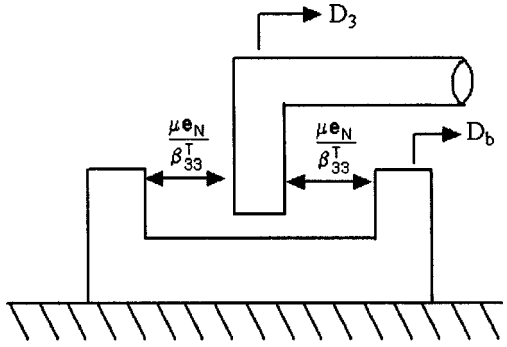


FIG. 11. Schematic representation backlash element.

The kth-order describing function of the backlash nonlinearity depicted in Fig. 11 can easily be derived and the result is as follows:

$$N_k(D_3)=\frac{j}{\pi D_3}\left(\int_0^{\pi/2}\left(D_3\sin(\psi)-\frac{\mu e_N}{\beta_{33}^T}\right)e^{-jk\psi}d\psi+\int_{\pi/2}^{\psi_1}\left(D_3-\frac{\mu e_N}{\beta_{33}^T}\right)e^{-jk\psi}d\psi\right.$$
$$+\int_{\psi_1}^{\pi}\left(D_3\sin(\psi)+\frac{\mu e_N}{\beta_{33}^T}\right)e^{-jk\psi}d\psi+\int_{\pi}^{3\pi/2}\left(D_3\sin(\psi)+\frac{\mu e_N}{\beta_{33}^T}\right)e^{-jk\psi}d\psi$$
$$\left.+\int_{3\pi/2}^{\psi_2}\left(D_3+\frac{\mu e_N}{\beta_{33}^T}\right)e^{-jk\psi}d\psi+\int_{\psi_2}^{2\pi}\left(D_3\sin(\psi)-\frac{\mu e_N}{\beta_{33}^T}\right)e^{-jk\psi}d\psi\right), \tag{55}$$

where

$$\psi_1=\pi-\arcsin\left(1-\frac{2\cdot\mu\cdot e_N/\beta_{33}^T}{D_3}\right), \tag{56a}$$

$$\psi_2=2\pi-\arcsin\left(1-\frac{2\cdot\mu\cdot e_N/\beta_{33}^T}{D_3}\right). \tag{56b}$$

The result of the analytical evaluation of Eq. (55) is quite long; however, the evaluation of the integral for the first-order describing function (for $k=1$) is given below

$$\text{For } D_3>\mu e_N/\beta_{33}^T,\quad N(D_3)=0, \tag{57a}$$

$$\text{For } D_3>\mu e_N/\beta_{33}^T,$$

$$N(D_3)=\frac{\frac{\pi}{2}+\arcsin\left(1-\frac{2\mu e_n}{\beta_{33}^T D_3}\right)+\left(1-\frac{2\mu e_n}{\beta_{33}^T D_3}\right)\sqrt{2\frac{2\mu e_n}{\beta_{33}^T D_3}-\left(\frac{2\mu e_n}{\beta_{33}^T D_3}\right)^2}}{\pi}-j\frac{2\frac{2\mu e_n}{\beta_{33}^T D_3}-\left(\frac{2\mu e_n}{\beta_{33}^T D_3}\right)^2}{\pi}. \tag{57b}$$

Using the describing function given above, the nonlinearity can be transformed into the frequency domain (in other words can be written as a function of sinusoidal functions). The accuracy of the approximation can be increased if higher-order describing functions are used ($k>1$). It is noted that the IM nonlinearity is odd and all the even-order describing functions are zero. In Fig. 12 the effect of the addition of the elasto-slide elements and higher harmonics is shown. One can use the describing function method for the solution of the nonlinear equations of motion (53)–(54) in the frequency domain.

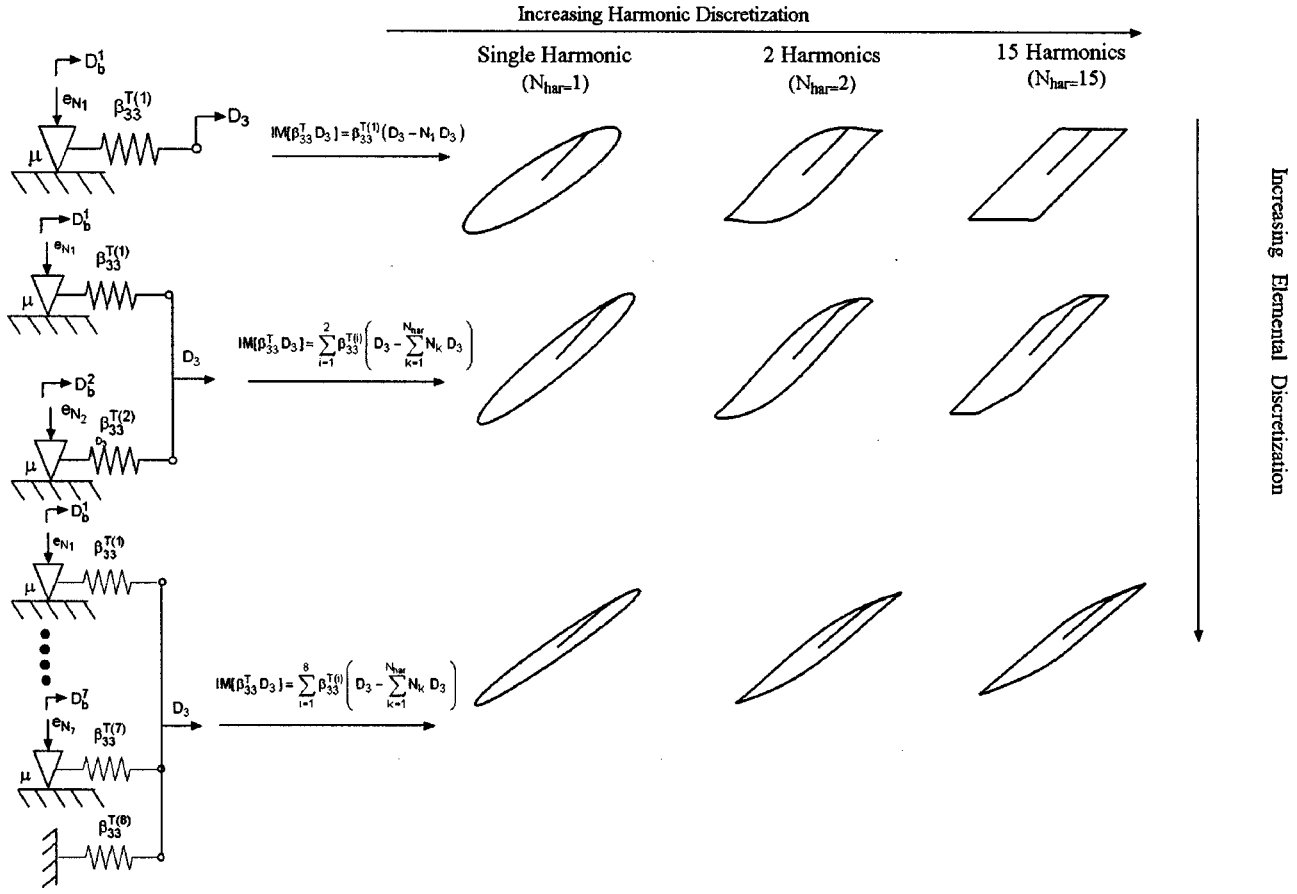


FIG. 12. Improvement of the hysteresis loop approximation with increasing number of harmonics.

C. Optimal inductance to minimize acoustic pressure accounting for hysteresis

The expression for inductance that minimizes the acoustic pressure is given in Eq. (40). However, that expression needs to be updated since the $(1/C)$ term is different in nonlinear and linear cases. Following Eq. (54), the $(1/C)$ term can be written as

$$\frac{1}{C}=h_{pz}(2g_{31}h_{31})\frac{1}{(x_2-x_1)(y_2-y_1)}+[\beta_{33}^{T}D_3]\frac{1}{Q}, \quad (58)$$

where

$$[\beta_{33}^{T}D_3]=\sum_{i=1}^{N}\beta_{33}^{T(i)}\left(D_3-\sum_{k=1}^{N_{\mathrm{har}}}N_kD_3\right). \quad (59)$$

In the above expression N determines the number of elasto-slide elements to be used and N_{har} determines the number of harmonics that is to be included in the nonlinear model. As discussed before, N_k represents the kth-order describing function. The expression for the kth-order describing function is given in Eq. (55).

Using the Eqs. (40), (58)–(59), one can write the inductance expression that will reduce the acoustic pressure (around a given mode and at a given point) radiating from a plate and a nonlinear PZT as follows:

$$L=\mathrm{Re}\left(\sum_{k=1}^{N}p_k\left(\frac{\Sigma_{i=1}^{N}\boldsymbol{\alpha}^{\langle k\rangle T}\mathrm{coups}\mathbf{q}^{\langle i\rangle}q_{\mathrm{open}_i}}{-p_{\mathrm{open}}}\right)-h_{pz}(2g_{31}h_{31})\frac{1}{(x_2-x_1)(y_2-y_1)}+\sum_{i=1}^{N}\beta_{33}^{T(i)}\left(D_3-\sum_{k=1}^{N_{\mathrm{har}}}N_kD_3\right)\frac{1}{Q}+\mathrm{coup}^{T}\boldsymbol{\alpha}\mathrm{coup}\right)\cdot\frac{1}{-\omega^2}. \quad (60)$$

It is observed that the inductance expression is a function of electrical charge. Therefore, direct calculation of its value in closed form is not possible. Iteration is required for the nonlinear system; Fig. 13 is a schematic representation of this iteration. First, an initial value for iteration is chosen (e.g., the electrical charge of the linear system). Then, output of the IM hysteresis operator is calculated, using describing functions and Eq. (60), and the first inductance estimate is determined. For resistance the value from the linear DH analysis is used. Using these inductance and resistance values one can calculate the electrical charge value by solving the nonlinear equations of motion of the plate–PZT system. The last calculated value of the electrical charge is fed back into the loop, and this is to be continued until the convergence criterion is satisfied. If the forcing level is low (i.e., nonlinearity is not dominant), the calculated inductance approaches the value obtained from the linear analysis.

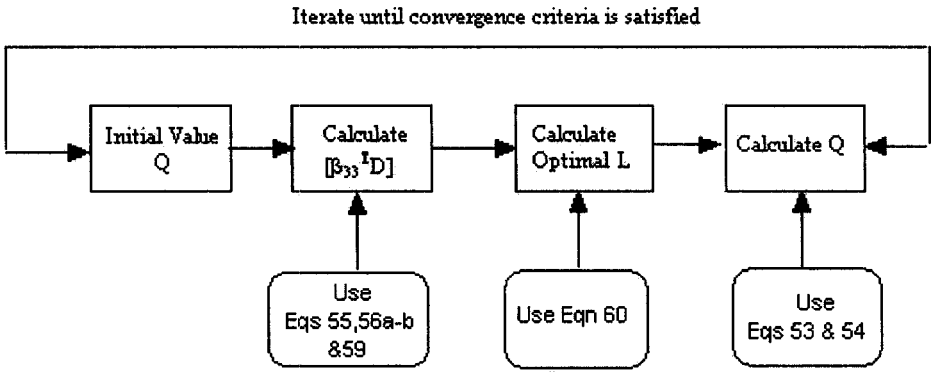


FIG. 13. The procedure for the inductor calculation with a nonlinear PZT model.

The effect of hysteresis on the inductor value is investigated via a case study. The system that will be used has the same geometry, boundary conditions, forcing, and measurement locations as in the previous case studies. However, a much larger forcing value is applied to show the effect of nonlinearity on the system and the inductance value. Table IV provides forcing information and some of the nonlinear properties used in the analysis. In the simulation five elasto-slide elements and a first-order harmonic describing function are used. It can be seen in Fig. 14 that the inductance value calculated using the nonlinear analysis results in significantly better performance than the response with the inductance value calculated using linear analyses.

VII. CONCLUSIONS

A new method for calculation of optimal shunt parameters of a passive circuit consisting of PZT, an inductor, and a resistor has been presented. The ''classic'' (DH) method, based on adapting den Hartog's invariant points principle, is incapable of accounting for the effects of neighboring (multiple) modes, damping in the main system, and nonlinearity; consequently, when any of these are significant, shunting circuit performance may be severely degraded. Additionally, the calculated shunt parameter values are not optimal in minimization of acoustic pressure or radiated sound power, which depend on weighted summations of multiple degrees of freedom.

The Sherman Morrison (SM) method introduced in this paper, which utilizes the SM matrix inversion formula, gives better estimates of optimal inductance, especially for systems with close modes and moderate to high damping. New inductance calculation procedures are introduced for minimization of acoustic pressure and acoustic power radiation. The damping value obtained from the SM method results in effective narrow-band attenuation. The damping value calculated using the DH method can be combined with the SM inductance value to obtain superior results in broadband attenuation around a given mode. The SM method provides an analytical closed-form expression for the inductance value for the linear system problem; combining this with the DH-based damping value is thus, still, a closed-form approach. A numerical search algorithm may find a more optimal damping value given an SM inductance value; however, the focus here for the linear system problem is to find an analytical closed-form alternative to the DH method, which itself is analytical.

Finally, the Ishlinskii hysteresis model of piezoceramic material, which was introduced previously by Lee and Royston,[25] is revisited and the describing function representation of it is given for efficient frequency-domain analysis. When the forcing levels are high, optimal inductance values calculated using a linear analysis result in poor performance. An iterative SM method is introduced, which calculates the optimal inductance values for the nonlinear problem. In numerical case studies it is shown that the attenuation performance of the shunting circuit is greatly enhanced with the shunting parameter estimates when nonlinearity is considered in the design.

TABLE IV. Some constants for the nonlinear system.

$\beta_{33}^{T(i)}$ values	$e_{rc}(i)$
$0.308\,68\times10^{7}$, $0.231\,88\times10^{7}$, $0.183\,56\times10^{7}$, $0.141\,80\times10^{7}$, $1.677\,96\times10^{7}$	$0.089\,96\times10^{5}$, $0.135\,15\times10^{5}$, $0.160\,48\times10^{5}$, $0.16\,529\times10^{5}$, ∞
Forcing amplitude	100 N

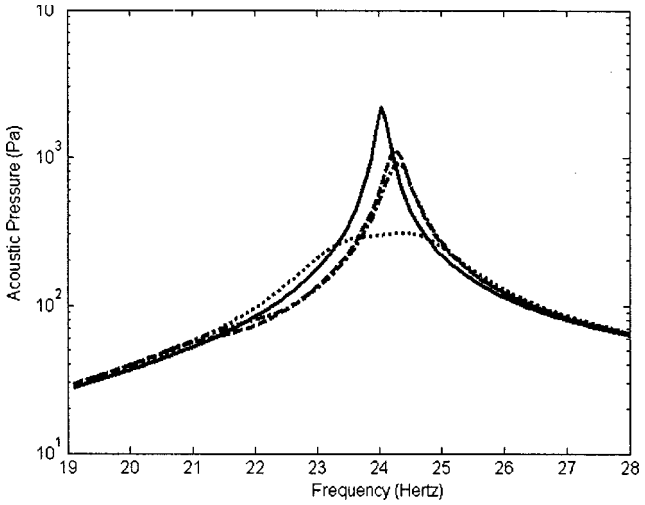


FIG. 14. Acoustic pressure near (1,1) mode resonance. Key: —— open circuit configuration ($L=0,R=\infty$), ––– using DH principle ($L=125.23H,R=1738\Omega$), — – — using SM linear method ($L=119.582,R=1738$), - - - using SM nonlinear method ($L=100.28$, $R=1738\Omega$).

ACKNOWLEDGMENTS

The financial support of the National Science Foundation (Grant No. 9733565) and the Office of Naval Research (Grant No. N00014-99-1-0342) is acknowledged.

[1] N. W. Hagood and A. von Flotow, ''Damping of structural vibrations with piezoelectric materials and passive electric networks,'' J. Sound Vib. **146**, 243–268 (1991).
[2] M. S. Tsai and K. W. Wang, ''On the structural damping characteristics of active piezoelectric actuators with passive shunt,'' J. Sound Vib. **221**, 1–22 (1998).
[3] J. Tang and K. W. Wang, ''Active-passive hybrid piezoelectric networks for vibration control: Comparison and improvement,'' Smart Mater. Struct. **10**, 794–806 (2001).
[4] G. Caruso, ''A critical analysis of electric shunt circuits employed in piezoelectric passive vibration damping,'' Smart Mater. Struct. **10**, 1059–1068 (2001).
[5] S-Y. Wu, ''Method for multiple mode piezoelectric shunting with single PZT transducer for vibration control,'' J. Intell. Mater. Syst. Struct. **9**, 991–998 (1998).
[6] L. R. Corr and W. W. Clark, ''Comparison of low-frequency piezoelectric switching shunt techniques for structural damping,'' Smart Mater. Struct. **11**, 370–376 (2002).
[7] J. Kim and J.-K. Lee, ''Broadband transmission and noise reduction of smart panels featuring piezoelectric shunt circuits and sound-absorbing material,'' J. Acoust. Soc. Am. **112**, 990–998 (2002).
[8] G. A. Lesieutre, ''Vibration damping and control using shunted piezoelectric materials,'' Shock Vib. Dig. **30**, 187–195 (1998).
[9] J. P. den Hartog, *Mechanical Vibrations*, reprint of 4th ed. (Dover, New York, 1984), pp. 87–104.
[10] M. B. Ozer and T. J. Royston, ''Effect of piezoceramic hysteresis in passive and hybrid structural control,'' in Proceedings of ASME Design Engineering Technical Conferences, DETC2001-VIB21465, 1–10, Pittsburgh, PA (2001).
[11] B.-T. Wang and C. R. Fuller, ''Near field pressure, intensity, and wavenumber distributions for active structural acoustic control of plate radiation: Theoretical analysis,'' J. Acoust. Soc. Am. **92**, 1489–1498 (1992).
[12] C. R. Fuller, S. J. Elliott, and P. A. Nelson, *Active Control of Vibration* (Academic, New York, 1996).
[13] S. J. Elliott and M. E. Johnson, ''Radiation modes and active control of sound power,'' J. Acoust. Soc. Am. **94**, 2194–2204 (1993).
[14] C. E. Wallace, ''Radiation resistance of a rectangular plate,'' J. Acoust. Soc. Am. **51**, 946–951 (1972).
[15] J. Sherman and W. J. Morrison, ''Adjustment of an inverse matrix corresponding to changes in element of a given column or a given row of the original matrix,'' Ann. Math. Stat. **21**(1), 124–127 (1950).
[16] H. Adrians, W. L. K. Koning, and R. Banning, ''Modeling the piezoelectric actuator,'' IEEE/ASME Trans. Mechatron. **5**, 331–341 (2000).
[17] H. T. Banks, A. J. Kurdila, and G. Webb, ''Identification of hysteretic control influence operators representing smart actuators. I. Formulation,'' Math. Probl. Eng. **3**, 287–328 (1997).
[18] R. C. Smith and C. L. Hom, ''A domain wall theory for ferroelectric hysteresis,'' J. Intell. Mater. Syst. Struct. **10**, 195–213 (1999).
[19] M. Goldfard and N. Celanovic, ''A lumped parameter electromechanical model for describing the nonlinear behavior of piezoelectric actuators,'' ASME J. Dyn. Syst., Meas., Control **119**, 478–485 (1997).
[20] S-H. Lee, M. B. Ozer, and T. J. Royston, ''Hysteresis models for piezoceramic transducers,'' J. Mater. Process. Manuf. Sci. **9**, 33–52 (2000).
[21] S-H. Lee, M. B. Ozer, and T. J. Royston, ''Piezoceramic hysteresis in adaptive structural vibration control problem,'' J. Intell. Mater. Syst. Struct. **13**(2), 117–124 (2002).
[22] M. B. Ozer and T. J. Royston, ''Optimal passive and hybrid control of vibration and sound radiation from linear and non-linear PZT-based smart structures,'' in Proceedings of SPIE Smart Structures and Materials Conference on Modeling, Signal Processing and Control, **4693**, 69–80, San Diego, CA (2002).
[23] S-H. Lee, T. J. Royston, and G. Friedmann, ''Modeling and compensation of hysteresis in piezoceramic transducers in vibration control,'' J. Intell. Mater. Syst. Struct. **11**, 781–790 (2000).
[24] S.-H. Lee and T. J. Royston, ''Modeling piezoceramic transducer hysteresis in structural vibration control problem,'' J. Acoust. Soc. Am. **108**(6), 2843–2855 (2000).
[25] A. Gelb and W. E. vander Velde, *Multiple Input Describing Functions and Non-linear System Design* (McGraw-Hill, New York, 1968), pp. 41–101.
[26] O. Tanrikulu, B. Kuran, H. N. Ozguven, and M. Imregun, ''Forced harmonic response analysis of non-linear structures using describing functions,'' AIAA J. **31**, 1313–1320 (1993).
[27] B. Kuran and H. N. Ozguven, ''A modal superposition method for non-linear structures,'' J. Sound Vib. **189**, 315–339 (1996).

Scale-model study of the effectiveness of highway noise barriers

Todd Busch[a] and Murray Hodgson[b]
Department of Mechanical Engineering, University of British Columbia, 3rd Floor, 2206 East Mall, Vancouver, British Columbia V6T 1Z3, Canada

Clair Wakefield
Wakefield Acoustics Ltd., 1818 Belmont Avenue, Victoria, British Columbia V8R 3Z2, Canada

(Received 28 June 2002; revised 20 June 2003; accepted 14 July 2003)

A scale-model facility was developed to test the insertion loss (IL) of highway noise barriers. Three model materials were utilized to simulate packed-earth berms and ground (expanded polystyrene), vertical walls (dense polystyrene), and roadways (varnished particleboard). Thirty-eight noise-barrier configurations were tested and used to compare how IL varied with changes to the barrier profile for walls, berms, and combinations of walls and berms for receivers at a representative, highway-adjacent location. The atmospheric conditions were assumed to be homogeneous and nonrefracting. Changes of barrier surface impedance were also assessed. A highway line source was simulated by positioning both an air-jet point source and a receiver microphone at a series of equally spaced points, in order to form an array of source-receiver measurement pairs making differing angles of propagation to the noise-barrier crest line. The IL measurement results are presented in unweighted third-octave bands. In addition, total A-weighted insertion losses (ILA) were obtained by applying an A-weighted, traffic-noise spectrum. When a berm was modeled with surface impedance closely matching that of packed earth, it was found that walls outperformed berms by 1 to 2 dBA. When the surface impedance of a berm was modeled to be acoustically soft, the ILA increased sufficiently to favor berms by about 2 dBA. The result for an acoustically soft berm does not support the long-standing practice of assuming that earth berms outperform walls by 3 dBA, but is consistent with the performance predicted by newer prediction algorithms. When the slopes of berms were made shallower, the IL generally decreased for a berm alone, but generally increased in cases with a wall atop the berm. [DOI: 10.1121/1.1605412]

PACS numbers: 43.50.Gf, 43.50.Lj, 43.50.Vt [DKW]

I. INTRODUCTION

Highway noise mitigation is increasingly achieved using highway noise barriers, commonly constructed using concrete, wood, metal, or plastic in the case of walls, and earth in the case of berms. For the work reported here—part of a scale-modeling study of highway noise barriers[1]—an acoustical scale-modeling facility was developed to study noise-barrier performance. The objective was to determine how changes to noise-barrier profile, height, and surface impedance affect barrier noise reduction at a representative, highway-adjacent receiver location. The noise barriers tested included walls, earth berms, and earth berms crested by a wall. The atmospheric conditions were assumed to be homogeneous and nonrefracting.

In British Columbia, for aesthetic reasons, the heights of walls adjacent to provincial highways are restricted to 3-m height—even when constructed atop a substantial berm.[2] Earth berms (with or without a crest wall of limited height) have no aesthetic height limit. Thus, it was of interest to compare the noise reduction of walls, berms, and berm/wall combinations. Furthermore, field tests have indicated that a soft-top correction of 2 dBA favoring earth berms is warranted in comparison to a wall of the same height. In practice, a 3-dBA correction has been applied in highway noise-modeling procedures[3] and derivative highway noise-prediction software, such as the U.S. Federal Highway Administration (FHWA) STAMINA model. The validity of this assumption merited evaluation, since there are a number of publications which report results that do not support the application of such an extreme correction and, in some cases, conclude the opposite, in favor of walls. A recently updated FHWA guidance manual[4] asserts that ''...a berm barrier will typically provide an extra 1 to 3 dBA of attenuation...'' over a wall of comparable height at the same offset location.

Noise reductions by barriers are quantified by the insertion loss (IL)—the difference in the sound-pressure levels before and after the barrier is put in place. A positive IL indicates that a noise reduction has occurred; a negative value indicates an increase of noise levels due to the addition of the barrier. IL measurement results are presented here in unweighted third-octave bands, and further references to IL in this paper refer to these results. In addition, the total A-weighted insertion losses (ILA) were obtained by applying the measured IL to an assumed A-weighted, traffic-noise spectrum before calculating the total levels with and without the barrier in place. IL varies with frequency, source/receiver geometry, ground-surface impedance, and barrier configura-

[a]Current address: Acentech, 1429 E. Thousand Oaks Blvd., Suite 200, Thousand Oaks, CA 91362.
[b]Author to whom correspondence should be addressed. Electronic mail: hodgson@mech.ubc.ca

tion. Another factor is whether the sound source effectively acts as either a point source (e.g., a single vehicle) or a line source (e.g., a continuous line of vehicles). Measurements of IL using a point source only consider a single cross-section of a barrier's profile; the cross-section tested is usually perpendicular to the barrier's centerline where the measured IL values are generally highest, leading to a higher ILA. For instance, at an angle of 63.5° to the perpendicular, an energy-based calculation indicates a 3-dB reduction in IL relative to that for the perpendicular cross section. For a line source, the IL is typically lower for any given cross section not perpendicular to the barrier crest, so the result of integration to calculate the line-source levels is an average IL that is lower in value than that for the perpendicular test.

As discussed fully in Ref. 1, the design and construction of the scale model assumed that the geometric dimensions are reduced by an assumed scale factor, while test frequencies are increased in proportion to that scale factor. Scale modeling required the design of an ultrasonic air-jet noise source, the selection of suitable test equipment, of optimal model scale-factor and model materials, and the specification of test configurations. The scale-model tests were carried out in an anechoic chamber, which was determined to be sufficiently anechoic over the ultrasonic test-frequency range used to measure IL.

II. LITERATURE REVIEW

Numerous studies have been undertaken on the noise reduction provided by highway barriers—using theoretical prediction, full-scale testing, and scale-model testing. The limits of noise-barrier performance are dictated by the extent to which sound diffracts into regions shielded from view by the barrier. Factors considered as part of the efforts to maximize noise-barrier IL and cost-effectiveness have included barrier profiles, dimensions and surface impedance, and the attenuation provided by ground surfaces with/without a barrier present.

Watts[5,6] made a practical and thorough review, with particular emphasis on the effects of barrier absorption and profile. Hayek[7] complemented an overview of barrier-performance theoretical issues with a summary of previous noise-barrier work, by categorizing publications with respect to the boundary conditions (presence/absence of ground, finite/infinite surface impedance) and barrier profiles considered.

Scale modeling has been utilized previously to study outdoor sound propagation and the subsidiary issue of highway noise control using barriers. Modeling has employed scales ranging from 1:6 to 1:100. Scale-model sound sources have included air-jets, electrical discharge (spark), and ultrasonic transducers. Scale-model-material selection has depended on the model's application and the chosen scale factor.

Delany, Rennie, and Collins[8] developed a 1:30 scale-modeling facility to evaluate predictive accuracy through comparisons to field measurements of traffic-noise propagation. They modeled roadways using aluminum and rigid plastic, absorptive ground using 11-mm-thick Insulite boards covered by a coarse weave nylon cloth, and buildings using the rough side of 3-mm-thick standard hardboard.

Jones, Stredulinsky, and Vermeulen[9] developed a scale-modeling facility to study urban noise propagation. They established that 1:80 scale was effective when using air as a scale-modeling medium. They utilized aluminum sheet to model roadways and wooden blocks to model buildings, and adopted sanded polystyrene with tissue paper glued to its surface to model grass. For 11 different source-receiver configurations the comparisons of full-scale to model-scale results were reported to show a rms deviation of 1.4 dBA.

Osman[10] studied scale-model ground materials for use at 1:16, 1:32, and 1:64 scale based on the attenuation of total, A-weighted sound-pressure level with distance. An impulsive spark source was used. Fiberboard was recommended to represent soft ground at 1:16 and 1:32 scales. A felt-material ''combination'' was recommended to represent soft ground at 1:64 scale. Vinyl sheet with a thickness of 10 mm was recommended for hard ground at all three scales. The author also summarizes the materials utilized in seven previous scale-model studies with reference to the materials used to simulate hard or soft ground, building surfaces, barriers, trees, and intervening vehicles.

The U.S. Federal Highway Administration (FHWA) guidelines past[3] and present[4] both reference field tests and analysis[11] that indicate that earth berms outperform walls by 2.4 dBA, based on linear-regression analysis of tests of seven wall installations and three berm installations. Berm/wall combinations were found to outperform walls by 1.3 dBA. Past FHWA design guidelines[3] recommended a ''soft-top correction'' of 3 dBA in favor of berms over a wall of the same height. Revised FHWA design guidelines[4] recommend a 1 to 3 dBA correction favoring berms.

Hajek[12] conducted model tests at 1:16 scale to compare walls, flat-top berms, and wedges with a crest wall. For 3- and 4.9-m-high barriers, the ''grass'' berm had a total ILA of about 1 dBA lower than thin walls of equal heights. For 3-m-high barriers, an unrealistically ''absorptive'' berm with a urethane top improved upon ''grass'' berm performance by about 2 to 3 dBA (1 to 2 dBA better than a wall of the same height). The addition of a wall of 0 to 1.9 m height to the top of a 3-m-high ''grass'' or ''absorptive'' berms had differing effects. Increased crest wall height from 0 to 1.9 m increased the total ILA for the ''grass'' berm but reduced the total ILA for the ''absorptive'' berm for wall heights from 0 to 1.2 m. For crest wall heights of 1.2 to 1.9 m, the performance of both types of berms was essentially identical and similar to that of a thin wall without a berm, respectively.

May and Osman[13] used 1:16-scale modeling and the materials described in Ref. 10 to measure the total, A-weighted IL for barriers of varying profiles. The barriers were assumed to be located next to a six-lane, divided highway with a representative vehicle mix. For 4.9-m-high barriers, the vertical reference wall had the lowest total ILA. T-top barriers of 2.4-m width increased total ILA by 4 dBA, whereas rectangular barriers of 2.4-m width increased total ILA by only 3 dBA. Absorptive tops increased total ILA for these two profiles by about 2 dBA.

Hutchins, Jones, and Russell[14] employed 1:80-scale

modeling to investigate changes to the IL as a function of frequency. Thin, highly reflective barriers were tested in the presence of both model asphalt- and grass-covered surfaces. High impedance asphalt was modeled using sheet aluminum. Grass was modeled using polystyrene covered by a layer of tissue. Barriers were modeled using 1-mm-thick aluminum sheet. The IL was strongly dependent on the ground surface. The IL of a barrier was typically a minimum for frequencies centered near 450 Hz.

Hutchins, Jones, and Russell[15] studied how different profiles of a 4.9-m-high barrier affect barrier IL for grass- and asphalt-covered ground. Source/receiver heights were 1.2 m, and the source and receiver distances to the barrier were 12.2 m. Changes in barrier profile exhibited greater variability over an asphalt surface. Increases in the effective barrier top width were beneficial in all cases. A T-shaped wall on a berm was among the most effective profiles.

Maekawa and Osaki[16] compared theoretically the reductions of sound by semi-infinite screens and wedge-shaped barriers. Compared to a thin barrier, the wedge IL was lower for all ranges of wedge angle and source and receiver positions. A progressively shallower wedge had progressively reduced insertion loss; up to 6 dB higher IL was predicted for the semi-infinite screen compared to a semi-infinite wedge.

Cremers, Fyfe, and Cremers[17] utilized 2-D boundary-element methods to compare the performance of T-shaped, semi-circular, and wedge-shaped barriers on reflective ground with that of a vertical wall. At a frequency of 500 Hz, the IL performance was found to be, respectively, 1.9 dB more effective, 1.4 dB less effective, and 2.6-dB less effective than a vertical wall.

Hothersall, Crombie, and Chandler-Wilde[18] utilized boundary-element methods to implement a theoretical comparison of T-shaped barriers to flat-top mounds, and arrow-shaped barriers to wedges. Results for solid barriers indicate a 1- to 2-dBA ILA degradation with the addition of sloped surfaces to the sides of either a T-shaped barrier or an arrow-shaped barrier.

III. EXPERIMENTATION

Scale modeling of outdoor ground surfaces and noise barriers at 1:31.5 scale was performed on a platform in an anechoic chamber. The chamber was anechoic over the test frequency range, up to 100 kHz, as long as source and receiver positions were not within 0.5 m of the chamber's surface.[1] The chamber's dimensions were 4.0 m by 4.7 m, with a height of 2.6 m. These dimensions correspond to full-scale equivalent dimensions of 126 mFS by 148 mFS by 82 mFS (throughout this article, 'FS' appended to a unit indicates the full-scale-equivalent value, as opposed to the model-scale value). To select scale-model materials, an improved procedure was developed for optimizing the choice of scale factor and associated scale-model materials, as described elsewhere.[19]

A. Model sound source

The ultrasonic noise source used was a crossing air-jet source patterned closely on the design by Novak.[20] As described more fully in Ref. 1, a design with a 6.5-mm- (20-cmFS-) diameter cavity was used, with six equally spaced, co-planar air-jets of 0.35-mm diameter directed inwards towards a common central point. Novak reported that the sound-power output increased, and the spectral shape smoothed out, as air-supply pressures increased to 500 kPa, without changing the directivity appreciably.

The air-jet was connected to a pressurized air supply having gauge pressures from 612 to 748 kPa, via 6-mm brass tubing, a moisture and particulate filter, a globe valve, a pressure regulator and flexible air hose. A pressure gauge with an operable range from 0 to 700 kPa (± 5 kPa) was used to monitor the regulated air-supply pressure. Tests of the source's radiation characteristics showed that the total sound-power output increased by 6 dB per doubling of air-supply pressure, and that the best combination of sound power and reproducibility of output was achieved at an air-supply pressure of 500 kPa.

B. Scale-model materials

Three model materials were selected[19] to simulate packed-earth berms and ground (effective flow resistivities of 4000 to 8000 $\mathrm{kN\cdot s\cdot m^{-4}}$),[21] vertical walls or roadways ($20\,000\ \mathrm{kN\cdot s\cdot m^{-4}}$), and an atypically soft surface such as a forest floor (20 to 80 $\mathrm{kN\cdot s\cdot m^{-4}}$). Roadways were simulated using varnished particleboard, walls using dense polystyrene, and packed-earth berms and ground using expanded polystyrene. The impedance of berm surfaces was made harder or softer using dense polystyrene or felt, respectively. Tests of barrier IL were conducted by aligning the air-jet source such that the air stream exited the source cavity horizontally; the source was oriented perpendicular to the direction of sound propagation for each individual source to receiver test.

C. Measurement equipment

A $\frac{1}{8}$-in. Brüel & Kjær Type 4138 microphone and Type 2619 preamplifier, powered by a Type 2804 Microphone Power Supply, were used for all measurements; the microphone was placed at a receiver height of 1.8 mFS. Measured signals were conditioned using a Brüel & Kjær Type 2607 Measuring Amplifier, and were sampled and analyzed using a Stanford Research Systems SR-770 FFT Network Analyzer. Spectral data were recorded in 400, 250-Hz bandwidth frequency bins, extending from 0 to 100 kHz. Individual tests employed 1500 nonoverlapping spectral averages, requiring 5 s per test. The fixed-bandwidth measurements were integrated into proportional third-octave bands with center frequencies from 80 to 2500 HzFS.

D. Line-source insertion-loss tests

Figure 1 shows a cross-section of the scale-model highway configuration modeled—that of a four-lane divided highway. The dimensions of the roadway were as follows: highway elevation=1 mFS; width of median=2.6 mFS; lane width=3.7 mFS; width of shoulder=3 mFS. A shoulder slope of 1:4 was utilized (the slope convention vertical rise: horizontal run is used here). The air-jet source was placed at a nominal height of 0.5 mFS, with the effective full-scale diameter of the air-jet source distributing the noise source over a vertical range of 0.4 to 0.6 mFS. This source height

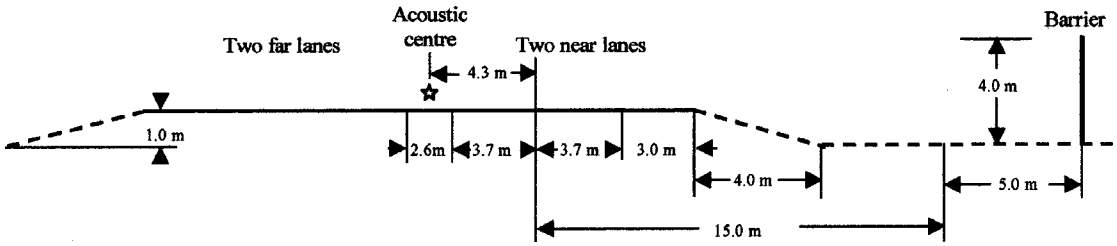


FIG. 1. Profile view of the scale-model-highway geometry (all dimensions are full-scale equivalent values).

was consistent with local practice for the mixture of heavy vehicles and speeds observed when measuring a traffic-noise spectrum, as discussed below.

Figure 2 shows plan and elevation views of the roadway, barrier, source, and receiver positions. The dividing line between the two near lanes of traffic was chosen to be 20 mFS from the barrier crest, so that there would be sufficient space for the shallow 1:3-slope berms. The acoustical center was estimated to be 4.3 mFS beyond this position. The microphone was located at 15 mFS beyond the berm crest. This distance is representative of first-row receiver locations that are the primary beneficiaries of noise barriers. The source and receiver rows consisted of points spaced apart by a convenient distance of 25.0 cm—a linear spacing equivalent to 7.88 mFS, and representative of very closely spaced vehicles traveling along a highway at a constant speed.

The minimum distance from the barrier ends required to minimize the contributions of end diffraction was identified by making sound-pressure-level measurements, with a barrier in place, at all perpendicular source/receiver positions noted in Fig. 2 (S1-R1,...,S15-R15). After examining the associated degradation of barrier performance, further tests of barrier IL were not performed at the latter two positions at either end.

To determine the IL for each individual source-receiver pair, sound-pressure levels were measured at 11 positions along a scale-model highway, with and without a noise barrier. The entire process was repeated with the source traversed in the opposite direction. For both the with-barrier and without-barrier measurements, the average and standard deviations were calculated for each equivalent source/receiver pair (e.g., S9-R4 and S7-R12). The resulting average sound-pressure levels were used to reconstruct the sound-pressure levels for a source length equivalent to 21 such source/receiver pairs which, when viewed from the receiver position, spanned a 127° angle of highway. This coverage is sufficient to generate stable line-source sound-pressure levels for both the open-ground and with-barrier tests. Outside of this angle of coverage, the combined reduction of noise levels between a road segment and a receiver, due to divergence and source-receiver distance, are estimated to be on the order of 10 dB or more in excess of the combined reduction for the perpendicular case.

For the calculation of IL, no corrections for air absorption were applied. The path-length differences involved in the open-ground and the with-barrier tests were not large enough over the ultrasonic test frequency range to skew the results. Also, since IL is calculated by subtraction of with-barrier from open-ground results, the increased attenuation at ultrasonic frequencies due to air absorption effectively cancels out, with the net effect calculated to be on the order of 0.1 dB.

The ILA was calculated after applying an A-weighted traffic-noise spectrum, as shown in Fig. 3, to the line-source pressure levels for both the open-ground and with-barrier measurements, summing each energetically, and taking the difference. The assumed spectrum was based on field measurements of traffic, comprised of automobiles, 14% to 18% medium trucks, and 5% to 7% heavy trucks, traveling at speeds of 75 to 80 km/h. The spectrum was essentially flat up to 200 HzFS, increased by about 18 dB to a maximum at 1000 HzFS, then decreased by about 4 dB to the highest test frequency of 2500 HzFS.

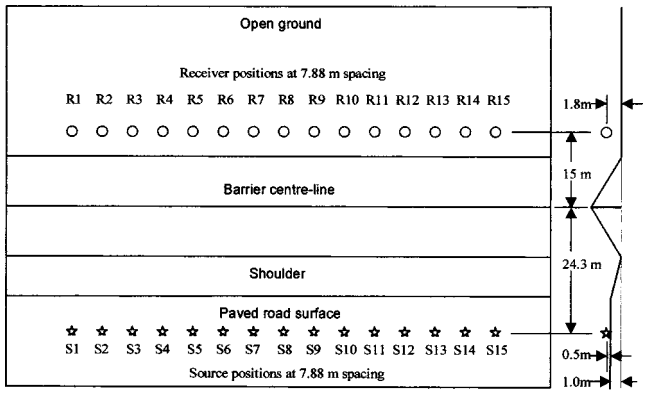


FIG. 2. Plan view of the scale-model-highway geometry (all dimensions are full-scale equivalent values).

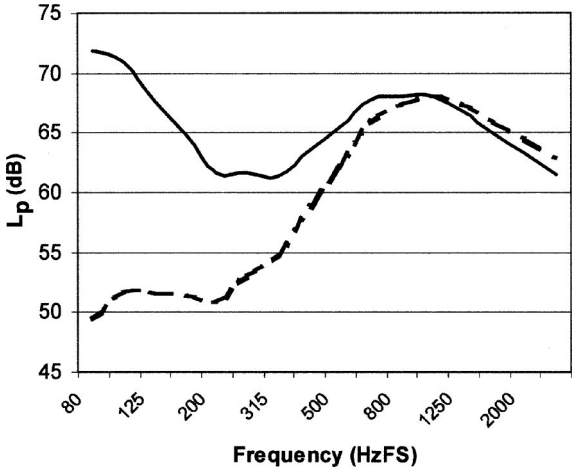


FIG. 3. Representative traffic-noise spectrum used in ILA calculation: (———) unweighted; (– –) A-weighted.

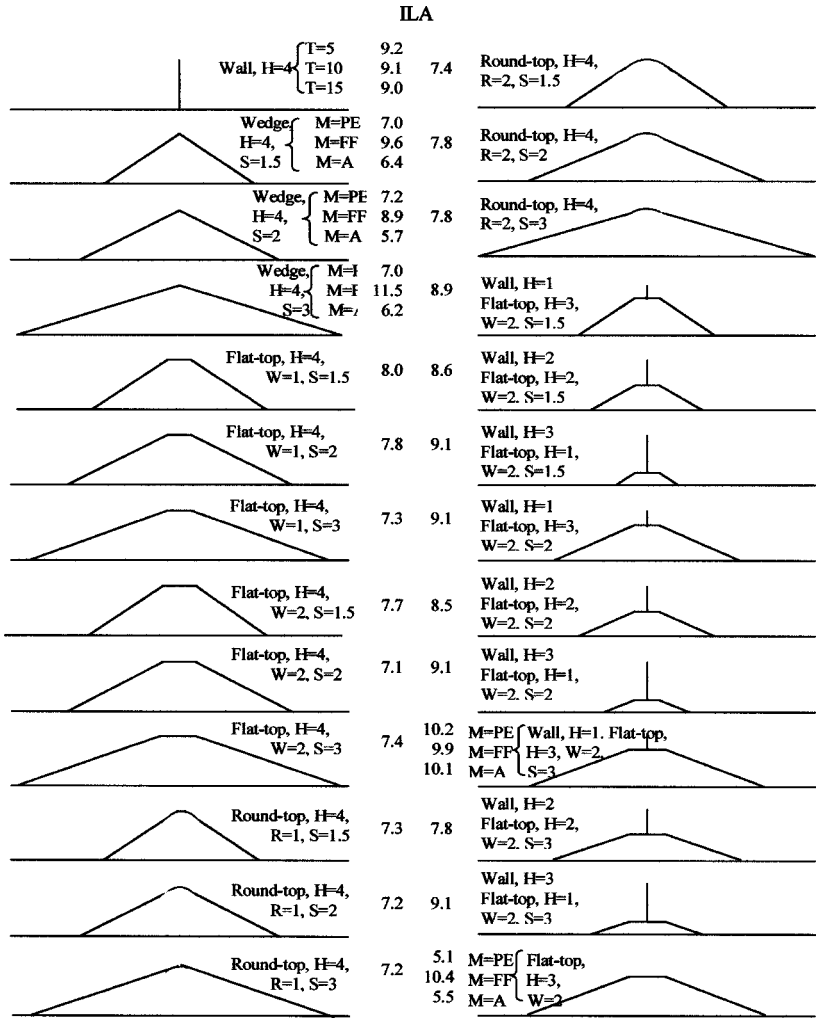


FIG. 4. Cross-sections, details and total, weighted insertion losses (ILA) of all barrier configurations tested. [H=height in m; T=wall thickness in cm=10 unless otherwise indicated; W=top width in m; R=top radius in m; S=slope; M=berm-surface material (PE=packed-earth unless indicated, FF=forest floor, A=asphalt).]

E. Noise-barrier profiles

Scale-model barriers were constructed to a height of 4 mFS, which equates to 3 mFS above the surface of the elevated highway. Tests were made with simulated asphalt and packed-earth ground surfaces. Barriers were either hard, vertical walls, berms modeled with packed-earth surfaces, or walls on top of berms with packed-earth surfaces. The ILs of these three configurations were studied. Furthermore, the effects of varying the following parameters were studied: wall thickness; barrier height; barrier profile; berm slope; berm top profile; berm top-width; and relative wall/berm heights. In subsequent tests, various berms, acting alone or with a crest wall, were made acoustically harder or softer than packed earth to determine the effect of surface composition and how this is affected by barrier configuration. As summarized in Fig. 4, 38 barrier configurations were studied and compared, to illustrate the effects of single-parameter variations.

IV. RESULTS OF LINE-SOURCE TESTS

A. General characteristics

In free-field conditions, the IL of a barrier is expected to increase with frequency. However, with ground reflections, frequency-dependent fluctuations in the IL occurred which were common to all of the scale-model results. At low frequency (80 to 125 HzFS), the IL was typically below 5 dB. From 160 to 630 HzFS, the IL increased gradually in some cases, but typically the IL peaked in the 200, 250 or 315-HzFS third-octave bands, sometimes exceeding 15 dB. For barriers with such IL peaks, as frequency increased towards the 800-HzFS band the IL generally decreased, typically below 5 dB. This behavior is due to the frequency dependent excess attenuation (EA) of the packed-earth open-ground. For every barrier tested, the IL increased with frequency above 800 HzFS, peaking to 10 dB or more in the 1600-HzFS band, with the IL then decreasing in the 2000- and 2500-HzFS bands.

B. Third-octave-band results

1. Wall thickness

Tests were done with a 4-mFS-high wall with a thickness of 5, 10, and 15 cmFS. Wall thickness did not significantly affect the ILA—which varied between 9.0 and 9.2 dBA.

2. Berm slope

The effect of varying the slope of a packed-earth berm was measured for 4-mFS-high wedge-shaped berms, berms with 1- or 2-mFS-wide flat tops, and for berms with round

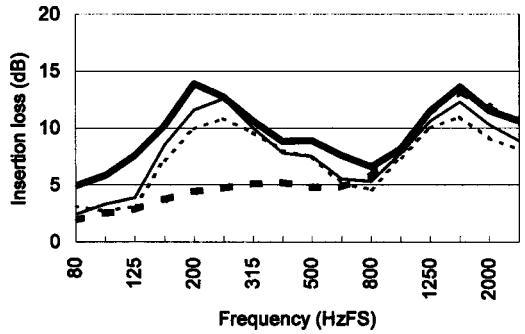


FIG. 5. Measured third-octave-band IL for packed-earth berms of 4-mFS height, with a 2-mFS-wide flat top, and with three slopes: (———) 1:1.5 slope, ILA=7.7 dBA; (···) 1:2 slope, ILA=7.1 dBA; (•••) 1:3 slope, ILA =7.4 dBA. Also shown (———) is the IL for a 4-mFS-high wall, ILA=9.1 dBA.

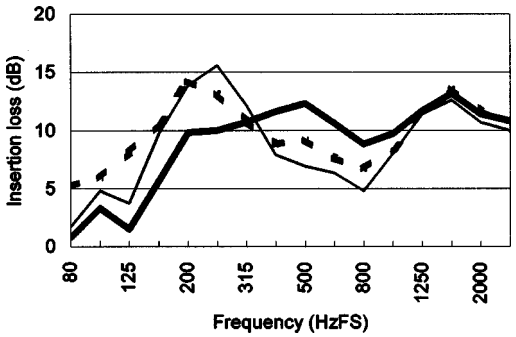


FIG. 6. Measured third-octave-band IL for 4-mFS-high barriers composed of flat-topped packed-earth berms of 1:3 slope, and 2-mFS top-width, topped by a wall, for three relative wall/berm heights: (———) 1-mFS wall on 3-mFS berm, ILA=10.2 dBA; (———) 2-mFS wall on 2-mFS berm, ILA=7.8 dBA; (•••) 3-mFS wall on 1-mFS berm, ILA=9.0 dBA.

tops having either a 1- or 2-mFS radius of curvature. Figure 5 shows, for example, the results for a berm with a 2-mFS-wide flat top. For all cases tested, the only significant effect measured was that, as the berm slope became shallower, the IL decreased in the third-octave bands from 160 to 500 HzFS. The highest effect typically occurred in the 250-HzFS third-octave band, amounting to a 5- to 8-dB reduction. The ILA varied insignificantly—between 7.1 and 7.7 dBA—for the flat-topped berm of 2-mFS width. Also shown in Fig. 5 is the measured IL for a wall, with a corresponding ILA of 9.1 dBA. Comparison of the results for berms with those for walls revealed the general result that walls consistently outperformed packed-earth berms, having an IL in the 200-HzFS third-octave band of up to 9 dB higher than the shallowest 1:3-slope berm, with a corresponding ILA approximately 2 dBA higher. Considering walls to be berms of infinite slope, it can be concluded that performance decreases, as packed-earth berms become shallower.

3. Berm-top width and radius

Tests were performed for flat-topped berms of 1- and 2-mFS top-width, and for round-top berms with a 1- and 2-mFS radius of curvature. Varying the berm top-width or radius for various berm profiles resulted in IL changes of less than 2 dB in all third-octave bands. The ILAs among all profiles considered varied from 7.0 to 8.0 dBA. It can be concluded that the effect of the different top shapes examined is not as significant as variations of slope, and is not of practical interest.

4. Relative berm/wall height

The effect of changing the relative wall and packed-earth-berm heights was measured for combined wall/berm barriers of 4-mFS total height, in the case of 2-mFS-wide, flat-topped berms with various slopes. The results were similar for the three slopes tested and were more complex than previous results. Figure 6 shows the IL for a 1:3-slope berm for which changes in the IL were most substantial. For a 1-mFS-high earth berm topped by a 3-mFS vertical wall, the results were indistinguishable from those of a vertical wall (Fig. 5). At low frequencies, a 1-mFS wall on a 3-mFS berm was worst, a 3-mFS wall on a 1-mFS berm was best. From 315 to 1000 HzFS, a 1-mFS wall on a 3-mFS berm had the best performance, a 2-mFS wall on a 2-mFS berm the worst. The effect of varying the wall height was small above 1000 HzFS. A 3-mFS-high berm topped by a 1-mFS-high wall was particularly effective at the mid-frequencies that most strongly affect the ILA. The results suggest that a shallower and higher berm is the best choice when a wall is present. Furthermore, as the berm becomes shallower, the impact of varying the crest-wall height becomes more significant.

The ILA of the test wall/berm combinations ranged from 7.8 to 10.2 dBA. For a given berm slope, varying the relative wall/berm height changed the ILA between 0.5 and 2.4 dBA. Importantly, the ILA results indicate that a wall/berm combination—in particular, a shallow, 1:3-slope packed-earth berm with a crest wall—can be as effective as, or up to 1 dBA more effective than, a vertical wall. This advantage did not extend to the 1:1.5- or 1:2-slope berms for which the range of ILA was from 8.5 to 9.1 dBA.

5. Slope of berm with crest wall

The above results also show how, for a 4-mFS-high wall/berm combination, changes in berm slope affect IL. The effect on IL resulting from changes in the berm slope when a wall is present is opposite to that for a berm alone—shallower earth-berm slopes resulted in increased ILA. For the 3-mFS-high earth berm topped by a 1-mFS wall, the ILA increased from 8.9 to 9.1 to 10.2 dBA as slope decreased, respectively, from 1:1.5 to 1:2 to 1:3. When berm height was the smaller proportion of the total 4-mFS height, the effect of slope on noise reduction was reduced. As the berm height increased, representing a larger proportion of the total 4-m barrier height, the effect of slope on noise reduction was more evident.

6. Increased barrier height

Figure 7 shows that the effect of increasing the height of a 3-mFS packed-earth berm of 1:3 slope, by making the berm 4-mFS high, differs significantly from the effect of adding a 1-mFS wall. The added berm height consistently increased IL above 400 HzFS (ILA increase of 2.3 dBA). A 1-mFS wall increased the IL at almost all frequencies (ILA increase of 5.1 dBA). The advantages of increasing barrier height using a wall were not so evident in the case of steeper

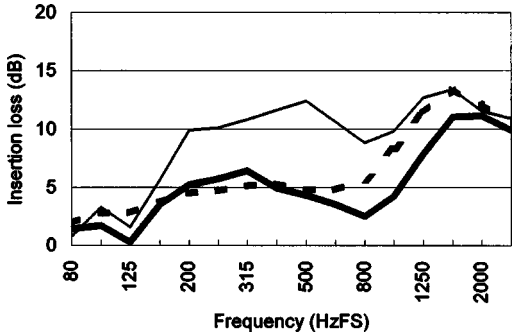


FIG. 7. Measured third-octave-band IL for flat-topped packed-earth berms with 2-mFS top width and 1:3 slope and heights of (———) 3-mFS (ILA =5.1 dBA) and (···) 4-mFS (ILA=7.4 dBA) heights, and for (———) a 3-mFS-high berm topped with a 1-mFS wall (ILA=10.2 dBA).

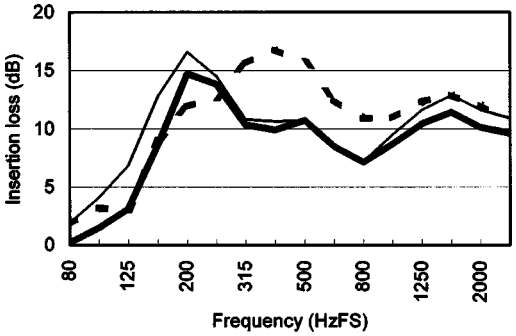


FIG. 9. Measured third-octave-band IL for wedge-shaped berms of 4-mFS height, covered by felt, and with three slopes: (———) 1:1.5 slope, ILA =9.6 dBA; (———) 1:2 slope, ILA=8.9 dBA; (···) 1:3 slope, ILA=11.5 dBA.

berms (1:1.5 and 1:2 slopes). Moreover, the increase in ILA was only 0.5 dBA when the berm was made acoustically softer with a felt material.

C. Alternative surface impedance

In the cases of various berms and combined wall/berm barriers, the surface impedance of the berm was made acoustically harder or softer by adding 2-cmFS-thick dense polystyrene, or a felt layer, respectively. At 1:31.5 scale, felt was found to have an effective flow resistivity of 50 to 80 $kN \cdot s \cdot m^{-4}$,[19] and was a good model of a forest floor.[21]

1. Berm surface impedance

Figure 8 shows, for wedge-shaped berms of 1:3 slope, that changes of surface impedance caused the largest variation of IL. In general, making the surface acoustically harder decreased IL by up to 1.5 dB. The effects of changing impedance are modest for berm surfaces when the effective flow resistivity is analogous to a packed-earth surface, or grounds considered acoustically harder. Making the surface softer, on the other hand, increased IL substantially—by up to 13 dB at middle frequencies centered at 400 HzFS. The ILA of this berm configuration was the highest of all barriers tested, at 11.5 dBA.

2. Berm slope

As shown in Fig. 9, for a berm with a felt surface, the 1:1.5-slope berm matched or outperformed the 1:2-slope berm in all third-octave bands; the ILA was 0.7 dBA greater.

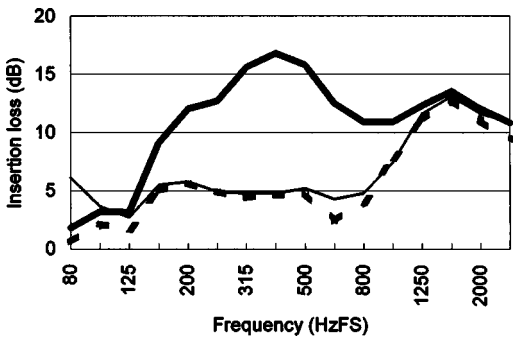


FIG. 8. Measured third-octave-band IL for wedge-shaped berms of 4-mFS height and 1:3 slope when (———) uncovered, ILA=7.0 dBA, and covered by (···) dense polystyrene, ILA=6.2 dBA, or (———) felt, ILA=11.5 dBA.

The 1:3-slope berm had much higher IL at mid frequencies from 250 to 1600 HzFS; the ILA was more than 2 dBA higher. This suggests that, at least for berms that are acoustically soft, there is a benefit to employing a shallow slope.

3. Surface impedance of berm with/without wall

The effect of changes to surface impedance on the performance of a wall/berm barrier was similar to that of a berm alone (see Fig. 8), with a soft surface producing increased performance. However, the effects on IL were not as great, and did not extend to as high a frequency. The barrier profile considered was a 1:3-slope berm with a height of 3 mFS and a flat-topped berm of 2-mFS width, with a 1-mFS-crest wall. The ILA ranged from 9.9 to 10.2 dBA. When a berm has a crest wall, softening the berm's surface does not significantly improve IL.

D. ILA results for all barrier profiles tested

The discussion above highlights the more interesting insertion-loss behavior for various single-parameter changes to barrier configurations. The ILAs for all test configurations are shown in Fig. 4.

V. DISCUSSION AND SUMMARY

Following are the main conclusions concerning the effect of parameter changes on a noise barrier's third-octave-band IL and overall ILA performance at representative, highway-adjacent locations, in the case of homogeneous, nonrefracting atmospheric conditions:

(1) Wall thickness, and berm-top width or radius, are not significant factors.
(2) As a packed-earth berm becomes steeper, the IL increases, becoming more like that of a wall. This is consistent with published results.[17,18]
(3) When a berm is combined with a crest wall, a shallower berm increases the IL. As berm height increases, representing a larger proportion of the total barrier height, the effect of slope on the noise reduction increases. A 1-mFS wall on a 3-mFS berm (with shallow 1:3 slope) has an ILA about 1 dBA greater than for a 4-mFS wall. This agrees well with previous publications that have suggested tuning berm slope to optimize performance,[15] since reflections from the sloped surface of the earth

berm can intersect coherently with other sound paths and diminish the sound-pressure level at the diffracting crest of a barrier.

(4) Barriers have a relatively low IL at mid-frequencies due to the ground effect. This agrees with published results.[14]

(5) Softening a berm surface produces substantial improvements in performance. The benefit of softening a berm is more significant for a 1:3-slope berm than for steeper berms. However, if a berm has a crest wall, softening the berm does not significantly improve the IL. This finding is consistent with previous work.[12]

(6) Performance of berms relative to walls depends strongly on berm surface impedance. For various berms with packed-earth surfaces (wedges, flat-topped, and round-topped) of varying slopes, the ILA was 1 to 2 dBA less than a comparable wall. This is consistent with most published results[12,16–18] but not all.[11] For a harder surface berm (dense polystyrene) of varying slopes, the ILA was 3 dBA less than a comparable wall. For a berm having an acoustically soft ''forest-floor'' surface, the ILA is 0 to 2 dBA higher depending on berm slope, with the shallowest 1:3 slope berm having the 2 dBA advantage. As such, a ''soft-top correction'' of 3 dBA represents a 1 to 5 dBA overestimate of the performance of a berm comprised of a realistic ground surface (packed earth or forest floor). Previous results using unusually ''absorptive'' tops to T-shaped or rectangular barriers found an advantage of 2 dBA over a wall.[13]

(7) Berm/wall combinations for packed-earth berms give an ILA change of -1 to $+1$ dBA relative to a comparable wall. For a softer surface (felt), a 3-mFS-high berm either with or without a 1-mFS wall is about 1 dBA better than a comparable 4-mFS-high wall.

(8) Degradation of noise-barrier performance can occur when walls with vertical and reflective surfaces are located on opposing sides of a highway. This effect was not quantified as part of this research program. Inclined walls and earth berms with sloped surfaces serve to reduce the significance of this degradation.

VI. CONCLUSION

The research objective, to evaluate the relative effectiveness of different types of highway noise barrier with the goal of reducing traffic-noise levels in communities, has been achieved. This involved the design and construction of a 1:31.5 scale-model facility and the selection of materials to simulate asphalt, walls, packed earth, and acoustically harder or softer variations. Scale-model barriers were constructed, and their ILs measured and reported in the case of an incoherent line source. The research has shown the importance to berm performance of crest walls and surface impedance, the possible effectiveness of tuning berm slope to maximize performance, and the lack of any evidence to support the assumption of a ''soft-top correction'' of 3 dBA. For a surface impedance ranging from an acoustically hard packed earth to an acoustically-soft forest floor, the measured range of barrier performance relative to a comparable wall was from -2 to $+2$ dBA.

ACKNOWLEDGMENT

The authors wish to acknowledge the collaboration of Michael Kent of the British Columbia Ministry of Transportation and Highways, which supported the project.

[1] T. Busch, ''Scale-Model Investigation of Highway Traffic-Noise Barriers,'' M.A.Sc. Dissertation, Department of Mechanical Engineering, University of British Columbia, Vancouver, Canada, 1997.
[2] ''Policy for Mitigating the Effects of Highway Traffic Noise from Freeways and Expressways,'' British Columbia Ministry of Transportation and Highways—Highway Engineering Branch, Victoria, February 1991.
[3] M. A. Simpson, ''Noise Barrier Design Handbook,'' Report No. FHWA-RD-76-58 Federal Highway Administration, Washington, February 1976.
[4] G. G. Fleming, H. S. Knauer, C. S. Y. Lee, and S. Pedersen, ''FHWA Highway Noise Barrier Design Handbook,'' Federal Highway Administration, Washington, February 2000.
[5] G. Watts, ''Acoustic Performance of Traffic Noise Barriers—A State of the Art Review, Part 1,'' Acoust. Bull. **18**(3), 13–18 (1993).
[6] G. Watts, ''Acoustic Performance of Traffic Noise Barriers—A State of the Art Review, Part 2,'' Acoust. Bull. **18**(6), 29–39 (1993).
[7] S. I. Hayek, ''Mathematical Modeling of Absorbent Highway Noise Barriers,'' Appl. Acoust. **31**(1-3), 77–100 (1990).
[8] M. E. Delany, A. J. Rennie, and K. M. Collins, ''A Scale Model Technique for Investigating Traffic Noise Propagation,'' J. Sound Vib. **56**(3), 325–340 (1978).
[9] H. W. Jones, D. C. Stredulinsky, and P. J. Vermeulen, ''An Experimental and Theoretical Study of the Modeling of Road Traffic Noise and its Transmission in the Urban Environment,'' Appl. Acoust. **13**(4), 251–265 (1980).
[10] M. M. Osman, ''MTC Scale Model Facility for Transportation Noise Problems: Materials Choice and Validation for Scale Modeling,'' Ontario Ministry of Transportation and Communications Report 77-AC-4, June 1977.
[11] M. A. Simpson, ''Noise Barrier Attenuation: Field Experience,'' Report No. FHWA-RD-76-54, Federal Highway Administration, February 1976.
[12] J. J. Hajek, ''Are Earth Berms Acoustically Better than Thin-Walled Barriers?'' Noise Control Eng. J. **19**(2), 41–48 (1982).
[13] D. N. May and M. M. Osman, ''Highway Noise Barriers: New Shapes,'' J. Sound Vib. **71**(1), 73–101 (1980).
[14] D. A. Hutchins, H. W. Jones, and L. T. Russell, ''Model studies of barrier performance in the presence of ground surfaces. Part I. Thin, perfectly reflecting barriers,'' J. Acoust. Soc. Am. **75**(6), 1807–1816 (1984).
[15] D. A. Hutchins, H. W. Jones, and L. T. Russell, ''Model studies of barrier performance in the presence of ground surfaces. Part II. Different shapes,'' J. Acoust. Soc. Am. **75**(6), 1817–1826 (1984).
[16] Z. Maekawa and S. Osaki, ''A Simple Chart for the Estimation of the Attenuation by a Wedge Diffraction,'' Appl. Acoust. **18**, 355–368 (1985).
[17] C. C. Cremers, K. R. Fyfe, and L. J. Cremers, ''Insertion Loss Characteristics of Barriers and Berms,'' Can. Acoust. **22**(3), 121–122 (1994).
[18] D. C. Hothersall, D. H. Crombie, and S. N. Chandler-Wilde, ''The Performance of T-Profile and associated Noise Barriers,'' Appl. Acoust. **32**, 269–287 (1991).
[19] T. Busch and M. R. Hodgson, ''Letters to the Editor: Improved Method for Selecting Scale Factors and Materials for Scale Modelling of Outdoor Sound Propagation,'' J. Sound Vib. **243**(1), 173–181 (2001).
[20] J. Novak, ''Technical Note: Sound Source for Scale Model Measurement of Traffic Noise,'' Appl. Acoust. **24**(1), 63–70 (1988).
[21] M. E. Delany and E. N. Bazley, ''Acoustical Properties of Fibrous Absorbent Materials,'' Appl. Acoust. **3**, 105–116 (1970).

Hearing protection: Surpassing the limits to attenuation imposed by the bone-conduction pathways[a)]

Elliott H. Berger[b)] and Ronald W. Kieper
E•A•R/Aearo Company, 7911 Zionsville Road, Indianapolis, Indiana 46268-1657

Dan Gauger
Bose Corporation, The Mountain, Framingam, Massachusetts 01701-9168

(Received 24 January 2003; revised 20 June 2003; accepted 14 July 2003)

With louder and louder weapon systems being developed and military personnel being exposed to steady noise levels approaching and sometimes exceeding 150 dB, a growing interest in greater amounts of hearing protection is evident. When the need for communications is included in the equation, the situation is even more extreme. New initiatives are underway to design improved hearing protection, including active noise reduction (ANR) earplugs and perhaps even active cancellation of head-borne vibration. With that in mind it may be useful to explore the limits to attenuation, and whether they can be approached with existing technology. Data on the noise reduction achievable with high-attenuation foam earplugs, as a function of insertion depth, will be reported. Previous studies will be reviewed that provide indications of the bone-conduction (BC) limits to attenuation that, in terms of mean values, range from 40 to 60 dB across the frequencies from 125 Hz to 8 kHz. Additionally, new research on the effects of a flight helmet on the BC limits, as well as the potential attenuation from deeply inserted passive foam earplugs, worn with passive earmuffs, or with active-noise reduction (ANR) earmuffs, will be examined. The data demonstrate that gains in attenuation exceeding 10 dB above the head-not-covered limits can be achieved if the head is effectively shielded from acoustical stimulation.
[DOI: 10.1121/1.1605415]

PACS numbers: 43.50.Hg, 43.66.Vt [DKW]

I. INTRODUCTION

When personnel are exposed to very high levels of noise, such as generated by current and new military weapons systems that approach and even exceed a continuous A-weighted sound pressure level (SPL) of 150 dB, the need for maximum hearing protection is obvious. But, maximum hearing protection may still not be enough, and, by the way, what is the maximum hearing protection that can be provided? Surprisingly, few authors have explored the limits to attenuation since von Békésy (1960) and also von Gierke and Warren (1953) addressed the question in the early 1950s. In this report we review the available data, update them as needed, and explore how the application of active noise reduction (ANR), or the use of an enclosure to shield the head, can affect those limits. The purpose is to provide hearing conservationists and hearing protection developers a reliable benchmark that defines the maximum levels of protection achievable for humans exposed to noise.

As Zwislocki (1957) observed in his landmark paper, direct measurement of such values was long overdue, and in that paper he provided values that have indeed withstood the test of time. In this current work, his data will be compared to a handful of estimates that others, including the first author, have published in the past half-century.

The concept of bone conduction (BC) limits implies that sound is transmitted via bony structures in the head that bypass the normal air-conduction mechanism of transmission through the earcanal. In effect these are flanking sound pathways that circumvent the noise-blocking features of the hearing protection device (HPD) that is covering or occluding the earcanal. The primary BC pathways as described in the literature are (a) vibration of the earcanal walls, (b) energy transmitted due to excitation of ossicular motion, and (c) direct mechanical excitation of the cochlea (Tonndorf, 1972; Khanna *et al.*, 1976). Additional discussions are contained in Berger and Kerivan (1983), Berger (1985), and Ravicz and Melcher (2001). In general terms, BC refers to any pathway other than that of conventional air conduction. For example, sound passing through the open mouth and the soft tissues of the eustachian tube excites no bones except the ossicles, but it is still included under the rubric of bone conduction. However, Zwislocki noted that it might be more appropriately called body conduction, though for purposes of adherence to common convention he chose to title his paper ''bone conduction.'' We will use that more common term herein as well.

The BC thresholds in a sound field can be measured similarly to the minimum audible field, except that sound must be prevented from being transmitted via the conventional air-conduction pathways that begin at the earcanal. This can be accomplished by sufficiently occluding the earcanal, or by canceling sound that is present at the eardrum.

[a)]Portions of this work were presented at the conference of the Survival and Flight Equipment Association (SAFE), Jacksonville, FL, October 2002, and at the joint conference of the 144th Meeting of the Acoustical Society of America, the 3rd Iberoamerican Congress of Acoustics, and the 9th Mexican Congress of Acoustics, Cancun, Mexico, December 2002.

[b)]Electronic mail: eberger@compuserve.com

The latter approach, employed by Schroeter and Els (1980), is uncommon and has its potential shortcomings since canceling sound at the eardrum, which is directly connected to the ossicles, might affect ossicular motion, and such motion is one of the BC pathways [i.e., pathway (b) as discussed above].

When one uses the method of sound blockage at the ear, the question will naturally arise whether sufficient sound has been excluded. Whatever technique is employed, artifacts are always possible. For example, Zwislocki used resonator earplugs whose tips were metal rods coated with latex and wax, that were inserted deeply into the bony meatus. These devices substantially eliminated vibration of the canal walls, which is the dominant path in the occluded ear below 2 kHz [pathway (a) as described above]. One might argue that with actual usable HPDs such reduction of this pathway is not feasible. This suggests that the low levels of BC that Zwislocki reported would likely not be observed in practice. Alternatively, in the procedure that Berger has used with success, a very heavy lead earmuff (more than 10 times the mass of a conventional earmuff) with unusually high band force (about twice the force of a conventional earmuff) was worn in conjunction with deeply fitted foam earplugs. A concern with this approach is that the inordinately high mass and force of the combination distorts the skull in a way unlike any actual HPD, potentially creating unrealistic BC limits. Thus, within our manuscript we have compared a variety of procedures from various reports to provide a range of estimates of the limits that are likely to prevail.

In studying this paper, the reader should keep in mind the following important caveat: all of the measurements in this paper are based upon optimum fitting of HPDs in a laboratory environment. In fact, the fitting might be termed ''hyper-optimum'' in that in some cases the devices used are fitted uncomfortably in ways that might not be feasible in practice for the sustained periods that would be anticipated in the real world. Thus, it is questionable whether such values of protection as the BC limits reported in this paper could ever be achieved for groups of users in occupational settings regardless of the degree of motivation, training, and supervision that was employed. For additional discussion of such matters see ANSI S12.6-1997, Berger *et al.* (1998), and Berger (2000).

II. PROCEDURES

The data in this report consist of published real-ear attenuation at threshold (REAT) values from Zwislocki (1957), Nixon and von Gierke (1959), Schroeter and Els (1980), Berger (1983), Ravicz and Melcher (2001), and the current research.[1] To the authors' knowledge, these few papers, together with the early work of von Békésy (1960), represent the sum total of the available data on direct measurement of the bone-conduction limits on human subjects in a free or diffuse sound field. Others such as von Gierke and Warren (1953) reported predictions based on BC thresholds derived from direct stimulation of the forehead via ''sound tubes,'' but did not conduct measurements with the entire body or head irradiated in a sound field. Their computations, however, provide additional support of the empirical values presented by the other authors. Békésy's work, though exceedingly clever, yielded empirical data at only a few low frequencies, and therefore are not included in the discussions that follow. However, his results, like those of von Gierke and Warren, are in accord with the findings of the five studies that are specifically reviewed. Brief descriptions of the empirical studies follow.

A. Zwislocki (1957)

These experiments consisted of the measurement of REAT in a free sound field on groups of six subjects wearing solid earplugs in combination with heavy earmuffs, three of those same subjects wearing only resonator earplugs tuned to frequencies from 300 to 600 Hz, and one subject for the frequencies below 125 Hz. The BC limits were determined by the resonator earplugs below 400 Hz, the better performing of the resonator earplugs or the dual combination from 400 to 1500 Hz, and the dual combination only above 1500 Hz.

B. Nixon and von Gierke (1959)

Measurements were conducted in a free sound field with frontally incident sounds, as well as incidence on the back of the head. Measurements were in conformance with the standard for hearing protection attenuation measurements that was in effect at the time (ANSI Z24.22-1957). The attenuation of five different commercially available earplug/earmuff combinations was measured on eight subjects. Additional measurements were conducted by covering the subjects' heads in part or in total with medical cotton wicks of 8-in. width that were wound around the head and fastened with tape until the desired thickness of 2–3 in. was achieved.

C. Schroeter and Els (1980)

REAT measurements were taken using very large custom-built (approximately 30 000 cm^3) sound attenuation enclosures that coupled to the head circumaurally with conventional earmuff cushions, and contained cancellation speakers for the low frequencies. At and above 2 kHz the attenuation of the enclosures was complemented by deeply inserted foam earplugs. The tests were conducted on ten subjects. The 1980 reference (in German) cited above is the original, but those results which are also summarized in Schroeter and Poesselt (1986), p. 512, Table II, are more easily accessed.

D. Berger (1983)

REAT values for three earplugs (one of which was worn with three different depths of insertion) and three earmuffs were evaluated both singly and in various combinations according to ANSI S3.19-1974. Thirteen subjects participated in the entire experiment, with seven common to all tests. For any one test, ten subjects were measured three times each. All of the earplugs were commercially available. The one of greatest interest for the purposes of this work was the foam earplug (E•A•R® Classic® plug with a length of approximately 19 mm) which was inserted partially (Pl; about

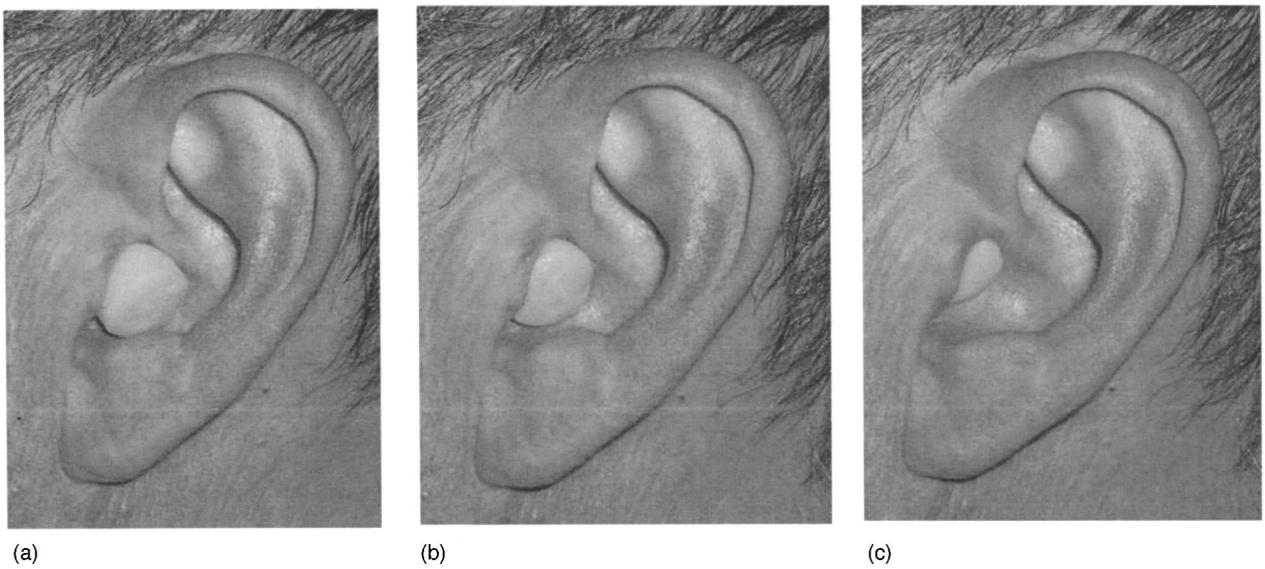


FIG. 1. (a)–(c) Partial (PI), standard (SI), and deep (DI) insertions of a Classic foam earplug in one subject's earcanal.

15%–20% of the plug in the earcanal), to a standard depth (Sl; about 50%–60% in the earcanal), or deeply (Dl; about 80%–100% in the earcanal). The DI was the maximum depth of insertion that a subject could tolerate before experiencing significant discomfort. The fit was measured by drawing black circumferential rings on the plugs at the intersection of the plug and the posterior entrance of the earcanal (i.e., the floor of the concha). All fitting was by the experimenter in conjunction with the subject. Figures 1(a)–(c) show a view of the plug inserted to the PI, SI, and DI depths.

Two of the earmuffs were commercially available and of substantially different size and mass. The third was a specially constructed damped lead earmuff with an extensional layer of a vinyl damping compound. Its total mass for both cups was 3300 g and volume (per earcup) 300 cm^3, with an inordinately high band force of 24 N. The lead earmuff was utilized for exploring BC limits. It is not a feasible product to wear outside the laboratory.

In combination with a small-volume earmuff Berger tested five different earplug conditions, each of which when worn individually provided four substantially different levels of protection. The dual combinations of the various plugs with the small earmuff also provided four substantially different values of protection for the frequencies below 1 kHz, corresponding to the increasing attenuation provided by the earplugs alone. At and above 2 kHz all combinations provided the same level of protection, which was also within 3 to 7 dB of the values reported by Zwislocki (1957) and by Nixon and von Gierke (1959) in their prior studies.

Berger's estimate of the BC limits to attenuation was taken to be the values found for the deeply inserted foam earplug worn in combination with the lead earmuff. Even though there were dramatically differing levels of attenuation for the three earmuffs in the study when used as a single hearing protector, when worn in combination with a DI foam earplug the earmuffs performed identically, i.e., results were essentially the same for all earplug-plus-earmuff combinations.

An ancillary observation that Berger made was that with one exception, the combined protection of a muff and a plug always exceed either of the individual devices at all test frequencies. The interesting exception was one premolded earplug (V-51R type) worn together with the large-volume earmuff. The combined attenuation of the plug plus muff was actually about 4 dB less at 1000 Hz than found for the earmuff alone. This could be attributed to the occlusion effect for this relatively short (and hence shallowly inserted) premolded earplug. Shallow fitting causes an amplification of the external-earcanal bone-conduction pathway (Berger and Kerivan, 1983), thus enhancing any vibrations of the earcanal walls caused by the sound field or by vibration of the earmuff itself.

E. Ravicz and Melcher (2001)

In the context of a study examining noise created by functional magnetic resonance imaging the authors explored the limits to attenuation. Their most protective condition consisted of foam plugs worn with conventional high-attenuation earmuffs, covered by a free-standing sound attenuating helmet large enough to fit loosely over the head of a subject wearing earmuffs. The helmet required a breathing tube and was sealed to the shoulders with a towel. Five subjects were tested using a REAT paradigm with unconventional stimuli consisting of tone bursts at frequencies from 500 to 2800 Hz. Microphone-in-real-ear (MIRE) data were also acquired, but those values are not reported herein.

F. Current research (this paper)

Similar procedures were employed as in Berger (1983) except that the subject count was increased to 16, with two instead of three measurements of each condition per subject. Fourteen of the subjects met the hearing sensitivity requirements of ANSI S12.6-1997, and the remaining two met the requirements at all but one frequency. With the exception of subject fitting and subject count (16 instead of 20) the pro-

(a) (b)

FIG. 2. (a) Gentex HGU-55/P flight helmet with one earcup removed, illustrating the Velcro attachment material on the back of the cup and the foam/Velcro spacers (partly beneath the cup) that are used for positioning. (b) The helmet on subject's head with visor and cover in position.

cedure conformed to Method A of ANSI S12.6-1997. All REAT measurements were conducted in the E•A•RCAL^{SM} acoustical laboratory of Aearo Company. MIRE measurements on the active earmuff were made at Bose Corporation on a separate group of subjects in conformance with ANSI S12.42-1995 with two exceptions: only six subjects were tested, and measurements were made in a medium-sized audiometric booth lined with loudspeakers. Between the three fittings of the muff, subjects rotated their position in the booth to vary the incidence of the sound field. Measurements done in this way have correlated well with measurements in the Bose reverberation room.

The important new features of this series of experiments consisted of the devices. The E•A•R® Classic® Plus earplug (a longer version of the E•A•R Classic—24 mm vs 19 mm) was utilized and fitted deeply, and was worn in conjunction with one of two types of earmuffs. One was a prototype ANR headset made by Bose for U.S. military evaluation. It utilized large yet conventional circumaural earmuff cups, a conventional headband, and was an experimental adaptation of ANR technology from Bose's present commercial products for use in higher noise levels. The other was a Gentex lightweight fighter/attack aircrew helmet (HGU-55/P) with thick edge roll and internal plastic earmuff cups, as illustrated in Fig. 2(a). We utilized the medium and extra-large sizes of the helmet, and carefully inserted foam spacers between the cups and helmet shell, as is the norm for proper positioning. Larger volume cups and cushions from the Navy HGU-84/P helmet, rather than those normally fitted in the USAF HGU-55, were used, and communications were removed (i.e., receivers and cables) to maximize the attenuation achieved. The chin and nape-of-the-neck straps were adjusted for optimum fitting. All tests were conducted with the visor down in the operational position and covered by its fabric protection cover [see Fig. 2(b)]. Obviously the helmet could not be worn in such a condition, but the purpose herein was to explore the limits to protection; if need be a clear visor with higher transmission loss could always be devised. Note the foam edge roll at the rim of the helmet, and the resulting tight fit to the sides of the face and below the ears. However, protection of the lower frontal face, i.e., the mouth and jaw, was missing.

Once subjects entered the chamber they generally stayed in place for the entire set of measurements lasting approximately 90 min. Occasionally they took a brief break between the first series and the repeat series. The testing sequence was intended to limit fitting variability as much as possible to allow for the best comparison between plug-only and plug-plus-muff conditions. The testing was as follows:

(i) Flight helmet positioned, wait 2 min for acclimatization, and take occluded threshold.
(ii) Occluded threshold with ANR headset, electronics off.
(iii) Open threshold.
(iv) Occluded threshold with DI foam earplug.
(v) Occluded threshold, leaving DI foam earplug in place for this and following occluded thresholds, with the flight helmet donned as well.
(vi) Occluded threshold with the ANR headset, turned off, in place of the flight helmet.
(vii) Occluded threshold with the ANR headset left in position, but now turned on.[2]
(viii) Remove earplugs and ANR headset and repeat entire series.

Test signals were $\frac{1}{3}$-octave-bands of noise spanning the range from 125 Hz to 8 kHz at octave-band center frequencies. For one subject (the first author) test data were also acquired at 80 Hz. Although those values are not included in the subsequent analyses they indicate that the BC limits and other attenuation values at 80 Hz are essentially the same as those found at 125 Hz.

III. RESULTS

The data from the current study as well as key results from the literature summarized in the prior sections of this paper are presented in Table I.

TABLE I. Mean real-ear attenuation and standard deviation values in dB from this study, and mean values from prior published data.

Device		Frequency (Hz)						
		125	250	500	1000	2000	4000	8000
DI foam	Mean	39.9	44.4	47.8	43.7	37.4	44.4	47.0
	SD	5.5	4.8	3.5	4.2	2.9	3.7	4.7
ANR Muff OFF	Mean	18.1	23.7	25.9	30.1	36.1	43.3	43.3
	SD	3.5	1.8	1.6	2.6	2.9	2.8	3.1
Flight Helmet	Mean	20.6	24.1	29.5	41.7	46.5	56.6	58.6
	SD	4.2	6.0	3.5	4.3	4.4	3.2	4.9
DI foam+ANR Muff OFF	Mean	47.1	57.4	62.0	49.5	40.5	50.1	49.7
	SD	4.8	3.1	4.4	5.8	3.8	5.3	4.6
DI foam+ANR Muff ON	Mean	50.4	57.3	61.5	49.2	40.8	50.5	50.1
	SD	5.1	3.2	3.6	6.1	4.3	5.6	4.7
DI foam+Flight Helmet	Mean	42.0	50.5	60.8	53.6	48.6	60.2	61.3
	SD	5.6	4.0	4.2	6.6	4.0	5.1	6.6
Foam+small volume earmuff, S12.6-1997 Method B	Mean	26.1	27.0	34.4	38.6	40.4	50.8	47.3
	SD	6.5	6.1	6.3	5.5	5.5	5.4	4.4
Estimate of ''best possible achievable protection''	Mean	50.4	57.3	61.5	55.0	65.0	75.0	75.0
	SD	5.6	4.0	4.2	6.6	4.0	5.1	6.6
Zwislocki (1957) BC	Mean	51	60	68	60	46	54	41
Nixon and von Gierke (1959) plug+muff	Mean	30	33	38	40	42	52	41
Nixon and von Gierke (1959) plug+muff+head covered	Mean	31	35	41	45	52	68	58
Schroeter and Els (1980)	Mean	52	51	48	46	49	54	44
Berger (1983) BC	Mean	47	51	57	47	39	49	49
Ravicz and Melcher (2001) plug+muff+box	Mean			58	55	66		
Hachey and Roberts (1983) real-world plug+muff	Mean	21	24	31	35	35	43	38

A. Attenuation values for single HPDs

The attenuation values for the individual plugs and muffs of the current study are presented in Figs. 3 and 4 where they are compared to the BC limits from Berger (1983). Note that at all frequencies the attenuation of the deeply inserted Classic Plus, a longer version of the Classic foam earplug, exceeds that of the shorter version by 2 to 4 dB (Fig. 3). Statistical tests (*t*-test assuming equal variance) yielded significance at 500 Hz ($p<0.01$), and ''near significance'' at 250 and 1000 Hz ($p<0.06$ and $p<0.08$, respectively). Other differences were not statistically significant. The increased attenuation in the low frequencies may be in part due to the longer plug combined with a deeper fit, but also note that the two subject populations (10 in the prior study and 16 in the current study) are completely different. The slightly increased attenuation of the DI plug in this study, together with the fact that the plug's performance has a large impact on the attenuation of the combined devices, must be kept in mind when comparing the current BC estimates to those from Berger (1983).

Also of note in Figs. 3 and 4 is the sharp minimum of attenuation at 2 kHz in all of the head-not-covered conditions. This is a commonly observed feature of real-ear attenuation data for high-attenuation hearing protectors.

In Fig. 4 the attenuation of the ANR earmuff is compared to the small-volume and lead earmuffs from Berger (1983). Because ANR electronics generate residual low-level noise, which can mask the threshold test signals, they are not

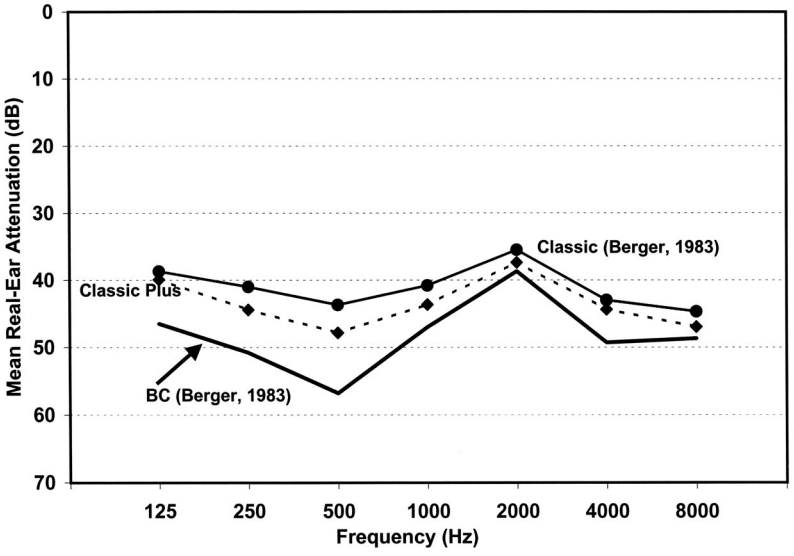


FIG. 3. Comparison of the real-ear attenuation of a deep insertion of the Classic Plus (this study) to the Classic, and to the BC limits, from Berger (1983).

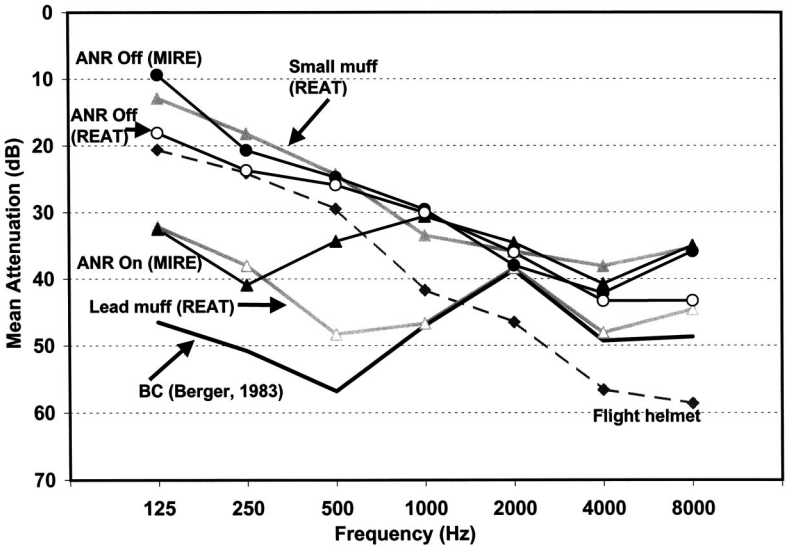


FIG. 4. Attenuation of the ANR earmuff and flight helmet, worn individually, from this study, compared to the muffs from Berger (1983). For the ANR earmuff both real-ear attenuation at threshold (REAT) and microphone-in-real-ear (MIRE) data are compared.

amenable to REAT testing. Hence for that earmuff's on-condition, MIRE data (measured at Bose's laboratory) are reported. To provide the best comparative data for the off-mode, the muff was tested with both MIRE and REAT procedures. The REAT measurement was possible for the ANR device in the off-mode since in that condition the electronics generate no residual noise. REAT measurement is also possible when an ANR device is turned on and worn over earplugs, as is reported later, since the earplugs attenuate the self-noise of the ANR device below the subject's hearing threshold.

Comparison of the dark sold lines with the filled and open circles illustrates the REAT vs. MIRE measurements for the ANR earmuff. The differences, due to low-frequency physiological-noise masking that elevates the occluded threshold (Berger and Kerivan, 1983), are of the pattern which are expected, though the magnitude of the effect, 9 dB at 125 Hz, is larger than anticipated for a muff of this size. A 5-dB REAT-MIRE difference at 125 Hz is more typical (Gauger, 2002). If it were possible to also produce REAT data for the ANR earmuff in the on position, they too would be expected to show increased low-frequency attenuation relative to MIRE values. Hence it is likely that the performance of the ANR device in the on-mode would have exceeded the REAT-measured attenuation of the lead earmuff at 125 and 250 Hz by a greater amount than shown in Fig. 4, if both were measured in the same way. In the off-mode the performance of the ANR earmuff was similar to the small-volume earmuff.

The flight helmet data are similar to earmuff attenuation values up through about 1 kHz, but quite different from the circumaural earmuffs in the high frequencies. Note that at and above 2 kHz, the flight helmet exceeds Berger's 1983 estimate of the BC limits for circumaural devices. This is not surprising since in the high frequencies the important BC pathway has been shown to be direct transmission to the cochlea bypassing the external earcanal. If the skull is shielded, less energy will be incident on the head and consequently less will be transmitted to the inner ear via this flanking pathway (Khanna *et al.*, 1976; Ravicz and Melcher, 2001).

B. Individual subject data

Before presenting the averaged data for the dual protection combinations from the current study, it will be illuminating to review the data for the 17 individual subjects in Figs. 5–7. All the subjects except for KLD were included in the subsequent analyses. KLD has had recurring ear infections since childhood and an abnormal tympanogram indicating poor tympanic mobility, but was included because his open-ear audiometric thresholds are near normal and he could be representative of a person required to wear dual protection. Also in the past, KLD has been well-fitted with foam earplugs and obtained representative attenuation data. Although his current data are aberrant, they might be experienced in the real world. Note that with respect to the foam earplug, as shown in Fig. 5, KLD experienced unusually low values of attenuation at 500, 1000, and 8000 Hz. The boxes indicate his retest data on a subsequent day. Even excluding KLD, there is a 15- to 20-dB range in data across frequencies. Thus, even though this well-fitted foam earplug delivered 25 or more decibels of attenuation at all frequencies for all subjects in this study, and more than 30 decibels at all frequencies for all *but one* subject, the actual range in attenuation values is large.

Figure 6 demonstrates the spread for the ANR earmuff, turned off, i.e., worn as a conventional earmuff. Note the reduced range in values compared to the results for the earplug. This is not surprising since earmuffs normally fit more uniformly across groups of listeners, producing less attenuation variability. The data for the flight helmet (not shown) exhibit variability greater than that of the earmuff and closer to the foam plug. Although the helmet contains circumaural cups, once the cups are ''hidden'' under the helmet, the positioning and adjustment is more problematic than with a conventional headband-mounted earmuff unit.

Finally, Fig. 7 shows the effect when the DI foam plug is combined with the flight helmet. As with the earplug alone,

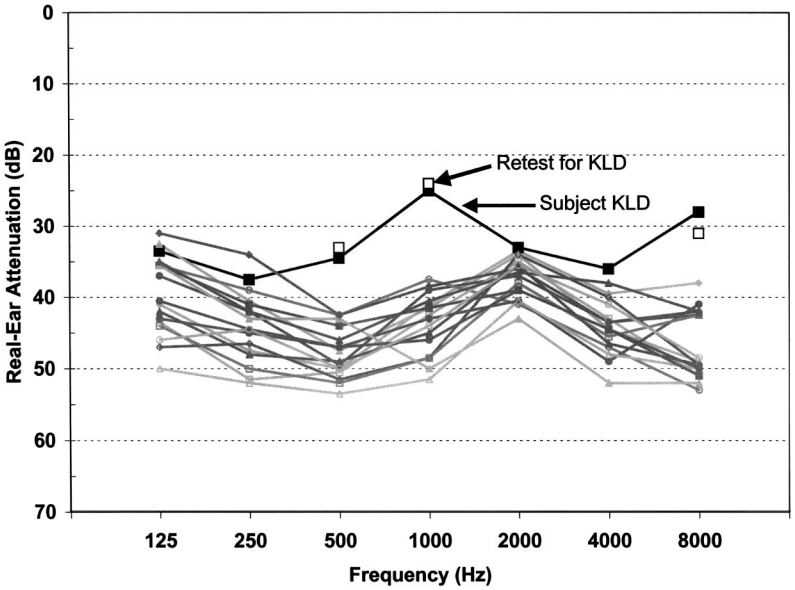


FIG. 5. Real-ear attenuation for foam plugs with a deep insertion (DI). Each line represents the average of two measurements on one subject. The ''aberrant'' subject KLD is specifically noted.

KLD is an outlier, as might be expected since the earplug strongly controls the combined attenuation of the devices. When the data for the flight helmet alone are examined, KLD is low, but only by a few decibels from 1000 Hz on upwards in frequency.

C. The value of ANR in a dual-protection scenario

Figure 8 demonstrates the performance of the DI foam earplug plus ANR earmuff, in its on- and off-modes, as compared to Berger's prior BC estimate. Turning first to the question of the value of the active noise reduction, note that the ANR-on and ANR-off values of attenuation in Fig. 8 are essentially identical at all frequencies except 125 Hz (3.3-dB difference, significant at $p<0.05$ using t-test for paired samples), the frequency at which ANR is typically most effective. Recall that the ANR earmuff was left in position on top of the earplug and tested turned off and then on, in immediate sequence. This eliminated fitting as a cause of potential variability between the two conditions. Although significant, the increase in protection at 125 Hz is of questionable importance due to its small magnitude and limited frequency range. The small increase provided by ANR in a dual-protection scenario with the DI plug is presumably due to the same reason that increasing muff size offers no dual-protection benefits. Transmission through a small-volume muff and DI plug, or transmission through the ANR earmuff in the off-mode when combined with the DI plug, is already equivalent to the BC limits at most frequencies, so the added attenuation of a more protective earmuff or condition is inconsequential.[3]

The current dual-protection data for the ANR earmuff demonstrate values that equal or exceed Berger's prior estimate obtained using a lead earmuff as seen in Fig. 8, with the differences achieving statistical significance at $p<0.05$ (t-test assuming equal variance) for the frequencies from 125 to 500 Hz. This is not surprising since the earplug is such an important component of the dual-protection system, and as shown in Fig. 3 the DI earplug fit in this study exceeds the

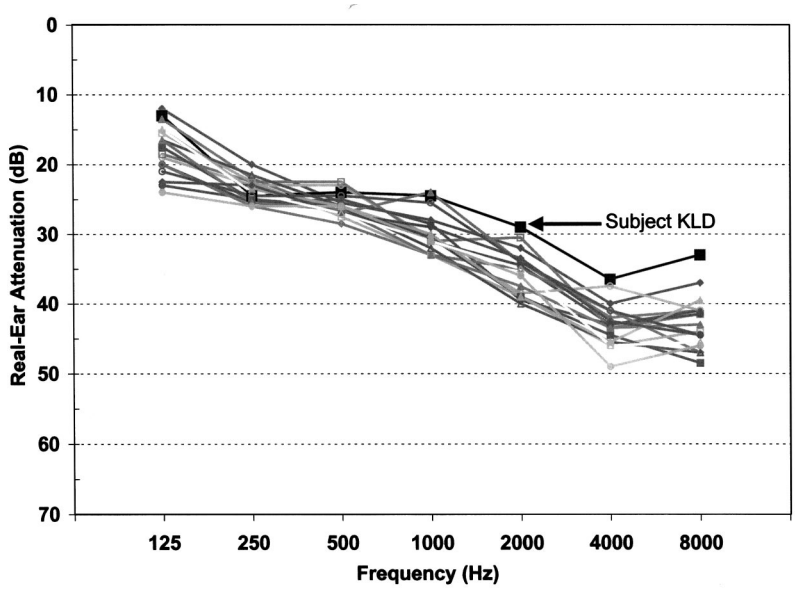


FIG. 6. Real-ear attenuation for ANR earmuff with the ANR off. Each line represents the average of two measurements on one subject. The ''aberrant'' subject KLD is specifically noted.

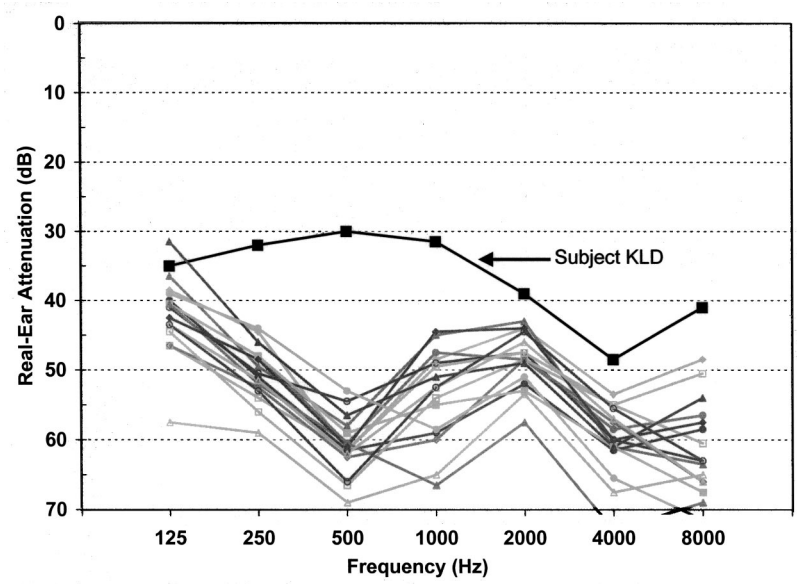


FIG. 7. Real-ear attenuation for foam plugs with deep insertion (DI) plus flight helmet. Each line represents the average of two measurements on one subject. The ''aberrant'' subject KLD is specifically noted.

performance of the one previously employed by Berger. Additionally, note that the difference between the studies is largest below 1 kHz. Khanna *et al.* (1976) concluded that the primary BC path at these frequencies is through the skull to the walls of the earcanal which vibrate and reradiate the sound; the deeper insertion resulting from the longer plug used in the present study may have changed the BC limit itself. Finally, the differences between the studies may also be explained by the attributes of the current subject pool, and the possibility that their inherent BC limits differ slightly from the group used nearly 20 years ago.

D. Limits to attenuation

Figure 9 presents the best estimates of the limits to attenuation from the current study, based on the DI foam plug worn with the ANR earmuff in the on-mode, and the DI foam plug worn with the flight helmet with the visor down and covered. The values are compared to the DI foam earplug worn alone and to one of the higher attenuation conventional earmuffs we have measured in our laboratory. Unless the head is shielded, switching from single to dual protection can only achieve modest gains of about 6 dB at and above 1 kHz before reaching the BC limits (compare plug+ANR muff on, to the single-HPD curves). However, at the low frequencies gains of 10 to 15 dB are realizable.

By covering the head, as accomplished in this study using the tight-fitting flight helmet with face plate (visor), additional gains above single protection of from 4 to 11 dB are possible from 1 kHz and up. However, the attenuation at 125 and 250 Hz suffers relative to the optimum low-frequency combination of a DI foam earplug plus circumaural earmuff.[4]

IV. LIMITS TO ATTENUATION: CURRENT VERSUS PRIOR DATA

The data in Fig. 10 compare the four previously published directly measured estimates of the BC limits to those from the current study, for the head-not-covered condition. The current data and the Berger (1983) values were measured in a diffuse field; the others were frontal incidence free field.

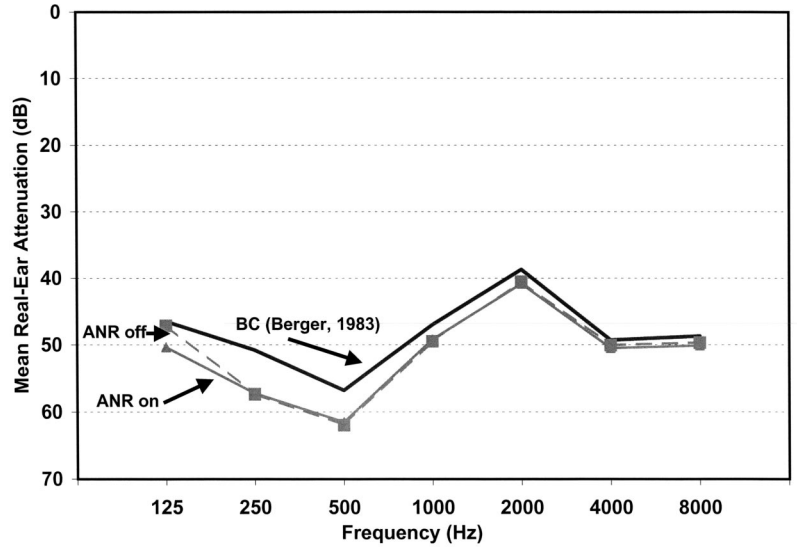


FIG. 8. Real-ear attenuation of the ANR earmuff worn in combination with a deeply inserted foam plug (DI), with the ANR turned on or turned off, as compared to Berger's (1983) prior BC estimates.

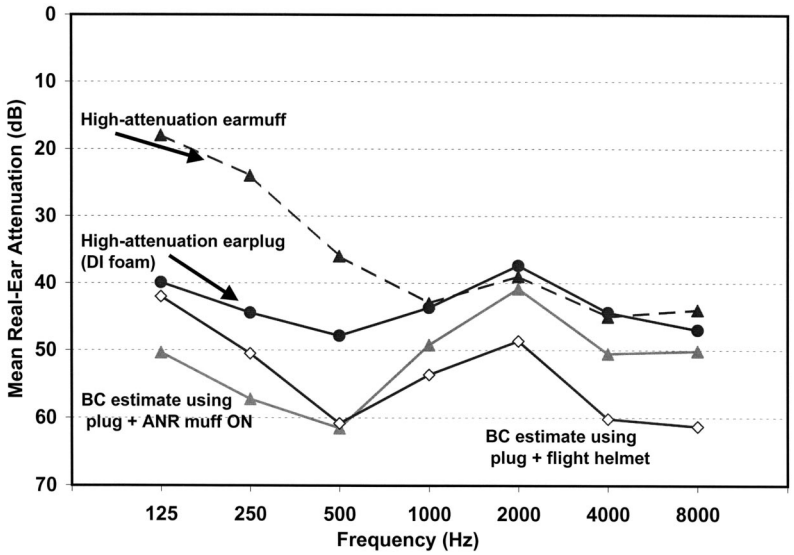


FIG. 9. Various estimates of the BC limits with head exposed versus covered, from the current study.

Zwislocki, who used very rigid deeply fitted plugs with muffs, obtained the greatest noise reduction, i.e., his are the highest estimates of the BC limits at most frequencies. However, Zwislocki relied upon only three to six subjects depending upon the test frequency and test condition. One could take the few subjects with the greatest attenuation in the current study and construct an estimate within a couple of decibels of the Zwislocki data at all frequencies. In all likelihood, the reason that the data of Nixon and von Gierke diverge from the other studies below 2 kHz is that the premolded earplugs available to them in the 1950s could not provide the high levels attenuation, especially in the low-frequencies, that are available using today's foam earplugs. (In fact Nixon and von Gierke also discounted the validity of their BC estimates below 2 kHz because of how their values compared to the prior work of von Gierke and Warren, and of Zwislocki.) The unusually low value for Schroeter and Els at 500 Hz may be attributed to their decision not to have the subjects wear an earplug underneath the attenuation boxes for the frequencies below 2 kHz. Berger's data and the current study present the highest values at 8 kHz, perhaps due to their sound field conditions that differed from the other studies.

Because of the extreme measures employed by Zwislocki, with different types of plugs used to acquire data at different frequencies, and the type of seal he achieved in the bony meatus, his data may indeed define the true BC limits with uncovered heads. However, in all likelihood his values are unachievable for groups of subjects with any single combination of wearable HPDs, regardless of how well they are fitted.

Figure 11 addresses the question of the gains to be made if we cover the head so that sound conduction pathways of the skull are blocked to varying degrees, thus limiting the flanking energy to transmission via the lower face, the neck, or the torso. In the current study we shielded the head by using a flight helmet and were able to measure data across the entire range of conventional test frequencies. This provided increases in the BC limits at and above 1 kHz, with substantial improvements of from 4 to 11 dB available at and above 1 kHz. Nixon and von Gierke accomplished the goal of shielding the head by wrapping it in ''medical cotton

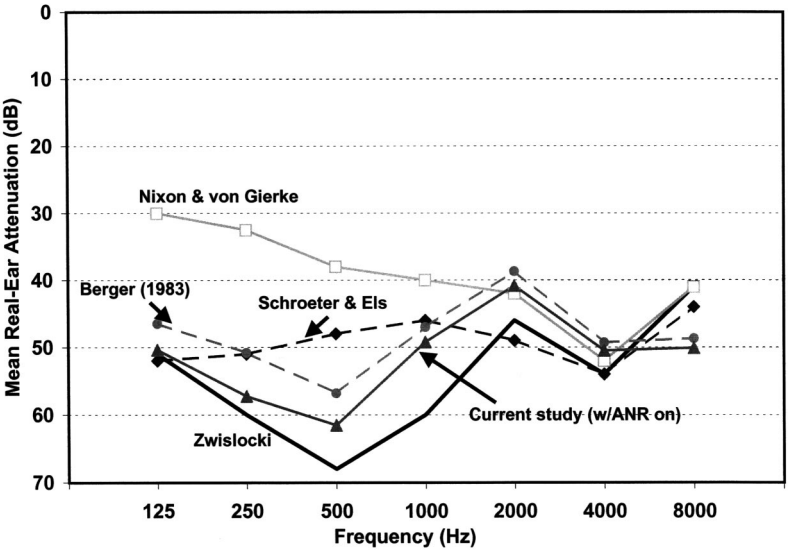


FIG. 10. Various estimates of the BC limits to attenuation with head exposed, from this study and the prior literature.

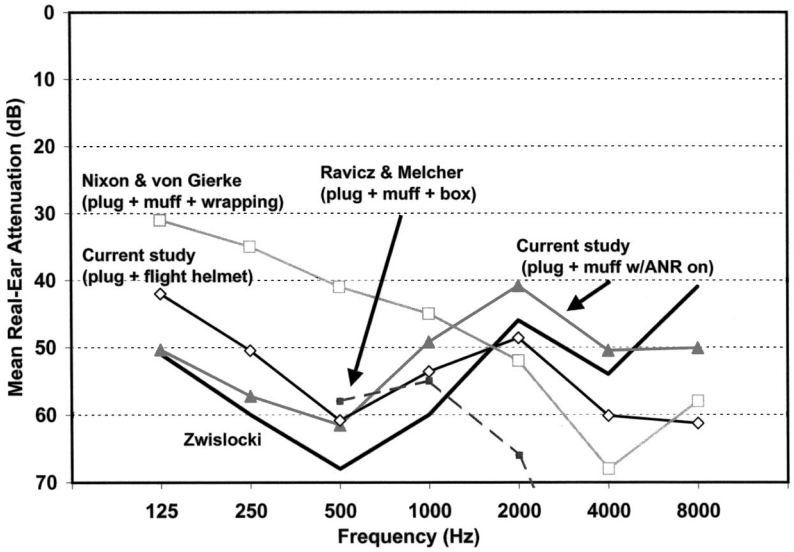


FIG. 11. Various estimates of the BC limits to attenuation with head exposed versus covered from this study and the prior literature.

wicks'' to a thickness of 2 to 3 in. As mentioned with regard to Fig. 10, below 2 kHz their data should be disregarded because of limitations of the earplugs they utilized. However, from 2 kHz and up, note in Fig. 11 the relatively close correspondence between the cotton-wrapping technique and the flight helmet used in this study, though the cotton appears to perform better at 4 kHz.

Another effect observed in the current study was the importance of covering the face (even the partial covering provided by the flight helmet) when listening in a diffuse sound field with a helmet shielding the remainder of the head. This was illustrated by testing the attenuation on one subject (the first author) with the DI foam plug plus helmet, with the visor in the up and down positions. A loss of attenuation of from 2–4 dB at and above 2 kHz was observed, suggesting that even with a helmet, facial shielding is important for maximum protection.

The most extreme approach to shielding the head was taken by Ravicz and Melcher (2001), who utilized a helmet completely separate from an earmuff worn beneath it. The helmet fully enclosed the head and chin, sealing to the neck. Even when the muff and helmet (box) were worn in combination with earplugs, the earplugs provided the dominant portion of the attenuation in the lower frequencies. Because the fit of the foam plug utilized by Ravicz and Melcher was closer to PI than DI, this led to estimates of attenuation at 500 and 1000 Hz that were unremarkable. However, at the very high frequencies where each of the three hearing protectors was an effective sound block, the noise reduction achieved by the combination became exceptional; 68 dB at 2 kHz and 82 dB at 2.8 kHz. The gain they demonstrated of about 20 dB, by ''completely'' shielding the head at 2 kHz, is in close agreement with the prediction of von Gierke and Warren (1953) which was based upon irradiating small segments of the forehead and sternum with sound from a metal tube in order to measure hearing thresholds.

A few additional points merit discussion. When maximum attenuation is desired, the concern may also arise regarding sound transmission through the nose and mouth into the eustachian tube. Would it be necessary to keep the mouth shut and plug the nose in extreme environments? Von Gierke and Nixon addressed that question by doing all testing with mouths shut, and then including a condition when the subjects heads were shrouded in cotton, of a threshold with nostrils pinched shut. The average effect was no more than about 2 dB at any frequency, although one subject experienced a 5-to-7-dB effect. We hypothesize that might have been due to a patent eustachian tube. To examine this, one subject (first author) was tested in a dual-protection mode with mouth closed and nose open, mouth closed and nose closed, and mouth open and nose open. No differences of more than 2 dB were observed at 500, 1000, 2000, and 4000 Hz.

Since the first author can also open his eustachian tubes at will, he listened in high-level broadband noise with mouth shut and nose open or pinched shut, while opening his tubes. Subjectively, attenuation was clearly affected in the high-frequency range when his tubes were open, while he alternately pinched or opened his nostrils. The effect seemed in the range of 5 dB. An attempt was made to measure this effect by tracking his $\frac{1}{3}$-octave band thresholds, but this was problematic due to an artifact. When he opened his eustachian tubes it also caused his acoustic reflex to trigger, thereby increasing his low-frequency physiological noise and to some extent masking and elevating his thresholds. The measured threshold shifts yielded an apparent 5-dB reduction in attenuation at 2 kHz and an increase at 4 kHz, but these results may have been contaminated due to the physiological-noise artifact. The effects were of the magnitude reported by Nixon and von Gierke for their ''worst case'' subject. Recall, however, that for maximum attenuation the head must be shielded, as in using a helmet, and under such conditions the nose would be shielded as well.

V. A WORD ABOUT SINGLE NUMBER ATTENUATION VALUES—NRRs AND SNRs

To place the BC limits in perspective relative to the type of attenuation factors that are often associated with HPDs, the noise reduction rating (NRR) and the single number rating (SNR) were computed for several of the measurements.

These ratings are weighted attenuation values averaged across frequencies and then adjusted to represent what 98% of the test subjects obtained in the case of the NRR (a 2-standard-deviation adjustment), and what 84% obtained (a 1-standard-deviation adjustment) in the case of the SNR (Berger, 2000). NRRs and SNRs are subtracted from C-weighted sound exposures to estimate the protected exposures in terms of an A-weighted value. For the head-not-covered conditions, the NRR for the BC limits is approximately 34 dB, and the SNR about 42 dB. With the head covered the values increase approximately 7 dB, to about 41 and 49 dB respectively.

VI. DISCUSSION

Although in most noise exposure scenarios, the protection afforded by an individual well-fitted earplug or earmuff will be sufficient, there are exposures in which more protection is either desired or required. For example, while shooting handguns, a single well-fitted high-attenuation earplug or earmuff is generally adequate, but for reasons of comfort or to reduce flinching and improve shooting, the additional protection gained from two products is desirable. In 1983 Berger explored various combinations of HPDs in detail and provided suggestions for the selection of each of the devices to be used in a dual protection scenario. In this paper we reexamined this issue to uncover the absolute maximum hearing protection that can be provided, in order to address current and new military systems as well as exceptional occupational situations with extreme noise exposures. And we also asked, ''if the dual protection isn't enough, is there anything that can be done about it?''

We have reviewed the handful of extant studies published during the past 50 years and added recent data of our own, to provide insight into the limits to attenuation. The answers are summarized in Figs. 10 and 11 and in Table I.

For a head-not-covered condition at frequencies from 125 Hz to 4 kHz the best estimate of the limits to protection is the range of values encompassed by the Berger (1983) data and the Zwislocki (1957) results, though it is unlikely that with any single set of plugs and muffs that the Zwislocki data can be achieved. For a realistic estimate, in a diffuse field with a single pair of well-fitted muffs and plugs, a better choice would be encompassed by the data of Berger (1983) and the current study. These values range from 45 to 60 dB, except at 2 kHz where a minimum is observed that hovers around 40 dB. At 8 kHz the Berger and recent estimates provide the highest measured BC limits, and also those most appropriate for diffuse sound fields.

The prior work of Nixon and von Gierke (1959) and the recent test on one subject by the current authors suggests that whether or not the nostrils are open or closed is unimportant. However, anecdotal evidence provided by McKinley (2002) raises another issue. He has observed in high-level steady-sound fields of approximately 140 dB SPL that those wearing dual protection notice a difference in the perceived sound level depending on whether they have their mouths shut, with jaws either relaxed or clenched. He observed that clenching the teeth increases the sound transmission, i.e., reduces the attenuation. The authors of the current paper were unable to confirm this finding based on two of our authors listening in a 90-dB SPL diffuse sound field, but the third author did experience a small effect of clenching his jaws, estimated to be in the range of 3 to 4 dB in the middle frequencies.

For the absolute maximum in attenuation, the head must be acoustically shielded in addition to wearing dual protection. Shielding with a lightweight acoustically leaky barrier such as a safety helmet has no effect. What is required is a tight fitting helmet, like a standard military flight helmet, that encloses the entire skull and face. In such cases, one can achieve the values in Fig. 11 as obtained in this study with the DI foam plug plus flight helmet. Note that this requires the visor be down, otherwise 2–4 dB of protection may be sacrificed in the upper frequencies. Figure 11 also shows that with the highest-attenuating plug and muff combination (head not covered), attenuation is lowest at 2 kHz. Since extreme noise levels tend to be dominated by energy at higher frequencies, increasing the attenuation of noise reaching the skull should be a primary objective in providing protection in these environments. The authors are not aware of much engineering effort to date to develop a head and face noise-shielding helmet that can be practically worn (i.e., while addressing comfort considerations). Clearly, better shielding by a more complete helmet can potentially achieve much improved protection in the high frequencies, as demonstrated by Ravicz and Melcher (2001).

An estimate of the ''best possible protection,'' i.e., with the ears plugged and the head and neck fully enclosed, can be made by combining the low-frequency limits (125–500 Hz) for the DI foam plug plus ANR earmuff in the on-mode, with the Ravicz and Melcher data for 1 and 2 kHz. To extend the estimate upwards in frequency, we make the conservative presumption that if 82 dB can be achieved at 2.8 kHz, then at least 75 dB can be obtained at 4 and 8 kHz. NRRs and SNRs computed from such data using the standard deviations for the DI foam plus flight helmet (see Table I) are 46 and 55 dB, respectively. This represents an increase of 12 to 13 dB over the DI foam plug plus ANR earmuff. In an actual noise spectrum typical of that to which crew are exposed while launching modern Navy jets on an aircraft carrier (a relatively flat spectrum with a difference of about +1 dB between the C-weighted and A-weighted sound levels), the noise reduction computed for 84% of the test population (a minus 1-standard-deviation adjustment) is about 41 dB for the DI foam plug plus ANR earmuff. The computed noise reduction for the best possible protection (see Table I) with a complete head enclosure is 51 dB. Whether or not such a ''theoretical'' protector could ever be fielded and worn in practice, it provides a benchmark for the limits to protection.

Finally, we turn to the thorny question of the real world. All of the forgoing measures were based upon compliant and well-trained subjects, willing to undergo some degree of discomfort while optimally fitting perfectly functioning hearing protection or special test devices, for short periods of time under pristine laboratory conditions. Temperatures were moderate and no exertion was required. When occupational situations or adverse military conditions, such as servicing jet aircraft with noise levels of 150 dB, are repeatedly en-

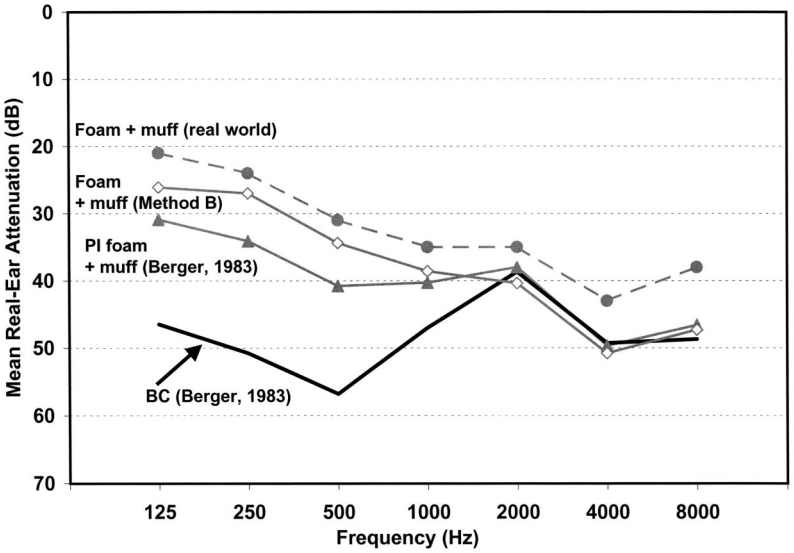


FIG. 12. Comparison of laboratory-estimates of the BC limits versus the actual protection in the real world for a foam plug plus small earmuff.

countered throughout a workshift, the situation is very different. Little indication of the real-world performance of maximum-attenuation dual hearing protection systems is available based on field studies, except for one report by Hachey and Roberts (1983) based on ten employees in a textile plant. Another indication is provided by testing to the new American National Standard S12.6-1997 under its Method-B procedure, which uses subject fit with naïve listeners to provide an estimate of field performance. Both of these data sets are compared in Fig. 12 to the PI foam plus small-volume muff data from Berger (1983) as well as to the Berger estimate of BC, which is the most conservative of the estimates provided in Fig. 10.

Both the real-world measurement and the real-world estimate (Method B) demonstrate attenuation up through 1000 Hz that is substantially less than even the PI foam plug under an earmuff, let alone the BC limits. Above 1000 Hz the disparity is somewhat reduced. Nevertheless the message is clear: approaching the laboratory BC limits in actual occupational situations is going to require very serious motivation, training, supervision, and enforcement. Whether it can be achieved is open to question.

As we move forward to examine new technologies we should be well grounded in the learnings from existing data and experiments. For example, ANR technology in earmuffs with communication systems provides certain advantages, but in terms of maximum protection when combined with a well-fitted earplug, there appears little to be gained. Efforts are moving forward to develop custom-molded earplugs with built-in transducers for communication, or ANR earplugs, or even helmets that might cancel skull vibrations. One should be aware of the limitations. Custom inserts may provide enhanced comfort and in some cases approach the attenuation of deeply inserted foam earplugs, but they will not exceed it. Also, alternative technology using foam eartips already exists and has been shown to afford excellent performance (Ribera *et al.*, 1996). ANR earplugs may extend the effectiveness of noise cancellation into higher frequencies than ANR headsets, but this is unlikely to have an important effect on the high-frequency BC limits, which Khanna *et al.* (1976) showed are due to flanking pathways that are largely outside the reach of earcanal-based systems. Helmet-based ANR systems that cancel bone vibration will probably be very complex to devise, and although they may produce some reductions in the BC-transmitted energy in the high frequencies, as can be seen in Fig. 4, even the best of helmets provide little attenuation at the frequencies below 1000 Hz. Furthermore, above 1000 Hz, the simple shielding that a flight helmet currently offers may be all that is needed, unless one takes the next step of fully encasing the chin and neck (see Fig. 11).

In the meantime, the best gains available to us today may well be in the motivation, training, and supervision of the users of hearing protection to make sure they get the most out of the devices that are currently available.

ACKNOWLEDGMENTS

The published works of the many authors cited have been instrumental in shaping our thinking, research, and the discussions in this paper. Additionally, reviews of early drafts of the manuscript and feedback from Mike Ravicz of the Massachusetts Eye and Ear Infirmary, Charles Nixon formerly of Wright–Patterson AFB, Mead Killion of Etymotic Research, Inc., and Armand Dancer of the Franco-German Institute of Saint-Louis (ISL) were helpful in crafting the final manuscript.

[1]REAT measurements are conducted at low sound levels that never exceed about 80 dB SPL even for HPDs with maximum values of attenuation. Thus, the BC pathways are stimulated within the range in which they have been shown to behave linearly (Khanna *et al.*, 1976). However, in the extreme 150-dB environments in which maximum dual protection is warranted, one might ask whether the BC response is still linear. We are aware of no data that bear on this question. In all likelihood, if the responses are nonlinear, the effect would be compressive since mechanical systems that are overdriven tend to produce less output, i.e., as the sound field excitation increases, the increase in the BC response does not keep pace. In this case, the cochlear excitation would be less than predicted by simply subtracting the REAT-measured attenuation of the HPD from the noise exposure. Therefore, any errors arising from application of the method of this report would tend in a conservative direction, meaning that noise-induced hearing loss associated with sound transmitted via the BC pathways, while hearing

protection is being worn, would be expected to be no more severe, but possibly less severe than otherwise anticipated. Experiments are underway to address these questions (McKinley *et al.*, 2003)

[2]For those familiar with threshold procedures such as REAT measurements it might seem problematic to conduct an REAT measurement on an ANR system, since active devices emit low residual noise levels even when operating in extremely low ambient noise levels. This was not a problem, however, in the measurement of our dual-protection system as discussed in Sec. III A.

[3]Subsequent to the experiments conducted for this paper an additional ten-subject study was completed to evaluate the effects of ANR when worn with a PI foam plug. This was to answer questions of ANR performance in dual-protection scenarios, not to address the focus of the current study. As anticipated due to the lower attenuation of the PI fitting in the low and middle frequencies, the benefit of ANR was more pronounced at 125 Hz (7 dB vs. 3 dB with the DI fitting) and was still statistically significant at 250 Hz showing a 2-dB effect. As in the DI condition no benefits were observed above 250 Hz.

[4]Note the unexpected observation that in the on-mode, the ANR muff-plus-plug combination outperforms the helmet-plus-plug combination at 125 and 250 Hz. The same applies in the off-mode for the ANR muff-plus-plug versus the helmet-plus-plug (see Table I). This is true, even though as can be noted in Table I, the attenuation of the ANR muff in the off-mode, and the helmet, are essentially identical at those two frequencies. Our hypothesis is that at the low frequencies, the helmet acts as an acoustic antenna that couples the sound-field vibrations more closely to the earmuff cup than is the case for conventional circumaural cups directly exposed to a sound field, and this bone-conducted energy flanks the earplug. At the frequencies above 500 Hz, the foam/Velcro pads that affix the cups to the helmet shell effectively decouple the helmet from the cup, allowing for increased attenuation.

ANSI (**1957**). Z24.22-1957 (R1971), ''Method for the Measurement of Real-Ear Attenuation of Ear Protectors at Threshold'' (American National Standards Institute, New York).

ANSI (**1974**). S3.19-1974 (ASA STD 1-1975), ''Method for the Measurement of Real-Ear Protection of Hearing Protectors and Physical Attenuation of Earmuffs'' (American National Standards Institute, New York).

ANSI (**1995**). S12.42-1995, ''Microphone-in-Real-Ear and Acoustic Test Fixture Methods for the Measurement of Insertion Loss of Circumaural Hearing Protection Devices'' (American National Standards Institute, New York).

ANSI (**1997**). S12.6-1997, ''Methods for Measuring the Real-Ear Attenuation of Hearing Protectors'' (American National Standards Institute, New York).

Berger, E. H. (**1983**). ''Laboratory Attenuation of Earmuffs and Earplugs Both Singly and in Combination,'' Am. Ind. Hyg. Assoc. J. **44**(5), 321–329.

Berger, E. H. (**1985**). ''Is real-ear attenuation at threshold a function of hearing level?'' J. Acoust. Soc. Am. **78**(5), 1588–1595.

Berger, E. H. (**2000**). ''Hearing Protection Devices,'' in *The Noise Manual*, 5th ed., edited by E. H. Berger, L. H. Royster, J. D. Royster, D. P. Driscoll, and M. Layne (Am. Ind. Hyg. Assoc., Fairfax, VA), pp. 379–454.

Berger, E. H., Franks, J. R., Behar, A., Casali, J. G., Dixon-Ernst, C., Kieper, R. W., Merry, C. J., Mozo, B. T., Nixon, C. W., Ohlin, D., Royster, J. D., and Royster, L. H. (**1998**). ''Development of a new standard laboratory protocol for estimating the field attenuation of hearing protection devices. Part III. The validity of using subject-fit data,'' J. Acoust. Soc. Am. **103**(2), 665–672.

Berger, E. H. and Kerivan, J. E. (**1983**). ''Influence of physiological noise and the occlusion effect on the measurement of real-ear attenuation at threshold,'' J. Acoust. Soc. Am. **74**(1), 81–94.

Gauger, D. (**2002**). ''Active Noise Reduction (ANR) and Hearing Protection: Where It's Appropriate and Why,'' NHCA *Spectrum* Supplement 1.

Hachey, G. A., and Roberts, J. T. (**1983**). ''Real World Effectiveness of Hearing Protection,'' Am. Ind. Hyg. Conf., Abstract #462, Philadelphia, PA.

Khanna, S. M., Tonndorf, J., and Queller, J. E. (**1976**). ''Mechanical parameters of hearing by bone conduction,'' J. Acoust. Soc. Am. **60**(1), 139–154.

McKinley, R. (**2002**). Personal communication.

McKinley, R., Dancer, A., and von Gierke, H. (**2003**). ''Bone and tissue conduction of high intensity acoustic energy to the human cochlea,'' J. Acoust. Soc. Am. **113**, 2239(A).

Nixon, C. W., and von Gierke, H. E. (**1959**). ''Experiments on the bone-conduction threshold in a free sound field,'' J. Acoust. Soc. Am. **31**(8), 1121–1125.

Ravicz, M. E., and Melcher, J. R. (**2001**). ''Isolating the auditory system from acoustic noise during functional magnetic resonance imaging: Examination of noise conduction through the ear canal, head, and body,'' J. Acoust. Soc. Am. **109**(1), 216–231.

Ribera, J. E., Mozo, B. T., and Murphy, B. A. (**1996**). ''Subjective Evaluation of the Communications Earplug with Flexible Harness (CEP/FH) among CH-47D Crewmembers,'' U.S. Army Aeromedical Res. Lab. Rept. No. 96-29, Fort Rucker, AL.

Schroeter, J., and Els, H. (**1980**). ''The Acoustic Properties of the Human Head (in German),'' Wirtschaftsverlag NW, Bremerhaven, Fed. Rep. of Germany, ISBN 3-88314-112-4.

Schroeter, J., and Poesselt, C. (**1986**). ''The use of acoustical test fixtures for the measurement of hearing protector attenuation. Part II: Modeling the external ear, simulating bone conduction, and comparing test fixture and real-ear data,'' J. Acoust. Soc. Am. **80**(2), 505–527.

Tonndorf, J. (**1972**). ''Bone Conduction,'' in *Foundations of Modern Auditory Theory, Vol. II*, edited by J. V. Tobias (Academic Press, New York), pp. 197–237.

von Békésy, G. (**1960**). *Experiments in Hearing* (McGraw-Hill, New York), pp. 177–181.

von Gierke, H. E., and Warren, D. R. (**1953**). ''Protection of the Ear From Noise: Limiting Factors,'' Benox Report, Contract N6 orl-020 Task Order 44, Univ. of Chicago.

Zwislocki, J. (**1957**). ''In search of the bone-conduction threshold in a free sound field,'' J. Acoust. Soc. Am. **29**(7), 795–804.

An angle-by-angle approach to predicting broadband high-frequency sound fields in rectangular enclosures with experimental comparison

Linda P. Franzoni
Department of Mechanical Engineering and Material Science, Duke University, Durham, North Carolina 27708-0300

Christopher M. Elliott
Caterpillar, Inc., Peoria, Illinois

(Received 9 October 2002; revised 24 July 2003; accepted 28 July 2003)

Experiments were performed on an elongated rectangular acoustic enclosure with different levels of absorptive material placed on side walls and an end wall. The acoustic source was a broadband high-frequency sound from a loudspeaker flush-mounted to an end wall of the enclosure. Measurements of sound-pressure levels were averaged in cross sections of the enclosure and then compared to theoretical results. Discrepancies between the experimental results and theoretical predictions that treated all incidence angles as equally probable led to the development of an angle-by-angle approach. The new approach agrees well with the experimentally obtained values. In addition, treating the absorptive material as bulk reacting rather than point reacting was found to significantly change the theoretical value for the absorption coefficient and to improve agreement with experiment. The new theory refines an earlier theory based on power conservation and locally diffuse assumptions. Furthermore, the new theory includes both the angle of incidence effects on the resistive and reactive properties of the absorptive material, and the effects of angle filtering, i.e., that reflecting waves associated with shallow angles become relatively stronger than those associated with steep angles as a function of distance from the source.
[DOI: 10.1121/1.1610455]

PACS numbers: 43.55.Br, 43.55.Ka [MK]

I. INTRODUCTION

Broadband high-frequency sound fields in regularly shaped rooms with evenly distributed absorption are often assumed to have spatially uniform reverberant fields. However, for rooms that are less cubic and more oblong, even if the absorptive material is distributed uniformly at the boundaries, the levels of mean-square pressure are not spatially uniform. Several recent studies[1,2] have derived formulas for the spatial variation of cross-sectional averages of mean-square pressure in these types of rooms. In Ref. 1, Franzoni and Labrozzi performed numerical simulations of an exact solution for the reverberant field in rectangular rooms (of different dimensions) excited by broadband high-frequency sources on an end wall. The rooms had different amounts and distributions of absorptive materials on their boundaries. By collapsing data from many numerical simulations, Franzoni and Labrozzi developed an empirical formula for the reverberant mean-square pressure averaged over a cross section.

In Ref. 2, Franzoni derived a theoretical formula for the cross-sectional average mean-square pressure as a function of distance, power, surface areas, and absorption coefficients of enclosure boundaries. This theoretical formula was based upon energy conservation applied to cross sections in the room. In that work, the theoretical formula was compared to the numerical simulations and the empirical formula of Ref. 1 and shown to agree reasonably well for small to moderate levels of absorption.

In this work, experiments were performed on a 16×24×60-in. rectangular acoustic enclosure; see Figs. 1 and 2. A small loudspeaker (dome tweeter) was placed on one of the end walls. The two side walls, and sometimes the end wall opposite the source, were covered with different types of absorbing foam. The loudspeaker was driven by white noise over a broad frequency range. Experimental measurements of the mean-square pressure at various locations inside the acoustic enclosure were made using a specially designed microphone probe tube.

The purposes of this experiment were to compare the measured sound-pressure levels with predictions from previously reported formulas and to recommend the most accurate formulas and the conditions for which they apply. For cases with light to moderate absorption, both of the previous theories agreed reasonably well with the experimental data. For cases with more significant amounts of absorption, however, the manner in which the theory and experiment did not agree led to insights that were then used to develop a more refined theory. The capability to predict the slow spatial variation of cross-sectionally averaged mean-square pressure in long, slender enclosures without having to perform extensive calculations is important for various interior noise applications.

II. EXPERIMENTAL DESCRIPTION

Sound-pressure levels in a rectangular enclosure excited by broadband high-frequency sound waves were measured experimentally. The experimental setup consisted of a rigid enclosure, a loudspeaker, two probe microphones, a two-

 0001-4966/2003/114(4)/1968/12/$19.00

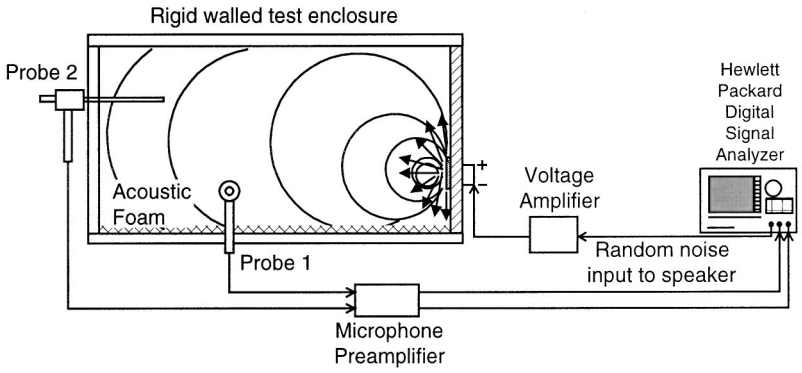


FIG. 1. Experimental setup.

channel digital signal analyzer which also functioned as a signal generator, a voltage amplifier, a microphone preamplifier, and various types of absorptive foam placed along certain enclosure walls. A schematic of the experimental apparatus is shown in Fig. 1.

The separate components of the experiment were characterized individually. For example, the directivity pattern of the loudspeaker sound source was obtained experimentally and found to be approximately omnidirectional in the frequency range of interest. A transfer function was obtained between the applied voltage and the emitted power of the loudspeaker. Similarly, another transfer function was obtained to relate the measurements of a bare microphone to those of a microphone with a probe-tube attachment. Normal incidence impedance measurements for the absorptive materials used were made using a standing wave tube device.

A. The enclosure

With the exception of the end wall in which the loudspeaker was mounted, all of the walls of the enclosure were made of 1-in.-thick Lucite. The dimensions of the rectangularly shaped enclosure were $16\times24\times60$ in. The source wall (16×24 in.) consisted of a $\frac{3}{4}$-in.-thick wooden baffle covered by 3/8-in.-thick Plexiglas. All walls were bolted together and their interfaces coated with silicone sealant.

In order to insert a pressure probe at various locations within the enclosure, holes were drilled and tapped (1/4–20) at hundreds of locations around the boundaries. While not in use, the holes were plugged with 1-in.-long hex screws. When a hole was used for taking measurements, the hex screw was replaced by a bolt with a through-hole drilled in the center. The hole's inner diameter was a ''slip fit'' for the probe tube's outer diameter. Figure 2 shows a photograph of the enclosure with a typical configuration of absorptive material. The hole locations are evident by the dark hex screws.

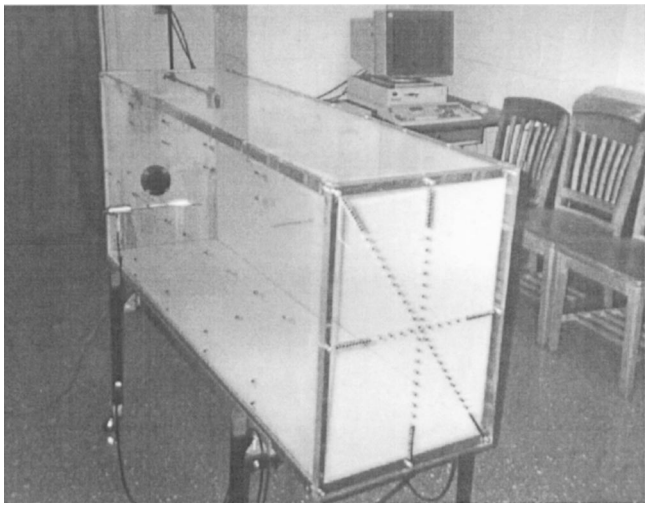


FIG. 2. Typical example of enclosure configuration for experiments. Damping configuration: material with medium absorption coefficient on the bottom and one side of the interior.

B. Broadband sound source

The loudspeaker used in these experiments was selected on the basis of size, directional properties, and frequency response. A small-diameter speaker cone was chosen in order to minimize the fraction of the end wall occupied by the loudspeaker and to optimize performance in the mid-to-high frequency range (roughly 4–8 kHz). The speaker had a fairly uniform frequency response (approximately constant level output for a constant input voltage) over the frequency range of interest. The speaker used was a 1-in. dome tweeter.

Sound-pressure levels were measured in an anechoic chamber at different angular orientations from the speaker and were found to be relatively constant with position, i.e., there was no particular directional preference of the speaker. An experimentally determined transfer function was obtained between the voltage across the lead wires of the speaker and the sound power emitted from the speaker. The power was deduced from measurements at fixed radii from the speaker using an omnidirectional theoretical model for the speaker in a free-field environment.

C. Probe-tube attachment

A special probe-tube attachment for a $\frac{1}{2}$-in. microphone was used in order to measure sound-pressure levels in the interior of the enclosure without being intrusive. The outer diameter of the probe tube was 1/8-in. The design of the probe-tube attachment has been described in Ref. 3. A special feature of this probe tube is its relatively flat transfer function with respect to frequency.

D. Impedance of absorbing materials

The normal incidence impedance was measured using a standing wave tube device at different frequencies for each absorbing material. The reactive and resistive components of

acoustic impedance were calculated from the standing wave pattern. These values were originally used to directly obtain random incidence absorption coefficients, assuming point-reacting behavior of the absorbing material. However, it was later determined that a better estimate of the absorption coefficients was needed. This issue is discussed further in Sec. III. As explained in Appendix A, better estimates of the absorption coefficients were produced when the normal incidence impedance measurements were used to curve fit a porous media model of the absorptive material, thereby allowing the inclusion of bulk-reaction effects.

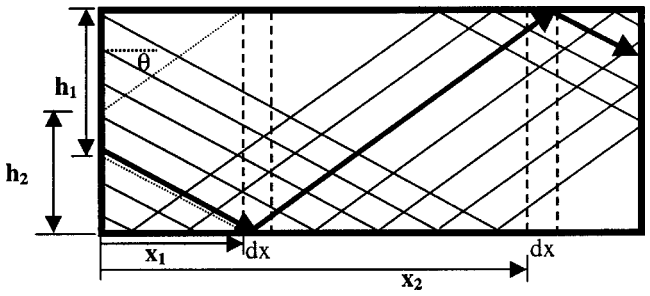


FIG. 3. Two-dimensional enclosure with sound field corresponding to a single angle θ.

III. THEORETICAL MODEL

In previous work (Refs. 1 and 2), theoretical models were developed to predict the cross-sectionally averaged sound-pressure levels in enclosures similar to the current apparatus. However, comparisons between the experimental results and the theoretical formulas for cross-sectional averages along trajectories from the source wall to the opposite end wall yielded some discrepancies. The agreement was improved by recalculating the random incidence absorption coefficient using a porous media model with bulk reaction effects, as explained in Appendix A, and suggested by Refs. 4 and 5. It was observed that the theoretical curves with the revised absorption coefficients still exhibited spatial decay that was too rapid away from the source for more absorptive cases and too slow for the least absorptive case.

A possible reason for this remaining discrepancy was that the theoretical model of Ref. 2 assumed that all waves were of equal strength from all directions in a given cross section of the enclosure and that all orientations were equally probable. In actuality, the shallower angles would be of greater strength farther away from the source than the steeper angles, since the steeper angles have undergone many more damped reflections before arriving at the same distant cross section. Also, the theory removed energy from each cross section using a *random-incidence* absorption coefficient, whereas at a given cross section not all waves are of equal strength, nor are all wave angles equally probable.

Thus, the original simple theory was refined in order to account for the effects of incidence angle on both the number of reflections and the effectiveness of the absorptive material. Certain features of the original theory were retained. Waves were still treated in an average sense and assumed to be equally distributed over a given cross section on a *per angle* basis. The original theory and its refined version predict *average* levels at any cross section of the enclosure. Since the sound field is a high-frequency broadband field, it was assumed that the right- and left-traveling wave fields were uncorrelated, i.e., they could be analyzed separately and superimposed. Waves having different incidence angles were also assumed to be uncorrelated.

The quantity of interest is the cross-sectional average of sound-pressure level at various stations in a given enclosure. Fine details regarding spatial variation within the cross section, including the effects of source position on an end wall, are not included in this refinement nor in the previous theory. Therefore, it was assumed that the source wall consists of a distribution of sources that cumulatively represents whatever actual sources are on the end wall. The distributed sources radiate energy equally at all angles, but their effects are considered on an angle-by-angle basis.

A. The general idea of an angle-by-angle theory

The theory will be developed for a three-dimensional enclosure; however, for the purposes of conveying the general idea of an angle-by-angle theory, consider the two-dimensional enclosure shown in Fig. 3. Only a representative sample of waves leaving the source wall in the downward direction is shown, but there would also be a comparable set of waves leaving the wall in the upward direction. The purpose of the theory is to predict the spatially averaged sound-pressure level in a cross section of an enclosure that is long relative to its cross-sectional dimensions and that has a sound source or sources on one end wall. Consistent with that motivation, the source or sources are considered to be equally distributed over the end wall and additionally, the power coming into any cross section is considered to be averaged over the cross section.

The left face of the enclosure in Fig. 3 consists of a continuous distribution of infinitesimal sources that send out sound waves at a prescribed angle θ. This face can be thought of as an end wall, or as any cross section of the enclosure, the only difference being the strength of the sources. At a distance x_1 from the sources, only those sources lying within the section labeled h_1 will result in direct sound into the cross section from the initially downward waves (the section labeled h_2 corresponds to the equivalent set of sources for the upward-leaving waves, which are not shown). At x_1, all other sources will contribute reflected, and therefore weaker, sound waves. The resulting variation in energy in the cross section is accounted for only through the change in cross-sectional average energy. Farther down the enclosure, at the cross section labeled x_2, all waves (at fixed θ) will have undergone at least one reflection before entering the cross section. These effects are included in the angle-by-angle theory. In the theoretical development that follows, an energy balance is performed for a generic cross section of width dx. A differential equation is derived from a power balance, where the power into the cross section is from the previous cross section (accounting for direct and reflected sound waves for a given θ) and the energy leaving is diminished from this amount by the absorbed power at the side wall. Note, this theory accounts for all reflections and is not just a first reflection theory.

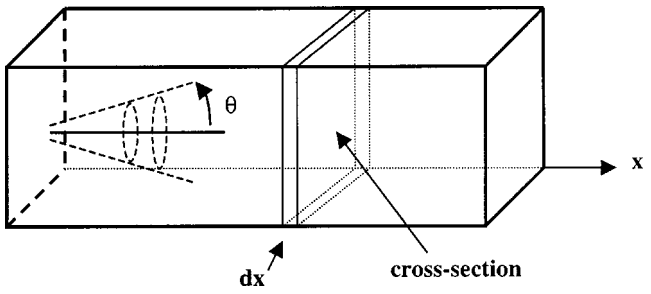


FIG. 4. Model for the enclosure showing a representative source.

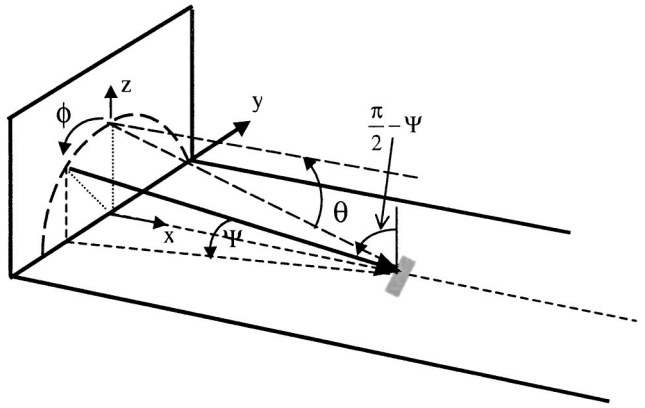


FIG. 5. Sources contributing to the mean-square pressure at a point on the sidewall for a given cone angle θ. Source may be located along an arc (ϕ). Angle of incidence with sidewall is Ψ.

B. Radiated power from the end wall

The end wall is modeled as being covered with a distribution of omnidirectional sources, replacing the original discrete source in the experiment. For a given source, there is a hemisphere of radius a through which the power travels, where $a \gg \lambda$, the nominal wavelength. By treating the waves as locally planar to calculate intensity and then integrating over the surface area of the wall in order to relate the entire power of the sources W to the mean-square pressure $\overline{p^2}$ (where $\overline{p}$ represents the root-mean-square)

$$W = \frac{\overline{p^2}}{\rho_o c} 2\pi a^2, \tag{1}$$

where $\rho_o c$ is the characteristic acoustic impedance of air. The differential power in an area element that forms a ring of the hemisphere is

$$dW = \frac{\overline{p^2}}{\rho_o c} 2\pi (a \sin\theta)(a\, d\theta), \tag{2}$$

where θ is the angle from the surface normal to the ring location in spherical coordinates. The derivative of the differential power with respect to angle is

$$\frac{dW}{d\theta} = \frac{\overline{p^2}}{\rho_o c} 2\pi (a \sin\theta)(a) = W \sin\theta \equiv W_\theta. \tag{3}$$

The symbol W_θ will be used in the subsequent derivation to denote power density per angle. For the theoretical derivation, the sound waves associated with cone-angle θ will be considered separately and superimposed. Recall that the end wall is modeled as a distribution of sources, even though only one is shown in Fig. 4. This modeling assumption is consistent with the approach of considering cross-sectional averages.

C. Basic relationships: Intensity components, mean-square pressure

The sound field can be analyzed as separate left- and right-traveling wave fields. The component of intensity that is associated with the right-traveling wave field and incident upon the vertical plane of a cross section shown in Fig. 4 is given by

$$\mathbf{I}^{(+)}_{x(\theta)} = \frac{\overline{p^{2(+)}_\theta}}{\rho_o c} \cos\theta, \tag{4}$$

where both the mean-square pressure per angle and the intensity per angle are only associated with right-traveling ("+") waves.

In order to perform the power balance, the component of intensity that is incident upon a side wall is also needed. To obtain the component of intensity that is associated with the right-traveling wave field and incident upon a side wall, it is more convenient to change reference frames and to look back at the source wall (or at an appropriate intermediate cross section after multiple reflections) from a position on the side wall, as in Fig. 5. From this vantage point, only certain sources on the source wall will contribute for a particular θ. In three dimensions, the angle of incidence of the sound wave impinging on an incremental surface area from a differential source is ψ. The amplitude of the intensity incident on the side wall, $\mathbf{I}_{sw(\theta,\phi)}$, is dependent upon the source position along the arc (angle ϕ) that is subtended on the source wall. This relationship is given by

$$\mathbf{I}^{(+)}_{sw(\theta,\phi)\text{''in''}} = \frac{\overline{p^{2(+)}_{\theta,\phi\text{''in''}}}}{\rho_o c} \sin\psi = \frac{\overline{p^{2(+)}_{\theta,\phi\text{''in''}}}}{\rho_o c} \sin\theta \cos\phi, \tag{5}$$

where the mean-square pressure and intensity are angular densities in both ϕ and θ and the subscript "in" indicates that both the intensity and the mean-square pressure components are right traveling and *into* the side wall corresponding to the downward direction in Fig. 5. The angle ψ is the incident angle measured from the surface, and is related to θ and ϕ.

The right-traveling component of mean-square pressure consists of both waves that are incident on the side wall and those that are reflected from the side wall

$$\overline{p^{2(+)}_{\theta,\phi}} = \overline{p^{2(+)}_{\theta,\phi\text{''in''}}} + \overline{p^{2(+)}_{\theta,\phi\text{''ref''}}}. \tag{6}$$

The mean-square pressure reflected ("ref") from an absorbing wall is of less amplitude than the mean-square pressure into the wall by a factor of $(1-\alpha_w)$, where the absorption coefficient, $\alpha_w = \alpha_w(\psi) = \alpha_w(\theta,\phi)$. The absorption coefficient relates the reflected to the incident pressure

$$\overline{p^{2(+)}_{\theta,\phi\text{''ref''}}} = (1-\alpha_w)\overline{p^{2(+)}_{\theta,\phi\text{''in''}}}. \tag{7}$$

Equation (6) becomes

$$\overline{p}^{2(+)}_{\theta,\phi}=(2-\alpha_w)\overline{p}^{2(+)}_{\theta,\phi\text{``in''}}. \tag{8}$$

This equation relates the average mean-square pressure density in the cross section to the component of mean-square pressure density incident on the side wall. Assuming that these are constant in ϕ for a given θ (consistent with cross-sectional averaging) and integrating over ϕ from $-\pi/2$ to $\pi/2$

$$\overline{p}^{2(+)}_{\theta}=\pi(2-\hat{\alpha}_w)\overline{p}^{2(+)}_{\theta,\phi\text{``in''}}. \tag{9}$$

When considering the relationship between an incident and a reflected wave, only one reflection needs to be analyzed. However, in the theory, multiple reflections are accounted for because the amplitude of the incident wave depends on the previous reflections that may have occurred, and this enters into the mathematics through the power balance on a differential cross-sectional element (developed in the next subsection). The resulting amplitude of mean-square pressure for the cross section is obtained by solving the forthcoming differential equation.

The absorption coefficient $\hat{\alpha}_w$ is the ϕ-averaged absorption coefficient for a given θ

$$\hat{\alpha}_w=\frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\alpha_w(\psi(\theta,\phi))d\phi. \tag{10}$$

The absorption coefficient and its functional dependence on the generalized impedance of the absorptive material will be discussed in more detail following this derivation.

Implicit in the above development is an assumption that the mean-square pressure near the absorbing surfaces can be approximated by the cross-sectional average mean-square pressure. This assumption was also made in the earlier work of Ref. 2, and the Appendix of that paper shows that the variation in mean-square pressure over the cross section is on the order of approximately $0.1\alpha^2$. Therefore, this approximation is accurate even for large absorption coefficients due to the small coefficient in front of α^2.

As a function of source cone angle θ, the relationship between the intensity into the side wall and the mean-square pressure is [from Eqs. (5) and (9)]

$$\mathbf{I}^{(+)}_{\text{sw}(\theta,\phi)\text{``in''}}=\frac{\overline{p}^{2(+)}_{\theta}}{\pi(2-\hat{\alpha}_w)\rho_o c}\sin\theta\cos\phi. \tag{11}$$

D. Power balance in a cross section

Considering a typical cross section of the enclosure, the power that leaves the cross section is equal to the power that arrives minus the power that is absorbed at the boundaries of the cross section. For a fixed angle, the differential equation in terms of power is

$$W^{(+)}_{\theta,\text{``in''}}-W^{(+)}_{\theta,\text{``out''}}=W^{(+)}_{\theta,\text{``in''}}-\left(W^{(+)}_{\theta,\text{``in''}}+\frac{d}{dx}W^{(+)}_{\theta,\text{``in''}}\Delta x\right)$$

$$=-\frac{d}{dx}W^{(+)}_{\theta,\text{``in''}}\Delta x=W^{(+)}_{\theta,\text{``abs''}}, \tag{12}$$

where the equation is applied on an angle-by-angle basis; thus, the powers are angular densities. Note that the power absorbed per angle $\theta(W_{\theta,\text{``abs''}})$ is related to the component of intensity ''into'' the side wall, namely

$$W^{(+)}_{\theta,\text{``abs''}}=\int_{-\pi/2}^{\pi/2}\mathbf{I}^{(+)}_{\text{sw}(\theta,\phi)}\beta L_p\Delta x\alpha_w(\psi(\theta,\phi))d\phi, \tag{13}$$

where β is the fraction of the perimeter L_p that is covered by the absorptive material and α_w is the absorption coefficient of the material, dependent upon the angle of incidence. In terms of the mean-square pressure density, this can be written as

$$W^{(+)}_{\theta,\text{``abs''}}=\int_{-\pi/2}^{\pi/2}\frac{\overline{p}^{2(+)}_{\theta}\sin\theta\cos\phi}{\pi(2-\hat{\alpha}_w)\rho_o c}\beta L_p\Delta x\alpha_w(\psi(\theta,\phi))d\phi. \tag{14}$$

The absorption coefficient that represents the fraction of power absorbed per cone angle θ is defined as

$$\tilde{\alpha}_w\equiv\frac{\int_{-\pi/2}^{\pi/2}\overline{p}^{2}_{\theta,\phi}\alpha_w(\psi)\cos\psi\,d\phi}{\int_{-\pi/2}^{\pi/2}\overline{p}^{2}_{\theta,\phi}\cos\psi\,d\phi}$$

$$=\frac{\overline{p}^{2}_{\theta,\phi}\sin\theta\int_{-\pi/2}^{\pi/2}\alpha_w(\psi)\cos\phi\,d\phi}{\overline{p}^{2}_{\theta,\phi}\sin\theta\int_{-\pi/2}^{\pi/2}\cos\phi\,d\phi}$$

$$=\frac{1}{2}\int_{-\pi/2}^{\pi/2}\alpha_w(\psi)\cos\phi\,d\phi, \tag{15}$$

assuming that $\overline{p}^{2}_{\theta,\phi}$ is constant with respect to ϕ. The power absorbed per angle θ can then be rewritten

$$W^{(+)}_{\theta,\text{``abs''}}=\frac{\overline{p}^{2(+)}_{\theta}\sin\theta}{\pi(2-\hat{\alpha}_w)\rho_o c}\beta L_p\Delta x 2\tilde{\alpha}_w. \tag{16}$$

The power into the cross section per angle θ is related to the component of intensity (per angle) normal to the cross section, $\mathbf{I}^{(+)}_{x(\theta)}$, which is written in terms of the mean-square pressure density [from Eq. (4)]

$$W^{(+)}_{\theta,\text{``in''}}=\mathbf{I}^{(+)}_{x(\theta)}\cdot S=\frac{\overline{p}^{2(+)}_{\theta}}{\rho_o c}\cos\theta\cdot S, \tag{17}$$

where S is the cross-sectional area. Substituting these last two relations into the power balance differential equation [Eq. (12)] and rearranging

$$\frac{d\overline{p}^{2(+)}_{\theta}}{dx}+\left(\frac{2\beta L_p\tilde{\alpha}_w}{S\pi(1-\hat{\alpha}_w/2)}\tan\theta\right)\overline{p}^{2(+)}_{\theta}=0. \tag{18}$$

Solving the above differential equation for the right-traveling mean-square pressure per angle

$$\overline{p}^{2(+)}_{\theta}=P^{(+)}e^{([(-\beta L_p\tilde{\alpha}_w)/S\pi(1-\hat{\alpha}_w/2)]\tan\theta)x}. \tag{19}$$

A similar analysis can be performed on the left-traveling wave field, which results in an expression for the left-traveling mean-square pressure per angle

$$\overline{p}^{2(-)}_{\theta}=P^{(-)}e^{([(\beta L_p\tilde{\alpha}_w)/S\pi(1-\hat{\alpha}_w/2)]\tan\theta)x}, \tag{20}$$

where the $P^{(+)}$ and $P^{(-)}$ are coefficients to be determined by the boundary conditions.

E. Boundary conditions at end walls

At the source wall, the power associated with the right-traveling waves is equal to the power added to the enclosure plus the power that is reflected from the left-traveling waves. Assuming that the source end wall is nonabsorptive

$$\mathbf{I}^{+}_{x\theta}S=\mathbf{I}^{-}_{x\theta}S+W_{\theta}, \tag{21}$$

where all quantities are on a per-angle basis.

At the absorbing end wall (opposite the source wall), the power balance relationship can be written as

$$\mathbf{I}^{-}_{x\theta}=(1-\alpha_b)\mathbf{I}^{+}_{x\theta}, \tag{22}$$

where α_b is the absorption coefficient of the end wall material, which is solely dependent on θ since the angle incident on the end wall is equivalent to θ. The formula for α_b is discussed in Sec. IV, and is given by Eq. (31) with γ replaced by θ.

F. Mean-square pressure per angle

Solving the differential equations, applying the boundary conditions, and superimposing the right- and left-traveling wave fields leads to the average mean-square pressure per angle in a given cross section (x location) as a function of W_{θ}, the power per angle

$$\overline{p^2_{\theta}}=\left(\frac{W_{\theta}\rho_o c}{S\cos\theta}\right)\frac{2\cosh c_{\theta}x-\alpha_b e^{c_{\theta}x}}{2\sinh c_{\theta}L_x+\alpha_b e^{-c_{\theta}L_x}}, \tag{23}$$

where

$$c_{\theta}=\left(\frac{\beta L_p\tilde{\alpha}_w}{S\pi(1-\hat{\alpha}_w/2)}\tan\theta\right).$$

Rewriting the mean-square pressure per angle in terms of the power into the enclosure rather than the power per angle by substituting for W_{θ} from Eq. (3)

$$\overline{p^2_{\theta}}=\left(\frac{W\rho_o c\tan\theta}{S}\right)\frac{2\cosh c_{\theta}x-\alpha_b e^{c_{\theta}x}}{2\sinh c_{\theta}L_x+\alpha_b e^{-c_{\theta}L_x}}. \tag{24}$$

The average mean-square pressure at a cross section is obtained by integrating the above formula over angle θ

$$\overline{p^2}(x)=\int_{\theta_{\min}}^{\pi/2}\left(\frac{W\rho_o c\tan\theta}{S}\right)\frac{2\cosh c_{\theta}x-\alpha_b e^{c_{\theta}x}}{2\sinh c_{\theta}L_x+\alpha_b e^{-c_{\theta}L_x}}d\theta. \tag{25}$$

On physical grounds, very shallow angles (≈ 0) are not permissible, as discussed in Appendix B, and therefore the lower limit of integration is the small angle $\theta_{\min}$ derived in the Appendix.

G. Comparison with previous results

Equation (24) [i.e., the integrand of Eq. (25)] bears a striking resemblance to the simple formula obtained in Ref. 2, which is repeated here for comparison

$$\overline{p^2}(x)=\frac{2\rho_o cW}{S}\times\left[\frac{2\cosh\left(\frac{\alpha_w\beta L_p}{(2-\alpha_w)S}x\right)-\alpha_b e^{+((\alpha_w\beta L_p/[(2-\alpha_w)S])x)}}{2\sinh\left(\frac{\alpha_w\beta L_p}{(2-\alpha_w)S}L_x\right)+\alpha_b e^{-((\alpha_w\beta L_p)/[(2-\alpha_w)S])L_x}}\right], \tag{26}$$

where the absorption coefficients in this formula are the classical (e.g., Ref. 6) three-dimensional random-incidence absorption coefficients (integrated in ϕ and θ) as explained in the next section. In the earlier work, the sound field was assumed locally diffuse in all directions; thus, the θ integration was carried out earlier [effectively after Eq. (11)] before the power balance [Eq. (12) onward]. As mentioned earlier, the current work retains the angular dependence on θ until the end. This inclusion allows the variation in wave strength as a function of angle to be retained, accounting more precisely on both the different reflection histories of different angles and the angular dependence of the absorption coefficients.

For reference, the formula for the predicted mean-square pressure of the *reverberant* field derived empirically from the numerical simulations of Ref. 1 is repeated here

$$\overline{p^2_{\text{rev}}}=\left(\frac{4\rho_o cW}{A}\right)\left[\frac{(1-\overline{\alpha}_{\text{total}})(1-\overline{\alpha}_{\text{total}}/2)}{(1+\overline{\alpha}_w S/2)}\right]e^{-(1/2)\overline{\alpha}_w\overline{S}\overline{x}}, \tag{27}$$

where $\rho_o c$ is the characteristic impedance of air, W is the source power, A is the total absorptivity in the room (in sabins), $\overline{x}=x/L_x$, and $\overline{S}$ is the ratio of side-wall surface area to the area of the cross section. The various absorption coefficient ratios are weighted averages of the three-dimensional random-incidence absorption coefficient and are defined as follows:

$$\overline{\alpha}_{\text{total}}=\alpha_{\text{random}}\frac{S_{\text{covered}}}{S_{\text{total}}}, \tag{28}$$

$$\overline{\alpha}_w=\alpha_{\text{random}}\frac{S_{\text{covered}}}{S_{\text{side wall}}}, \tag{29}$$

where S stands for the surface area of the subscripted region. The absorption coefficients that are used in the various theories will be discussed in detail in the next section.

The three theoretical results that will be compared to experiment are given by Eq. (25), Eq. (26), and Eq. (27). They are, respectively: the angle-by-angle refined power conservation theory, the original power conservation theory where all incidence angles are treated as equally probable and of equal strength in a given cross section, and the formula based on numerical simulations.

IV. THE ABSORPTION COEFFICIENT

To compare experimental results with theoretical results from Refs. 1 and 2, a random-incidence absorption coefficient was needed for the various absorptive materials used on

the side walls and/or the end wall of the enclosure. Morse and Ingard derived a formula (Ref. 7, p. 580, Eq. 9.5.8) for the random-incidence absorption coefficient of point-reacting absorptive materials in terms of their normal incidence admittance components. Using the experimentally determined real and imaginary parts of the normal incidence impedance, relating them to admittance, and inserting in the Morse and Ingard formula, the values of random-incidence absorption coefficient were found. These coefficient values were then used in the theoretical formulas of Ref. 1 [Eq. (27) above] and Ref. 2 [Eq. (26) above] and compared to the corresponding experimental values for the cross-sectionally averaged mean-square pressure at a given cross section.

These previous theoretical results overestimated the amount of loss (i.e., the spatial decay away from the source was too rapid) for cases with larger absorption. For cases with less absorption, the theoretical curves showed much better agreement but tended to slightly underestimate the degree of spatial decay. These original comparisons suggested that a more complete treatment of the absorption coefficient including bulk-reaction effects would be necessary. However, the only available experimental measurements for the properties of the absorbing materials were the normal incidence impedance measurements. Appendix A presents an approximate method for determining porous media properties from normal incidence impedance measurements. This method was used to approximate the porous material properties from which generalized impedance components were computed. The generalized impedance components were needed in order to obtain the appropriate absorption coefficients.

The random-incidence absorption coefficient is a ratio of the power absorbed by a differential surface element from incident sound waves in all directions to the total power incident on that surface element. Namely

$$\alpha_{\text{random}}=\frac{\int_0^{2\pi}\int_0^{\pi/2}\alpha(\gamma)\frac{\partial^2\overline{p^2}}{\partial\gamma\partial\eta}\cos\gamma\sin\gamma\,d\gamma\,d\eta}{\int_0^{2\pi}\int_0^{\pi/2}\frac{\partial^2\overline{p^2}}{\partial\gamma\partial\eta}\cos\gamma\sin\gamma\,d\gamma\,d\eta}, \quad (30)$$

where γ, η are general solid angles in spherical coordinates with γ measured down from the axis perpendicular to the surface, and η is the azimuth angle around the normal. These angles are not the same as ϕ and θ used previously. The mean-square pressure density per solid angle $\partial^2\overline{p^2}/\partial\gamma\partial\eta$ is assumed to be constant, and the absorption coefficient in the integrand is given by

$$\alpha(\gamma)=\frac{4\overline{r_g}(\gamma)\cos\gamma}{(\overline{r_g}(\gamma)\cos\gamma+1)^2+(\overline{x_g}(\gamma)\cos\gamma)^2}, \quad (31)$$

where $\overline{r_g}$ is the normalized real part of the general impedance $(\overline{z_g})$ for a porous layer and $\overline{x_g}$ is the imaginary component, which are derived in Appendix A and depend on incidence angle.

The absorption coefficient pertaining to the side wall, $\tilde{\alpha}_w$, in the angle-by-angle theoretical formula [Eq. (25)] is a function of θ only and was defined in Eq. (15). The absorption coefficient in the integrand of Eq. (15) $\alpha_w(\psi)$ is simply

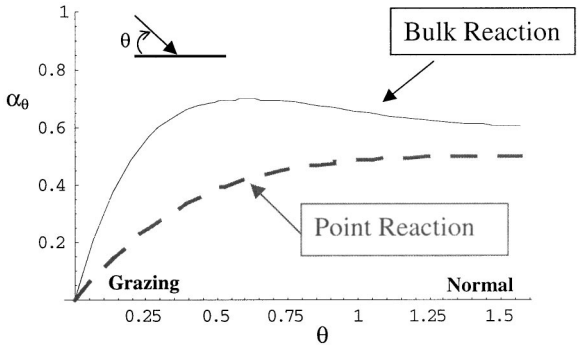


FIG. 6. Comparison of absorption coefficient versus angle for two cases: general impedance (bulk reaction) and normal incidence impedance (point reaction).

Eq. (31) with $\gamma=\pi/2-\psi$. Therefore, $\cos\gamma$ in Eq. (31) becomes $\sin\psi$ and $\sin\psi=\sin\theta\cos\phi$ from geometry (see Fig. 5). The resulting equation for the absorption coefficient associated with side-wall absorption as a function of θ is

$$\tilde{\alpha}_w(\theta)=\int_{-\pi/2}^{\pi/2}\frac{2\overline{r_g}\sin\theta\cos^2\phi}{(\overline{r_g}\sin\theta\cos\phi+1)^2+(\overline{x_g}\sin\theta\cos\phi)^2}\,d\phi, \quad (32)$$

where the components of the impedance, $\overline{r_g}$ and $\overline{x_g}$, are dependent upon incidence angle $\psi(\theta,\phi)$, referring to Fig. 5. The corresponding coefficient for the end wall, α_b, is essentially Eq. (31) with γ replaced by θ, since there is no ϕ dependence.

In Fig. 6, $\tilde{\alpha}_w(\theta)$ is plotted versus θ for a typical material used in the experiment. $\tilde{\alpha}_w(\theta)$ is computed two ways for comparison: one, where the values of $\overline{r_g}$ and $\overline{x_g}$ come from the general impedance including bulk-reaction effects; and two, where $\overline{r_g}$ and $\overline{x_g}$ are the real and imaginary parts of the point-reacting impedance. This plot shows the significant difference on a per-angle basis between including bulk-reaction effects and assuming point-reaction behavior. A similar conclusion was reached by Bliss in Refs. 4 and 5.

This difference between using the generalized impedance components, $\overline{r_g}$ and $\overline{x_g}$, versus using the normal-incidence impedance components leads to considerably different values of $\alpha_{w\,(\text{random})}$ [Eq. (30)] used in the theories of Refs. 1 and 2 and of $\tilde{\alpha}_w$ used in the present theory. In both cases, including bulk reaction substantially increases the absorption coefficient.

V. COMPARISON BETWEEN EXPERIMENT AND THEORY

Predictions of the average mean-square pressure in a cross section were compared to experimental measurements averaged in a cross section. The predicted values were based on earlier theories and the angle-by-angle theory. Typical results are presented for different levels of absorption and different center frequencies. All data were taken in one-third-octave bands.

In Ref. 1 the theoretical prediction is based on an empirical curve fit to numerical simulations of the *reverberant* field. In that study, the direct field was removed mathematically before calculating the mean-square pressure. In order to

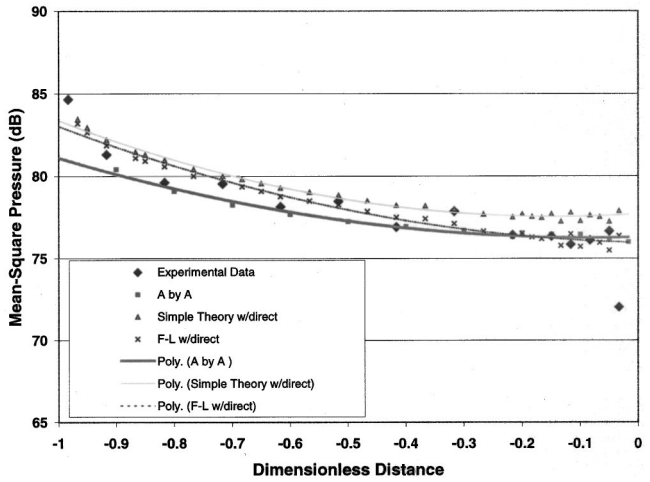


FIG. 7. Least absorptive case at center frequency of 4000 Hz. Absorptive material on two side walls only, end wall opposite the source is hard. Experimental data versus various theoretical results. Second-order polynomial curve fits have been run through the theoretical points for ease of comparison.

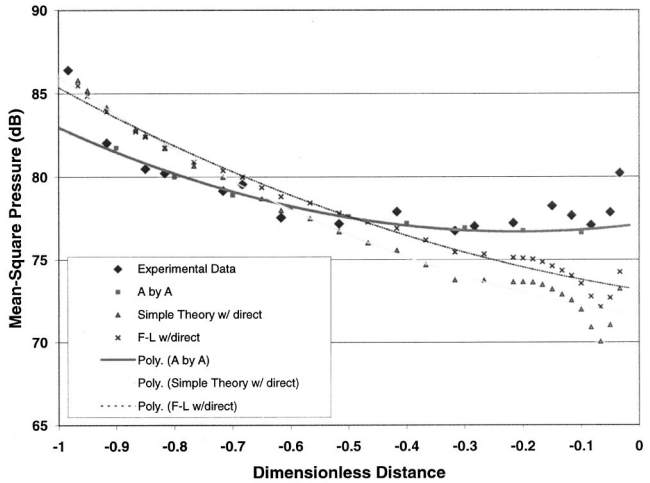


FIG. 8. Most absorptive case at center frequency of 6300 Hz. Absorptive material on two side walls only, end wall opposite the source is hard. Experimental data versus various theoretical results. Second-order polynomial curve fits have been run through the theoretical points for ease of comparison.

compare the results of Ref. 1 with the experimental data, which include the direct field as well as the reverberant field, an estimated direct field was added to the predictions. An analytical model of the enclosure with the hard-wall surfaces represented by image surfaces and a monopole-type source on the source wall was used to compute the direct field. The predicted sound-pressure levels from the formula given in Ref. 1 are designated as ''F-L w/direct'' on the plots. Although the predictions were computed at certain discrete cross-section locations, a least-squares curve fit has been drawn through the data for ease of comparison.

The predictions of Ref. 2 compared reasonably well with the numerical simulations of the reverberant field from Ref. 1 when the source power was adjusted to remove the power that went into the first reflection. Therefore, for comparison purposes, the reverberant field prediction from Ref. 2 with the direct field added was used as the ''simple theory'' result on the subsequent plots and was labeled as ''Simple Theory w/direct.'' Again, for ease of comparison, a least-squares curve fit has been drawn through the discrete points.

The first two cases presented for comparison are the least and the most absorptive cases that were tested. In Fig. 7, predictions from the previous theoretical formulas and the refined angle-by-angle (A-by-A) theory are shown compared to the experimental data for the least absorptive case. The center frequency is 4000 Hz. In this case, all theoretical predictions work moderately well, although the older theories slightly overpredict the levels by roughly 2 dB at the far end of the enclosure (farthest from the source wall). In this case, the random incidence absorption coefficient (including bulk-reaction effects) is $\alpha_{\text{random}}=0.37$, and two of the four side walls are covered, so there is an overall effective absorption coefficient of approximately half that value, 0.19. Close to the source wall ($\bar{x}=-1$), the new theory does not predict the mean-square pressure as well as the other two theoretical results. This difference is due to the fact that the other theories consist of a reverberant field estimate for the cross-sectional average plus a more precise estimate of the direct field. In the A-by-A theory, the source wall is represented as a distribution of sources, and the prediction of sound-pressure level includes both the reverberant and direct fields as a cross-sectional average. In this case, both of the previous theories predict the sound-pressure levels within a few decibels, so there is little advantage in invoking the angle-by-angle method.

In Fig. 8, the case with the most absorptive material covering two side walls is shown at a center frequency of 6300 Hz. Here, the random-incidence absorption coefficient is 0.81, but since only two side walls are covered there is an effective side-wall absorption for a cross section of about 0.4. The previous theories overpredict by almost 3 dB near the source (around $\bar{x}\approx-0.9$) and underpredict by as much as 5 dB at the far end of the enclosure. The new theory that accounts for angular dependence tracks the data very closely. Erratic behavior at the end wall is due to hard-wall reflections since there is no absorbing material on the end wall and the fact that only the waves with the shallowest of angles (grazing incidence) remain. The theoretical formulas exhibit this behavior due to the addition of the direct field (the hard end wall is an image plane in the analytical model), but this trend is also apparent in the experimental data. In the previous case, there was also no end-wall absorption, but since there was less absorption on the side walls, the reflections from the end wall were not as pronounced. Again, the angle-by-angle theory does not predict the value near the source due to the way that the source wall is approximated. Note that the angle-by-angle theory could be refined to better represent the field near the source, but this was not done in the results shown.

The next case presented has an absorbing end wall. These results are plotted in Fig. 9 for the 5000-Hz center frequency data. The random-incidence absorption coefficient of the side wall material is 0.71 and 0.78 for the end wall. Only two of the side walls and the end wall opposite the source wall were covered. The source wall itself is hard, so again the effective absorption coefficients are about half of

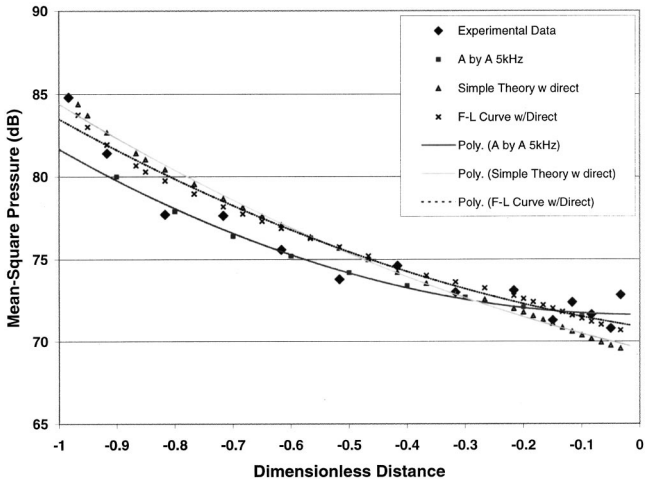


FIG. 9. Case with end wall and side wall absorption at center frequency of 5000 Hz. Experimental data versus various theoretical results. Second-order polynomial curve fits have been run through the theoretical points for ease of comparison.

those values. Although the predicted values are within a few decibels for the previous theoretical results, the shape of the decay curve is poorly predicted. The angle-by-angle theory not only predicts the general shape of the spatial decay curve, but the average mean-square pressure levels predicted are much closer than the other two predictions. This is the most absorptive case, since both the two side walls and an end wall are absorptive. It is important to note that the angle-by-angle theory works well, even for this case.

In Fig. 10, an intermediate case with a hard end wall and absorptive side walls is shown. The absorption coefficient for this case is 0.70, and the overall area-averaged absorption is approximately 0.35. In this case, the center frequency is 6300 Hz. Once again, there are indications that the end-wall reflections are very strong, as seen in both the theoretical (from the direct field) and experimental data. The previous theoretical predictions decay too rapidly as a function of distance from the source, but the angle-by-angle (A-by-A) theory agrees very well over the entire length of the enclosure (excluding the region closest to the source).

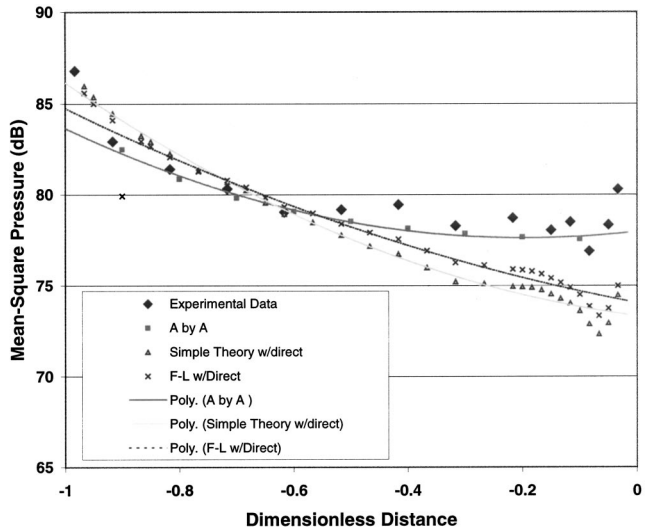


FIG. 10. An intermediate case without end wall absorption at a center frequency of 6300 Hz. Experimental data versus various theoretical results. Second-order polynomial curve fits have been run through the theoretical points for ease of comparison.

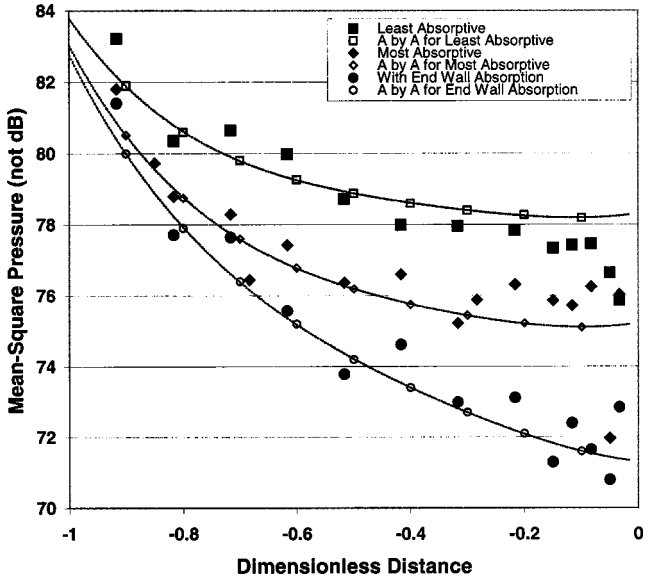


FIG. 11. Comparison of angle-by-angle theoretical results with experimental results for different cases with 5000-Hz center frequency.

Comparing the two previously derived theories to each other, it appears that the ''F-L curve'' from Ref. 1 is a better prediction of the experimental data than the ''Simple Theory'' from Ref. 2. This advantage is most likely because the F-L curve is based on exact numerical simulations that include the angular effects of angle of incidence and absorption coefficient. Also, at least half of the simulations were performed on a long, rectangular-shaped enclosure similar to the one in this experiment, furthering the likelihood that the simulations would agree with the experiment. The primary reason that the simple theory predictions were not in better agreement is that they ignore angular effects by treating all incidence angles with equal probability at every cross section of the enclosure.

On the previous plots, three theoretical results were compared to the experimentally measured mean-square pressures that had been averaged over the cross sections. In Fig. 11, only the angle-by-angle theoretical predictions versus experimental values are shown for three different cases. The three cases are the same as previously discussed, but at different center frequencies from those previously shown. Here, all results are for the 5000-Hz center frequency. A least-squares polynomial curve fit has been drawn through each of the angle-by-angle predictions in order to visualize the comparison between theory and experiment. The experimental data are shown as discrete data points: solid squares, diamonds, and circles, corresponding to the cases of least absorption on the side walls only, most absorption on the side walls only, and absorption on side walls and the end wall, respectively. The curves closest to the data points and drawn through the corresponding symbols (open, rather than solid) are the angle-by-angle theoretical predictions. Agreement is shown to be within a decibel for all cases.

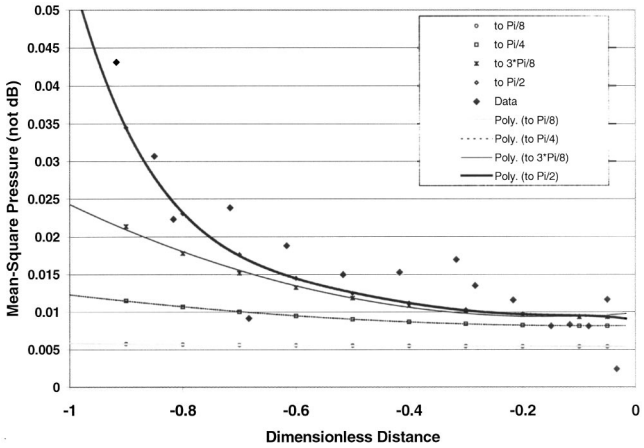


FIG. 12. Plot showing the angular build-up of energy for the case with the most absorptive material on two side walls and a hard end wall at a center frequency of 4000 Hz.

VI. ANGULAR BUILD-UP OF ENERGY

Since the refined theory looks at the mean-square pressure on an angle-by-angle basis, it is noteworthy to look at the angular content of the acoustic energy in the enclosure. Figure 12 shows the build-up of energy (mean-square pressure) as a function of angle. The mean-square pressure per angle [Eq. (25)] is integrated over different ranges of angles from approximately zero to different upper limits as indicated on the legend of the plot. The result corresponding to the curve labeled ''Pi/2'' is the actual mean-square pressure predicted by the angle-by-angle theory for this case, which is the most absorptive material on two side walls with a hard end wall at 4000 Hz.

The plot shows the mean-square pressure versus distance from the source wall. The energy at the end of the enclosure opposite the source is almost entirely due to sound waves that have the shallowest angles, whereas the energy nearest the source wall has a large contribution from the sound waves whose incidence angles are the steepest. This effect of angular filtering is due to the fact that the shallower angles undergo fewer reflections by the time they reach the end of the enclosure and are therefore not well absorbed. Distinguishing between different angles, which is most important for higher levels of absorptivity, was not taken into account in the earlier version of the theory (i.e., the simple theory, Ref. 2).

VII. CONCLUSIONS

A refinement of a simple theory based on conservation of power in cross sections of an enclosure was presented. The refinement conserved power in cross sections, but on an angle-by-angle basis. It was found that the bulk-reaction effects of the absorbing materials are important in computing the absorption coefficients even if random-incidence coefficients were going to be used.

The results comparing theoretical and experimental data showed better agreement for the refined theory than for the previously reported predictions in all cases that were explored. However, the predictions from the simple theory and numerical-simulation-based formula agree very well with the experimental data for cases that are more reverberant, reducing the need for the angle-by-angle approach in those cases. For a very accurate prediction when the enclosure is very absorptive, an angle-by-angle approach is recommended.

ACKNOWLEDGMENT

This research was sponsored by the National Science Foundation.

APPENDIX A: CURVE FIT FOR POROUS-MEDIA MODEL

The generalized impedance ($\bar{z}_g$) for a porous material will be derived in this Appendix following the analysis of Refs. 4 and 5. A means of extracting the porous material properties from normal-incidence impedance measurements will also be explained.

Consider oblique waves in the porous material of thickness L as shown in Fig. 13. Here, the x component of wave number $k_x = k \sin\theta$ must be the same inside and outside the porous layer. The pressure in the material is governed by the wave equation, and the solution for complex pressure for a right-traveling wave is

$$p = P \cos k_{ey}(y-L) e^{i(\omega t - k_x x)}, \tag{A1}$$

where P is the complex amplitude, and k_{ey} is the effective wave number in the y direction given by the dispersion relation

$$k_{ey} = \sqrt{\left(\frac{\omega}{c_e}\right)^2 - (k \sin\theta)^2}, \tag{A2}$$

where c_e is the effective wave speed in the porous material.

The y component of velocity, from the momentum equation, is

$$v = \frac{-iP}{\rho_e c_e} \sqrt{1 - \left(\frac{k c_e}{\omega} \sin\theta\right)^2} \sin k_{ey}(y-L) e^{i(\omega t - k \sin\theta x)}. \tag{A3}$$

Therefore, the general impedance at the surface between the porous material and the air is

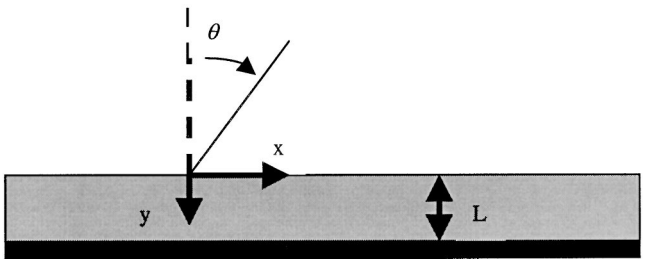


FIG. 13. Porous layer with oblique waves.

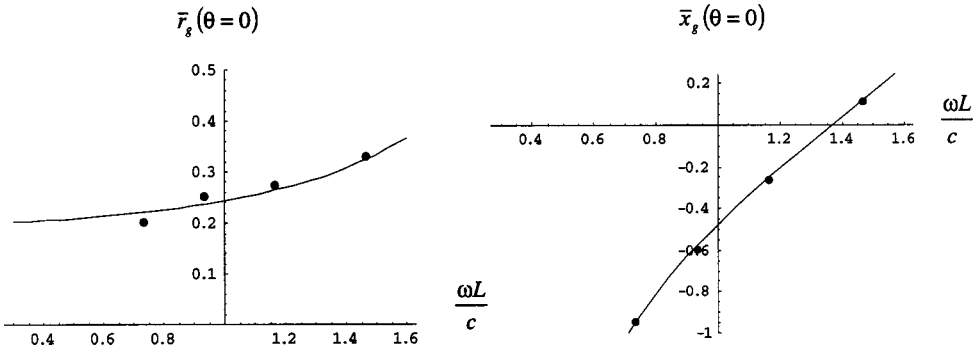


FIG. 14. Curve fit of the components of the normal-incidence impedance using the porous media parameters $\Phi=0.54$, $\rho_p=1.25$, $c_p=0.86$ compared to experimental measurements at various frequencies for a representative sample of absorptive material.

$$z_g=\left.\frac{p}{v}\right|_{y=0}=\frac{-i\rho_e c_e}{\sqrt{1-\left(\frac{kc_e}{\omega}\sin\theta\right)^2}}\cot\sqrt{\left(\frac{\omega L}{c_e}\right)^2-(kL\sin\theta)^2}. \tag{A4}$$

When $\theta=0$, the general impedance reduces to the normal incidence case

$$z_g(\theta=0)=-i\rho_e c_e\cot(\omega L/c_e), \tag{A5}$$

where ρ_e and c_e are the complex effective density and sound speed in the porous material, respectively. From Refs. 4 and 5, ρ_e and c_e are related to the actual density ρ_p and sound speed c_p by

$$c_e=\frac{c_p}{\sqrt{\left(1-i\left(\frac{\Phi}{\rho_p\omega}\right)\right)}}\quad\text{and}\quad\rho_e=\rho_p\left(1-i\left(\frac{\Phi}{\rho_p\omega}\right)\right), \tag{A6}$$

where Φ is the flow resistance of the porous material. While ρ_p, c_p, and Φ are unknown, the real and imaginary parts of $z_g(\theta=0)$ are known, in this case, from the normal incidence impedance measurements at several different frequencies.

Using the definitions for ρ_e and c_e and expanding the cotangent in a Taylor Series for small $\omega L/c_e$

$$z_g(\theta=0)=-i\rho_p c_p\sqrt{1-i\frac{\Phi}{\rho_p\omega}}\left[\frac{1}{\omega L/c_e}-\frac{1}{3}\frac{\omega L}{c_e}-\frac{1}{45}\left(\frac{\omega L}{c_e}\right)^3+\cdots\right]. \tag{A7}$$

Rewriting in terms of new dimensionless parameters

$$\bar{z}_g(\theta=0)=-i\frac{\overline{\rho c}}{\bar{\omega}}+i\frac{1}{3}\bar{\omega}\bar{\rho}\left(1-i\frac{\bar{\Phi}}{\bar{\omega}\bar{\rho}}\right)+i\frac{1}{45}\frac{\bar{\omega}^3}{\bar{c}^2}\bar{\rho}\left(1-i\frac{\bar{\Phi}}{\bar{\omega}\bar{\rho}}\right)+\cdots, \tag{A8}$$

where the following definitions have been applied:

$$\bar{\omega}=\frac{\omega L}{c_o},\quad\bar{\rho}=\frac{\rho_p}{\rho_o},\quad\bar{c}=\frac{c_p}{c_o},$$

$$\bar{\Phi}=\frac{\Phi L}{\rho_o c_o},\quad\text{and}\quad\bar{z}_g=\frac{z_g}{\rho_o c_o}.$$

Separating the real and imaginary parts of the truncated series such that $\bar{z}_g=\bar{r}_g+i\bar{x}_g$

$$\bar{r}_g(\theta=0)\cong\frac{1}{3}\bar{\Phi}+\frac{1}{45}\left(\frac{\bar{\omega}}{\bar{c}}\right)^2\bar{\Phi}, \tag{A9}$$

$$\bar{x}_g(\theta=0)\cong\frac{-\overline{\rho c}}{\bar{\omega}}+\frac{1}{3}\bar{\omega}\bar{\rho}+\frac{1}{45}\frac{\bar{\omega}^3}{\bar{c}^2}\bar{\rho}. \tag{A10}$$

Note that the flow resistance variable only appears in the real term.

Solving for the flow resistance variable

$$\bar{\Phi}=\left(\frac{45}{15+(\bar{\omega}/\bar{c})^2}\right)\cdot\bar{r}_g(\theta=0), \tag{A11}$$

where the measured normal impedance data can be used for $\bar{r}_g(\theta=0)$.

Looking at the imaginary term

$$\bar{x}_g(\theta=0)=\bar{\rho}\left(-\frac{\bar{c}}{\bar{\omega}}+\frac{1}{3}\bar{\omega}+\frac{1}{45}\frac{\bar{\omega}^3}{\bar{c}^2}\right), \tag{A12}$$

which crosses zero when $\bar{c}/\bar{\omega}=(1/3)\bar{\omega}+(1/45)(\bar{\omega}^3/\bar{c}^2)$. From the experimental data for different frequencies, the zero crossing can be estimated to predict a value for $\bar{c}$. Knowing $\bar{c}$ and the measured $\bar{x}_g(\theta=0)$, $\bar{\rho}$ can be determined. An example curve fit is shown in Fig. 14.

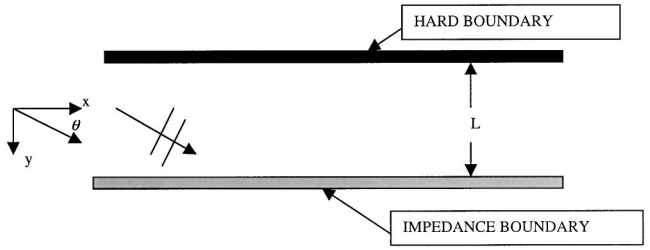


FIG. 15. Duct with oblique traveling plane waves.

APPENDIX B: SMALLEST ALLOWABLE ANGLE

To estimate the smallest allowable angle, consider an infinite two-dimensional duct, as shown in Fig. 15. The upper wall is hard; the lower wall has an impedance z_b. Oblique waves in this duct will have wave number components k_x and k_y, where $k_x^2+k_y^2=k^2$. The dispersion relation for waves in this duct is given by

$$k_y L \tan k_y L = -i\frac{kL}{\overline{z_b}}, \tag{B1}$$

where $\overline{z_b}$ is the normalized impedance $z_b/\rho_o c$.

In the high-frequency limit $kL \gg 1$. Assuming $\overline{z_b}$ is not large, the resulting magnitude of the right-hand side of the dispersion relation will be very large. In this case, the solution for $k_y L$ is very near $[(2n-1)\pi]/2$.

Since $k_y/k \approx \sin\theta$ and θ is small (approximating grazing incidence waves), $kL\,\theta \approx (2n-1)(\pi/2)$. Thus, the smallest allowable angle is $\theta_{\min} \approx \pi/2kL$.

Considering a three-dimensional duct does not change the result since the smallest angle will occur when one of the two cross-sectional wave numbers is approximately zero.

[1] L. P. Franzoni and D. S. Labrozzi, "A study of damping effects on spatial distribution and level of reverberant sound in a rectangular acoustic cavity," J. Acoust. Soc. Am. **106**, 802–815 (1999).

[2] L. P. Franzoni, "A power conservation approach to predict the spatial variation of the cross-sectionally averaged mean-square pressure in reverberant enclosures," J. Acoust. Soc. Am. **110**, 3055–3063 (2001).

[3] L. P. Franzoni and C. M. Elliott, "An innovative design for a probe tube microphone attachment," J. Acoust. Soc. Am. **104**, 2903–2910 (1998).

[4] D. B. Bliss, "Study of bulk reacting porous sound absorbers and a new boundary condition for thin porous layers," J. Acoust. Soc. Am. **71**, 533–545 (1982).

[5] D. B. Bliss and S. E. Burke, "Experimental investigation of the bulk reaction boundary condition," J. Acoust. Soc. Am. **71**, 546–551 (1982).

[6] A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, New York, 1991).

[7] P. M. Morse and U. K. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, 1986).

Behavioral criterion quantifying the edge-constrained effects on foams in the standing wave tube

Dominic Pilon and Raymond Panneton
GAUS, Department of Mechanical Engineering, Université de Sherbrooke, Sherbrooke, Quebec J1K 2R1, Canada

Franck Sgard
LASH-DGCB URA CNRS 1652, ENTPE, 2 Maurice Audin, 69518 Vaulx-en-Velin Cedex, France

(Received 9 October 2002; revised 19 May 2003; accepted 16 June 2003)

The influence of mounting conditions on the measurement of the sound absorption coefficient is investigated. More specifically, the effects of a circumferential edge constraint applied on elastic foams in a standing wave tube are studied. The objective is to identify the foams for which it is possible to measure the theoretical absorption coefficient using the tube. This identification relies on the evaluation of a new parameter, the Frame Acoustical Excitability (FAE), based on the foams' physical properties. In order to quantify the difference between the measured and theoretical absorptions, a linear correlation coefficient is evaluated for a wide range of foams with various sample sizes. The calculated correlation coefficients are then sorted in terms of the FAE. It is shown that as the value of FAE increases, the difference between the measured and theoretical absorption coefficients becomes more significant. A critical value of FAE, below which the theoretical absorption can be efficiently measured using the tube, is established. Through the use of this elasto-acoustic criterion, an experimenter can make an educated guess as to which absorption is measured with the tube: the theoretical absorption or one that is affected by either the edge constraint or the sample's dimensions. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1598193]

PACS numbers: 43.55.Ev, 43.20.Mv, 43.20.Jr [LLT]

I. INTRODUCTION

The standing wave tube (swt) is commonly used to measure the normal incidence sound absorption coefficient of acoustical materials. Standards ASTM C384-01, E1050-98, and ISO 10534-2 describe methods classically used to measure this coefficient. Ideally, the objective is to measure the theoretical absorption coefficient (α_{th})—defined here as the absorption of a material with laterally infinite dimensions, as shown in Fig. 1(a). What is actually measured with the swt is the absorption coefficient of a sample with finite lateral dimensions (α_m). Figure 1(b) shows that this sample can have various mounting conditions within the tube. Theoretically, amongst these conditions, only the sliding condition will lead to α_{th}.[1] In reality, the sliding condition is hardly likely to be encountered compared to the other ones. It is therefore reasonable to assume that the swt measurements do not always lead to α_{th} as the vibro-acoustic behavior of a sample can be influenced by the different lateral mounting conditions in the tube.

Two distinct questions arise from this situation. First, do the mounting conditions inevitably impact on the swt measurements for all acoustical materials? Second, how far from the theoretical absorption (α_{th}) will the measured absorption (α_m) be? The main objective of this paper is to answer these questions in the particular case of ''elastic foams'' and ''circumferential edge constraint,'' i.e., the bonded condition.

The bonded condition is often encountered during experimentation. Generally speaking, it is obtained by having a sample diameter which is slightly larger than the internal diameter of the tube. The resulting lateral compression, usually less than 1%, creates sufficiently high friction forces, between the sample circumference and the tube wall, to nearly constrain the edge from axial motion. This mounting condition is typically used to prevent the presence of air gaps around the sample circumference as they can have a great influence on the absorption coefficient. This was studied experimentally and analytically by Cummings.[2] It was shown that the effects of air gaps on the surface impedance of porous materials was quite significant as the frequency decreases and the flow resistivity increases. There also exists a method to evaluate the influence of these air gaps on the measurement of the absorption coefficient in the swt.[3]

The sample's sensitivity to the edge constraint condition is an experimental reality that has been known for some time. The influence of such a condition on poroelastic samples has been investigated by several authors. Early work by Donato[4] underlined the sensitivity to mounting conditions of acoustical measurements in the swt. He attributed the cause of certain anomalies in his measurements to the circumferential edge constraint on the samples. By comparing free field and swt measurements, Allard and Delage[5] noted that the $\lambda/4$ resonance frequency of the frame-borne wave was higher when measured with the swt. Through finite element modeling, Kang and Bolton[6] noted that at low frequencies the edge constraint had the effect of increasing the amplitude of the material's normal surface impedance imaginary part. Furthermore, Vigran *et al.*[7] also concluded that the edge constraint had the effect of stiffening the foam samples they studied. This phenomenon was shown to occur espe-

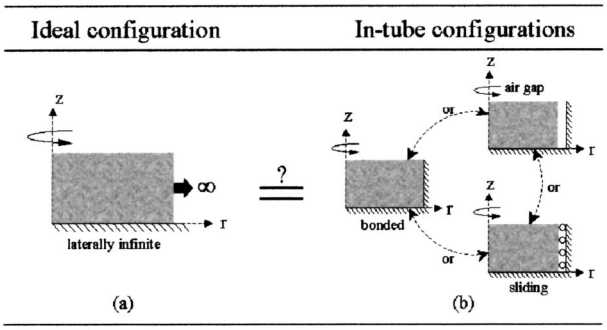


FIG. 1. Various types of lateral conditions on a disk-shaped porous sample. (a) Ideal or theoretical configuration: sample with infinite radial dimension. (b) In-tube configurations: the bonded, the free (with a circumferential air gap), and the sliding conditions. Under normal acoustic excitations, the in-tube sliding condition is equivalent to the radial infinite condition; however, in reality, this condition is hardly likely to happen.

cially when the elastic behavior of the material was predominant. Recently, Song *et al.*[8] confirmed that one effect of sample edge constraint in a *swt* was to stiffen the sample at low frequencies. This effect was found to have a significant impact on both the transmission loss and absorption coefficient. The main conclusion of the combined previous works is that the principal effect of the edge constraint is to stiffen the material at low frequencies, therefore shifting the λ/4 resonance of the frame-borne wave higher in frequency, as shown in Fig. 2(a).

Most of these researches were carried out to study the effects of a circumferential edge constraint on acoustical measurements in the *swt*. No information was given in order to identify which materials would be influenced by this mounting condition. Furthermore, these researches have only been conducted on a small number of materials, usually two or three. It is therefore difficult to give a general assessment of the effects of that condition on the vibro-acoustic behavior of other materials, i.e., which ones will be affected and which ones will not. This is where the present research aims to bring new developments.

As stated previously, the main objective of this paper is to devise a method that allows for an efficient identification of the elastic foams for which α_{th} can be measured using the *swt*—in the case where the sample is mounted with a circumferential edge constraint in the tube. This method must be relatively easy to carry out at any point during the characterization process. As a result, a new parameter based on the foams' physical properties, the Frame Acoustical Excitability (FAE), is developed and used as the identification tool. A secondary objective is to quantify the difference between α_m and α_{th} when the latter cannot be measured using the tube. This is achieved through a linear correlation coefficient. The investigation is performed on elastic foams, since this type of poroelastic material is especially sensitive to the circumferential edge constraint. Also, in view of drawing general conclusions on the effect of the edge constraint on the foams, a large number of foam samples are studied. As for fibrous materials, they are not used in this study because of the fiber motion that the circumferential edge constraint may induce in them.[9] This phenomenon is relatively complex

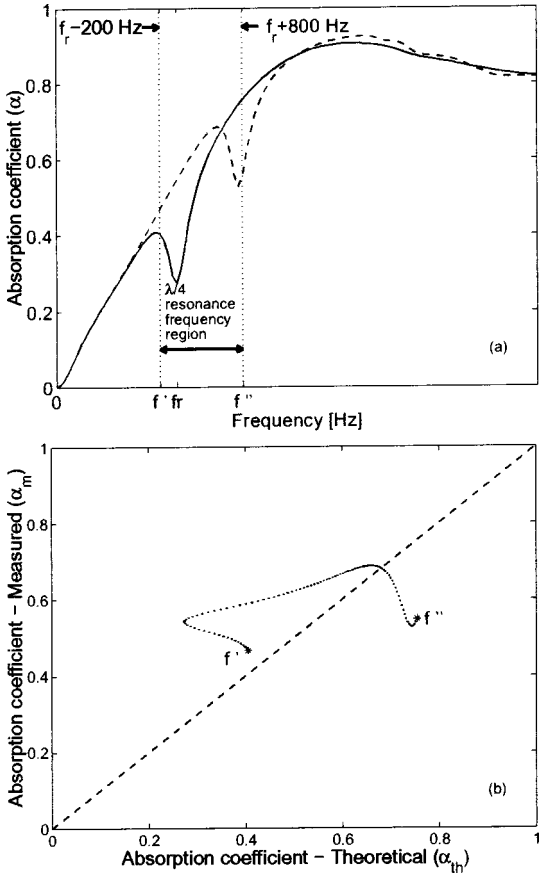


FIG. 2. Typical effects of the edge constraint on the sound absorption coefficient and representation of the selected frequency region of influence for the calculation of r_{xy}. (a) Dip in the absorption curve, related to frame resonance, shifted upwards in frequency; ——: α_{th} (unconstrained), — — : α_m (constrained). (b) ···: Scatter plot of α_m vs α_{th} in the λ/4 resonance frequency region ($r_{xy}=0.26$), — — —: Theoretical curve for a perfect linear correlation between two variables ($r_{xy}=1$).

and cannot be taken into account by a simple ratio of the physical properties. Hence, the results obtained in this paper should not be applied to those materials.

This paper is structured as follows. In the first section, α_{th} and α_m are briefly introduced to set the bases of the forthcoming analysis. In the second section, the linear correlation coefficient depicting the effect of the edge constraint is presented. In the third section, the properties and dimensions of the studied foams are given. In the last section, the development of the elasto-acoustical behavioral criterion is detailed and discussed.

II. EXPRESSION OF α_{th} AND α_m

To identify how the sound absorption coefficient measured with the *swt* (α_m) compares to the expected theoretical absorption coefficient (α_{th}) of a material, both coefficients need to be defined.

Following Biot's theory,[10,11] three waves may propagate simultaneously in an air-saturated open-cell poroelastic material; two compressional waves and one shear wave. If the material is assumed to be of infinite lateral extent and excited by a normally incident acoustical plane wave, no shear wave propagates. Under these circumstances, the normal-incidence

sound absorption coefficient of the poroelastic material has been analytically derived by Allard[12] and writes:

$$\alpha_{th}=1-|R_x|^2, \tag{1}$$

with R_x, the complex reflection coefficient, given by

$$R_x=\frac{Z_n-\rho_0 c_0}{Z_n+\rho_0 c_0}, \tag{2}$$

where ρ_0 and c_0 are the ambient density and velocity of sound in the air, and Z_n is the normal-incidence surface impedance of the poroelastic material. The expression of Z_n is detailed in the Appendix.

Since it is difficult to master correctly the boundary conditions applied on a poroelastic sample in the experimental setup of the *swt*, numerical simulations of the *swt* are used to evaluate the ''measured'' absorption. The simulations use high order axisymmetric finite elements[13] based on the mixed $(u\text{-}p)$ Biot poroelastic formulation.[14] In this formulation, 5 acoustical parameters (porosity, tortuosity, static airflow resistivity, and the viscous and thermal characteristic lengths), 3 elastic parameters (Young's modulus, Poisson's ratio, and the structural damping), and the bulk density describe the poroelastic foam. A detailed description of these parameters can be found elsewhere.[12] As for the boundary conditions that need be applied on the mixed $(u\text{-}p)$ formulation, they have been detailed by Debergue *et al.*[15] They are only briefly described here.

The effect of the normally incident plane wave is simulated by imposing a uniform pressure on the excited surface of the poroelastic sample. To simulate the rigid end termination of the tube, a sliding condition is applied on the rear surface of the sample. Hence, on this surface, the axial nodal displacements of the solid phase are constrained to zero, the flux through it is null and the tangential displacements of the solid phase are free (there is no tangential stress on the surface). As for the circumferential edge constraint, a bonded condition is applied on the circumference of the sample. Consequently, the nodal displacements of the solid phase are fixed to zero over the circumference and the flux through the latter is null. Finally, using this model, the absorption coefficient ''measured'' with the *swt* is obtained using Eqs. (1) and (2); however this time, the normal-incidence surface impedance is given by the pressure to the normal velocity ratio:

$$Z_n=\frac{\overline{p}}{\overline{V}_n}, \tag{3}$$

where $\overline{p}$ and $\overline{V}_n$ are averaged over the excited surface. While $\overline{p}$ is fixed to unity in the numerical model, $\overline{V}_n$ is deduced from

$$\overline{V}_n=j\omega\overline{U}_{\text{tot}}, \tag{4}$$

where $j=\sqrt{-1}$, ω is the angular frequency, and $\overline{U}_{\text{tot}}$ is the mean total normal displacement on the excited surface of the modeled sample. $\overline{U}_{\text{tot}}$ is obtained through the evaluation of the reaction forces on the finite element system.

III. LINEAR CORRELATION COEFFICIENT

A linear correlation coefficient is evaluated to quantify the difference between the two absorption coefficients, namely α_{th} and α_m. It measures the strength of the linear relationship between the two. For α_{th} and α_m, which are denoted here by X and Y respectively so as to simplify the writing, the correlation coefficient is defined as:[16]

$$r_{xy}=\frac{|\sigma_{xy}|}{\sqrt{\sigma_{xx}\sigma_{yy}}}, \tag{5}$$

where

$$\begin{aligned}
\sigma_{xy}&=\sum_{s=1}^{n}(X_sY_s)-\frac{1}{n}\left(\sum_{s=1}^{n}X_s\right)\left(\sum_{s=1}^{n}Y_s\right),\\
\sigma_{xx}&=\sum_{s=1}^{n}(X_s^2)-\frac{1}{n}\left(\sum_{s=1}^{n}X_s\right)^2,\\
\sigma_{yy}&=\sum_{s=1}^{n}(Y_s^2)-\frac{1}{n}\left(\sum_{s=1}^{n}Y_s\right)^2,
\end{aligned} \tag{6}$$

are the total sum of the squares about the mean, s is the frequency at which the absorption is taken, and n is the number of frequencies in the frequency sampling.

The value of r_{xy} is always between 0 and 1. $r_{xy}=1$ points to a perfect linear correlation, as all of the (x,y) points in a scatter plot are connected by a single straight line with a slope of $+1$ [see dashed line in Fig. 2(b)]. For this particular value of r_{xy}, the circumferential edge constraint has no effects on the vibro-acoustic behavior of the material since α_m and α_{th} are the same. On the other hand, as r_{xy} moves away from 1 toward 0, it implies that the material behavior becomes more sensitive to the mounting condition, meaning that the difference between α_m and α_{th} increases.

The frequency range on which the calculation of r_{xy} will apply is chosen to be in the vicinity of the λ/4 resonance frequency of the frame-borne wave, f_r, as defined further in Eq. (8). This is motivated by the fact that, under the acoustical excitation in the *swt*, the effects of the frame vibration on the absorption curve of a material will mostly impart this frequency range [see Fig. 2(a)]. The bounds f' and f'' defining this frequency range in the vicinity of f_r are shown in Fig. 2(a). The material used in Figs. 2(a) and (b) is a typical open cell elastic foam whose properties are contained within the prescribed values in Table I.

Following numerous simulations on different materials, it was found that $f'=f_r-200$ Hz and $f''=f_r+800$ Hz lead to representative values of the r_{xy} coefficient for the purpose of this work. The fact that f'' is farther away from f_r than f' is due to the stiffening effect of the edge constraint shifting f_r upwards in frequency.[8]

Using the bounds previously defined, the calculation of r_{xy} will be a measure of the correlation between the theoretical and measured sound absorptions over a 1000 Hz wide frequency range in the vicinity of f_r. This is always true except when $f_r<200$ Hz. In this case, the lower bound f' is fixed to 0 Hz. This will typically occur when the thickness of the sample is large or the material's Young modulus is low.

TABLE I. Range of values for the properties of the 30 elastic foams.

Properties			
Name	Symbol	Range of values	Units
Young's modulus	E	30 000–320 000	N/m^2
Bulk density	ρ_1	9–150	kg/m^3
Poisson's ratio	ν	0–0.46	
Loss factor	η	0.05–0.25	
Static airflow resistivity	σ	5 000–85 000	Ns/m^4
Porosity	ϕ	0.87–0.995	
Tortuosity	α_∞	1–3.2	
Viscous length	Λ	21–213	μm
Thermal length	Λ'	50–525	μm

The different pairs of absorption data (α_{th}, α_m) between f' and f'' are plotted in Fig. 2(b). Each dot represents a single pair of absorption data at a given frequency in between the two frequency bounds. As it can be seen, the scatter plot deviates from the theoretical curve for a perfect linear correlation (dashed line). This is the deviation that is evaluated, or quantified, by the r_{xy} coefficient. The theoretical curve is obtained by using $\alpha_{th}=\alpha_m$ over the entire frequency range. For this open cell elastic foam, $r_{xy}=0.26$. This implies a very poor linear correlation between α_{th} and α_m. In other words, the two absorption coefficients are quite different from each other in the vicinity of f_r, as shown in Fig. 2(a).

IV. SAMPLE SIZE AND MATERIAL PROPERTIES

The measurements in a *swt* are usually performed over three different frequency ranges, the high (up to 6900 Hz), medium (up to 3100 Hz), and low (up to 2000 Hz). They correspond to three different tube diameters which are respectively 29 mm, 64 mm, and 99 mm for each frequency range.

To represent a wide variety of experimental configurations, each material sample is modeled with the three different tube diameters and with three different thicknesses (25 mm, 50 mm, and 100 mm). A total of nine sample sizes are studied for each material. The evaluation of α_{th} and α_m is carried out over 30 different open-cell elastic foams. Consequently, a total of 270 numerical simulations are accomplished to reach the objectives of this work. The range of values for the properties of the 30 elastic foams are presented in Table I. The mechanical properties (E, ν, and η) of the foams were evaluated using the method described by Langlois *et al.*[17] The flow resistivity (σ), the porosity (ϕ) and the bulk density (ρ_1) are measured using non-acoustical methods.[18,19] As for the tortuosity (α_∞) and the viscous and thermal lengths (Λ and Λ'), they are measured using both acoustical[20] and nonacoustical[21,22] methods. The acoustical method is used in the high frequency range, i.e., over the decoupling frequency,[23] to make sure that the elastic effects do not influence the measurements.

V. THE FRAME ACOUSTICAL EXCITABILITY

In the following, the Frame Acoustical Excitability (FAE) concept is proposed. The FAE is a parameter that will weight the sensitivity of a disk-shaped foam sample to the circumferential edge constraint in the *swt*. The development of the elasto-acoustic criterion is based on critical values of this parameter. The FAE should take into account the main parameters of a tested foam that may influence the measurements. Generally speaking, one can expect that to be efficient, the FAE should take into account the three following aspects of the foam: its solid/fluid visco-inertial coupling, its specific stiffness, and its geometry or shape factor.

A. The visco-inertial coupling—σ/ρ_1

Above a certain frequency, a partial decoupling between the solid and fluid phases of an open-cell porous medium exists. An expression for this frequency is given by Zwikker and Kosten:[23]

$$f_d=\frac{1}{2\pi}\left(\frac{\phi^2\sigma}{\rho_1}\right), \tag{7}$$

where ϕ, σ, and ρ_1 are the porosity, static airflow resistivity, and bulk density of the porous medium, respectively. For frequencies higher than f_d, the visco-inertial coupling between the two phases is weak enough so that an acoustical wave propagating in the fluid phase would not exert a sufficient force to generate vibrations in the solid phase.

To keep the devised criterion as simple as possible, only the σ/ρ_1 ratio from the decoupling frequency is kept in the following. Moreover, since the porosity is usually close to unity for typical sound absorbing materials, σ/ρ_1 is the most significant ratio for quantifying the visco-inertial coupling between the two phases. For small values of σ/ρ_1 (a low static airflow resistivity combined with a high bulk density), the solid phase tends to be motionless under an acoustic excitation. Oppositely, for large values of σ/ρ_1 (high static airflow resistivity and low bulk density), the visco-inertial coupling is strong. It is then relatively easy for the fluid phase to induce motion in the solid phase.

As discussed previously, the mounting conditions will have a greater influence on the acoustical behavior of the foam if its solid phase vibrates. Hence, a foam with a high σ/ρ_1 ratio will be more sensitive to the edge constraint.

B. The specific stiffness of the frame—E/ρ_1

Under *in vacuum* conditions, the square of the first $\lambda/4$ resonance frequency of the frame-borne wave (f_r) is proportional to E/ρ_1. An approximation of f_r is given by Allard *et al.*[24] It is defined as:

$$f_r\approx\frac{1}{4t}\sqrt{\frac{E}{\rho_1}}, \tag{8}$$

where t is the sample thickness and E the foam's Young modulus. The E/ρ_1 ratio can therefore be interpreted as a parameter that takes into account the stiffness of the solid phase or frame.

For a given specific stiffness, the frame may show a f_r below, or in the vicinity, of f_d. In this case, the interaction between the solid and fluid phases is strong and the motion of the frame will thus be important. Nevertheless, even if $f_r>f_d$, the motion of the frame might still be important as

only a partial decoupling occurs at frequencies higher than f_d. Consequently, E/ρ_1 is clearly an important ratio to take into account in the devised criterion.

C. Shape factor—*t*/*R*

In the previous analysis of the frame's stiffness, only the foam properties were involved. It is obvious that for an edge-constrained disk-shaped porous sample, the stiffness will also be strongly governed by its geometry, that is its radius R and thickness t. As a result, the shape factor t/R is introduced to account for the geometry effects.

For absorption coefficient measurements in the swt, the foam is backed by the rigid end of the tube. As the shape factor decreases (increasing radius and/or decreasing thickness), the foam tends to behave more as an unconstrained (or laterally infinite) sample; the influence of the edge constraint decreases. Oppositely, as the shape factor increases (decreasing radius and/or increasing thickness), the foam is more sensitive to the edge constraint.

Consequently, independently from the two above ratios (σ/ρ_1 and E/ρ_1), the shape factor is clearly one of great importance to qualify the effects of the edge constraint on the swt measurements; the higher the value of the shape factor, the more sensitive the material sample. Experimentations and numerical simulations have shown that, for the forthcoming results, the square of the shape factor was more appropriate to identify the sensitive foams.

D. Physical interpretation of the *FAE*

From the previous analysis, the FAE is now built by multiplying the three aforementioned ratios together. This yields

$$FAE = \frac{\sigma E t^2}{\rho_1^2 R^2} \ [\mathrm{W/kg}]. \tag{9}$$

By looking at the units of FAE, it can be seen that it represents a characteristic amount of power per unit of mass (W/kg). In other words, the FAE can be interpreted as a measurement of the rate at which the acoustical energy is transferred from the air to a specific amount of porous sample. As FAE increases, a greater amount of acoustical energy is transferred to the solid phase, which induces motion in the latter as the magnitude of the frame-borne compressional wave becomes larger. In that situation, the influence of the edge constraint on the acoustical behavior of the foam becomes important.

VI. RESULTS

In the previous section, the FAE parameter was introduced and defined. The next step in the development of the elasto-acoustic criterion is to identify the values of FAE for which a foam is either going to be affected or not affected by the circumferential edge constraint. Hence, in the following, the different results leading to these critical values are presented. A detailed analysis is carried out for samples having a thickness of 25 mm. As for the other thicknesses, that is to say 50 and 100 mm, general results are presented. Finally, the identification of the critical values of FAE, or the development of the elasto-acoustic criterion per say, is accomplished.

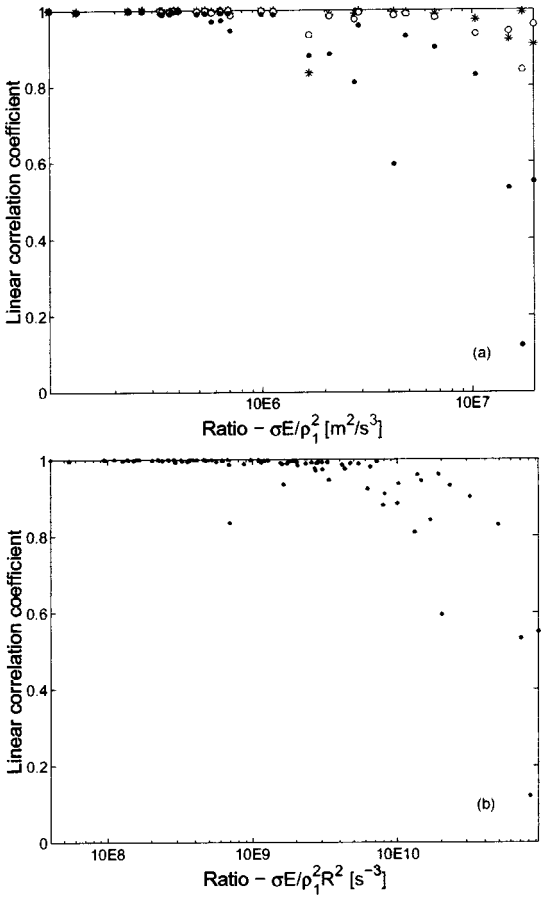


FIG. 3. Evolution of the linear correlation coefficient in the f_r region for a sample thickness of t=25 mm. (a) r_{xy} for the three different sample radii (the ratio does not take into account the shape factor); ●: R=14.5 mm, ○: R=32 mm, and *: R=49.5 mm. (b) r_{xy} for all the samples of thickness t =25 mm (the ratio only takes into account the effect of the radius).

A. Sample thickness of 25 mm

Figure 3(a) presents the linear correlation coefficient r_{xy} as a function of the $\sigma E/\rho_1^2$ ratio for a sample thickness of 25 mm. Here, the shape factor is not included in the ratio in order to see the evolution of r_{xy} with the dimensions of the samples. In a way, this allows to illustrate some of the effects that were explained in the previous section.

The first noticeable tendency is that the value of the correlation coefficient decreases with increasing value of $\sigma E/\rho_1^2$. This confirms what was said previously about the effects of σ/ρ_1; high values of the ratio result in a material that is more sensitive to the edge constraint. Furthermore, the r_{xy} coefficients for a sample radius of 14.5 mm are significantly lower than for the other two radii. In other words, the difference between α_m and α_{th} is the greatest with the smaller sample radius. To a lesser extent, the same can be said for the sample radius of 32 mm as its r_{xy} coefficients are slightly smaller than the ones obtained with the radius of 49.5 mm. Once again, this is in accordance with the previous discussion regarding the effect of the shape factor on the sensitivity of the materials.

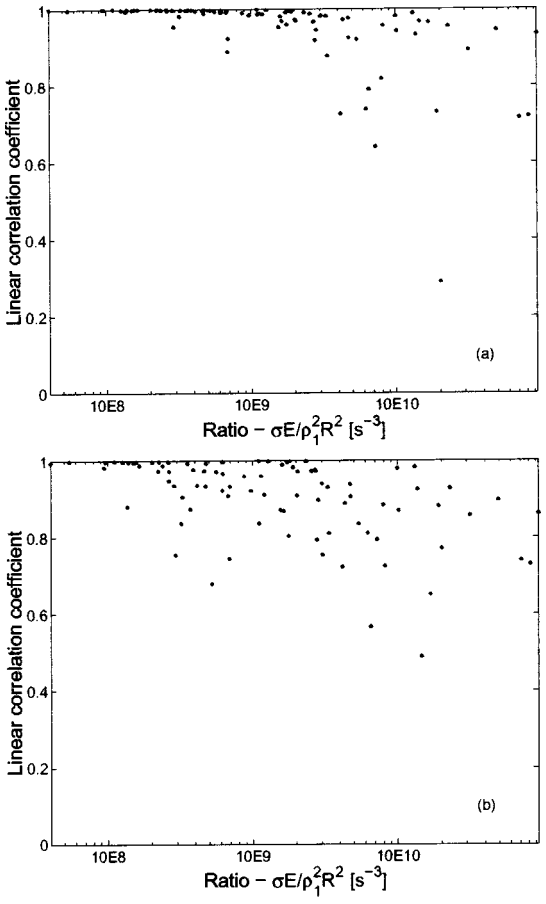


FIG. 4. Evolution of the linear correlation coefficient in the f_r region as a function of the ratio $\sigma E/\rho_1^2 R^2$. (a) r_{xy} for all the samples of thickness t $=50$ mm. (b) r_{xy} for all the samples of thickness t$=100$ mm.

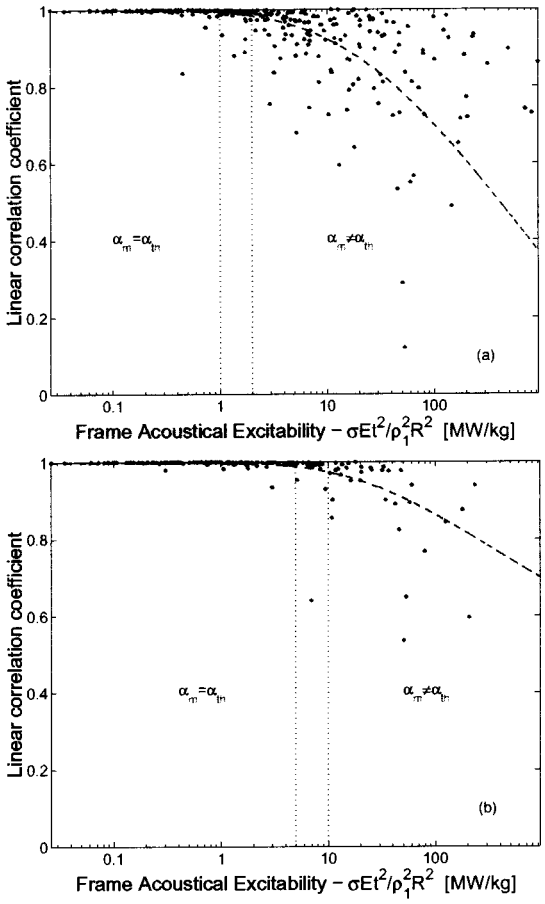


FIG. 5. Critical value of the specific structural capacity $\sigma E t^2/\rho_1^2 R^2$. (a) r_{xy} for all the samples in the f_r frequency region. (b) r_{xy} for all the samples in the high frequencies. Legend: — — —: Polynomial regression of the different r_{xy} coefficients.

Figure 3(b) presents the same data as in Fig. 3(a) except that the various r_{xy} coefficients are sorted with the $\sigma E/\rho_1^2 R^2$ ratio. Unlike in Fig. 3(a), now a clearer tendency can be observed in Fig. 3(b); the higher $\sigma E/\rho_1^2 R^2$, the greater the difference between α_m and α_{th} (r_{xy} decreases). For high values of the ratio, there are no more r_{xy} coefficients close to 1, as in Fig. 3(a).

The critical value of $\sigma E/\rho_1^2 R^2$, for a sample thickness of 25 mm, is of about 2×10^9 s^{-3}. As a result, if $\sigma E/\rho_1^2 R^2 < 2\times10^9$ s^{-3} for a given elastic foam sample, then the absorption coefficient measured with the *swt* will not differ from the theoretical absorption ($\alpha_m=\alpha_{th}$). However, if $\sigma E/\rho_1^2 R^2 > 2\times10^9$ s^{-3}, then α_m departs from α_{th}.

B. Sample thicknesses of t=50 mm and t=100 mm

In a similar manner to what is done in Fig. 3(b), Figs. 4(a) and (b) present the r_{xy} coefficients for the samples with a thickness of 50 mm and 100 mm, respectively. Compared to the results obtained for t$=25$ mm, it is evident that the r_{xy} coefficients for t$=50$ mm and t$=100$ mm start to decrease with inferior values of $\sigma E/\rho_1^2 R^2$. For t$=50$ mm, the critical value is approximately 9×10^8 s^{-3}, while for t$=100$ mm it is of about 2×10^8 s^{-3}. One can thus conclude that the thicker the sample, the more sensitive it is to the edge constraint.

Regarding the dimensions of a given material sample, it is now clear that it is always preferable to use a thin sample with an important diameter, i.e., to have a small shape factor. That way, the acoustical measurements in the *swt* are less likely to be influenced by the edge constraint.

C. All sample thicknesses

In the previous analyses, critical values in term of $\sigma E/\rho_1^2 R^2$ were obtained for the different thicknesses. For practical reasons, it is preferable to have only one critical value which is suitable for any thicknesses. Consequently, the final step is to find a unique critical value of the *FAE* parameter. To do so, all the r_{xy} coefficients presented in Figs. 3(b), and 4(a) and (b) are now plotted, in Fig. 5(a), as a function of the Frame Acoustical Excitability.

Figure 5(a) presents the r_{xy} coefficients evaluated in the vicinity of the f_r frequency region for the 270 studied samples. The r_{xy} coefficient are sorted in terms of the *FAE* parameter; the effect of the thickness is now taken into account. The critical value of *FAE* is of 1 MW/kg. Therefore, when $FAE<1$ MW/kg, the elastic foam sample is not affected by the edge constraint. In this situation, it is possible to measure the theoretical absorption coefficient α_{th} with the *swt*. On the other hand, if $FAE>2$ MW/kg, the elastic foam may become notably sensitive to the edge constraint. In other

words, the measured *swt* absorption departs from the theoretical absorption (unconstrained or laterally infinite).

When the value of the ratio is between these two values, the experimenter must show extreme caution when performing the acoustical measurements. The lateral compression must be as minimal as possible to make sure that the edge constraint effect is minimized. This situation falls in a gray zone, where it cannot be asserted that a sample is sensitive or not to the edge constraint.

The previous analyses and discussion can now be summarized in the following elasto-acoustic criterion:

$$\left.\begin{array}{rl} FAE<1 & \text{MW/kg}\rightarrow \alpha_m=\alpha_{th} \\ 1\ \text{MW/kg}<FAE<2 & \text{MW/kg}\rightarrow \alpha_m\simeq\alpha_{th} \\ FAE>2 & \text{MW/kg}\rightarrow \alpha_m\neq\alpha_{th} \end{array}\right\} \quad \text{elasto-acoustic criterion.} \tag{10}$$

To conclude the analysis, Fig. 5(b) presents the r_{xy} coefficients evaluated in the high frequency region, more specifically from f'' up to the cut-off frequency of tube for the 270 studied samples. These results show that evaluating the r_{xy} coefficients in the vicinity of the f_r frequency is sufficient to give a general idea of the material behavior. In this case, the critical value of FAE is 5 times greater than the one identified in Fig. 5(a); as explained in Sec. III, this result is logical.

VII. CONCLUSION

The objective of this paper was to develop a practical method to identify the elastic foams that are sensitive to the circumferential edge constraint. To achieve this objective, an elasto-acoustic criterion was developed. The criterion relies on the evaluation of a new parameter, the Frame Acoustical Excitability (FAE), based on easily measured properties of the material sample; $FAE=\sigma E t^2/\rho_1^2 R^2$. The FAE takes into account the specific stiffness of the frame, the visco-inertial coupling between the two phases and the size, or shape factor, of the sample. The developed criterion consists in comparing this ratio to critical values which have been determined from a large number of numerical experiments. For any sample sizes of a given elastic foam, the absorption coefficient measured with the *swt* (α_m) will be equal to the theoretical absorption (α_{th}) only if $FAE<1$ MW/kg. On the other hand, if $FAE>2$ MW/kg, then $\alpha_m\neq\alpha_{th}$. In between those two critical values, there is some sort of a gray zone where $\alpha_m\simeq\alpha_{th}$. It is then up to the experimenter to decide whether or not an accurate measurement of α_{th} is necessary. The use of this criterion is valid for elastic foams whose physical properties are contained within the prescribed values in Table I.

ACKNOWLEDGMENTS

The authors wish to thank N.S.E.R.C. Canada, F.Q.R.N.T. Quebec and C.Q.R.D.A. Quebec for their financial support. This study was also supported in part by the ''IMPACT'' research project of the Rhône-Alpes Region in France.

APPENDIX: ANALYTICAL EXPRESSION OF Z_n

In making the assumption that the poroelastic material has infinite lateral dimensions and that it is excited by a normally incident plane wave, one can thus neglect the effect of the shear wave in the material. Hence, the normal surface impedance only depends on the two compressional waves and is defined as:[12]

$$Z_n=-j\frac{(Z_1^s Z_2^f \mu_2 - Z_2^s Z_1^f \mu_1)}{D}, \tag{A1}$$

where D is

$$D=(1-\phi+\phi\mu_2)[Z_1^s-(1-\phi)Z_1^f\mu_1]\tan(\delta_2 l) + (1-\phi+\phi\mu_1)[Z_2^f\mu_2(1-\phi)-Z_2^s]\tan(\delta_1 l), \tag{A2}$$

with

$$\left.\begin{array}{l} \mu_i=\dfrac{\tilde{P}\delta_i^2-\omega^2\tilde{\rho}_{11}}{\omega^2\tilde{\rho}_{12}-\tilde{Q}\delta_i^2} \\ Z_i^f=(\tilde{R}+\tilde{Q}/\mu_i)\dfrac{\delta_i}{\phi\omega} \\ Z_i^s=(\tilde{P}+\tilde{Q}/\mu_i)\dfrac{\delta_i}{\omega} \end{array}\right\} \quad \forall\ \ i=1,2 \tag{A3}$$

and

$$\delta_1^2=\frac{\omega^2}{2(\tilde{P}\tilde{R}-\tilde{Q}^2)}[\tilde{P}\tilde{\rho}_{22}+\tilde{R}\tilde{\rho}_{11}-2\tilde{Q}\tilde{\rho}_{12}-\sqrt{\Delta}], \tag{A4}$$

$$\delta_2^2=\frac{\omega^2}{2(\tilde{P}\tilde{R}-\tilde{Q}^2)}[\tilde{P}\tilde{\rho}_{22}+\tilde{R}\tilde{\rho}_{11}-2\tilde{Q}\tilde{\rho}_{12}+\sqrt{\Delta}], \tag{A5}$$

$$\Delta=(\tilde{P}\tilde{\rho}_{22}+\tilde{R}\tilde{\rho}_{11}-2\tilde{Q}\tilde{\rho}_{12})^2-4(\tilde{P}\tilde{R}-\tilde{Q}^2) \times(\tilde{\rho}_{11}\tilde{\rho}_{22}-\tilde{\rho}_{12}^2). \tag{A6}$$

In the previous equations, ϕ is the porosity of the material, μ_1 and μ_2 are the ratios of the velocity of the air over the velocity of the frame for the given compressional wave. They indicate the phase of the material in which the waves mainly propagate. Z_1^s and Z_2^s are characteristic impedances related to the propagation of the compressional waves in the frame, and Z_1^f and Z_2^f are related to their propagation in the air. δ_1 and δ_2 are the complex wave numbers. $\tilde{P}$, $\tilde{Q}$, $\tilde{R}$ and $\tilde{\rho}_{11}$, $\tilde{\rho}_{12}$, $\tilde{\rho}_{22}$ are, respectively, Biot's complex elastic coefficients and complex effective densities. The expressions of these parameters are detailed elsewhere.[12]

[1] R. Panneton and N. Atalla, ''An efficient finite element scheme for solving the three-dimensional poroelasticity problem in acoustics,'' J. Acoust. Soc. Am. **101**, 3287–3298 (1997).

[2] A. Cummings, ''Impedance tube measurements on porous media: The effects of air-gaps around the sample,'' J. Sound Vib. **151**, 63–75 (1991).

[3] D. Pilon and R. Panneton, ''Effects of circumferential air gaps on the measurement of the absorption coefficient of poroelastic materials,'' J. Acoust. Soc. Am. **112**, 2383(A) (2002).

[4] R. J. Donato, ''Model experiments on surface waves,'' J. Acoust. Soc. Am. **63**, 700–703 (1978).

[5] J.-F. Allard and P. Delage, ''Free field measurements of absorption coef-

ficients on square panels of absorbing materials,'' J. Sound Vib. **101**, 161–170 (1985).
[6] Y. J. Kang and J. S. Bolton, ''Finite element modeling of isotropic elastic porous materials coupled with acoustical finite elements,'' J. Acoust. Soc. Am. **98**, 635–643 (1995).
[7] T. E. Vigran, L. Kelders, W. Lauriks, P. Leclaire, and T. F. Johansen, ''Prediction and measurements of the influence of boundary conditions in a standing wave tube,'' Acta Acust. (Beijing) **83**, 419–423 (1997).
[8] B. H. Song, J. S. Bolton, and Y. J. Kang, ''Effect of circumferential edge constraint on the acoustical properties of glass fiber materials,'' J. Acoust. Soc. Am. **110**, 2902–2916 (2001).
[9] M. D. Dahl, E. J. Rice, and D. E. Groesbeck, ''Effects of fiber motion on the acoustic behavior of an anisotropic, flexible fibrous material,'' J. Acoust. Soc. Am. **87**, 54–66 (1990).
[10] M. A. Biot, ''Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range,'' J. Acoust. Soc. Am. **28**, 168–178 (1956).
[11] M. A. Biot, ''Theory of propagation of elastic waves in a fluid-saturated porous solid. II. Higher frequency range,'' J. Acoust. Soc. Am. **28**, 179–191 (1956).
[12] J.-F. Allard, *Propagation of Sound in Porous Media. Modeling Sound Absorbing Materials* (Elsevier Applied Sciences, New York, 1993).
[13] D. Pilon, R. Panneton, and F. Sgard, ''Convergence of Nth order Biot poroelastic finite elements,'' Proc. 9th International Conference on Sound & Vib., Orlando, July 8–11 (2002).
[14] N. Atalla, R. Panneton, and P. Debergue, ''A mixed displacement-pressure formulation for poroelastic materials,'' J. Acoust. Soc. Am. **104**, 1444–1452 (1998).
[15] P. Debergue, R. Panneton, and N. Atalla, ''Boundary conditions for the weak formulation of the mixed (u,p) poroelasticity problem,'' J. Acoust. Soc. Am. **106**, 2383–2390 (1999).
[16] H. Wadsworth, *Handbook of Statistical Methods for Engineers and Scientists* (McGraw-Hill, New York, 1990).
[17] C. Langlois, R. Panneton, and N. Atalla, ''Polynomial relations for quasi-static mechanical characterization of isotropic poroelastic materials,'' J. Acoust. Soc. Am. **110**, 3032–3040 (2001).
[18] M. R. Stinson and G. A. Daigle, ''Electronic system for the measurement of flow resistance,'' J. Acoust. Soc. Am. **83**, 2422–2428 (1988).
[19] Y. Champoux, M. R. Stinson, and G. A. Daigle, ''Air-based system for the measurement of the porosity,'' J. Acoust. Soc. Am. **89**, 910–916 (1990).
[20] R. Panneton and Y. Atalla, D. Blanchet, and M. Bloor, ''Validation of the inverse method of acoustic material characterization,'' Noise & Vibration Conference and Exhibition, available from SAE Technical Paper Series (ISSN 0148-7191, www.sae.org), paper 2003-01-1584 (2003).
[21] J.-F. Allard, B. Castagnède, M. Henry, and W. Lauriks, ''Evaluation of tortuosity in acoustic porous materials saturated by air,'' Rev. Sci. Instrum. **65**, 754–755 (1994).
[22] P. Leclaire, L. Kelders, W. Lauriks, M. Melon, N. R. Brown, and B. Castagnede, ''Determination of the viscous and thermal characteristic lengths of plastic foams by ultrasonic measurements in helium and air,'' J. Appl. Phys. **80**, 2009–2012 (1996).
[23] O. C. Zwikker and C. W. Kosten, *Sound-Absorbing Materials* (Elsevier, Amsterdam, 1949).
[24] J.-F. Allard, C. Depollier, P. Guignouard, and P. Rebillard, ''Effects of a resonance of the frame on the surface impedance of glass wool of high density and stiffness,'' J. Acoust. Soc. Am. **89**, 999–1001 (1991).

Blind deconvolution applied to acoustical systems identification with supporting experimental results

Michael J. Roan,[a] Mark R. Gramann,[b] Josh G. Erling,[c] and Leon H. Sibul[d]
Applied Research Laboratory, The Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804

(Received 20 December 2001; revised 16 July 2003; accepted 22 July 2003)

Many acoustical applications require the analysis of a signal that is corrupted by an unknown filtering function. Examples arise in the areas of noise or vibration control, room acoustics, structural vibration analysis, and speech processing. Here, the observed signal can be modeled as the convolution of the desired signal with an unknown system impulse response. Blind deconvolution refers to the process of learning the inverse of this unknown impulse response and applying it to the observed signal to remove the filtering effects. Unlike classical deconvolution, which requires prior knowledge of the impulse response, blind deconvolution requires only reasonable prior estimates of the input signal's statistics. The significant contribution of this work lies in experimental verification of a blind deconvolution algorithm in the context of acoustical system identification. Previous experimental work concerning blind deconvolution in acoustics has been minimal, as previous literature concerning blind deconvolution uses computer simulated data. This paper examines experiments involving three classical acoustic systems: driven pipe, driven pipe with open side branch, and driven pipe with Helmholtz resonator side branch. Experimental results confirm that the deconvolution algorithm learns these systems' inverse impulse responses, and that application of these learned inverses removes the effects of the filters. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1610465]

PACS numbers: 43.60.Cg, 43.60.Mn, 43.60.Np [JCB]

I. INTRODUCTION

Many cases arise in acoustics in which an unknown filtering function corrupts a desired signal. In these cases, the only possible measurements are of the corrupted signal. Such a process is often modeled as an input signal convolved with a filter having an unknown impulse response. One example involves a speech signal that propagates from speaker to listener through multiple paths inside a room. This would be modeled as the speech signal being convolved with a finite impulse response (FIR) filter. In a simplified sense, this filter would consist of taps that have time domain spacings corresponding to the various path length differences. The process of removing the effect of the filtering function and thus restoring the signal to its uncorrupted state is called deconvolution. This paper is concerned with the process of blind deconvolution (BDC). In BDC, neither the filtering function nor the input signal is known deterministically. Rather, a hypothesis is made regarding the form of the *a priori* probability density function (PDF) of the input signal. Based on this hypothesis, a nonlinear, adaptive, tap-weight updating algorithm learns the inverse filtering function. The learned inverse filter can then be applied to the data to remove the effects of the filter and thus restore the desired signal to its unfiltered state. In the room acoustics example, this process would result in the removal of the echoes and reverberation from the signal received by the listener. This process differs significantly from classical deconvolution, which requires prior knowledge of the inverse filter, and hence an accurate detailed model of the room and the precise locations of the speaker and listener. Although BDC recently has received extensive attention in the signal processing literature,[1–9] very little experimental BDC work has been done in acoustics. Current published results commonly use measured signals, such as voice signals,[1–5,7] that are artificially convolved with a filtering function. Broadhead and Pflug[8] considered the blind deconvolution of the underwater propagation channel and analysis of sparseness criteria. Here again they resorted to simulated data for confirmation of the algorithm. Roginsky[9] proposed the use of an algorithm based on least squares comparison between several sensors. Although Roginsky did provide experimental data, his approach differs significantly from the BDC approach used in this paper in that (a) multiple sensors are required and (b) second-order criterion such as least squares measures are employed rather than the higher order statistics of the maximum entropy approach.

This paper addresses the application of BDC to acoustical systems with supporting experimental results. The significant contributions of this paper are

(1) experimental evidence that BDC is capable of learning acoustical system impulse response functions,
(2) verification that application of the learned inverse impulse response restores the original input signal,
(3) demonstration that a controlled acoustical experiment with well-known theoretical system response functions provides an excellent platform for illustrating the operation of the BDC algorithm, and

[a]Electronic mail: mjr110@psu.edu
[b]Electronic mail: mrg227@psu.edu
[c]Electronic mail: jge107@psu.edu
[d]Electronic mail: lhs2@psu.edu

(4) identification of practical shortcomings of current BDC algorithms.

The organization of this paper is as follows: Section II provides a review of Bell and Sejnowski's derivation[1,2] of the blind deconvolution algorithm based on entropy maximization and Torkkola's[5] modification to the learning rules. Section III presents the experimental setup and parameters. Section IV describes the data processing and presents experimental results. Section V contains a discussion of practical issues in BDC, possible solutions to these problems, and conclusions.

II. BLIND DECONVOLUTION ALGORITHM

The objective in BDC is to recover the original input signal using only output measurements of a linear filter which has an unknown impulse response. Thus "blind" refers to the fact that neither the filter impulse response nor the input signal is known. The algorithm, as shown in Fig. 1, must simultaneously learn the impulse response of the system and apply it to the measured data, with the desired net effect of recovering the input signal. This section contains an outline for derivation of the iterative learning rules for blind deconvolution of a causally filtered signal. In keeping with convention and to preserve generality, the elements of all vectors in this section are considered to be dimensionless. This can be attained in practice by simply dividing the measured variable by a unit variable prior to processing.

Consider a system with an unknown input signal vector $\mathbf{s}=[s_0,s_1,\ldots,s_{N-1}]^{\mathrm{T}}$ that is passed through a causal filter, which can be described by an impulse response $\mathbf{h}=[h_0,h_1,\ldots,h_{L-1}]^{\mathrm{T}}$. The result is a system output vector $\mathbf{x}=[x_0,x_1,\ldots,x_{N+L-2}]^{\mathrm{T}}$ created as

$$\mathbf{x}=\mathbf{s}*\mathbf{h} \tag{1}$$

or

$$x_n=\sum_{k=0}^{L} h_k s_{n-k}, \tag{2}$$

where "*" denotes convolution, as defined in Eq. (2) for the case of finite length **h**. An example of such an acoustical system is a discrete time sampled version of the multipath propagation problem described in Sec. I. There **h** is a discrete time FIR estimation of the system's continuous time filtering function. Given the measured signal **x**, the goal is to reverse the effects of the filtering through convolution of **x** with an estimated inverse impulse response $\mathbf{w}=[w_0,w_1,\ldots,w_{L-1}]^{\mathrm{T}}$. The desired result is then to find $\mathbf{w}\approx\mathbf{h}^{-1}$ such that

$$\mathbf{s}\approx\mathbf{u}=\mathbf{w}*\mathbf{x}, \tag{3}$$

where **u** takes the form $\mathbf{u}=[u_0,u_1,\ldots,u_{M+L-1}]^{\mathrm{T}}$. Assuming that **h** does not consist merely of a single scaled and delayed delta function, it is apparent from Eq. (2) that the convolutive filtering introduces some redundancy into the measured signal. Therefore, one approach to reversing the filtering effect is to use **w** to eliminate redundancy between time points u_j and u_k for all pairs where $j<k$. Because this operation is conducted over all pairs (j,k), it is convenient to consider subsets of **u** of arbitrary length Q, $\mathbf{u}_j=[u_j,\ldots,u_{j+Q}]^{\mathrm{T}}$ and $\mathbf{u}_k=[u_k,\ldots,u_{k+Q}]^{\mathrm{T}}$. Then, eliminating the redundancy can be seen as reducing the mutual information between vectors $\mathbf{u}_j$ and $\mathbf{u}_k$. Defining entropy as

$$H(\mathbf{u}_j)=-E[\log(f(\mathbf{u}_j))], \tag{4}$$

where $E[\cdot]$ denotes the expectation operator and $f(\mathbf{u})$ denotes the PDF of **u**, and mutual information for a single pair is defined as

$$I(\mathbf{u}_j,\mathbf{u}_k)=H(\mathbf{u}_j)+H(\mathbf{u}_k)-H(\mathbf{u}_j,\mathbf{u}_k), \tag{5}$$

where $H(\mathbf{u}_j,\mathbf{u}_k)$ denotes the joint entropy of $\mathbf{u}_j$ and $\mathbf{u}_k$.

Therefore, the goal becomes to minimize $I(\mathbf{u}_j,\mathbf{u}_k)$, which in general can be accomplished by maximizing the joint entropy $H(\mathbf{u}_j,\mathbf{u}_k)$. Unfortunately, this raises the issue that $H(\mathbf{u}_j,\mathbf{u}_k)$ can be increased indefinitely by merely increasing the variance of $\mathbf{u}_j$ or $\mathbf{u}_k$. Introducing an appropriate nonlinear transformation $g(\cdot)$, which acts as a "squashing" or limiting function, addresses this problem by limiting the maximum variance. Therefore, it becomes useful to consider the transformed vector $\mathbf{y}=g(\mathbf{u})$ in place of **u**, noting that $g(\cdot)$ is memoryless and acts elementwise, so that $y_i=g(u_i)$. For the case of super-Gaussian (Pearson kurtosis >3) input, Bell and Sejnowski[1] suggest that $g(\cdot)=\tanh(\cdot)$ is a suitable, although not necessarily optimal, choice of nonlinear function. It is convenient to consider the convolution **u** in Eq. (3) as a matrix operation:

$$\begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_{M+L-1} \end{bmatrix} = \mathbf{W}*\mathbf{x}$$

$$= \begin{bmatrix} w_0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ w_1 & w_0 & 0 & 0 & 0 & \ddots & \vdots \\ \vdots & w_1 & w_0 & 0 & 0 & \ddots & 0 \\ w_{L-1} & \cdots & w_1 & w_0 & 0 & \ddots & 0 \\ 0 & \ddots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & w_1 & w_0 & 0 \\ 0 & \cdots & 0 & w_{L-1} & \cdots & w_1 & w_0 \end{bmatrix}$$

$$\times \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{M-1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{6}$$

and then **y** is given by

$$\mathbf{y}=g(\mathbf{u})=g(\mathbf{W}*\mathbf{x}), \tag{7}$$

where "*" now denotes matrix multiplication. Introducing the restriction that **w** is causal, **W** then takes the form of a lower triangular matrix of dimension $(M+L)\times(M+L)$, as shown in Eq. (6). Because the vector **y** is a function of the

vector **x**, the joint PDF of **y**, $f(\mathbf{y})$ is a function of the joint PDF of **x**, $f(\mathbf{x})$,

$$f(\mathbf{y}) = \frac{f(\mathbf{x})}{|J|}, \tag{8}$$

where $|J|$ is the Jacobian of the transformation from the vector **x** to the vector **y**, defined as

$$|J| = \det\left[\frac{\partial y_i}{\partial x_j}\right]_{ij} = \det(\mathbf{W}) \cdot \prod_{n=0}^{L+M-1} y_n'. \tag{9}$$

The joint entropy of **y**,

$$H(\mathbf{y}) = -E\left\{\log\left[\frac{f(\mathbf{x})}{|J|}\right]\right\}, \tag{10}$$

can then be maximized with respect to **W** by maximizing $\log|J|$, evaluated as

$$\log|J| = \log(w_0^{M+L-1}) + \sum_{n=0}^{M+L-1} \log(y_n'). \tag{11}$$

The updating rules for the inverse impulse response **w** are then determined by differentiation of Eq. (11) as

$$\Delta w_n = \frac{\partial}{\partial w_n}(\log|J|). \tag{12}$$

Given the selected nonlinear transformation $g(\cdot) = \tanh(\cdot)$, the resulting updating rules are described by

$$\Delta w_0 = \sum_{n=0}^{M+L-1} \left(\frac{1}{w_0} - 2x_n y_n\right), \tag{13}$$

$$\Delta w_k = -\sum_{n=k}^{M+L-1} 2x_{n-k} y_n. \tag{14}$$

This derivation and result are based in a large part on the work of Bell and Sejnowski,[1] with minor changes in notation and in the dimension of **W**. Their work suggests partitioning **x** into blocks of length *M*, and calculating and applying the updating weights $\Delta\mathbf{w}$ once per block. Torkkola[5] suggests that rather than operating on blocks of data in this manner, the updating weights $\Delta\mathbf{w}$ should be calculated and applied at each output sample, beginning with the *L*th sample. This eliminates the dependence of $\Delta\mathbf{w}$ on terms calculated using only part of the inverse impulse response. The resulting simplified updating rules for a given sample *n* are

$$\Delta w_0 = \frac{1}{w_0} - 2x_n y_n, \tag{15}$$

$$\Delta w_k = -2x_{n-k} y_n. \tag{16}$$

In practice these values in Eqs. (13)–(16) will be multiplied by a scalar, time-varying, learning rate function $\mu(n)$. This function generally decreases with increasing *n* in order to aid in convergence of **w**. Values for $\mu(n)$ can generally range in the order of 0.01 to 5×10^{-6}, although this can vary based on the length of the specified inverse filter (smaller values are generally used for longer filters). Using this scalar multiplier, **w** is then updated as

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu(n)\Delta\mathbf{w}_n. \tag{17}$$

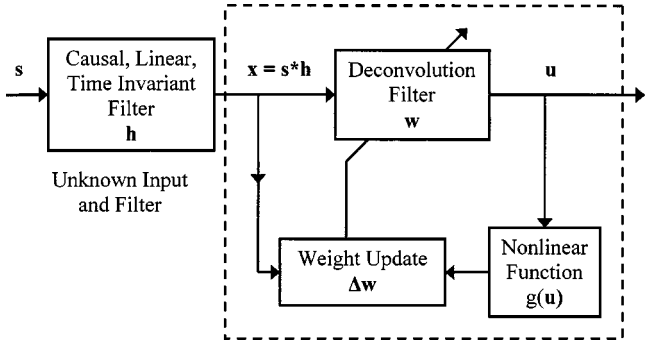


FIG. 1. Blind deconvolution algorithm.

At the beginning of the learning process, **w** must be initialized to a nonzero value. In this case, **w** is initialized to match a situation equivalent to no filtering, so $w_0 = 1$ and all remaining w_k are set to zero. In some implementations, a scaling factor is included in the input to the nonlinear function. In this case, because it is assumed that the input signal is stationary and processing is done offline, the required scaling is applied by dividing measured vector **x** by two times the standard deviation of **x** prior to processing.

III. EXPERIMENTAL VERIFICATION OF THE BLIND DECONVOLUTION ALGORITHM

Several simple experiments allow verification of the BDC algorithm and analysis of its performance when applied to actual acoustic systems. These experiments consist of filtering an acoustic signal through three relatively simple acoustic systems, and in each case then applying the BDC algorithm to the measured system output. This section contains a description of the experimental setup and provides estimated theoretical responses of the specific acoustic systems.

A. Experiment parameters

The three acoustical systems consist of cylindrical waveguides constructed with poly vinyl chloride (PVC) pipe with inner diameter of 7.62 cm and length of 1 m. One end of the pipe has an open termination. A moving coil driver is securely fit to the opposite end of the pipe. The size of the driver is such that the diameter of rubber cone surround is just smaller than the inner diameter of the pipe, resulting in a driving piston of area nearly equal to the inner cross-sectional area of the pipe. The first system consists of only the pipe and driver with no side branches. The second adds an open side branch at half the length of the pipe, and the third adds a Helmholtz resonator side branch at half the length of the pipe. These systems are diagrammed in Figs. 2(a)–(c), respectively, and will henceforth be referred to as system 1, system 2, and system 3.

A voltage source of super-Gaussian white noise with a bandwidth of approximately 1 kHz is applied to the driver. Given a zero mean Gaussian random vector **z** of length 8192 this input signal is calculated as

$$s_n = \text{sign}[z_n] \cdot \text{abs}[z_n]^{(1.5)} \tag{18}$$

where

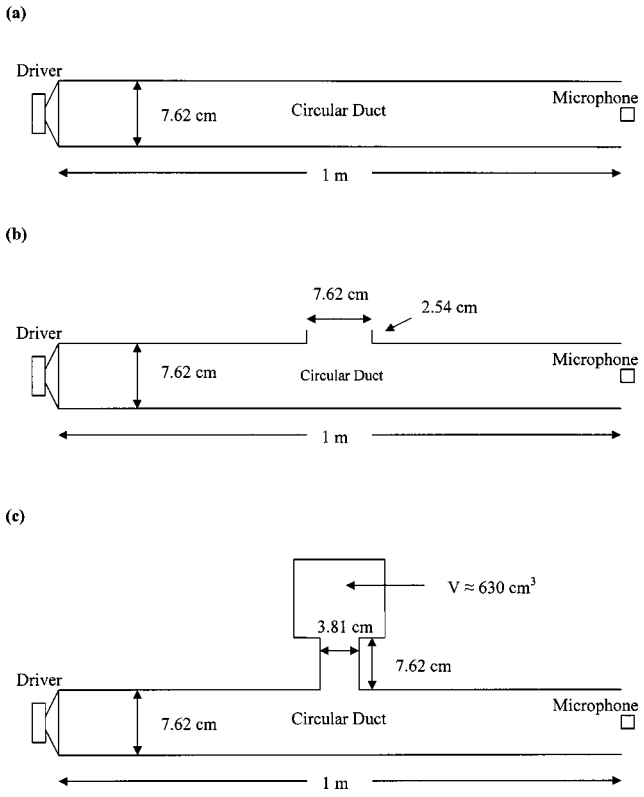


FIG. 2. Experimental setup—three classical filters. (a) system 1: driver, open pipe. (b) System 2: driver, open pipe, open side branch. (c) System 3: driver, open pipe, Helmholtz resonator side branch.

$$\text{sign}(x)=\begin{cases} 1, & x\geqslant 0, \\ -1, & x<0. \end{cases} \tag{19}$$

The -6-dB bandwidth of the input signal and the frequency range of interest are limited to fall below the first non-plane-wave mode of the pipe that occurs above approximately 1.3 kHz. The input signal is produced by repeating the vector **s**, as defined in Eq. (18), using a DSPT Siglab interface at a sample rate of 2560 Hz. The resulting analog signal is then amplified using a standard solid state amplifier before being applied to the driver. The output of each acoustical system is measured with a $\frac{1}{4}$-in. Brüel & Kjaer lab microphone centered approximately 3 cm outside the open end of the pipe. The input signal and output signal are recorded at a sample rate of 5120 Hz using the Siglab interface. The estimated PDF and power density spectrum of the recorded input signal are found in Figs. 3 and 4, respectively.

B. Theoretical filter response

The BDC algorithm updates a set of filter taps that estimate the inverse impulse response of the acoustical filter. Estimated theoretical magnitude responses for the given systems are provided for comparison with the response learned by the BDC algorithm. There is some uncertainty in the parameters of the system, including but not limited to driver properties and viscous losses associated with the pipe. Therefore, the predicted system response may not precisely match the actual system response.

The magnitude power response of each system is estimated by

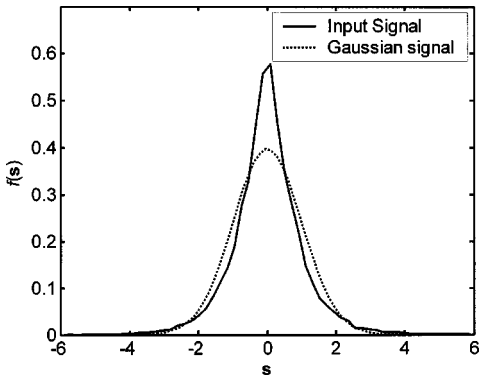


FIG. 3. Estimated PDF of recorded input signal (solid line) and Gaussian signal (dotted line).

$$\Pi(f)\approx|u(f)|^2\cdot\text{real}\{Z_{m0}(f)\}\cdot T_{\Pi\text{Branch}}(f). \tag{20}$$

Here $u(f)$ is the driver cone velocity, $Z_{m0}(f)$ is the mechanical input impedance at the driver, and $T_{\Pi\text{Branch}}(f)$ is the estimated power transmission coefficient past the given side branch. Calculations of $u(f)$, $Z_{m0}(f)$, and $T_{\Pi\text{Branch}}(f)$ for each system are based on the driver, waveguide, and side branch properties. Values of these properties and detailed descriptions of these calculations are provided in the Appendix. The resulting predicted magnitude responses of these systems are shown in Fig. 5.

IV. RESULTS

A. Processing

Approximately 12 s of data were collected for each system. Average power spectral density curves were calculated using a 2048-point window with 50% overlap for both the input and output signals. From these calculations, the measured system frequency responses were calculated. Plots of the measured response for each system, given in Fig. 6, show strong similarity in peak location, relative magnitude, and overall shape to the predicted response shown in Fig. 5.

After using the input and output signals to verify that the systems behaved as expected, the input signals were disregarded and the BDC algorithm was applied to the measured output of each system. The algorithm was run with a filter length of 150 taps and with the values of the step size function $\mu(n)$ given in Table I. With the convergence tolerance set at 0.000 005 and the given $\mu(n)$, the learned inverse filter

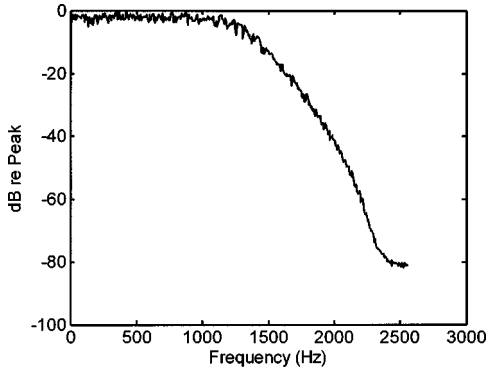


FIG. 4. Power spectral density of input signal **s**.

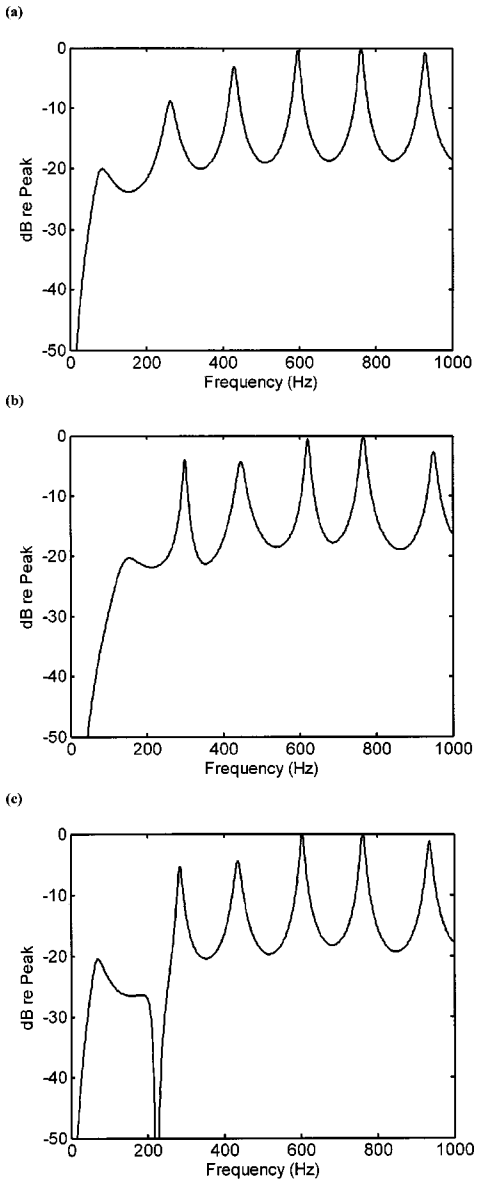


FIG. 5. Predicted magnitude response of (a) system 1, (b) system 2, and (c) system 3.

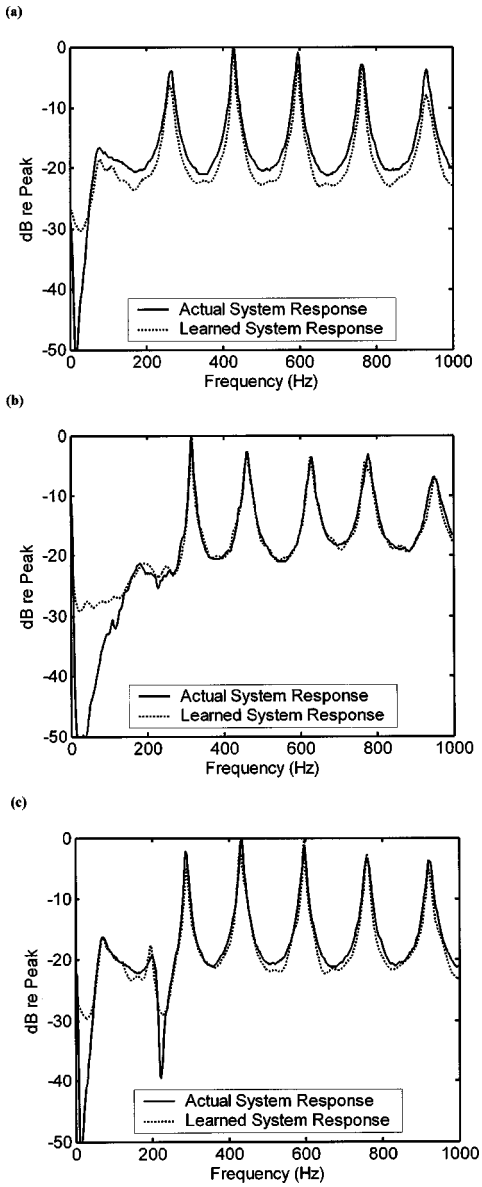


FIG. 6. Actual (solid line) and learned (dotted line) magnitude responses for (a) system 1, (b) system 2, and (c) system 3.

taps were allowed the entire 12 s of data to converge. Note that this convergence time is highly dependent on the values of the tolerance and $\mu(n)$. For the simpler systems 1 and 2, very good results can be obtained using a filter of 75 taps and only 4 s of data.

B. Frequency domain analysis

The BDC algorithm outputs the learned FIR inverse impulse response **w** of the system. This **w** allows determination of the learned frequency response of the system. Overlaid plots of the actual and learned system frequency responses, given in Fig. 6, show the high degree of similarity present. This demonstrates the ability of the BDC algorithm to accurately learn the response of these simple systems without deterministic knowledge of the input signal. Having learned the estimated inverse impulse response of the system, **w**, this FIR filter can then be convolved with the measured output, **x**, to estimate the original signal, as shown in Eq. (3). For all three systems, the frequency domain transfer functions between the input **s** and the inverse filtered data **u** are shown in Fig. 7. Comparing this to the frequency domain transfer function between the input **s** and the measured data **x** for systems 1 and 2, as shown in Figs. 6(a) and (b), it is apparent that the filtering effects are reduced from roughly 20 dB to roughly 3 dB. In the case of system 3, the peaks are reduced

TABLE I. Step size function values.

Sample n	$\mu(n)$
<1500	0.01
<5000	0.005
$<10\,000$	0.001
$<20\,000$	0.0005
$<30\,000$	0.0001
$<40\,000$	0.000 05
$<50\,000$	0.000 01
$>50\,000$	0.000 005

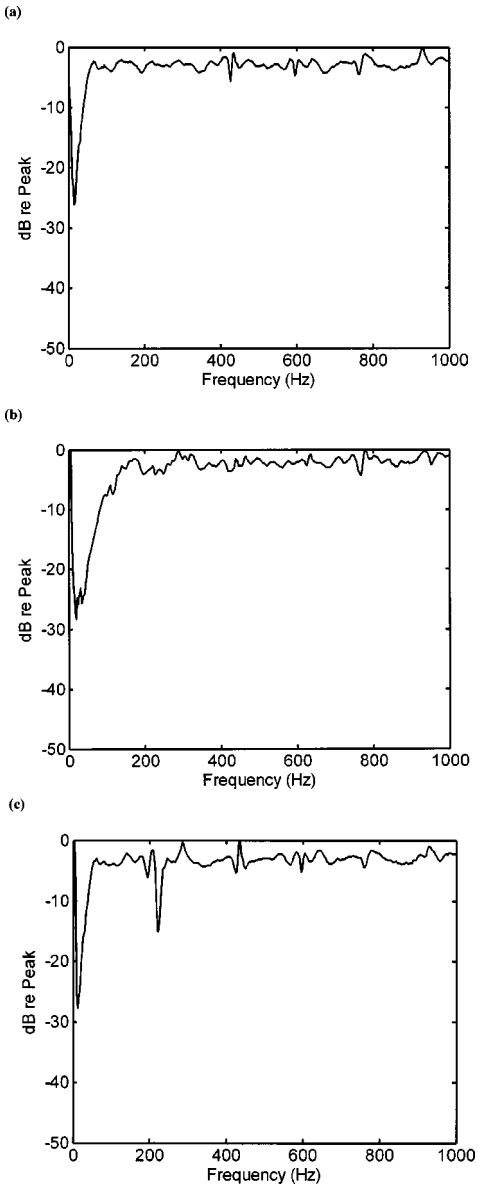


FIG. 7. Magnitude response of BDC filtered data **u** for (a) system 1, (b) system 2, and (c) system 3.

from roughly 20 dB to roughly 4 dB, with the exception of the 20-dB notch at 222 Hz that is reduced to a 10-dB notch.

C. Time domain analysis

It is important to note that this performance in reducing apparent filter effects in spectral response does not necessarily indicate performance in recovery of the original signal. Similar effects can be obtained by various means of equalization which do not actually recover the original signal. Thus to consider the actual performance in signal recovery, the time series of the measured and BDC filtered data are compared to the input time series. In this section, the vectors **s**, **x**, and **u** are assumed to be normalized to have zero mean and a zero lag autocorrelation peak of 1. In all three systems, cross correlation of the measured time series **x** and input time series **s** reveals a significant alteration of the signal. In these cases, the maximum cross correlation coefficient magnitude is approximately 0.5. In contrast, cross correlation of the inverse filtered time series **u** and input time series **s** results in maximum coefficient magnitude values between 0.8 and 0.9. These specific values are given in Table II. Overlaying the inverse filtered times series **u**, measured time series **x**, and input time series **s** also demonstrates the ability of the BDC algorithm to recover the input signal. Plots of short segments of the overlaid time series, normalized to unity variance and time aligned by maximum correlation coefficient to account for propagation delay, are shown in Figs. 8 and 9.

TABLE II. Maximum magnitude normalized cross-correlation coefficient with input time sequence.

System	Input	Time series measured	BDC measured
1	1	0.483	0.897
2	1	0.556	0.805
3	1	0.458	0.877

V. DISCUSSION

A. Practical issues

While a filter length of 150 taps was specified and apparently sufficient in these experiments, a longer filter may be necessary for adequate characterization of more complex systems. Selection of the step size function $\mu(n)$ is also an area in which further research is warranted. While the particular given values for $\mu(n)$ are sufficient here, it is almost certain that they are not optimal, resulting in higher than necessary convergence time and hence excess processing time.

Another concern regarding the performance of the algorithm is linked to the fact that in real world applications, not all system inputs are necessarily super-Gaussian or white in character. While the specifics are not given here, preliminary experimentation has shown that the given algorithm functions for a variety of sub-Gaussian and Gaussian unimodal source distributions in the three systems. This is not without cost, however, as performance is degraded in the form of increased convergence time and slightly lower maximum cross-correlation coefficient. This degraded performance is to be expected, as the nonlinearity is not optimized for these distributions, and one would expect that better knowledge of the input distribution would allow improved performance. It has been proposed[1] that input signals with other non-Gaussian distributions may also be recovered with appropriate selection of nonlinear transformation function. Bell and Sejnowski[1] have also raised the point that even with appropriate selection of nonlinear transform, the algorithm may fail to converge when applied to certain signals having a sub-Gaussian PDF. Amari *et al.*[3] have addressed these issues with an alternate tap updating algorithm. Douglas *et al.*[4] and Lee *et al.*[7] have presented automated processes for selection of appropriate nonlinear transform. Given significantly non-white input signals, such as human speech, one concern has been the over-whitening of the input signal during the recov-

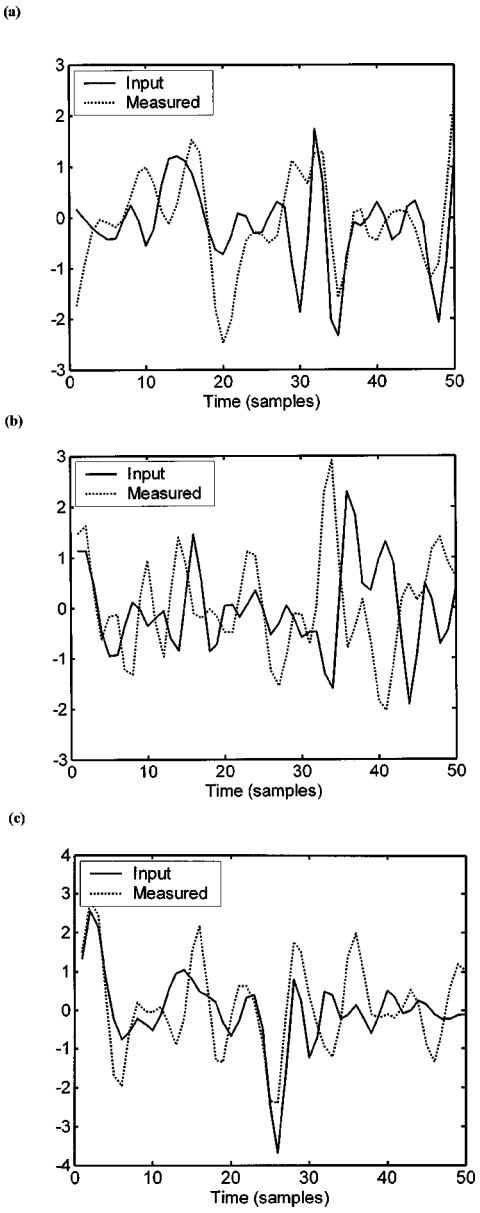


FIG. 8. Overlaid segment of input time series **s** (solid line) and measured time series **x** (dotted line) for (a) system 1, (b) system 2, and (c) system 3.

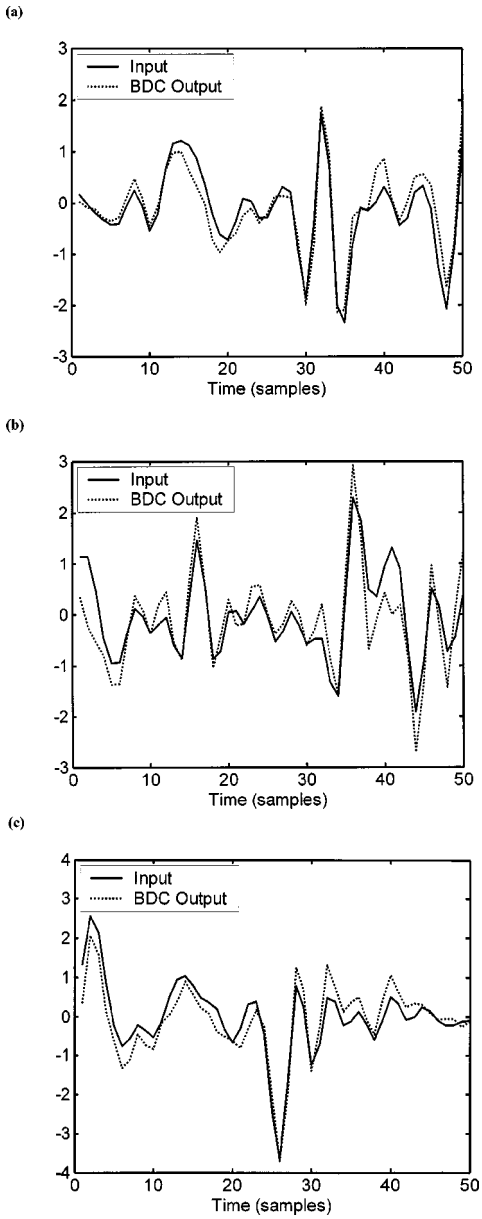


FIG. 9. Overlaid segment of input time series **s** (solid line) and BDC filtered time series **u** (dotted line) for (a) system 1, (b) system 2, and (c) system 3.

ery process. Torkkola[5] has proposed a method using a short length ''prewhitening'' filter to address this issue.

In Sec. I, the problem is framed using an example of room acoustics. While illustrative in nature, this problem would become slightly more complicated in real world applications and would pose significant challenges not seen in the experimentation with simple systems presented here. First, it would be important to note that the filter length must be at least several times the number of samples equivalent to the longest path delay. Assuming that human speech would be the source, prewhitening would probably become necessary. In addition, considering the variations inherent in human speech, the source might violate the stationarity assumption made regarding the source statistics. Improvements in the algorithm would be desirable to improve convergence time and to track source and channel nonstationarity. Future work would be useful in examining the application of blind deconvolution to such complex systems.

B. Conclusions

Filtering systems, and particularly unknown filtering systems, play an important role in acoustics. BDC can be used as a powerful tool for system identification in cases where the input signal to the system is not deterministically known. One field where this ability could prove useful is structural health monitoring, where the excitation is not controllable but instead is an environmentally induced random process. In addition to system identification, the algorithm can be used as a powerful tool for recovery of an input signal when the acoustical filtering can be described by convolution with an unknown impulse response. This feature is desirable in many tasks involving acoustic echo cancellation and signal recovery following multipath propagation in both underwater and audio systems.

The BDC algorithm has been applied to three experimental acoustical filters with known filter response func-

TABLE III. Parameter descriptions and values when applicable.

Symbol	Description	Value/Units
f	Frequency	Hz
ω	Angular frequency ($2\pi f$)	rad/s
k	Wave number (ω/c)	m
ρ_0	Density of air	1.21 kg/m^3
c	Speed of sound in air	343 m/s
S	Cross sectional area of pipe	0.0046 m^2
S_d	Effective driver cone area	0.0032 m^2
R_m	Mechanical resistance of driver	0.072 kg/s
s	Stiffness of driver cone suspension	950 N/m
m	Moving mass of driver	0.002 kg
R_0	dc electrical resistance of voice coil	6.6 Ω
L_0	Inductance of voice coil	0.2 mH
V	Voltage applied to driver	Volts
Φ_M	Transformation factor$=Bl$	5 T·m
a	Radius of pipe	3.81 cm
L_{b2}	Length of system 2 side branch	3 cm
L_{b3}	Length of system 3 side branch neck	7.6 cm
S_b	Area of system 3 side branch neck	0.0011 m^2
V_b	Volume of system 3 cavity	630 mL

tions. The results have shown that the adaptive BDC algorithm not only can learn these filter response functions, but also can reverse these filtering effects to recover the input signal. For example, in the case of system 1, the 20-dB resonance peaks in the frequency response are reduced to approximately 3 dB. In addition, BDC filtering the measured system output for system 1 improved the maximum cross-correlation magnitude with the input time series from 0.483 to 0.897. Unlike previous works[1–5] that have relied entirely on digital simulations, these results experimentally show the ability of the BDC algorithm to characterize real, physical, acoustical systems.

ACKNOWLEDGMENTS

This work was supported by Dr. David Drumheller, ONR Code 333, Contract No. N00014-00-G-0058.

APPENDIX: THEORETICAL FILTER RESPONSE CALCULATIONS

The calculations and equations in this appendix are based largely on equations found in Kinsler *et al.*[10] Table III contains a key to parameters and their values when applicable. Plots of the resulting predicted radiated power for the three systems can be found in Fig. 5.

1. General equations

From Eq. (20) of Sec. III B, it is apparent that the velocity magnitude of the driver cone $|u(f)|$ and the input impedance at the driver of the pipe system $Z_{m0}(f)$ are needed for each system.

In general for a pipe system, the input mechanical impedance at the driver is

$$Z_{m0}=\rho_0 c S_d \frac{(Z_{mL}/\rho_0 cS)+j\tan kL}{1+j(Z_{mL}/\rho_0 cS)\tan kL}, \tag{A1}$$

where Z_{mL} is the mechanical input impedance at a point a distance L from the driver. Assuming that the wavelength is long compared to the radius of the pipe, the mechanical radiation impedance from an unflanged pipe can be estimated by

$$Z_r=\rho_0 cS[\tfrac{1}{4}(ka)^2+j(0.6)ka]. \tag{A2}$$

Neglecting radiation load, the mechanical input impedance of the driver cone is

$$Z_{md}=R_m+j(\omega m-s/\omega). \tag{A3}$$

The blocked electrical impedance of the driver, Z_{EB}, the electrical impedance of the driver due to mechanical motion, Z_{MOT}, and the total electrical impedance of the driver, Z_E, are as follows:

$$Z_{EB}=R_0+j\omega L_0, \tag{A4}$$

$$Z_{MOT}=\frac{\Phi_M^2}{Z_{m0}+Z_{md}}, \tag{A5}$$

$$Z_E=Z_{EB}+Z_{MOT}. \tag{A6}$$

The velocity of the cone can then be calculated as

$$u=\frac{\Phi_M V}{Z_E(Z_{m0}+Z_{md})}. \tag{A7}$$

2. System specific equations

a. System 1

In the case of system 1, $L=1$ m, $Z_{mL}=Z_r$, and $T_{\Pi Branch}=1$. The estimated magnitude response can then be calculated as

$$\Pi(f)\approx|u(f)|^2\cdot\text{real}\{Z_{m0}(f)\}. \tag{A8}$$

b. System 2

System 2 adds an open branch at $L=0.5$ m with radius equal to the radius of the pipe a. Therefore the mechanical impedence Z_{mL} must be calculated at $L=0.5$ m, which is then

$$Z_{mL/2}=\left[\frac{1}{Z_{ab}}+\frac{1}{Z_{ah}}\right]^{-1}\cdot(S_{b2}+S)^2, \tag{A9}$$

where Z_{ab} is the acoustic impedance of the open branch given by

$$Z_{ab}=\rho_0 ck^2/4\pi+j\omega[\rho_0((L_{b2}+1.4a)/\pi a^2)] \tag{A10}$$

and Z_{ah} is the acoustic impedance of the remainder of the pipe from the branch to open end, as determined from Eq. (A1), given by

$$Z_{ah}=\rho_0 cS\frac{(Z_r/\rho_0 cS)+j\tan kL}{1+j(Z_r/\rho_0 cS)\tan kL}\cdot\frac{1}{S^2}, \tag{A11}$$

where $L=0.5$ m. Then from Eq. (A1), noting that $S_{b2}=S$, the mechanical input impedance at the driver is found to be

$$Z_{m0}=\rho_0 cS_d\frac{(Z_{mL/2}/\rho_0 c2S)+j\tan kL}{1+j(Z_{mL/2}/\rho_0 c2S)\tan kL}, \tag{A12}$$

where $L=0.5$ m. The power transmission coefficient past the side branch is estimated by the generalized lumped element high-pass transmission coefficient for an infinite pipe,

$$T_{\Pi\text{Branch}}=\frac{1}{1+(\pi a^2/2S(L_{b2}+1.4a))^2}. \quad \text{(A13)}$$

Because this coefficient is calculated based on an infinite pipe, this serves only as a rough estimation which is sufficient for these purposes. The radiated power for system 2 is then estimated as

$$\Pi(f)\approx|u(f)|^2\cdot\text{real}\{Z_{m0}(f)\}\cdot T_{\Pi\text{Branch}}. \quad \text{(A14)}$$

c. System 3

Similar to system 2, system 3 has a branch at $L=0.5$ m. Therefore the mechanical impedance $Z_{mL/2}$ at the branch must be calculated. Making the estimation that the branch has no resistance, the acoustic impedance of the branch is calculated as

$$Z_{ab}=j\rho_0[\omega(L_{b3}+1.7a)/S_b-c^2/\omega V_b]. \quad \text{(A15)}$$

The acoustic impedance of the remainder of the pipe is then the same as for system 2, given by Eq. (A11). Then substituting S_{b3} for S_{b2}, the mechanical impedance at $L=0.5$ m is then given by Eq. (A9), and the mechanical impedance at the driver is given from Eq. (A1) to be

$$Z_{m0}=\rho_0 c S_d\frac{(Z_{mL/2}/\rho_0 c2(S_{b3}+S))+j\tan kL}{1+j(Z_{mL/2}/\rho_0 c2(S_{b3}+S))\tan kL}. \quad \text{(A16)}$$

The power transmission coefficient past the side branch is estimated by the generalized lumped element band stop transmission coefficient for an infinite pipe,

$$T_{\Pi\text{Branch}}=\frac{1}{(c/2s/[w(L+1.7a)/S_b-c^2/\omega V_b])^2}. \quad \text{(A17)}$$

Noting that like the case in system 2, this is a rough estimate based on an infinite pipe, and the radiated power can be estimated by

$$\Pi(f)\approx|u(f)|^2\cdot\text{real}\{Z_{m0}(f)\}\cdot T_{\Pi\text{Branch}}, \quad \text{(A18)}$$

where the appropriate values from this section are substituted. It is noted that the given resonator, when baffled by the interior of the main duct, has a resonance frequency of approximately 223 Hz.

[1] A. J. Bell and T. J. Sejnowski, "An Information Maximization Approach to Blind Separation and Blind Deconvolution," Neural Comput. **7**, 1129–1159 (1995).

[2] A. J. Bell and T. J. Sejnowski, "Blind Separation and Blind Deconvolution: An Information Theoretic Approach," IEEE Int. Conf. on Acoust. Speech, Signal Process. **5**, 3415–3418 (1995).

[3] S. Amari, S. Douglas, A. Cichocki, and H. H. Yang, "Novel On-line Algorithms for Blind Deconvolution Using Natural Gradient Approach," Proceedings 11th IFAC Symposium on System Identification (1997), pp. 1057–1062.

[4] S. C. Douglas, A. Cichocki, and S. Amari, "Multichannel Blind Separation and Deconvolution of Sources with Arbitrary Distributions," IEEE Workshop on Neural Networks for Signal Processing (1997), pp. 436–445.

[5] K. Torkkola, "Blind Deconvolution, Information Maximization and Recursive Filters," IEEE Int. Conf. Acoust. Speech, Signal Process. **4**, 3301–3304 (1997).

[6] J. F. Cardoso, "Blind Signal Separation: Statistical Principals," Proc. IEEE **86**, 2009–2025 (1998).

[7] T. W. Lee, M. Girolami, and T. J. Sejnowski, "Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources," Neural Comput. **11**, 417–441 (1999).

[8] M. K. Broadhead and L. A. Pflug, "Performance of some sparseness criterion blind deconvolution methods in the presence of noise," J. Acoust. Soc. Am. **107**, 885–893 (2000).

[9] J. Roginsky, "Time-domain shallow water environmental inversion using unknown broadband sources," J. Acoust. Soc. Am. **99**, 2450–2457 (1996).

[10] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982), pp. 225–242.

Sensor array beamforming using random channel sampling: The aggregate beamformer

David I. Havelock[a)]
Institute for Microstructural Sciences, National Research Council, Ottawa, Ontario K1A 0R6, Canada

(Received 15 February 2003; revised 16 July 2003; accepted 22 July 2003)

The aggregate (AGG) beamformer can provide significant savings in hardware and software cost and complexity relative to conventional beamformers while retaining equivalent directional performance. The key to achieving the savings is collecting data from array channels in a random sequence, rather than simultaneously or sequentially. This allows the AGG beamformer to convert unwanted off-beam signals into wideband noise that can then be reduced by filtering. The total sampling rate is chosen to obtain the desired residual noise level and signal bandwidth. It is independent of the number of sensors. Analog anti-alias filters normally are not required, since the Nyquist frequency is determined by the total sampling rate, not the (lower) per-channel sampling rate. Beamformer delay quantization and steering resolution are greatly improved relative to conventional beamforming without the need for data interpolation. The beamformed signal, prior to decimation filtering, is obtained without arithmetic operations on the data. Low front-end hardware-complexity makes the AGG beamformer suitable for highly integrated systems. The number of sensors in the array can be altered without re-configuring buffers or altering the sampling rate. The principles of the AGG beamformer are introduced, and simulation and experimental results demonstrate performance. [DOI: 10.1121/1.1610464]

PACS numbers: 43.60.Fg, 43.60.Dh, 43.30.Wi [JCB]

I. INTRODUCTION

Arrays have been developed for many acoustical applications. Major uses include tracking sound sources in rooms and auditoria,[1–4] underwater,[5] and outdoors,[6,7] as well as ultrasonic imaging.[8] With the objective of minimizing the number of sensor channels, and thereby minimizing the requirements for hardware and data processing, sparse,[6,9–11] nested,[4] and synthetic aperture arrays[12,13] have been developed. Even though these techniques reduce requirements, each sensor and data channel normally requires its own analog anti-alias filter (AAF) and analog to digital converter/conversion (ADC).

While the requirement for AAF's can be eliminated by using delta-sigma analog to digital converters ($\Delta\Sigma$-ADC),[14] and the number of ADC's required can be reduced by multiplexing channels (Fig. 1), it is not possible to do both simultaneously, since a $\Delta\Sigma$-ADC cannot process multiplexed data. Thus, conventional beamformers require for each channel either an anti-alias filter or a dedicated ADC.

In this paper we present a beamforming technique, called here the *aggregate* (AGG) *beamformer*, that reduces significantly the cost of obtaining results equivalent to those obtained by conventional beamformers. The AGG beamformer does not require anti-alias filters for each channel, and permits sharing an ADC among many channels (Fig. 2). The technique is based on random sampling of array elements. It is particularly useful for arrays with many elements, since electronic component costs and computational complexity do not increase significantly with the number of sensors. The technique is also advantageous for applications with a narrow bandwidth or where component integration is important. The AGG beamformer provides higher beam steering resolution and less delay quantization error than conventional beamforming without the need for interpolation.[15]

Unlike conventional beamforming, the total sampling rate is determined by a combination of the signal bandwidth and the desired residual noise level, rather than by the product of the signal bandwidth and the number of sensors in the array. The number of sensors in the array can be changed easily without modifying the beamforming structure. The reduced number of required components simplifies the integration of sensors and processors. In spite of its relative simplicity, the AGG beamformer uses design principles similar to those for conventional beamformers, and it has the same directional response characteristics.

To demonstrate the operation and advantages of the AGG beamformer, its performance is compared to that of a delay and sum (DS) beamformer, which is one form of conventional time-domain beamforming. Beamforming with a DS beamformer is performed as a weighted sum of sensor (channel) signals after applying delays to compensate for propagation delays between sensors.[16] The delays ensure that signals arriving from the desired location (the beam steering direction) will sum coherently, and signals arriving from other locations will experience partial or complete signal cancellation. Weights may be applied in the DS beamformer to adjust the directional response pattern. A more generalized form of time-domain beamforming, called filter and sum beamforming, combines multiple delays from each channel, and can be used to improve control over the directional response pattern. This and other extensions to the basic beamforming principles, such as frequency domain

a)Electronic mail: david.havelock@nrc.ca

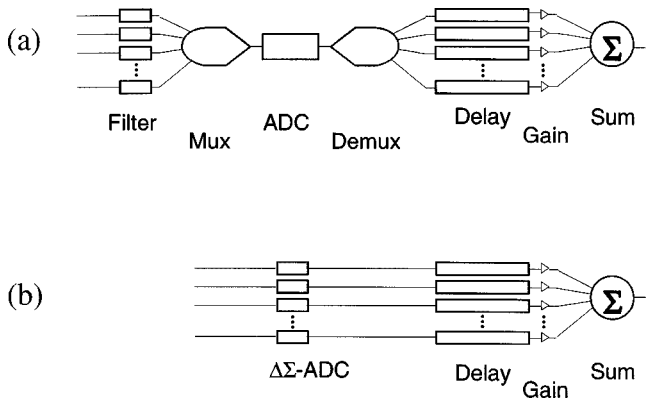


FIG. 1. Configurations for DS (delay and sum) beamforming. A single analog to digital converter (ADC) can be shared by using a multiplexer/demultiplexer pair (a) or the bank of anti-alias filters can be replaced with a bank of ΔΣ-type ADC (b).

beamforming[16,17] and adaptive beamforming,[18] are not addressed here.

The technique of random channel selection, which is essential to the AGG beamformer, has also been used to create a directional sound field with a "direct digital loudspeaker," [19,20] consisting of an array of small, digitally (pulse) actuated transmitters. While the AGG beamformer has principles in common with the direct digital loudspeaker, it differs in that it uses microphones rather than loudspeakers, requires an ADC, creates no harmonic distortion, and noise can be removed by filtering.

Another form of randomness that has been used in beamforming receivers is to dither the steering direction to achieve spatial averaging. This can reduce coherent interference,[21] provide a more constant frequency response near the main beam direction,[22] and reduce feedback effects,[23] but beam dithering is distinct from the use of randomness in the AGG beamformer.

In this paper we present the basic principles of the AGG beamformer. It extends earlier brief descriptions.[24,25] A simple demonstration is given to illustrate its implementation. Section II compares the general structures of the AGG (aggregate) and DS (delay and sum) beamformers, and describes the principles of operation. In Sec. III we present the underlying theory of the AGG beamformer and compare it with the DS beamformer. The principal characteristics are described, and *collisions* and *voids* are explained. In Sec. IV, AGG and DS beamformers are compared using experimental results obtained with a real-time system. In Sec. V, characteristics of the AGG beamformer are discussed. Conclusions are presented in Sec. VI.

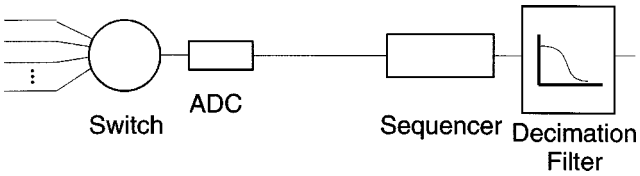


FIG. 2. Configuration for the AGG (aggregate) beamformer. The switch selects channels randomly.

II. STRUCTURE OF THE AGG BEAMFORMER

The structure of the AGG (aggregate) beamformer is illustrated in Fig. 2. Signals from individual sensors are multiplexed in a random sequence by a switch and are digitized at regular time intervals by a single analog to digital converter (ADC). The digital data samples are organized in an *alignment buffer* by a sequencer. The beamformer signal output is obtained by low-pass filtering and decimating the data.

The alignment buffer is a key feature of the AGG beamformer. The AGG beamformer works by assembling all the digitized samples from an array into this single buffer. The samples are arranged in the alignment buffer using the beamforming delays so that signals arriving from a source in the array steering direction are properly reconstructed. Signals from sources that are not in the array steering direction appear in the alignment buffer with samples out of sequence, reducing them to random, broadband noise. The sequenced data in the alignment buffer generally have a sampling rate many times greater than the sampling rate required to represent the desired bandwidth. When these data are filtered and decimated to obtain the desired frequency band and sampling rate, the decimation filter removes most of the residual noise, and thereby eliminates the unwanted off-beam signal power.

The AGG beamformer requires only a single alignment buffer, while the conventional DS beamformer requires a bank of delay lines (compare Figs. 1 and 2). The AGG beamformer beam is steered in the same way as the conventional beamformer, by applying sample delays, but the higher sampling rate allows greater time delay resolution.

For multiple simultaneous beams, the AGG beamformer requires a separate alignment buffer and decimation filter for each, while the DS beamformer requires a separate summing network for each simultaneously computed beam. With an AGG beamformer, no gain is applied and no sum is calculated. The only computation on data required by an AGG beamformer is for decimation filtering. Where an analog output is desired, the computational effort required for the decimation filter is comparable to that of the signal reconstruction filter required by a DS beamformer.

A. Collisions and voids

The key sequencing process of the AGG beamformer is illustrated in Fig. 3. Each digitized data sample is placed in the alignment buffer with a delay (or advance) in position equal to the beamforming sample delay for the channel from which the data sample was digitized. Two phenomena unique to the sequencing action are *collisions* and *voids*. A collision occurs when there is an attempt to place two or more data samples in the same entry of the alignment buffer. A void occurs when an entry in the alignment buffer receives no data samples. An important aspect of AGG beamformer design is the handling of collisions and voids (see Secs. III C and III D).

B. Example of C code segment for main processing step

The following segment of **C** code illustrates implementation of the AGG beamformer. The buffer **Xin** contains

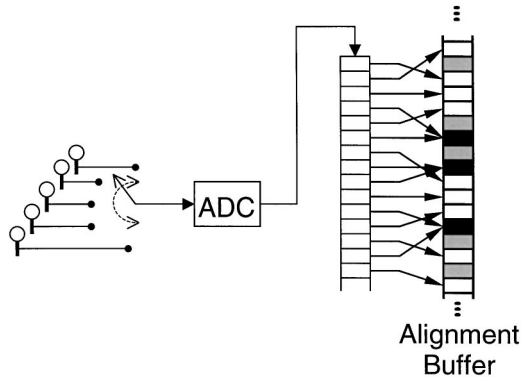


FIG. 3. The aggregate beamformer sequencer places data into the alignment buffer according to the beamforming time delays. Voids (gray boxes) and collisions (black boxes) may occur.

the digitized data, **S** contains the pre-computed random index sequence, and **D** contains the beamforming delays in unit samples. The output buffer **Z** must be low-pass filtered and decimated (code not shown.) The constant NODATA is an unused data value reserved to flag the absence of valid data. A random number generator (rand), which is uniformly distributed between 0 and 1, is used to do a coin-toss. Buffer end effects and continuity between successive buffers are handled by additional code (not shown):

```
/* Code segment for a buffer of data for the AGG*/
for (ii=start; ii<start+length; ++ii) {
    iput=ii+D[S[ii]];
    /* Collisions are managed by a coin-toss*/
    if (Z[iput]==NODATA) Z[iput]=Xin[ii];
    else if (rand()<0.5) Z[iput]=Xin[ii];
}
/* Voids are filled by repetition of the previous
value*/
for (ii=start; ii<start+length; ++ii) {
    if (Z[ii]==NODATA) Z[ii]=Z[ii-1];
}
```

III. THEORY

The AGG (aggregate) beamformer and the conventional DS (delay and sum) beamformer can be compared algebraically by introducing an intermediary expression for the data collected in buffers prior to computing beamformer output. When a natural index sequence is applied to these data, the conventional DS beamformer is shown to be a weighted block-sum operator on the resulting data sequence [Eq. (5)], while the AGG beamformer is a filtering operation [Eq. (12)].

A. Formulation of the beamformers

1. The delay and sum (DS) beamformer

DS beamforming aligns the signals from each sensor of an array so that a signal arriving with the chosen time delays will add coherently. Signals arriving from other directions sum incoherently and, normally, are partially cancelled. DS beamforming is performed in the time domain. It is the simplest form of conventional beamforming and is defined as

$$X_B(t)=\sum_{m=0}^{M-1} w_m x_m(t+\tau_m), \tag{1}$$

where $X_B(t)$ is the beamformer output at time t, M is the total number of sensors in the array, w_m is a weight factor, or array shading, for sensor m, x_m is the signal received on sensor m, and τ_m is the beamforming time delay applied to sensor m. It is convenient to normalize the weight factors,

$$\sum_{m=0}^{M-1} w_m=1. \tag{2}$$

Weight factors are used to adjust the shape of the directional response pattern of the beamformer. Sensor gain or sensitivity calibration adjustments may also be required.

The time delays τ_m are determined from the array geometry and the shape of the arriving wave front. For example, the time delays for a plane wave with normalized wave vector $\vec{v}$ arriving at an array with elements located at positions $\vec{x}_m$ are given by the inner product $\tau_m=(\vec{x}_m-\vec{x}_0)\cdot\vec{v}$, where $\vec{x}_0$ is a convenient reference point, such as the array center. For implementation without interpolation, the beamforming time delays are quantized to the unit sampling interval Δt.

We define the *sampling set* S_B for beamforming in Eq. (1) as the sequence of ordered pairs of sample times t_n and channel numbers m_n,

$$S_B=\{(t_n,m_n)\}_n. \tag{3}$$

If the signal is regularly sampled on all channels simultaneously, then $t_n=\lfloor n/M\rfloor\,\Delta t$, and $m_n=n \bmod M$, where $\lfloor ...\rfloor$ indicates the floor function (integer truncation). If the sensor channels are multiplexed through a single ADC, sampling is often done sequentially so that $t_n=n\Delta t/M$, and the beamforming delays are adjusted to compensate for the slightly skewed sample times.

We define the *alignment sequence* Z_B for the sampling set of the DS beamformer as the ordered sequence of data samples,

$$Z_B(n)=x_{m_n}(t_n+\kappa'_{m_n}\Delta t/M), \tag{4a}$$

or equivalently,

$$Z_B(n-\kappa'_{m_n})=x_{m_n}(t_n), \tag{4b}$$

where $\kappa'_m\equiv M\lfloor\tau_m/\Delta t\rfloor$ is the quantized beamformer sample delay for sensor m. The DS beamformer output [Eq. (1)] is obtained by applying a weighted block average to its alignment sequence,

$$X_B(n\Delta t)=\sum_{k=0}^{M-1} w_{m_k} Z_B(nM-k). \tag{5}$$

2. The aggregate (AGG) beamformer

The AGG beamformer differs from the conventional DS beamformer by sampling data from sensors randomly, rather than simultaneously or sequentially. The sampling set S_{AGG} of the AGG beamformer is

$$S_{\mathrm{AGG}}=\{(n\Delta t_{\mathrm{os}},\ \sigma(n))\}_n, \tag{6}$$

where $\sigma(n)$ is a random sequence of channel indices, and Δt_{os} is the sampling interval. The sampling set has samples uniformly spaced in time, just as for DS beamforming, but the sampling interval Δt_{os} is typically less than that of the DS beamformer. The ratio

$$\frac{\Delta t_{\mathrm{os}}}{\Delta t} = K_{\mathrm{os}} \tag{7}$$

is the AGG beamformer *over-sampling factor*. When $K_{\mathrm{os}} = M$, the AGG and DS beamformers acquire data at the same total rate. The random index sequence $\sigma(n)$ is independently and identically distributed over the index set $\{0 \leqslant m < M\}$ with probabilities $\{w_m\}$; that is, for m, m_1, and m_2 in the index set,

$$\Pr\{\sigma(n) = m\} = w_m \tag{8a}$$

and

$$\Pr\{\sigma(n_1) = m_1 \text{ and } \sigma(n_2) = m_2\} = \Pr\{\sigma(n_1) = m_1\} \cdot \Pr\{\sigma(n_2) = m_2\}, \tag{8b}$$

for all n_1, n_2, and n. The alignment sequence for the AGG beamformer can be expressed in a form similar to Eq. (4b) for the DS beamformer,

$$Z_{\mathrm{AGG}}(n - \kappa_{\sigma(n)}) = x_{\sigma(n)}(n \Delta t_{\mathrm{os}}), \tag{9}$$

where $\kappa_m \equiv \lfloor \tau_m / \Delta t_{\mathrm{os}} \rfloor$ are the beamforming sample delays quantized to the AGG beamformer sampling interval Δt_{os}. This formulation of the alignment sequence for the AGG beamformer is referred to as *input-driven* sequencing because the chronological index n is sequential in the input sampling. The essential difference between the alignment sequence for DS beamforming [Eq. (4b)] and that of the AGG beamformer [Eq. (9)] is that the systematic indexing scheme m_n is replaced by a random index $\sigma(n)$.

Consider the mapping $\Phi_{\mathrm{AGG}}: \{(n \Delta t_{\mathrm{os}}, \sigma(n))\} \mapsto \{Z_{\mathrm{AGG}}(n - \kappa_{\sigma(n)})\}$ from the sample set S_{AGG} to the alignment buffer Z_{AGG}, as suggested by Eq. (9). The mapping may not be *onto* (covering the entire alignment buffer) because, where voids occur, some entries in the alignment buffer are empty. Furthermore, the mapping may fail to be *one-to-one* (no two elements of the sample set are mapped to the same entry in the alignment buffer) because, where collisions occur, there are $n_1 \neq n_2$ satisfying

$$n_1 - \kappa_{\sigma(n_1)} = n_2 - \kappa_{\sigma(n_2)}. \tag{10}$$

In contrast, for conventional DS beamforming, whether channel sampling is done simultaneously or sequentially, the mapping $\Phi_{\mathrm{DS}}: \{(t_n, m_n)\} \mapsto \{Z_B(n - \kappa'_{m_n})\}$ is both *one-to-one* and *onto*.

An alternative to the input-driven formulation of the AGG beamformer [Eq. (9)] is

$$Z_{\mathrm{AGG}}(n) = x_{\sigma(n)}((n + \kappa_{\sigma(n)}) \Delta t_{\mathrm{os}}). \tag{11}$$

This formulation is called *output-driven* sequencing because the chronological index n is sequential in the output buffer Z_{AGG}. The *input-driven* and *output-driven* formulations of the AGG beamformer lead to somewhat implementations. Both can be implemented equally easily in software simulation, but input-driven sequencing is more easily implemented in hardware. This is because output-driven sequencing requires the presence of an auxiliary buffer to assemble the index sequence for data sampling in time-chronological order, whereas, for input-driven sequencing, the index sequence is already in time-chronological order [*viz.*, the right hand side of Eq. (9)]. An advantage of output-driven sequencing is that collisions and voids can be avoided completely by careful assembly of the index sequence in the auxiliary buffer and allowing a variable number of simultaneous samples to be taken. Where comparable strategies for collisions and voids are used, the two sequencing methods result in the same statistical characteristics in the aggregate beamformer output.

The AGG beamformer final output is obtained by low-pass filtering and decimating the alignment sequence by the over-sampling factor K_{os} to obtain the same output data rate as the conventional DS beamformer,

$$X_{\mathrm{AGG}}(n) = \sum_k h(k) Z_{\mathrm{AGG}}(n K_{\mathrm{os}} - k), \tag{12}$$

where h is the low-pass filter impulse response. Efficient decimation filters, such as cascaded integrator-comb filters, can be designed using the principles developed for $\Delta\Sigma$-ADC (Ref. 29; Chap. 13).

B. Basic characteristics of the aggregate (AGG) beamformer

The signal x_m arriving at each sensor can be expressed as the sum of a beam component X_B and an incoherent (off-beam) component η_m,

$$x_m(t) = X_B(t - \tau_m) + \eta_m(t), \tag{13a}$$

where [Eqs. (1), (2)]

$$\sum_{m=0}^{M-1} w_m \eta_m(t + \tau_m) = 0. \tag{13b}$$

For a signal arriving from a single source at a direction different from the beam steering-direction, the level of the beam component corresponds to the side-lobe response of the beamformer. Using Eq. (13) in Eq. (9), taking expected values $\langle \cdot \rangle$, and using the probability distribution of the random index sequence [Eq. (8)] yields (where n is neither a collision nor a void)

$$\langle Z_{\mathrm{AGG}}(n) \rangle = X_B(n \Delta t_{\mathrm{os}}) \tag{14}$$

If we ignore collisions and voids, the expected value of the AGG beamformer alignment sequence is the same as that of the DS beamformer operating at the sampling rate $1/\Delta t_{\mathrm{os}}$.

If the signals x_m at the sensors are within the pass-band of the decimation filter [h in Eq. (12)], then the filter rejects only the noise, and the expected value of the final output of the AGG beamformer will be identical to the output of the DS beamformer,

$$\langle X_{\mathrm{AGG}}(n) \rangle = X_B(n \Delta t). \tag{15}$$

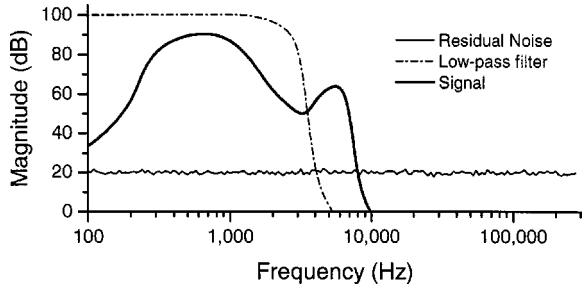


FIG. 4. A schematic representation of spectral components of the AGG beamformer output. The decimation filter eliminates out-of-band signals and reduces the residual noise power.

The residual noise power N^2_{AGG} of the aggregate beamformer is

$$N^2_{\mathrm{AGG}}=\langle |X_{\mathrm{AGG}}(n)-X_B(n\Delta t)|^2\rangle. \tag{16}$$

If we define the power N_B^2 in the off-beam component as

$$N_B^2(t)=\sum_{m=0}^{M-1} w_m \eta_m^2(t+\tau_m), \tag{17}$$

and if $\sigma(n)$ is an independent and identically distributed random variable [Eq. (8)], then N^2_{AGG} is white noise and

$$N^2_{\mathrm{AGG}}=\frac{N_B^2}{K_{\mathrm{os}}}. \tag{18}$$

The residual noise level of the AGG beamformer is determined by the level of the off-beam signal (including components in the stop-band of the decimation filter), and does not depend upon the level of the output signal. The AGG beamformer eliminates the off-beam signal by reducing it to random noise, whereas, the DS beamformer eliminates it by cancellation.

The spectral distribution of the residual noise can be changed by modifying the joint probability distribution of $\sigma(n)$. In particular, by prohibiting repeated values [so that $\Pr\{\sigma(n)=m \text{ given } \sigma(n-1)=m\}=0$], the spectrum of N_B^2 can be made to increase with frequency, and the residual noise after decimation will be less than that given in Eq. (18). Simulations using noise shaping show improvements beyond Eq. (18) of about 10 dB.

Figure 4 illustrates (schematically) the signal and residual noise components in the AGG beamformer. The noise in the unfiltered alignment sequence spans the entire bandwidth, which may extend beyond the range of the horizontal axis shown. The noise is substantially eliminated by the low-pass decimation-filter, leaving only the residual noise within the desired frequency band. If the signal spectrum extends beyond the Nyquist frequency, as does the second hump of the signal spectrum in the figure, then the decimation filter rejects this part of the spectrum and prevents aliasing.

The directional response of the AGG and the DS beamformer for simulated data are compared in Fig. 5(a). They are essentially identical, except that the valleys in the AGG response are filled by the residual noise. The data for this figure were obtained by simulating a pure tone signal incident at various angles upon a linear array of 8 elements, regularly spaced at half a wavelength, steered to a direction of 40°.

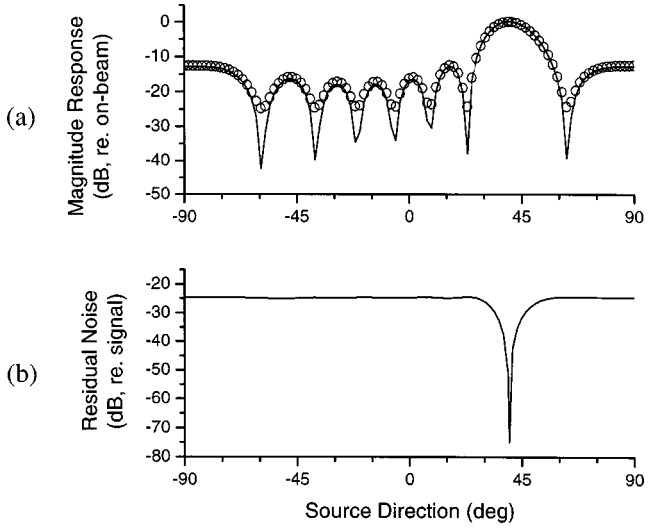


FIG. 5. Simulated directional responses: (a) DS (—) and AGG (○) beamformer outputs for an 8-element linear array and a pure tone source at the array design frequency; (b) AGG beamformer residual noise with an over-sampling factor of 256.

Uniform weighting ($w_m \equiv 1$) and an over-sampling factor of $K_{\mathrm{os}}=256$ were used. The residual noise power, relative to the response to an on-beam signal, is shown in Fig. 5(b) as a function of the signal incidence direction. The noise is constant except near the beam steering direction, where it declines dramatically. The off-beam residual noise level is approximately 24 decibels below the incident signal level, which is consistent with the level predicted by Eq. (18), and is below the side-lobe peak levels. There is very little residual noise for the on-beam signal.

In theory, the AGG beamformer may fail to detect a short pulse traversing the array. In practice, due to over-sampling, the probability of such an event is vanishingly small since a pulse that is band-limited by the Nyquist frequency f_{N} will span many AGG beamformer samples. In Fig. 6, a typical time-domain response of the AGG beamformer to a band-limited pulse is compared with the response of the DS beamformer to the same pulse. The output sample rate for this example is 11025 Hz, the pulse is band limited to 500–3400 Hz, and the over-sampling factor is 64 (705.6 kHz). The pulse arrives from the beam steering direction of 90° (end-fire). The AGG beamformer handles collisions by a fair-coin toss and handles voids by repetition of the previous

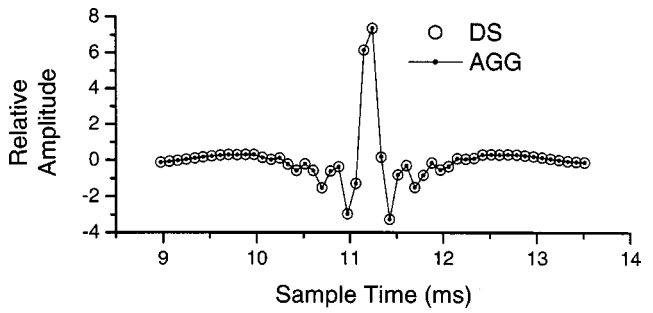


FIG. 6. The response of the DS and AGG beamformers to an end-fire, band-limited (500–3400 Hz) pulse on an 8-element linear array with 5 cm spacing. The AGG beamformer over-sampling factor is 64.

data value. As this example illustrates, the AGG and DS beamformers may equally well handle impulsive signals.

C. Managing collisions

Collisions require special management strategies, such as using the first or the last data values. Simple strategies that require little or no data computation are preferred. The choice of collision management strategy will affect the statistics of the AGG beamformer output. To illustrate this we compare how two strategies affect array shading. The first strategy (a) keeps the latest data sample, so that each new colliding data sample overwrites (supercedes) the older value. The second strategy (b) uses a fair-coin toss at each collision to choose whether to replace the old data value with the new one or to keep the old data value. Each strategy is represented by a modified sequencing index, σ_a and σ_b, respectively, based on input-driven sequencing [Eq. (9)]. Collisions can be resolved by assigning the modified index $\sigma_a(n)$ or $\sigma_b(n)$ a marker value, such as -1, for each data sample n that is overwritten, to indicate that no input channel was used for that sample and to remove it from the sample set.

We define the normalized weights $\overline{w}_m^x$ resulting after applying the collision management strategies as

$$\overline{w}_m^x = \frac{w_m^x}{\sum_{m=0}^{M-1} w_m^x}, \tag{19a}$$

where

$$w_m^x = \Pr\{\sigma_x(n) = m\}, \tag{19b}$$

and x indicates either strategy (a) or (b). [The denominator of Eq. (19a) may be less than unity due to the marking of collisions.] Using Eq. (19) in Eq. (14), it follows that the expected output of the AGG beamformer with this collision management strategy will correspond (except where voids occur) to the output of a (over-sampled) DS beamformer with weights $\{\overline{w}_m^x\}$. For convenience, and without loss of generality, we assume that the channels are numbered so that the beamforming delays κ_m are indexed in ascending order,

$$\kappa_{m_1} \leqslant \kappa_{m_2} \quad \text{for} \quad 0 \leqslant m_1 \leqslant m_2 < M. \tag{20}$$

Now consider strategy (a), where newer data supercedes older data. We define the set of channel indices $S^+(m) = \{m' | \kappa_{m'} > \kappa_m\}$ so that, for each m, $S^+(m)$ is the set of channels where the data can supercede data from channel m. For example, $S^+(M-1)$ is empty since channel $M-1$ has the largest delay, and data from this channel is never superceded. On the other hand, $S^+(0)$ includes all channel numbers greater than zero since data from channel 0 can be superceded by all other channels. From this it can be shown by induction that the (un-normalized) modified weights are obtained from the recursion relation

$$w_m^a = w_m \left(1 - \sum_{m' \in S^+(m)} w_{m'}^a \right), \tag{21}$$

with $w_{M-1}^a = w_{M-1}$.

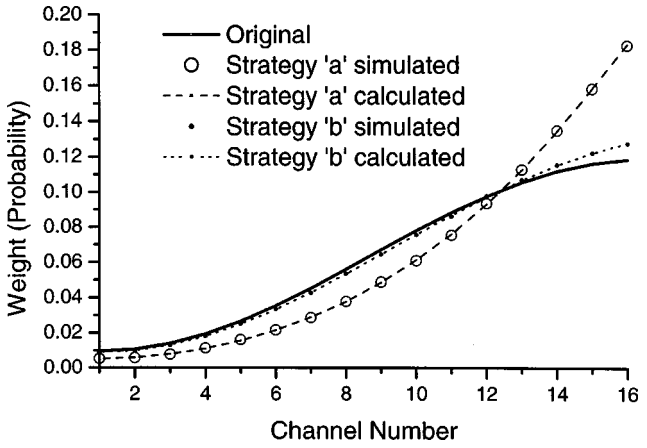


FIG. 7. The effect of collision management on array shading. The original array shading function (solid curve) is modified when collisions are managed by retaining older samples (strategy "a") or a coin toss (strategy "b"). The calculated results are obtained from Eqs. (21) and (22).

Figure 7 shows a comparison of the original weights, the (normalized) modified weights computed according to Eq. (21) and Eq. (19a), and the weights estimated by numeric simulation using the collision management strategy (a), for a uniformly spaced, linear, 16-element array. The original 16 weights were obtained from the first half of the sequence of Hamming weights for a 32-element array. This is an unusual choice for a beamforming array, but it is a convenient one for demonstrating the effects of collision management on array shading. The modified weights deviate considerably from the original weights for this collision management strategy. The original probability distribution of the random channel index sequence, $\sigma(n)$, could be pre-compensated so that the weights resulting from collision management [Eq. (21)] are the desired values.

Now consider strategy (b), where a fair-coin toss is used at each collision to determine whether to replace the older data value with the current data value. For this strategy, we define another set of channel indices $S^-(m) = \{m' | \kappa_{m'} < \kappa_m\}$. The un-normalized modified weights are

$$w_m^b = \frac{1}{2} w_m \left(1 + \prod_{m' \in S^-(m)} (1 - w_{m'}) \right) \times \prod_{m' \in S^+(m)} \left(1 - \frac{1}{2} w_{m'} \right). \tag{22}$$

Despite the rather complicated appearance of Eq. (22), the (normalized) modified weights for the fair-coin toss case closely resemble the original weights (Fig. 7). This example shows that strategy (b) may be preferred to strategy (a) when pre-compensation of the array weights is not convenient.

D. Managing voids

The probability of a void in the aggregate sequence is $\Pi(1-w_m)$, which has an approximate value of 1/3 for a considerable range of (normalized) array shading weight values. Notwithstanding their frequency of occurrence, voids have a relatively minor impact on the final AGG beamformer output. This is because the voids are randomly distributed in

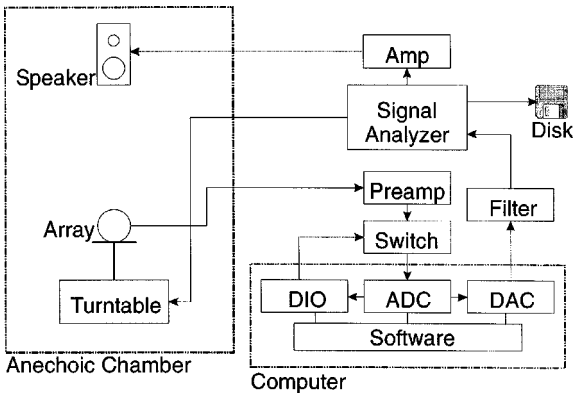


FIG. 8. Experimental setup for array response measurements, providing a common test bed with real-time processing for both the DS and AGG beamformers.

the alignment buffer, and many methods of filling the voids result in white noise that is substantially reduced by the low-pass decimation filtering.

As for collisions, it is desirable to manage voids without requiring numeric operations on data. One method for handling voids is to zero-fill the void entries. This generates white noise with a noise power equal to about one-third the total signal power. For an ideal decimation filter, the noise due to zero filled voids is approximately $(X_B^2/3+N_B^2)/K_{\mathrm{os}}$. This method of filling voids has the disadvantage that noise is generated even when the signal arrives from the beam steering direction.

An alternative method for handling voids is to replicate the previous value in the alignment sequence. For a tone of frequency f and unit amplitude arriving from the beam steering direction, this method introduces an error noise power of $\pi f/\sqrt{2}f_{\mathrm{N}}K_{\mathrm{os}}^2$, which is always less than the error noise generated by zero-filling. This noise power decreases by 6 dB per doubling of the over-sampling factor K_{os}, whereas the noise from zero-filling decreases at a rate of 3 dB per doubling. A disadvantage of this method is that it affects the conditional probability distribution of the index sequence σ and can disrupt noise shaping [which was discussed shortly following Eq. (18)].

IV. DEMONSTRATION

In this section we present experimental data comparing characteristics of the AGG and DS beamformers. Both beamformers were implemented in real-time using the same hardware and software configurations. For the AGG beamformer, collisions were handled by the fair-coin toss method (Sec. III C), voids were handled by repetition of previous values, and the decimation filter [Eq. (12)] was implemented with an analog low-pass filter.

The equipment setup for the demonstration is illustrated in Fig. 8. Various arrays were used, including a 16-element circular array with a diameter of 42 cm, a 32-element spherical array (described further in Ref. 2) with a diameter of 16 cm, and a 16-element linear array with 8 cm element spacing. The demonstrations were prepared with equipment at hand, and the configuration was not intended to illustrate an optimal or recommended approach.

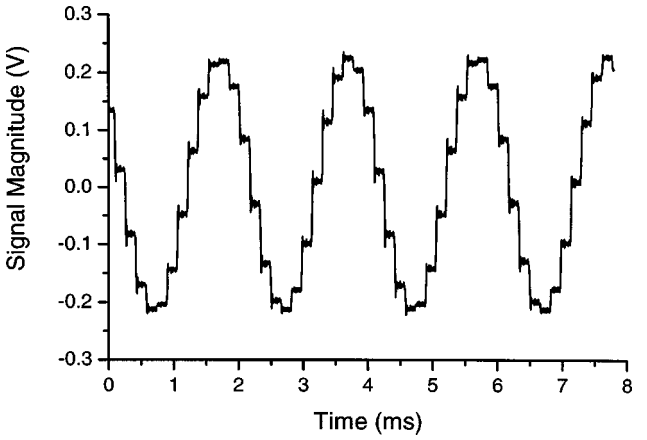


FIG. 9. A typical raw time-series output of the DS beamformer, showing characteristic quantization steps at the sampling rate.

A. Output time series

In Figs. 9 and 10 the DS and AGG beamformer output time series (before decimation) are exemplified using data from the circular array. Decimation filtering was omitted, and the source was located in a direction different from the beam steering direction to make the presence of the AGG beamformer residual noise evident. The total sampling rate was 100 kHz (6.25 kHz per channel) and the source was a 500 Hz tone. The DS beamformer output time series (Fig. 9) is a stepped sinusoid with magnitude determined by the side-lobe level for this particular source direction. The quantization steps could be reduced by processing with a reconstruction filter. The AGG beamformer output time series (Fig. 10) is a noisy sinusoid of the same magnitude as that of the DS beamformer. It does not exhibit any steps.

B. Output spectra

Figure 11 shows the AGG and DS beamformer output power spectra from the 32-element spherical array with a source signal consisting of tones at 2.5 kHz (A) and 5.0 kHz (B). The total sampling rate was 200 kHz (6.25 kHz per channel). In the AGG beamformer power spectrum, both tones are correctly positioned in the AGG beamformer out-

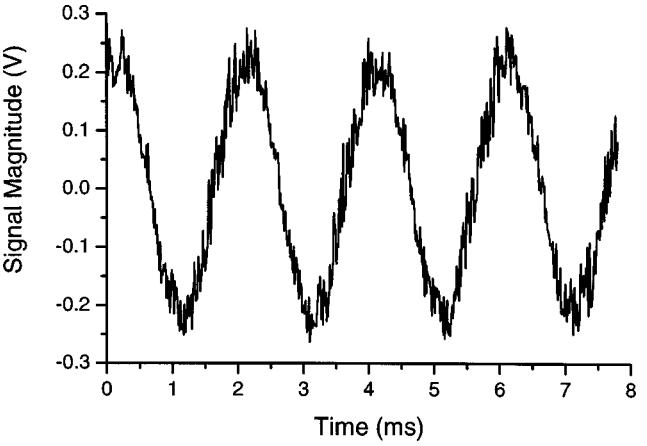


FIG. 10. A typical raw time-series output of the AGG beamformer, showing residual noise (before decimation filtering) added to the sinusoidal signal. There are no quantization steps.

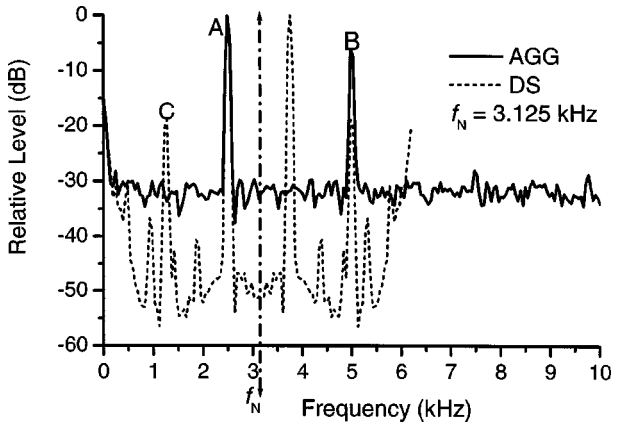


FIG. 11. The Raw output spectra of AGG and DS beamformers with a 32-element spherical array sampled at 200 kHz, and a two-tone source signal. No anti-alias filters were used. Tone A (2.5 kHz) is below the DS beamformer Nyquist frequency $f_N = 3.125$ kHz. Tone B is above f_N and is aliased by the DS beamformer at peak C. The DS beamformer spectrum repeats at 2 f_N (not shown). The AGG beamformer spectrum repeats at 200 kHz. Minor peaks, visible in the DS beamformer spectrum, are due to system noise and background room noise.

put spectrum since they lie below its Nyquist frequency, which, prior to decimation, is 100 kHz. (Only a portion of the spectrum is displayed in the figure.)

The AGG beamformer has a higher spectral noise floor than the DS beamformer. Increasing the over-sampling factor will lower the AGG beamformer spectral noise floor, and decimation filtering will reduce its total noise power. This example uses an over-sampling rate of 32.

The 5.0 kHz component (B) is aliased to a lower frequency (C) in the DS beamformer output since anti-aliasing filters were not present. The DS beamformer spectrum is mirrored symmetrically about the Nyquist frequency $f_N = 3.125$ kHz. (Periodic repetition of the digital signal spectrum at the channel sampling frequency $f_s = 6.25$ kHz is not shown in the figure.) Signals above the Nyquist frequency are normally removed by a signal reconstruction filter for analog output. The minor peaks in the DS beamformer output are due to system noise and background noise in the reverberant room where the measurements were done.

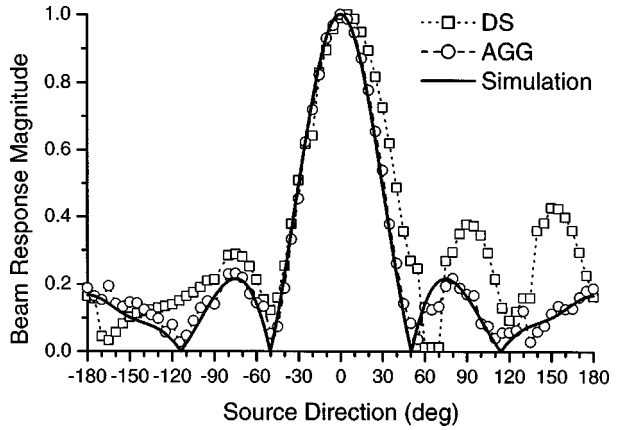


FIG. 12. The experimental directional response of a 16 cm diameter, 32-element spherical array to a 2.5 kHz source signal.

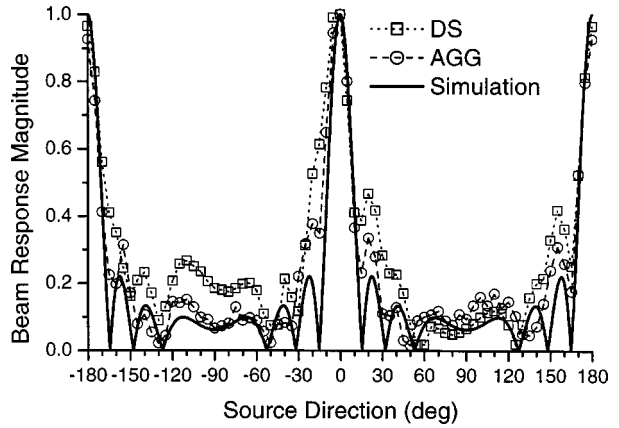


FIG. 13. The experimental directional response of a uniformly spaced 16-element linear array, with uniform array shading, to a 1 kHz source signal.

C. Directional response

Figure 12 shows the directional response of the 32-element spherical array for the DS beamformer, AGG beamformer, and a simulation of the DS beamformer. Directional responses were measured in an anechoic chamber with the source loudspeaker approximately 3.5 m from the center of the array. Each array was mounted on a turntable and rotated in 10-degree increments. Uniform array weights are used in each case. The DS beamformer data deviate from the simulation, and lack the expected symmetries, due primarily to beamformer delay quantization. This effect could be reduced with data interpolation or a higher sampling rate. The AGG beamformer data follow the simulation data more closely.

Figures 13 and 14 shows how the application of Hamming weights can reduce side-lobes of the beam response pattern. Figure 13 shows the response pattern for the 16-element linear array with uniform weighting. Figure 14 shows the corresponding response patterns with Hamming weights applied. The weights were applied to the sensor data of the simulation and DS beamformer [Eq. (1)], and to the probability distribution of the random channel-index sequence [Eq. (8)] of the AGG beamformer. While neither measured beamformer output achieves the low side-lobe levels of the Hamming weighted simulation, the side-lobes are similarly reduced in both beamformers. Gain compensation

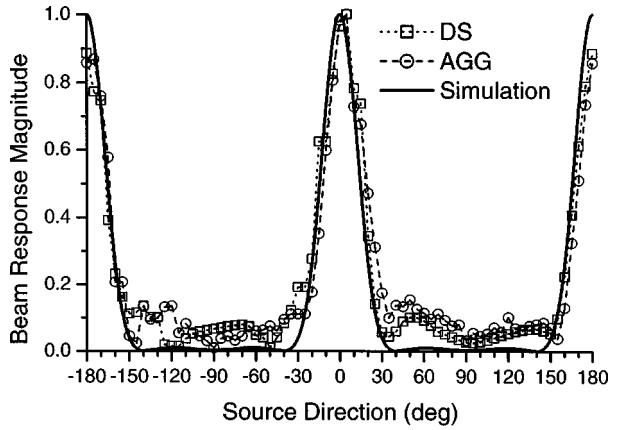


FIG. 14. The experimental directional response of a uniformly spaced 16-element linear array, with a Hamming window, to a 1 kHz source signal.

could be used to adjust for variations in microphone sensitivities and would improve the side-lobe reduction of both beamformers.

V. DISCUSSION

The AGG beamformer constructs an output signal from the total (aggregate) collection of data samples from all sensor channels without the need for an explicit digital representation of the signal on each individual sensor channel. This has the advantage that the Nyquist criterion for the AGG beamformer is determined by the total sampling rate across all channels. In contrast, the Nyquist criterion for conventional beamforming is determined by the (lower) sampling rate of the individual channels. The higher Nyquist frequency of the AGG beamformer often exceeds the upper frequency of either the source signal spectrum or the frequency response of the sensors. This gives the AGG beamformer the advantage of not requiring anti-alias filters. The AGG beamformer also has the advantage that it can operate at a sampling rate less than that of the conventional beamformer if the number of elements in the array exceeds the over-sampling factor required to obtain the desired residual noise level reduction.

The AGG beamformer samples data regularly in time but the channel from which each sample is taken is selected randomly. Consequently, the ABF exhibits some of the properties observed in randomly sampled sequences. The problem of reconstructing signals from irregular or random sampling has been studied extensively.[26] Various extensions to the Shannon sampling theorem for randomly sampled sequences,[27] and procedures to reconstruct the original signals[28] have been developed elsewhere. The reconstruction of the individual channel signals is not essential to the AGG beamformer unless related signal processing requires the array signal covariance matrix. In fact, the AGG beamformer can operate properly even though the sampling rate on each channel may not be sufficient for accurate signal reconstruction of individual channel signals, provided that the *total* sampling rate allows an accurate signal representation in the alignment buffer.

Much of the theory on signal reconstruction for randomly sampled sequences can be put aside since the alignment buffer data is presumed to be regularly sampled. Instead, the output data of the AGG beamformer can be treated as a regular, over-sampled digital representation of the desired signal with white noise included. For example, in the design of $\Delta\Sigma$-ADC, a signal can be recovered by low-pass filtering and decimating if the over-sampling rate is sufficiently high.[14] Efficient decimation filtering techniques have been developed for this purpose (Ref. 29; Chap. 13). Although techniques for noise shaping commonly used to improve the performance of $\Delta\Sigma$-ADC are not directly applicable to the aggregate beamformer, it is possible to achieve a similar affect by modifying the joint probability distribution function of the random channel selection sequence.[25]

The AGG beamformer is most advantageous relative to the DS beamformer when the number of array elements is large. The threshold above which the AGG beamformer is preferred depends upon the desired residual noise reduction, the need for component integration, the cost of anti-alias filtering, the importance of minimizing numeric operations on data, and the number of simultaneous beams that need to be formed. Each application requires consideration of these costs and requirements but, as a rule of thumb, when the number of sensors is greater than the required over-sampling factor, the AGG beamformer will be advantageous because it has both a lower sampling rate and a simpler structure than the DS beamformer.

The residual noise of the AGG beamformer is perceived by a human listener as a distortion on sounds (e.g., speech) arriving from directions other than the beam steering direction. This is because the noise envelope is the same as the signal envelope of these sounds. Signals arriving from the beam steering direction will cause little residual noise. For example, if a microphone array is steered to one of several people speaking, that person's speech will be received clearly while the speech from the other people, away from the beam steering direction, will be attenuated to the beamformer side-lobe level, and distorted by the residual noise. The distortion may be advantageous where it reduces the intelligibility of interfering speech relative to desired speech.

Since the residual noise is apparent only as a distortion to sound sources not in the beam steering direction, it may be hidden or obscured by reducing it below the level of the side-lobes. For example, the side-lobes of the 32 element spherical array used here (Fig. 12) are about 15 dB below the main lobe. An over-sampling rate of at least 64 would reduce the residual noise level to at least 18 decibels below the off-beam signal level. This would be below the side-lobe levels, and thus the residual noise would be hidden as a small distortion on the interfering signals.

VI. CONCLUSIONS

The aggregate (AGG) beamformer has the same directional response as the delay and sum (DS) beamformer, and has several advantages relative to the DS beamformer:

(i) The cost, complexity, and total sampling rate are independent of the number of array sensors;
(ii) hardware costs are reduced by eliminating anti-alias filters;
(iii) computational costs are reduced by eliminating numerical operations on data;
(iv) the time delay resolution is improved, giving greater beam steering precision without the need for data interpolation;
(v) reduced front-end complexity facilitates the integration of sensors and processors.

These advantages are achieved at the ''cost'' of a residual noise that arises from off-beam signals. On-beam signals do not generate residual noise. Residual noise can be reduced by increasing the over-sampling factor of the AGG beamformer, defined as the ratio of half the total sampling rate to the desired signal bandwidth. With adequate over-sampling, the residual noise may be hidden within the side-lobe response of the beamformer. As a rough rule, when the number of sensors is greater than the over-sampling factor,

the AGG beamformer will have both a lower sampling rate and a simpler structure than the DS beamformer. Since the AGG beamformer is most advantageous relative to the DS beamformer for arrays with many elements, it will find an application in 2-D arrays such as for ultrasonic imaging or sound pickup in large rooms.

ACKNOWLEDGMENTS

The 32-element spherical array was developed and tested by Bradford N. Gover. The author is grateful to the Associate Editor for many helpful suggestions to improve the manuscript.

[1] T. B. Hughes, H.-S. Kim, J. H. DiBiase, and H. F. Silverman, ''Performance of an HMM speech recognizer using a real-time tracking microphone array as input,'' IEEE Trans. Speech Audio Process. **7**, 346–349 (1999).
[2] B. N. Gover, J. G. Ryan, and M. R. Stinson, ''Microphone array measurement system for analysis of directional and spatial variations of sound fields,'' J. Acoust. Soc. Am. **112**, 1980–1991 (2002).
[3] H. F. Silverman, W. R. Patterson III, and J. Sachar, ''Factors affecting the performance of large-aperture microphone arrays,'' J. Acoust. Soc. Am. **111**, 2140–2157 (2002).
[4] J. L. Flanagan, J. D. Johnston, R. Zhan, and G. W. Elko, ''Computer-steered microphone arrays for sound transduction in large rooms,'' J. Acoust. Soc. Am. **78**, 1508–1518 (1985).
[5] B. J. Sotirin and J. A. Hildebrand, ''Acoustic navigation of a large-aperture array,'' J. Acoust. Soc. Am. **87**, 154–167 (1990); or B. J. Sotirin and J. A. Hildebrand, ''Large aperture digital acoustic array,'' IEEE J. Ocean. Eng. **OE-13**, 271–281 (1988).
[6] D. I. Havelock, Xiao Di, G. A. Daigle, and M. R. Stinson, ''Spatial coherence of a sound field in a refractive shadow: Comparison of simulation and experiment,'' J. Acoust. Soc. Am. **98**, 2289–2302 (1995).
[7] D. K. Wilson, C. R. Tate, D. C. Swanson, and K. M. Reichard, ''Acoustic scintillations and angle-of-arrival fluctuations observed outdoors with a large planar vertical microphone array,'' J. Acoust. Soc. Am. **106**, L24–L29 (1999).
[8] A. Fenster and D. B. Downey, ''Basic principles and applications of 3-D ultrasound imaging,'' in *Advanced Signal Processing Handbook*, Electrical Engineering and Signal Processing Series, edited by S. Stergioloulos (CRC Press LLC, 2001), Chap. 14.
[9] H. F. Silverman, W. R. Patterson III, and J. L. Flanagan, ''The huge microphone array,'' IEEE Concurrency, October–December 1998, 1998, pp. 36–46.
[10] G. S. Garrett, J. Naze Tjøtta, and S. Tjøtta, ''Nearfield of a large acoustic transducer, Part I. Linear radiation,'' J. Acoust. Soc. Am. **72**, 1056–1061 (1982).
[11] J. Hald and J. J. Christensen, ''A novel beamformer array design for noise source location from intermediate measurement distances,'' J. Acoust. Soc. Am. **112**, 2448 (2002).
[12] J. A. Fawcett, ''Synthetic aperture processing for a towed array and a moving source,'' J. Acoust. Soc. Am. **94**, 2832–2837 (1993).
[13] M. Inerfield, G. R. Lockwood, and S. L. Garverick, ''A sigma-delta-based sparse synthetic aperture beamformer for real-time 3-D Ultrasound,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 243–254 (2002).
[14] J. C. Candy, ''An overview of basic concepts,'' in *Delta–Sigma Data Converters: Theory, Design, and Simulation*, edited by S. R. Norsworthy, R. Schreier, and G. C. Temes (IEEE Press, Piscataway, NJ, 1997), Chap. 1; also available with more detail in J. C. Candy and G. C. Temes, ''Oversampling methods for A/D and D/A conversion,'' in *Oversampling Delta-Sigma Data Converters* (IEEE Press, New York, 1992), pp. 1–275.
[15] R. G. Pridham and R. A. Mucci, ''A novel approach to digital beamforming,'' J. Acoust. Soc. Am. **63**, 425–434 (1978).
[16] R. A. Mucci, ''A comparison of efficient beamforming algorithms,'' IEEE Transactions on Acoustics Speech and Signal Processing **32**, 548–558 (1984).
[17] P. Rudnick, ''Digital beamforming in the frequency domain,'' J. Acoust. Soc. Am. **46**, 1089–1090 (1969).
[18] B. Widrow and S. E. Stearns, *Adaptive Signal Processing* (Prentice–Hall, Englewood Cliffs, NJ, 1985), Chap. 13.
[19] W. E. Stinger, Jr., ''Direct digital loudspeaker,'' US Patent 4515997, 1985.
[20] Y. Huang, S. C. Busbridge, and D. S. Gill, ''Distortion and directivity in a digital transducer array loudspeaker,'' J. Audio Eng. Soc. **49**, 337–352 (2001).
[21] M. W. Hoffman, K. M. Buckley, M. J. Link, and S. Soli, ''Robust microphone array processor incorporating headshadow effects,'' *IEEE Proc. ICASSP-91*, 1991, Vol. 5, pp. 3637–3640.
[22] M. M. Goodwin and G. W. Elko, ''Constant beamwidth beamforming,'' *IEEE Proc. ICASSP-93*, 1993, Vol. 1, pp. 169–172.
[23] G. W. Elko and M. M. Goodwin, ''Beam dithering: acoustic feedback control using a modulated-directivity loudspeaker array,'' in Ref. 22, pp. 173–176.
[24] D. I. Havelock, ''The aggregate beamformer,'' Proceedings of Acoustics Week in Canada 2002, in Can. Acoust. **30**, 104–105 (2002).
[25] D. I. Havelock, ''Residual noise in the aggregate beamformer,'' *Proceedings of the 2003 IEEE Workshop on Applications of Signal Processing in Audio and Acoustics*, 19–22 October 2003, New Paltz, NY, 2003.
[26] *Nonuniform Sampling: Theory and Practice*, edited by F. Marvasti (Kluwer Academic/Plenum, New York, 2001).
[27] A. J. Jerry, ''The Shannon sampling theorem—its various extensions and applications: a tutorial review,'' *IEEE Proceedings*, November 1977, pp. 1565–1596.
[28] F. Marvasti and T. J. Lee, ''Analysis and recovery of sample-and-hold and linearly interpolated signals with irregular samples,'' IEEE Trans. Signal Process. **40**, 1884–1891 (1992).
[29] S. R. Norsworthy and R. E. Crochiere, ''Decimation and interpolation for delta–sigma conversion,'' in *Delta–Sigma Data Converters, Theory: Design, and Simulation*, edited by S. R. Norsworthy, R. Schreier, and G. C. Temes (IEEE Press, Piscataway, NJ, 1997), Chap. 13.

A phenomenological model for the responses of auditory-nerve fibers. II. Nonlinear tuning with a frequency glide

Qing Tan
Boston University Hearing Research Center, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215

Laurel H. Carney[a)]
Boston University Hearing Research Center, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215 and Department of Bioengineering and Neuroscience, Institute for Sensory Research, 621 Skytop Road, Syracuse University, Syracuse, New York 13244[b)]

(Received 31 December 2002; accepted for publication 22 July 2003)

A computational model was developed to simulate the responses of auditory-nerve (AN) fibers in cat. The model's signal path consisted of a time-varying bandpass filter; the bandwidth and gain of the signal path were controlled by a nonlinear feed-forward control path. This model produced realistic response features to several stimuli, including pure tones, two-tone combinations, wideband noise, and clicks. Instantaneous frequency glides in the reverse-correlation (revcor) function of the model's response to broadband noise were achieved by carefully restricting the locations of the poles and zeros of the bandpass filter. The pole locations were continuously varied as a function of time by the control signal to change the gain and bandwidth of the signal path, but the instantaneous frequency profile in the revcor function was independent of sound pressure level, consistent with physiological data. In addition, this model has other important properties, such as nonlinear compression, two-tone suppression, and reasonable Q_{10} values for tuning curves. The incorporation of both the level-independent frequency glide and the level-dependent compressive nonlinearity into a phenomenological model for the AN was the primary focus of this work. The ability of this model to process arbitrary sound inputs makes it a useful tool for studying peripheral auditory processing.
 [DOI: 10.1121/1.1608963]

PACS numbers: 43.64.Bt, 43.64.Pg [WJS]

I. INTRODUCTION

The auditory nerve (AN) transfers the information of sound stimuli from the cochlea to the cochlear nucleus, which projects to higher levels of the auditory nervous system. Detailed knowledge of the firing pattern of AN fibers is necessary to understand how sounds are encoded at the input stage of the auditory system. The goal of this study was to improve a previous nonlinear phenomenological model for the response patterns of AN fibers to different sound inputs. The computational AN model presented here includes the level-independent instantaneous frequency glide and the level-dependent compressive nonlinearity. These properties interact and influence both the rate and timing of AN responses. This model is a useful tool for the study of sound encoding in the peripheral auditory system, and it provides realistic responses that can be used as inputs to models of higher levels of the auditory system.

A system's impulse response can be estimated by the cross correlation of the response of the system to a wideband noise with the noise input waveform. This technique has been used as an indirect estimate of the basilar membrane (BM) impulse responses while the click response is a direct estimate (de Boer and Nuttall, 1997). The reverse-correlation (revcor) function is an extension of the cross-correlation method and is used as an indirect estimate of the AN impulse response (de Boer and de Jongh, 1978). A frequency modulation, or ''glide'' in the instantaneous frequency, has been reported in the impulse responses of BM (Robles *et al.*, 1976; de Boer and Nuttall, 1997; Recio *et al.*, 1997) and AN fibers (Lin and Guinan, 2000; Carney *et al.*, 1999). An upward frequency glide indicates that the early part of the impulse response is dominated by lower frequency components and the later part is dominated by higher frequency components (i.e., frequency increases as a function of time). A downward glide indicates the opposite trend (i.e., frequency decreases as a function of time).

Upward frequency glides were observed in BM and AN responses with relatively high characteristic frequencies (CF $>$1500 Hz), constant frequency glides were seen in AN fibers with medium CFs (CF=750–1500 Hz), and downward frequency glides were seen in low-CF AN fibers (CF$<$750 Hz) (Carney *et al.*, 1999). These frequency glides are consistent with the level-dependent peak-frequency shifts observed in auditory peripheral transfer functions [AN: Møller (1977); inner hair cell (IHC): Cheatham and Dallos (1999)]. Shera (2001a, b) explored instantaneous frequency glides in BM click responses and suggested that the slope of the normalized instantaneous frequency is independent of cochlear location for CFs above 1.5 kHz and strongly dependent on cochlear location for lower CFs.

The frequency glide pattern not only affects the fine

a)Author to whom correspondence should be addressed. Electronic mail: lacarney@syr.edu
b)Address for correspondence.

structure of AN responses in the time domain but also is related to the best-frequency[1] (BF) shift as a function of sound pressure level (SPL). This level-dependent BF shift can be qualitatively explained by the combination of the instantaneous frequency trend in the impulse response and the change in the shape of the impulse-response envelope at various input SPLs (Carney, 1999; Shera, 2001b). Due to the compressive nonlinearity, as SPL increases, the bandwidth of peripheral filters increases and the group delay of the filters decreases. Associated with the decreases in group delay is a shift of the envelope of the impulse response to smaller latencies. In fibers with CFs higher than 1500 Hz, the instantaneous frequency has an upward glide, which means that the beginning of the response has a relatively low instantaneous frequency. When the input SPL is increased, the group delay of the impulse response decreases; because earlier (lower-frequency) energy in the impulse response is emphasized as SPL increases, the best frequency at high SPLs decreases. For low-CF fibers, as SPL increases the group delay of the impulse response still decreases, but in this case the shift in the impulse-response envelope emphasizes the early (high-frequency) energy in the downward glide for the low-CF fiber. Therefore, the BF of low-CF fibers shifts upwards as the SPL increases. [See Fig. 1 in Carney (1999) for schematic diagram illustrating the inter relationship between the glide and the BF shift.]

An instantaneous-frequency glide in the impulse response of a filter is reflected in the asymmetry of its transfer function. This asymmetry indicates that there are poles with different damping coefficients within the filter. The difference in damping coefficients is associated with the frequency glide in the filter's impulse response (see below for more detail). Because the middle ear affects the asymmetry of cochlear filters (Cheatham and Dallos, 2001), a simple middle-ear function consisting of a linear band-pass filter (Rosowski, 1996) was used to model this aspect of the frequency glide. The contribution of the middle-ear filter to the frequency glide is most important at low CFs.

One goal of this study was to simulate the frequency-glide phenomenon in the AN fiber's impulse response. Another focus was the inclusion of the compressive nonlinearity, which is the decrease in the gain of the BM response for mid- to high-level sound inputs (Rhode, 1971; Ruggero *et al.*, 1997). The compressive nonlinearity causes broadened tuning of AN responses and shifted phase responses with increased SPLs. Two-tone suppression, which is the reduction of the response to a tone at CF when a second tone is presented at a frequency other than CF, is associated with the compressive nonlinearity (Ruggero and Rich, 1991).

Recent reports of phenomenological AN models have focused on various aspects of fiber responses. The responses of the auditory periphery, whether recorded from single AN fibers or single sites on the basilar membrane, are characterized as level-dependent bandpass filters. A nonlinear AN model that was developed by Carney (1993) and extended by Zhang *et al.* (2001) and Heinz *et al.* (2001c) includes a fourth-order gamma-tone filter with level-dependent bandwidth and gain. The most recent versions of this model included level-dependent phase responses, compression, and two-tone suppression; however, the frequency glide observed by Carney *et al.* (1999) in the reverse-correlation (revcor) functions of cat AN fibers was not included.

Meddis and colleagues' (2001) BM model consists of a dual resonance nonlinear (DRNL) filter with two parallel branches, one linear and the other nonlinear. This model and Goldstein's (1990, 1995) multiple bandpass nonlinear (MBNL) model are extensions of Pfeiffer's (1970) bandpass nonlinear (BPNL) model. These models successfully reproduce many physiological phenomena related to basilar membrane motion. However, they do not address the level independence of the instantaneous-frequency glide, which is a focus of the study presented here. Meddis *et al.* (2001) qualitatively describe the instantaneous-frequency glide in the impulse response of their model, but they do not quantify the frequency glide or demonstrate its level independence. The level dependency of this instantaneous frequency in the impulse response of the DRNL model will be compared to that of the model presented here.

Irino and Patterson (1997, 2001) proposed a gammachirp auditory filter to account for peripheral auditory processing. The gammachirp filter is an extension of the gammatone filter and was the first model to include the frequency glide property in its impulse response. Although this model includes frequency glides in the impulse responses, the trends of the BF shifts as a function of SPL in Irino and Patterson's (2001) model are not consistent with physiological data. The response areas of Anderson *et al.* (1971, their Fig. 8) and the revcor filters of Møller (1977), Evans (1977), and de Boer and de Jongh (1978) had monotonic shifts in BF to lower frequencies with increased SPLs for fibers with high CF. However, the BF of one example in Irino and Patterson (2001, their Fig. 7) shifts to higher frequencies for mid-level sounds and back to lower frequencies for high-level sounds. The BF of another example (their Fig. 10, CF=1800 Hz) shifts to higher frequencies with increased SPLs. An accurate representation of the level-dependent shifts in BF was one of the focuses of the model presented here, as it is associated with the frequency glide.

Robert and Eriksson's (1999) cochlear model is based on a filter bank of all-pole gamma-tone filters (APGFs). Each branch of the filter bank consists of a passive and an active (nonlinear) bandpass filter in series. These two filters are tuned to different center frequencies, and, therefore, a BF shift can be observed when input SPL changes. This level dependency of BF suggests that their filter-bank model may have an instantaneous frequency glide in the impulse response. However, Robert and Eriksson (1999) did not show the instantaneous frequency profile or test its level independence. In addition, Robert and Eriksson's (1999) gammatone filters do not include any zeros, which are important for producing downward instantaneous-frequency glides (see Sec. II for detail). These downward glides are observed in low-CF AN fibers (Carney *et al.*, 1999).

de Boer and Nuttall (2000) also investigated the instantaneous-frequency glide in the response of the basal BM in a modeling study based on the EQ-NL theorem (de Boer, 1997). In their model, the impedance of the BM is described by two linear components, with a signal level-

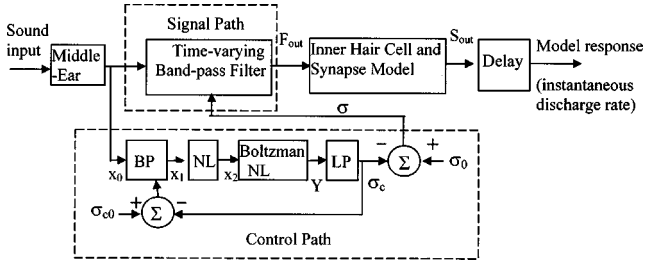


FIG. 1. Schematic diagram of the AN model. The model included a middle-ear model, a signal path, a control path, the IHC and synapse model, and a time delay. See text for more detail.

dependent coefficient that introduces the level dependence of bandwidth, gain, and peak frequency of the BM response in the frequency domain, along with the level-invariance of the glide in the impulse response. They only modeled responses in the base, and thus their model only described an upward frequency glide, but the general approach could be adapted to describe the downward glide at low frequencies. Their level-dependent coefficient varied the relative contributions of the two impedances; this coefficient did not vary with time, but was set to a single value to explain responses at a fixed stimulus noise level. The level-dependent parameter in the de Boer and Nuttall (2000) study plays a similar role in their model as the output of the control path in the model presented here. However, in the model described here, the parameter that controls the tuning at the level of the BM is both time and level dependent. The temporal variation of the nonlinearity is presumably more important at low frequencies because the dynamics of the compressive nonlinearity are fast enough that the properties of the system vary within a stimulus period at low frequencies. In addition, using a time-dependent coefficient to control the nonlinear system allows the model to be applied to a broader range of complex stimuli.

The present study combined the level-independent frequency glide with level-dependent features (e.g., the gain and bandwidth of peripheral tuning) in a simple manner to create a model that can process arbitrary stimuli. Previous studies (Shekhter and Carney, 1997; Tan and Carney, 1999) showed that careful selection of the locations of poles and zeros in the complex plane made it possible to design filters with realistic instantaneous frequency glides in the impulse responses of the filter. The model described here extended the model of Tan and Carney (1999) by combining the pole-zero approach with a feed-forward control path, thereby modeling the compressive nonlinearity of the auditory periphery.

II. MODEL DESCRIPTION

A. Model overview

The basic model components are shown in the block diagram in Fig. 1. The model consisted of four parts: a middle-ear model, a time-varying bandpass filter as the signal path, a nonlinear control path, and an IHC and synapse model.[2] The middle-ear model was a linear bandpass filter based on the middle-ear frequency response properties described by Rosowski (1996). This linear bandpass filter had two pairs of poles and one second-order zero in control space. The locations of the poles and zeros are specified in Table I. The low-frequency zeros of the middle-ear filter improved the downward frequency glide in the model's impulse response at low CFs.

Basilar membrane tuning was modeled with a time-varying bandpass filter, and the compressive nonlinearity of the BM was achieved with the nonlinear control path. The IHC and synapse model was based on that in Zhang *et al.* (2001). A 0.5-ms delay was added to the model output to match model and neural latencies. The output of this model was the instantaneous firing rate as a function of time.

This section describes the major components of the signal path and the control path of the model. This is followed by a description of how the parameters for model fibers across a range of CFs were estimated from AN recordings. The values of all model parameters are listed in Table I.

B. The signal path

The signal path was configured to produce a frequency glide in its impulse response. To illustrate mathematically how manipulation of pole-zero locations generates frequency glides in impulse responses, we consider a fourth-order linear filter with two complex-conjugate pole pairs[3] at $p_1(-x_1,2\pi f_1)$, $p_2(-x_2,-2\pi f_1)$, $p_3(-x_1,2\pi f_2)$, and $p_4(-x_2,-2\pi f_2)$, where $x_1>0$ and $x_2>0$. The transfer function of this simplified linear filter is

$$H(s)=\frac{a}{(s-p_1)(s-p_2)(s-p_3)(s-p_4)}. \tag{1}$$

The right side of Eq. (1) can be transformed into the sum of four first-order fractions:

$$H(s)=\frac{a_1}{(s-p_1)}+\frac{-a_1}{(s-p_2)}+\frac{a_2}{(s-p_3)}+\frac{-a_2}{(s-p_4)}, \tag{2}$$

where a_1 and a_2 are gains derived by factoring the right side of Eq. (1).

In the time domain, the impulse response of this linear filter [the inverse Laplace transform of Eq. (2)] is

$$h(t)=2a_1e^{-x_1t}\sin(2\pi f_1t) + 2a_2e^{-x_2t}\sin(2\pi f_2t), \quad \text{for } t\geqslant 0; \tag{3}$$

x_1 and x_2 determine how quickly the envelopes of the first and second terms in Eq. (3) reach their peak values, respectively. If it is assumed that $x_1>x_2$ and the values of a_1 and a_2 are carefully adjusted ($a_1>a_2$), then the first term has a larger amplitude and dominates at the beginning of the impulse response $h(t)$. The second term dominates the latter part of $h(t)$ because the first term decays faster than the second term. Thus, the instantaneous frequency is closer to f_1 at the beginning of the impulse response and is closer to f_2 at the end of the impulse response. Addition of poles to the filter provides increased control over the frequency shifts as a function of time in $h(t)$.

A fifth-order zero was placed on the real axis in the complex plane. For lower CFs, the zero is pushed closer to

TABLE I. Parameter values.

Parameter CF	Description Characteristic frequency	Value
Middle ear model		
Pm1_real	Real part of pole 1 in middle ear model (Hz)	−250
Pm1_img	Imaginary part of pole 1 in middle ear model (Hz)	400
Pm2_real	Real part of pole 2 in middle ear model (Hz)	−2000
Pm2_img	Imaginary part of pole 2 in middle ear model (Hz)	6000
Zm	Location of the zero on real axis for middle ear model (rad/s)	−200
Control path		
F_{cwb}	Center frequency of the wide-band filter (Hz)	1.2 mm basal to fiber CF
BW_{wb}	Bandwidth of the wide-band filter (Hz)	CF/4
Acp	Parameter in the first nonlinear function	100
Bcp	Parameter in the first nonlinear function	2.5
Ccp	Parameter in the first nonlinear function	0.6
S_0	Parameter in the second nonlinear function	8.0
S_1	Parameter in the second nonlinear function	3.0
T_0	Parameter in the second nonlinear function	0.85
T_1	Parameter in the second nonlinear function	5.0
Fclp	Cut-off frequency of the low-pass filter (Hz)	800
G_{control}	Gain in control path	See Eq. (18)
σ_c	Output of the control path	
Signal path		
P_a	Relative locations of poles, real part in the bandpath filter of the signal path	See Eq. (9)
P_b	Relative locations of poles, imaginary part	See Eq. (10)
P_ω	Imaginary part of the pole closest to the imaginary axis	See Eq. (12)
σ	Real part of the pole closest to the imaginary axis	$\sigma=\sigma_0-\sigma_c$
σ_0	Real part of the pole closest to the imaginary axis at quiet	See Eq. (11)
Z_0	Parameter for the location of the zeros in signal path	0.9
Z_1	Parameter for the location of the zeros in signal path	−1.5
Inner hair cell and synapse model See Zhang *et al.* (2001)		

the origin, which makes the low-frequency side of the filter transfer function steeper and the high-frequency side shallower than that for high CFs.

The combined influence of zeros and poles on the frequency glide can be illustrated as follows: For a simple system with one pair of poles (in conjugate) and one zero on the negative real axis

$$H(s)=\frac{(s+c)}{(s+a)^2+b^2}. \tag{4}$$

For the convenience of the inverse Laplace transform, Eq. (4) can be rewritten as

$$H(s)=\frac{(s+a)}{(s+a)^2+b^2}+\frac{(c-a)}{(s+a)^2+b^2}. \tag{5}$$

In the time domain, the impulse response is

$$h(t)=e^{-at}\left[\cos(bt)+\frac{(c-a)}{b}\sin(bt)\right], \quad \text{for } t\geqslant 0, \tag{6}$$

or, more conveniently,

$$h(t)=e^{-at}\sqrt{1+\left(\frac{(c-a)}{b}\right)^2}\cos\left(bt-\arctan\left(\frac{(c-a)}{b}\right)\right), \quad \text{for } t\geqslant 0. \tag{7}$$

For a slightly more complicated system with two pairs of poles and two zeros, as shown in Fig. 2, the poles and zeros can be divided into two groups, each having one zero and one pair of conjugate poles. The location of the zeros affects the coefficient in Eq. (7).

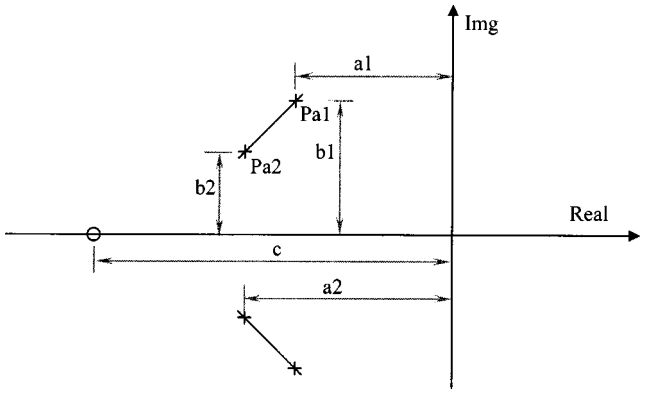


FIG. 2. A simple example of pole-zero locations for a bandpass filter, which has two pairs of poles (P_{a1}/P_{a2} and their conjugates) and one zero (z) in control space. The relative locations of the poles and zeros affect the trend of the instantaneous frequency profile in the filter's impulse response. In this example, P_{a2} has a larger damping coefficient and smaller resonance frequency than P_{a1} does. Therefore, the beginning part of the impulse response is dominated by relatively lower frequency (b_2) and the later part is dominated by higher frequency (b_1).

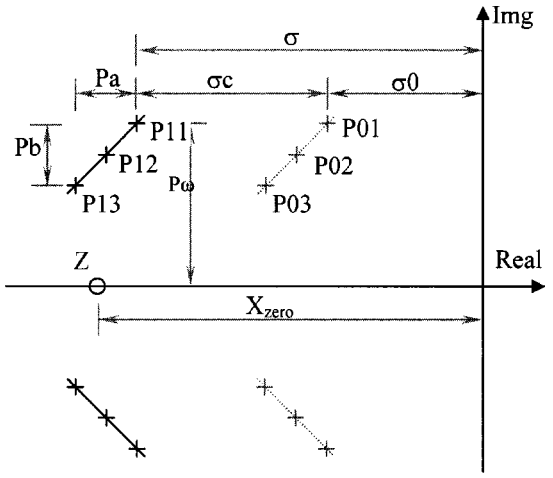


FIG. 3. Pole-zero locations for the bandpass filter in the model's signal path. Ten pairs of poles (P11 and P13 were fourth-order poles and P12 was a second-order pole) and a tenth-order zero were included. P01, P02, and P03 are the pole locations when the input sound intensity is zero for the poles P11, P12, and P13, respectively.

The envelope ratio between the two groups of poles and zeros, i.e.,

$$R=\frac{e^{-a_1 t}\sqrt{1+((c-a_1)/b_1)^2}}{e^{-a_2 t}\sqrt{1+((c-a_2)/b_2)^2}}, \tag{8}$$

determines the relative dominance of each group of poles when determining the instantaneous frequency at time t.

The model presented here had ten pairs of conjugate poles. The relative locations of the poles and the locations of ten zeros determined the instantaneous frequency glide (Fig. 3). The locations of the poles and zeros were set to be functions of model CF based on fitting revcor functions for a population of AN fibers, as described later. The model-CF dependence of the pole locations is described as

$$\log_{10}(P_a)=\log_{10}(\mathrm{CF})\times 1.0230+0.1607; \tag{9}$$

$$\log_{10}(P_b+1000)=\log_{10}(\mathrm{CF})\times 1.4292-1.1550; \tag{10}$$

$$\log_{10}(\sigma_0)=\log_{10}(\mathrm{CF})\times 0.4+1.9; \tag{11}$$

$$P_\omega=1.0854\times\mathrm{CF}-106.0034; \tag{12}$$

where P_a specifies the relative real parts of the poles, P_b specifies the relative imaginary parts of the poles, and P_ω is the imaginary part of the pole closest to the imaginary axis (see Fig. 3). All zeros were at the same location on the real axis, X_{zero}. The distance between the zeros and the origin was a function of CF on a log-log scale:

$$\log_{10}(X_{\mathrm{zero}})=Z_1\log_{10}(\mathrm{CF})+Z_0. \tag{13}$$

The zeros move away from the origin to negative infinity as CF increases. This definition of X_{zero} emphasizes the dominance of the poles with higher frequency at the beginning of the impulse response, especially for low CFs.

The signal-path filter had two eighth-order poles and one fourth-order pole, their complex conjugates, and a tenth-order zero on the real axis. This was the minimum number of poles and zeros required to generate realistic frequency glides in the time domain and realistic sharpness of tuning in the frequency domain.

C. The control path

The compressive nonlinearity is an important property of cochlear tuning in the healthy ear. This property was achieved by including the control path, which continuously changed the bandwidth and gain of the signal path. The control path included four segments in series (Fig. 1).

A nonlinear wideband filter determined the frequency range of the stimulus that affects the bandwidth and gain of the signal path. The bandwidth of the control-path wideband filter was set to twice the bandwidth of the signal path when there was no input signal. The center frequency of the wideband filter was set to a frequency corresponding to the place on the BM approximately 1.2 mm basal to the place that corresponded to the model CF. The bandwidth and the basal shift of the control path were chosen to achieve the appropriate shape of AN suppression tuning curves (e.g., Sachs and Kiang, 1968; Arthur *et al.*, 1970; Delgutte, 1990). The gain of the wideband filter was normalized to one at the model CF. A feedback signal derived from the output of the control path increased the bandwidth of the wideband filter with larger input sound intensity. This bandwidth control and the normalization of the gain resulted in different slopes in the two-tone suppression growth functions for suppressor frequencies above or below model CF (see below, Fig. 13).

A symmetric nonlinear function adopted from Zhang *et al.* (2001) followed the wideband filter:

$$X_2(t)=\mathrm{sgn}[X_1(t)]B_{cp}\log(1+A_{cp}|X_1(t)|^{Ccp}). \tag{14}$$

In Eq. (14), $X_1(t)$ is the output of the wideband filter in Pascals and $X_2(t)$ is the output of the symmetric nonlinear function. This compressive function made it easier to control the shape of the BM velocity-intensity function (see below, Fig. 7).

An asymmetric second-order Boltzmann function followed the symmetric logarithmic nonlinear function. This Boltzmann function corresponded to the membrane potential-displacement function of the outer hair cell, as suggested by Mountain and Hubbard (1996):

$$Y[X_2(t)]=B[X_2(t)]-B(0), \tag{15}$$

where x_2 was the output of the asymmetric function described by Eq. (14), and $B(x)$ was the second-order Boltzmann function:

$$B[x(t)]=\frac{1}{1+\exp[(T_0-x(t))/S_0]\times(1+\exp[(T_1-x(t))/S_1])}. \tag{16}$$

$B(0)$ was subtracted from $B[x(t)]$ to guarantee that $Y(0)$ was zero.

The parameters T_0, T_1, S_0, and S_1 were chosen such that the asymmetry of this control-path nonlinearity had a 7:1 ratio, as suggested by the responses of outer hair cells (OHCs) (Mountain and Hubbard, 1996).

The last component of the control path was a second-order low-pass filter with an 800-Hz cutoff frequency. The cutoff frequency of this filter was estimated from the results of Recio *et al.* (1998), which showed that the time course of the onset of compression has a time constant of approximately 0.2 ms, which corresponds to an 800-Hz cutoff frequency. This filter was chosen to be second order for simplicity; the effect of filter order will be explored in future work.

The control signal (the output of the control path), $\sigma_c(t)$, changes the real part of the locations of the poles of the band pass filter in control space [i.e., a positive $\sigma_c(t)$ makes the pole locations move further away from the imaginary axis and a negative $\sigma_c(t)$ makes the pole locations move closer to the imaginary axis]:

$$\sigma_i(t) = \sigma_{i0} - \sigma_c(t), \tag{17}$$

where $\sigma_i(t)$ is the damping coefficient of the ith pole, σ_{i0} is the damping coefficient of the ith pole when the input is zero, and $\sigma_c(t)$ is the control signal. Note that $\sigma(t)$ in Fig. 1 is a vector representing the real parts of all of the poles. The vector σ_0 contains the values of σ when no stimulus is present.

D. The IHC and synapse model

The IHC and synapse models were the same as in Zhang *et al.* (2001). The IHC model consisted of a logarithmic saturating function followed by a seventh-order low-pass filter. The IHC-AN synapse model was a time-varying three-store diffusion model [Westerman and Smith (1988); adapted into a time-varying model by Carney (1993) and Zhang *et al.* (2001)]. The model described here included only AN fibers with high spontaneous rates.

E. Parameter estimation

The relative positions of the poles in the signal path (P_a and P_b in Fig. 3) were estimated by fitting the model response to revcor functions of low-frequency cat AN fibers (Carney and Yin, 1988). The damping coefficient of the poles, σ_{80} (the average value of σ for responses to an 80-dB SPL noise), was initially estimated on the basis of cat revcor functions computed for AN responses to 80-dB SPL (rms) noise stimuli. Using 80-dB SPL responses for the initial parameter estimation had two advantages over using lower-level data. First, when the input SPL is high, the signal-to-noise ratio in the revcor function is relatively high. Second, revcor functions for 80-dB SPL responses were available for most fibers in the data set used (Carney and Yin, 1988). A linearized model (with the control signal set to zero and thus with level-independent pole locations) was first used to fit the 80-dB SPL revcor functions. The imaginary part of the pole that was closest to the imaginary axis was first set to the peak frequency of the revcor function's magnitude spectrum. The Marquardt (1963) method was then used to estimate the locations of the poles and zeros in the control space. For simplicity, the zeros were set so that they were always on the real axis. The target function of the parameter estimation was to minimize the rms value of the difference between revcor data and model revcor functions. Several values were tested as the initial imaginary part of the pole closest to the imaginary axis so that the fitting performance would not be limited by chance selection of a value.

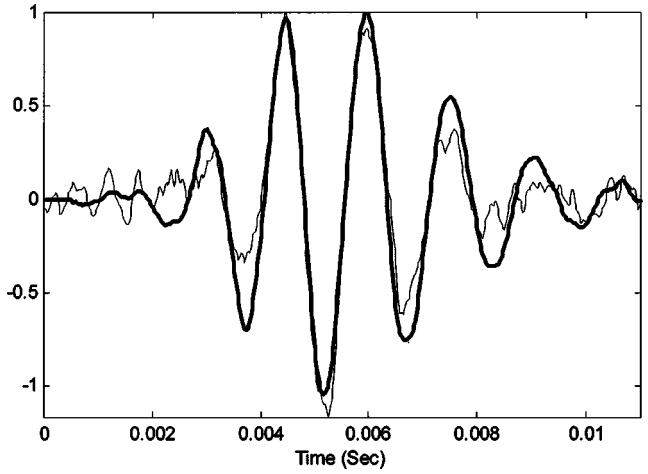


FIG. 4. An example of the model revcor fitted to cat data. The thin line is a revcor function from cat [unit 15 from cat 86166 (Carney *et al.*, 1999)] and the bold line is the corresponding model result. The CF of this fiber is 650 Hz. A downward frequency glide is apparent in these impulse responses, as the zero crossings are increasingly separated at later times in the responses.

An example of optimized locations for poles and zeros is shown in Fig. 3. (The poles for the high-SPL filter are on the solid short line and are P11, P12, and P13) The parameters estimated with the high-level AN revcor functions were P_a, P_b, P_ω and σ_{80}, where P_a and P_b were the relative positions of the poles as illustrated in Fig. 3; P_ω was the imaginary part of the pole with the largest imaginary part; and σ_{80} was the real part of the poles with the largest imaginary part and corresponds to the average value of σ when the stimulus is white noise at 80 dB SPL. Figure 4 shows an example of a model revcor function fit to cat revcor data [fiber U15-C86166 from Carney *et al.* (1999)]. This AN fiber with CF of 650 Hz has a downward frequency glide in its impulse response.

Revcor data (Carney and Yin, 1988) are available mostly for high SPLs (>40 dB SPL), which makes it difficult to estimate the tuning properties of AN fiber for low SPLs for a wide CF range based on revcor data (Carney and Yin, 1988). Q_{10} data (Miller *et al.*, 1997) are based on the threshold tuning curves of AN fiber responses. Thus Q_{10} data can be used to estimate the tuning properties of AN fibers at low SPLs. The Q_{10} data set was used to determine the locations of the poles in the resting state (when there was no sound input), specifically the value of σ_0 (σ in quiet). The CF-dependent gain of the control signal, G_{control} was adjusted such that an input of 80-dB SPL noise resulted in an average control signal equal to the difference in the real parts of P_{11} and P_{01}. The Q_{10} value was then measured for the nonlinear model, and P_{01}, P_{02}, P_{03}, and the gain of the control signal were adjusted until the model's Q_{10} value matched experimental Q_{10} data for the fiber's CF. After σ_0 and σ_{80} were set as functions of CF, the control signal was adjusted to produce an appropriate value to control the locations of the poles in

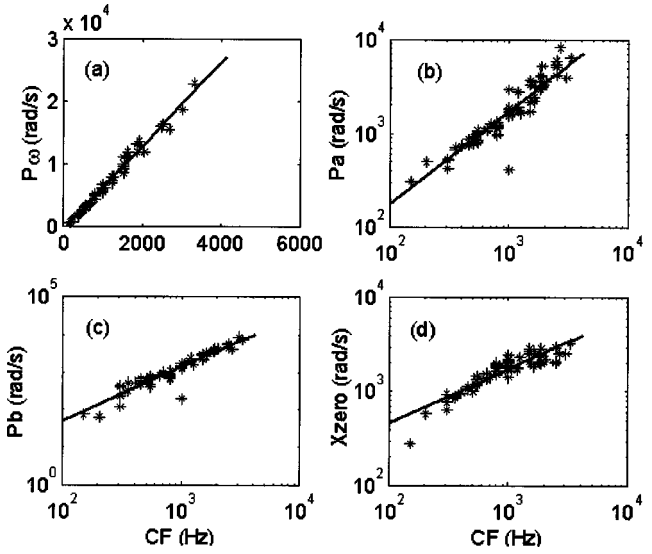


FIG. 5. Parameter estimation results for the pole and zero locations in the signal path for (a) P_ω, (b) P_a, (c) P_b, and (d) X_{zero} (see Fig. 3 for a description of the parameters). Parameter values fit to each fiber's revcor function are shown as stars. Simple expressions for each of the parameters, illustrated by the black lines, were set up and fit to the population results.

quiet and in 80-dB SPL noise. The CF-dependent gain was described by the following equation:

$$G_{\text{control}} = \frac{10^{0.5732CF+1.522} - 10^{0.4CF+1.9}}{0.3357}. \tag{18}$$

The parameter fits for the entire set of fibers were pooled for each parameter (Fig. 5). Simple expressions for the values of these parameters were established as functions of CF, as shown in the text above [Eqs. (9)–(13)], enabling the model to simulate AN fiber responses for any CF up to approximately 3500 Hz. The CF range was limited by the reduction in quality of revcor functions for high CF fibers (Carney and Yin, 1988). A total of 139 cat AN revcor functions from Carney and Yin (1988) were used for parameter estimation. In Fig. 5(a), it is clear that P_ω, the imaginary part (i.e., frequency) of the pole closest to the imaginary axis, is simply related to CF, as expected. P_a and P_b [Figs. 5(b) and (c)] are differences in the real and imaginary parts, respectively, of the poles at each CF (as illustrated in Fig. 3). Figure 5(d) shows the distance of the zeroes from the origin along the negative real axis (note that this is plotted as a positive distance, so that log axes can be used.) Appendix A provides CF-dependent model parameter values for three example CFs, one low-CF fiber with a downward glide, one with CF=1000 Hz which has essentially no glide, and one high-CF fiber with an upward glide.

III. RESULTS

This section illustrates several response properties of the model to tones and other stimuli. It begins with a description of the model's frequency glide, since that was the primary goal in the development of this model. Other fundamental response properties to tones at CF and to other stimuli are then shown. Nonlinear aspects of average rate and temporal response properties were of particular interest and are discussed below.

A. Instantaneous frequency glide

The primary goal of this effort was to incorporate a glide in the instantaneous frequency (IF) of the model's impulse response. The IF glide in this model's revcor function was very close to that reported in the data. An important feature of the IF glide is the constant slope at different noise levels. The model possesses this property because the relative positions of the poles do not change at different sound intensities (i.e., all the poles move in the same direction and have the same amount of displacement). Since the IF glide is determined by the relative positions of the poles, this model had a level-independent IF profile in its revcor function.

Figure 6(a) shows revcor functions for an AN fiber with CF equal to 2060 Hz at six noise levels from 30 to 80 dB SPL, alongside model revcor functions for a fiber with the same CF [Fig. 6(b)]. The zero-crossing points of the revcor functions were almost identical at different SPLs (as indicated by the vertical dotted lines) throughout the timecourse of the revcor functions; therefore, the revcor function's instantaneous frequency is independent of input SPL. Figures 6(c) and (d) show the instantaneous-frequency profiles for three AN fibers (CF=2500, 1600, and 550 Hz from top to bottom), at three levels (40, 60, and 80 dB SPL), alongside instantaneous-frequency profiles for three model AN fibers with the same CFs. To calculate the IF of a revcor function, the envelope of the revcor function was calculated by taking the absolute value of the Hilbert transform of the revcor function. IF was then calculated over the time period where the envelope was more than one quarter of the peak value, using the zero-crossing method (see Appendix B for details about calculation of instantaneous frequency). The overlap of the IF trajectories for the same revcor function at different levels verified the level independence of IF. Generally, the slopes of the IF trajectories increased as a function of CF. The slopes were usually positive (upward) for CFs greater than 1.5 kHz and were negative (downward) for CFs less than 0.75 kHz (Carney *et al.*, 1999). For the same AN fiber, or for the model at the same CF, the duration of the IF trajectory was shorter at higher levels because the time duration of the revcor function was shorter at higher levels [Figs. 6(a) and (b)]. This phenomenon reflects the increasing bandwidth of the revcor function at higher levels due to the compressive nonlinearity. Note that the model revcor functions have longer duration impulse responses than this particular example fiber [Figs. 6(a) and (b)]; this reflects the fact that the tuning for the model, which was based on Q_{10} values for a large population of fibers, was slightly sharper than the tuning for this particular fiber.

B. Rate-level curves

At low sound intensity, the control signal of this model was small (i.e., σ_c is almost zero and σ is near σ_0), and the filter in the signal path behaved like a linear bandpass filter with relatively narrow bandwidth and high gain. The filter output was compressed by the nonlinear control mechanism

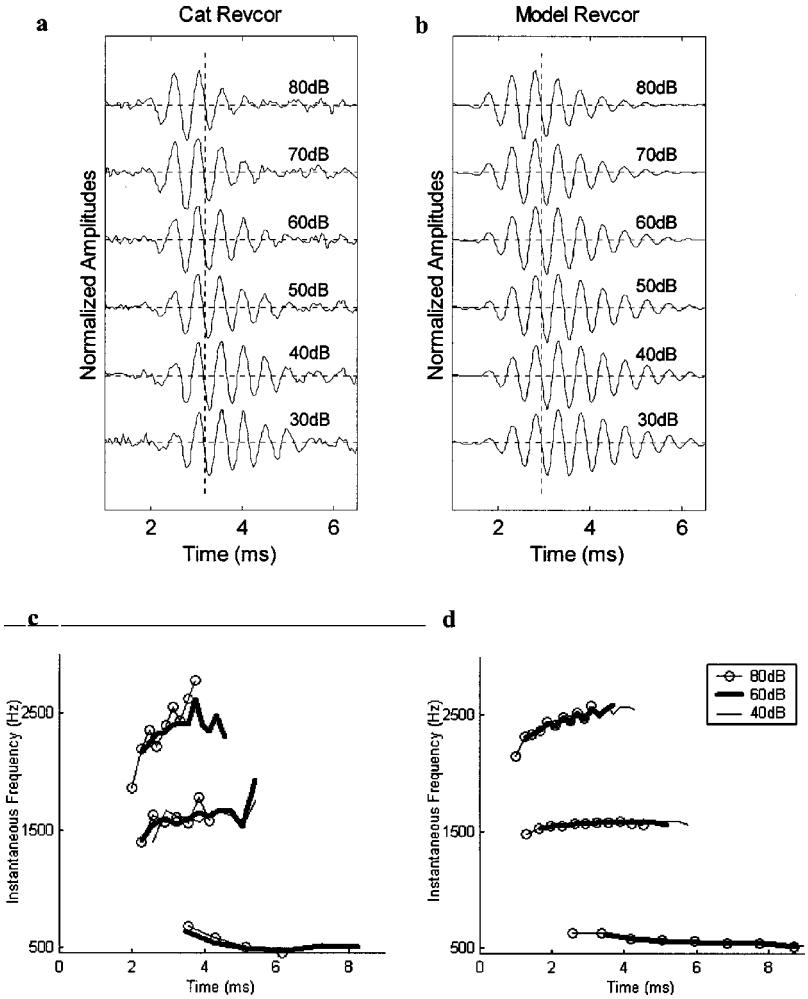


FIG. 6. Revcor functions and instantaneous-frequency profiles. (a) Measured revcor functions for AN fiber with CF=2060 Hz at six levels (30 to 80 dB SPL) (unit 86100-25 from Carney and Yin, 1988), (b) Model revcor functions for a fiber with matching CF. All revcor functions were normalized to their peak amplitude. (c) Measured AN instantaneous-frequency glides calculated based on zero-crossings from revcor functions. These examples show overlying instantaneous-frequency profiles computed from responses to 40, 60, and 80 dB SPL noise stimuli. CFs were 550 Hz (unit 86100-26), 1600 Hz (86100-7), and 2500 Hz (86100-2). (d) Model AN instantaneous-frequency profiles for fibers with CFs matched to the measured fibers. The overlap of the measured and model instantaneous-frequency profiles for different levels indicates that the instantaneous-frequency glides are level independent.

(i.e., σ_c and σ are larger than their values corresponding to sound intensity below threshold) when CF-tone levels were greater than 20 dB SPL. At very high SPLs, the control signal was nearly saturated, which made the filter output behave more linearly.

The compressive nonlinearity is illustrated in Fig. 7, which shows the root mean square (rms) value of the signal path output (F_{out} in Fig. 1) as a function of the input sound pressure level for several CFs. The input is a 50-ms duration pure tone at the model's CF with 2.5-ms onset and offset times. The compressive nonlinearity is stronger for higher CFs (Rhode and Cooper, 1996; Ruggero *et al.*, 1997). Figure 8 illustrates (a) the level-dependent onset rate, (b) sustained rate, and (c) synchronization coefficient of the model's responses to a pure tone at CF. (The left and right columns are results for model CFs at 1100 and 4000 Hz, respectively.) Both the onset rate and the sustained rate increase as the input SPL increases. The dynamic range of the onset rate is larger than that of the sustained rate (about 40 dB), which is appropriate for AN fibers (Smith, 1988).

C. Tuning curves and Q_{10} values

Tuning curves represent the excitation threshold of an AN fiber to tones at different frequencies (Kiang *et al.*, 1965) and thus quantify the relative sensitivity of the AN fiber to various tone frequencies. The stimulus used to measure the model tuning curves was a 50-ms tone followed by 50 ms of silence. The threshold is defined as the sound pressure level at which the average discharge rate during the 50-ms tone is 10 spikes/second greater than the average discharge rate dur-

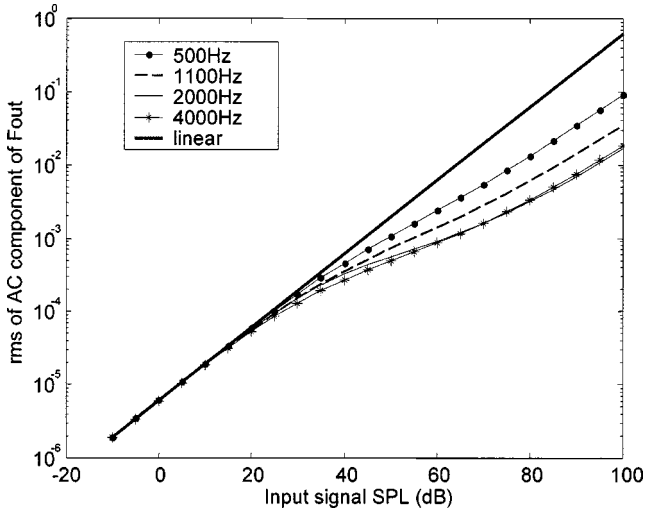


FIG. 7. The rms value of the signal path output, F_{out} (at steady state), in response to CF tones at different SPLs for model CFs at 500, 1100, 2000, and 4000 Hz, respectively. The response patterns demonstrate the compressive nonlinear nature of the signal path. The rms decreases as the model CF increases at a certain tone level, indicating a greater compression with higher CF.

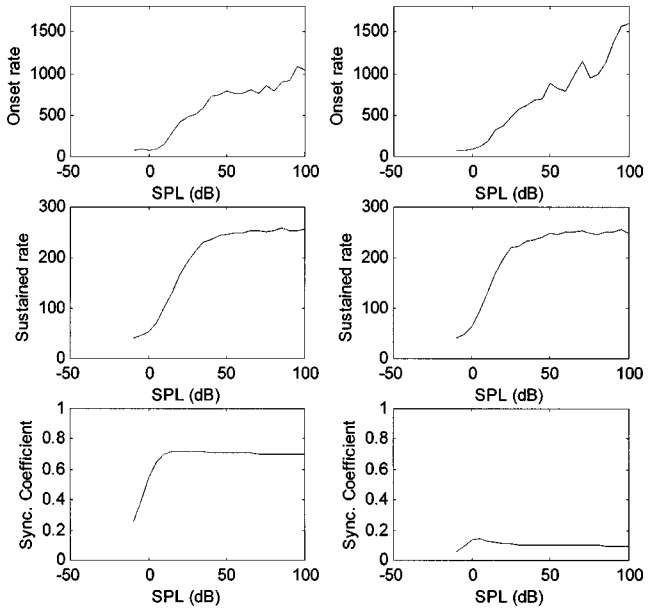


FIG. 8. Response rate and synchronization coefficient to CF-tone input with CF of 1100 Hz (left column) and 4000 Hz (right column). Onset rate is the maximum discharge rate during the first 10 ms and is calculated using 0.5-ms bins. The sustained rate and the synchronization coefficient were calculated in the 10- to 45-ms time window for 400 repetitions.

ing silence. Figure 9(a) shows tuning curves for the AN model at different CFs. The thresholds at CF were set between 0 and 10 dB by adjusting the gain in the IHC model. This threshold can also be adjusted by changing the gain in the middle-ear model. The width of model tuning curves depends on the bandwidth of the bandpass filter in the signal path. The model tuning curves lacks explicit ''tails'' on the low-frequency side as observed in physiological data (Kiang *et al.*, 1965; Kiang and Moxon, 1974; Liberman, 1978). However, this model is relatively sensitive to frequencies near subharmonics of CF due to the nonlinear filter in the signal path.

Q_{10} value (CF divided by the tuning-curve bandwidth 10 dB above threshold) is a standard measurement of the sharpness of tuning curves. The tuning of this AN model was adjusted to match model Q_{10} values to AN fibers with tuning sharper than 75% of the population data. Model Q_{10}'s, shown for a range of CFs in Fig. 9(b), were comparable to Q_{10} values measured from normal cat AN fibers (Miller *et al.*, 1997).

D. Reverse-correlation filters and level-dependence of BF

Threshold tuning curves provide a description of tuning in terms of changes in average discharge rate; however, reverse-correlation filters (the Fourier transform of the revcor function) provide a description of tuning that combines rate and temporal response properties. Figure 10 shows model and measured revcor filters based on noise responses across a range of sound levels. The peak of the revcor filter is an estimate of the fiber's BF, and shifts in the BF with level are seen for both measured [Fig. 10(a)] and model [Fig. 10(b)] revcor filters. The downward shift of BF as level increases is consistent with reports based on revcor filters (e.g., Evans,

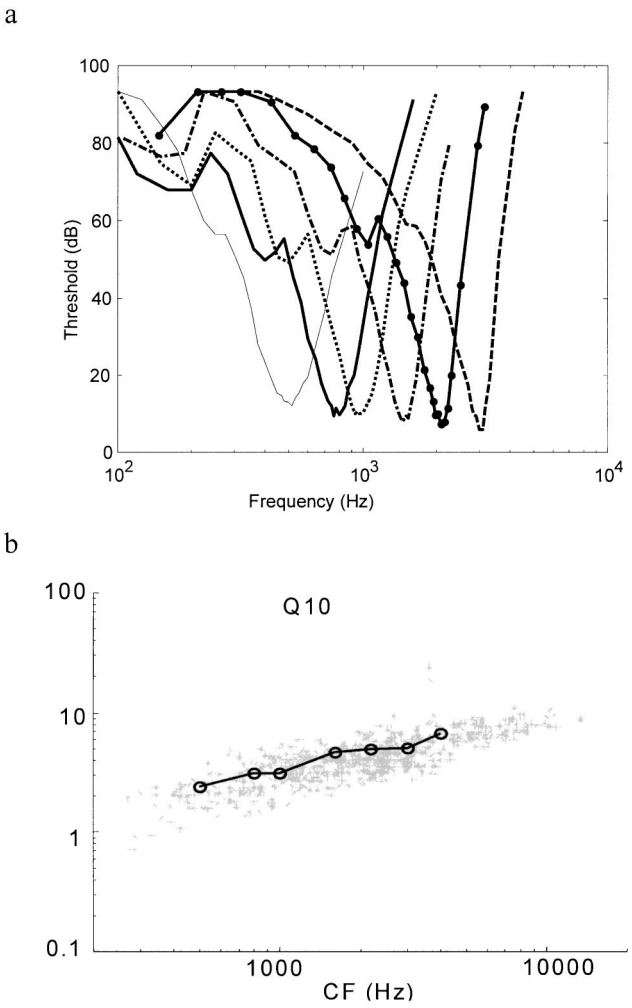


FIG. 9. (a) Model tuning curves for different CFs. The model threshold is defined as the pure-tone SPL that results in a rate response that is 10 spike/s greater than the spontaneous rate. (b) Q_{10} values measured from model tuning curves and compared with physiological data from Miller *et al.* (1997, their Fig. 3). Q_{10} data were used to set the locations of the poles of the bandpass filter in the signal path. Q_{10} values quantitatively described the sharpness of the model's tuning curves as a function of CF.

1977; Møller, 1977; de Boer and de Jongh, 1978) and with the level dependence of BM responses at high-CF places (e.g., Rhode, 1971; Ruggero *et al.*, 1997). As described above, this BF shift is also consistent with the level-independent upward instantaneous-frequency glide in the impulse responses of high-CF fibers. Low-CF fibers that have a downward glide have an upward shift in BF as level increases (not shown).

The shift in BF of the model AN fiber is less than that of the measured AN fiber [Figs. 10(a) and (b)]. This difference between the responses of the model and of this particular AN fiber suggests that they differ slightly either in the frequency range of the instantaneous-frequency glide and/or in the amount of compression, and thus in the temporal extent of the latency shifts as a function of level. The model parameters were based on the responses of a population of AN fibers and were not adjusted to match the responses of individual AN fiber examples.

In addition to the changes in BF and bandwidth that can be observed in the magnitudes of the revcor filters [Figs.

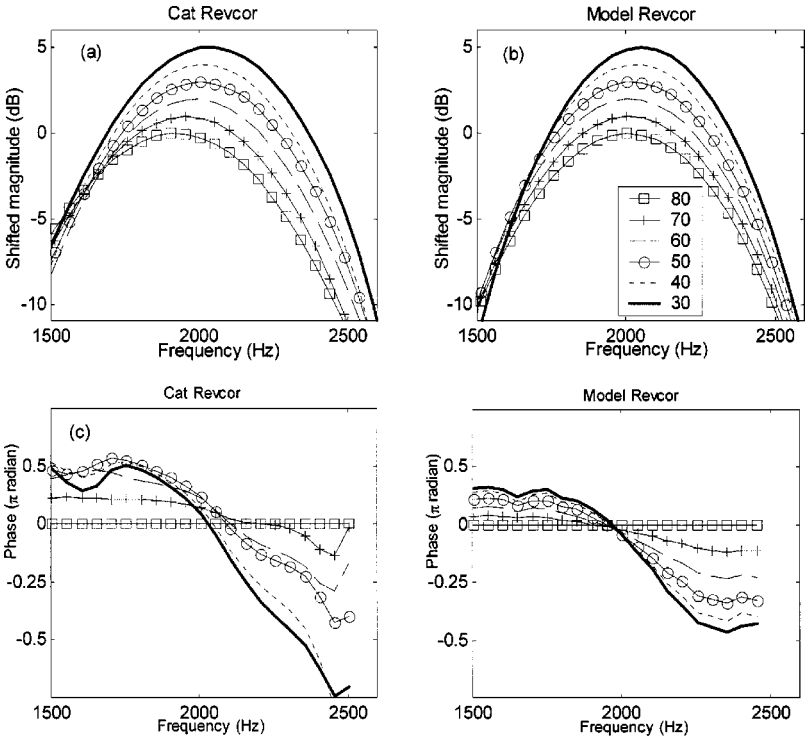


FIG. 10. Measured and model reverse-correlation filters. The top panels (a, c) show the magnitudes of the revcor filters computed in response to wideband noise at several stimulus levels. Each revcor filter was normalized by its peak magnitude; for clarity, a 1-dB shift was introduced between filters computed at different noise levels. The lower panels (b,d) show the level-dependent phases of the revcor filters, plotted relative to the phase at CF. The phases at CF were forced to be between $-\pi$ and $+\pi$. The measured AN responses are from unit 86100-25 from Carney and Yin (1988).

10(a) and (b)], the phases of the revcor filters change systematically with stimulus level [Figs. 10(c) and (d)]. As bandwidth increases at higher stimulus levels, the phase functions get shallower. These changes in bandwidth and phase are seen at both low and high CFs. The level-dependent phase properties of the revcor filters are consistent with the level dependence of phase-locked responses of AN fibers to tones (Anderson *et al.*, 1971). This property of the model was a focus of our modeling efforts because the level dependence of the timing of low-CF AN responses may be important for level coding at low frequencies (Carney, 1994; Heinz *et al.*, 2001b; Colburn *et al.*, 2003).

E. Response areas with phase responses

The upper panel of Fig. 11 shows a response area, or average discharge rates, for an AN model fiber (CF=2200 Hz) in response to tonal stimuli at several frequencies and levels. Each curve corresponds to responses to tones at a constant sound pressure level (iso-level contours). The peaks of the curves at low stimulus levels, or the center of gravity of the saturated responses at high stimulus levels, shift to lower frequencies as SPL increases; however, due to rate saturation, the shift in BF is not as apparent in the response area as it is in the revcor filter (Fig. 10). This BF shift is not seen in the Zhang *et al.* (2001) AN model (their Fig. 9) because the gammatone filter is essentially a symmetric filter in the frequency domain.

The frequency range over which this model responds at high stimulus levels is more limited here than for the Zhang *et al.* (2001) AN model. This narrower frequency range is more appropriate based on physiological descriptions of AN response areas (e.g., Anderson *et al.*, 1971). The difference in the models is due to the higher number of poles used in this model. This property of the higher-order filter can also be seen in the tuning curves (Fig. 9), which are narrower at higher levels for this model than for the Zhang *et al.*, (2001) model.

The lower panel of Fig. 11 shows the level-dependent phase shift for model fiber responses to pure tones with frequencies above and below model CF, referenced to the phase in response to tones at 90 dB SPL [following the plotting convention used in Anderson *et al.* (1971)]. Thus, any non-zero relative phase indicates that the response phase changes with level. The opposite phase change above and below CF is consistent with physiological data (Anderson *et al.*, 1971). However, the maximum negative phase change at frequen-

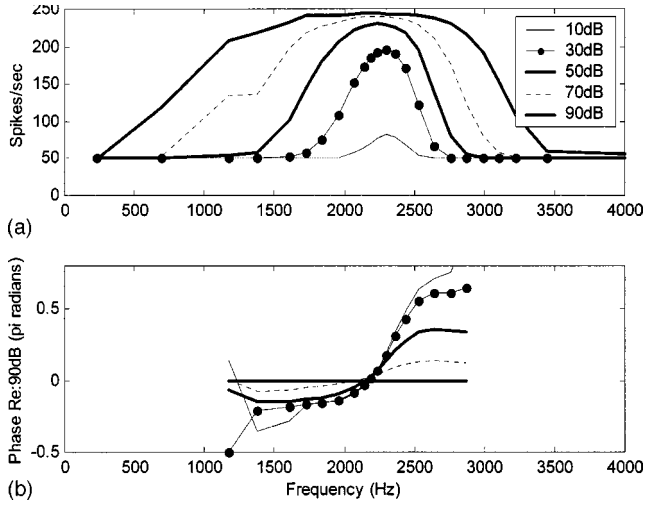


FIG. 11. Response area for model fiber with CF=2200 Hz. (a) Rate response (b) Phase response. Each line corresponds to an input tone SPL as indicated in the figure. The phase responses were referenced to the phase in response to that frequency at 90 dB SPL.

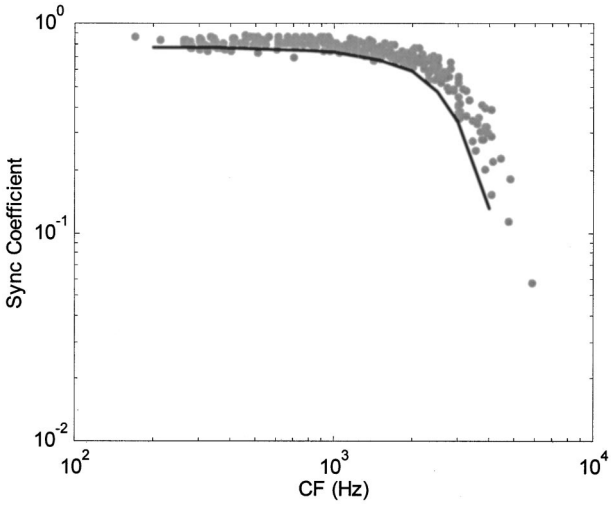


FIG. 12. Maximum synchronization coefficient. The results were measured by taking the maximum synchronization coefficient of model response to CF-tone inputs with SPLs from 0 to 100 dB (i.e., the maximum value from the curves in the bottom panels in Fig. 8). The stimuli were the same as in Fig. 8. Physiological data (Johnson, 1980) from a population of cat AN fibers are indicated with crosses.

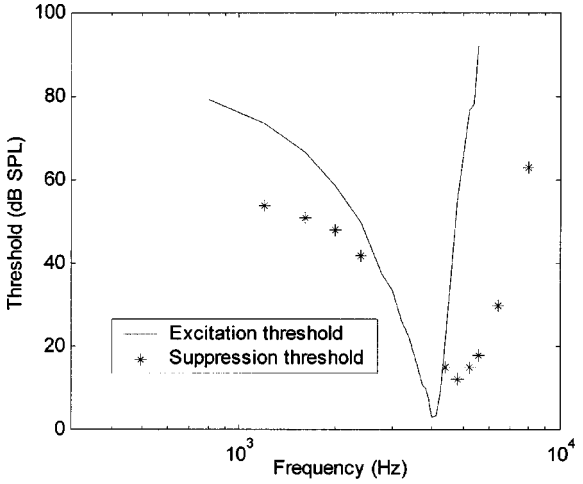


FIG. 13. Suppression threshold. The solid line illustrates the model's tuning curve with CF at 4000 Hz. The stars indicate the suppression threshold, which is defined as the suppressor tone SPL that decreases the response to CF tone by 10 spike/s.

cies below model CF is about $\pi/4$, which is smaller than the value of $\pi/2$ seen in AN fibers (Anderson *et al.*, 1971). This is a limitation of this model, as discussed below.

F. Synchronization coefficient of model AN fibers to CF tones as a function of CF

The synchronization coefficient measures how well the AN response is synchronized to the input pure tone in the time domain. A synchronization coefficient of one means that the AN response is perfectly phase-locked to the input pure tone. Figure 12 illustrates the maximum synchronization coefficient of this model's response to pure tones as a function of CF. The model response's synchronization is an important temporal property that indicates how well the AN response preserves the fine structure of input sound in the time domain. This ability of the AN model to phase lock to the fine structure of input sound is limited by the low-pass filtering in the IHC model. The parameters of the low-pass filter in the IHC and synapse model (Zhang *et al.*, 2001) were chosen to achieve the low-pass roll-off seen in AN data (Johnson, 1980). The model's synchronization coefficient (Fig. 12) is slightly smaller than that reported for cat (Johnson, 1980), due to the limitations of the synapse model. This limitation is also shown in Zhang *et al.* (2001), which has the same synapse model.

G. Two-tone suppression and suppression growth functions

Two-tone suppression (Nomoto *et al.*, 1964; Arthur *et al.*, 1970; Sachs and Kiang, 1968; Delgutte, 1990) is a nonlinear phenomenon of AN fiber responses, in which a stimulus away from CF can act to reduce the response to a stimulus near CF. Suppression was included in the AN model by making the bandwidth of the control path wider than that of the signal path, as suggested by Geisler and Sinex (1980). The suppressor passes through the relatively wideband control path and decreases the gain of the signal path. This feature is especially important for the simulation of AN responses to complex sounds, such as speech signals, where more than one frequency component is present. This suppression can be quantified by the suppression tuning curve (Fig. 13), which measures suppression threshold as a function of suppressor frequency. The suppression threshold is defined as the SPL of the suppressor when the response to a CF tone is reduced by ten spikes/second (Delgutte, 1990). The tips of suppression tuning curves are shifted toward the high-frequency side of excitatory tuning curves (Delgutte, 1990). To implement this tip shift, the center frequency of the wideband filter in the control path was set to be higher than the model CF. This selection of the wideband filter's center frequency (a higher value than CF) is also in agreement with the suggestion that the outer hair cells responsible for the nonlinearity at a given CF are located basally (i.e., tuned to a higher frequency) (Patuzzi, 1996).

While the suppression tuning curves describe the frequency tuning of the onset of suppression, it is also important to examine how suppression grows with level for frequencies above and below CF. The suppression growth function measures the amount of suppression at a particular suppressor frequency as a function of the SPL of the suppressor. The SPL of the CF tone was adjusted such that the response was constant as the suppressor SPL increased. Different slopes are shown for suppression growth functions at different suppressor frequencies for the same CF (Fig. 14): The suppressors with frequencies higher than CF show slower growth (shallower slope) than suppressors with frequencies below CF do. This asymmetrical growth in two-tone suppression has been observed in physiological experiments (Delgutte, 1990). For example, in Fig. 14 (CF=3500 Hz) the slope for growth of suppression by a 1550-Hz tone is 0.8 dB/dB, which is smaller than the averaged physiological data [about 1.3 for suppressor frequency/CF=0.44, Fig. 9 of Delgutte (1990)]. The slope for growth of suppression by a 4400-Hz tone is 0.4 dB/dB, which agrees with the slope

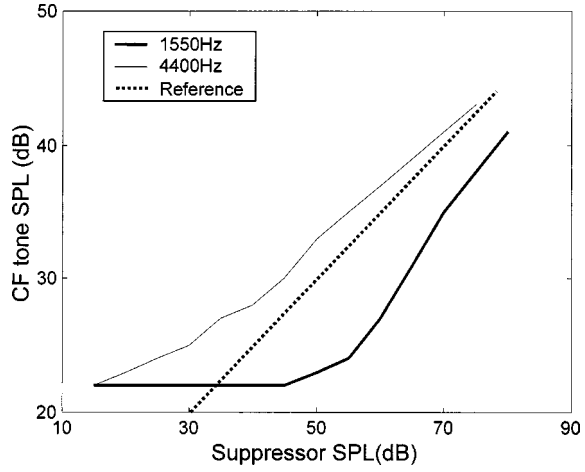


FIG. 14. Suppression growth functions measured for a CF of 3500 Hz, with suppressors at 1550 Hz (below CF) or 4400 Hz (above CF). The CF tone SPL was adjusted to maintain a response rate that was two-thirds of the maximum response rate at each suppressor SPL. The dotted line indicates a growth with slope of 1 (dB/dB).

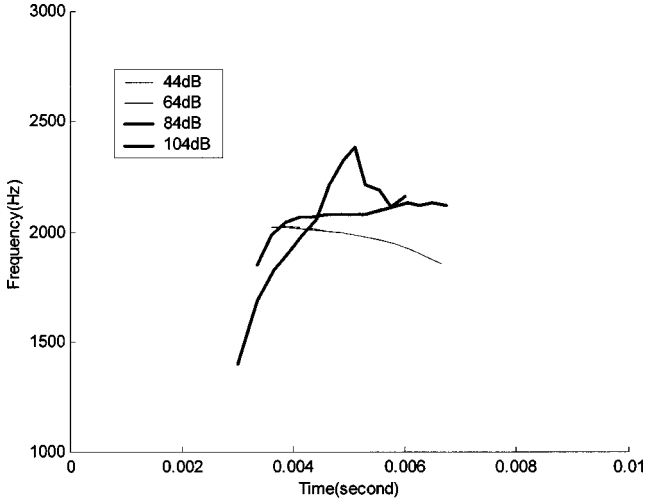


FIG. 15. Instantaneous frequencies calculated for the DRNL model (Meddis *et al.*, 2001) impulse responses for a CF of 2000 Hz at different sound pressure levels (40 to 100 dB in 20-dB steps) based on the zero-crossing method (see Appendix B). A level dependency in the instantaneous profile is demonstrated by the changes in the direction of the glides at different levels. Note that the instantaneous profiles for the two lowest levels are nearly identical, so the curves lie atop one another.

observed physiologically for a suppressor frequency 1.26 times CF (Delgutte, 1990).

IV. DISCUSSION

This report describes a computational AN model that has a level-independent frequency glide and nonlinear compression. AN models based on gamma-tone filters (Carney, 1993; Zhang *et al.*, 2001) do not have frequency glides in their impulse responses because the frequency response of the gamma-tone filter is symmetrical. Pfeiffer's (1970) BPNL model and its extension, Goldstein's (1990, 1995) MBNL model, did not address the level independence of the instantaneous-frequency glide. Meddis *et al.* (2001) qualitatively described the frequency glide in their model's impulse responses, but they did not quantify this response property or examine its level dependence. We explored the level dependence of the frequency glide in the DRNL model by producing impulse responses for a 2000-Hz CF fiber at several input sound pressure levels and calculating the corresponding instantaneous frequencies as a function of time (Fig. 15). Because the DRNL model achieves its changes in bandwidth by changing the overall shape (and thus symmetry) of its frequency response with level, we anticipated that the instantaneous-frequency glide of this model would be level dependent. Indeed, the instantaneous-frequency profile of the DRNL model changes considerably as a function of level (Fig. 15), which is inconsistent with the level-independent frequency glide reported for AN fibers [e.g., Fig. 3 in Carney *et al.* (1999), and cf. Fig. 6 for the model presented here]. The instantaneous frequency of the DRNL model has the same downward glide at the two lowest levels tested (44 and 64 dB SPL) and has an upward frequency glide at higher SPLs.

Irino and Patterson (2001) demonstrated an instantaneous-frequency glide in the impulse response of their gammachirp-based model. However, the trend of this model's best-frequency shift, which is a feature associated with the frequency glide in the impulse response, is not consistent with physiological results. In Fig. 7 of Irino and Patterson (2001), best frequency (CF=2000 Hz) first shifts to a relatively higher frequency and then shifts down as the input SPL increases from 30 to 80 dB. In Fig. 10 of Irino and Patterson (2001), the best frequency increases as the input SPL changes from 30 to 60 dB. However, Anderson *et al.* (1971, their Fig. 8) and Møller (1977) show that the best frequency in responses of AN fibers with mid- to high-frequency CFs shifts to lower frequencies as the input SPL increases. The best-frequency shift can be qualitatively explained by the interaction between the instantaneous frequency glide in the impulse response and the change in the shape of the impulse response envelope at various input SPLs due to the compressive nonlinearity (Carney, 1999). The monotonicity of the IF profiles of the impulse response is also consistent with the monotonic BF shifts that have been observed as a function of SPL.

A. Limitations and future work

The goal of this study was to provide a computational phenomenological AN model with more complete response features than those of previous AN models, and efforts focused on modeling the level-independent frequency glide. The instantaneous frequency glide and compressive nonlinearity were successfully included in this model. This model does have some limitations, however. Only AN fibers with high spontaneous rate have been implemented to date; other spontaneous rate fibers will be considered in future work. The major changes related to modeling low and medium spontaneous rate fibers would be mostly within the IHC and synapse model, and it is anticipated that the configuration of the signal path and the control path described here will not be changed.

The values of this model's parameters were estimated on the basis of revcor functions recorded from low-frequency AN fibers (Carney *et al.*, 1999, below 4000 Hz), which limits

the application of this model in processing signals with relatively high frequency components. Revcor data based on BM measurements are available for higher CFs (de Boer and Nuttall, 1997). These BM revcor data could be used for parameter estimation at high CFs by calculating model revcor functions from the signal path output and then fitting these model revcor functions to BM data.

Another limitation is the relatively small phase shift below CF in the intensity-dependent phase responses (Fig. 11). The limited phase shift is due to the limited number of poles in the signal-path filter and may also be due to the level independence of the zeros' locations. These limitations can be overcome with a more complicated AN model; however, the results presented here focused on a relatively simple model structure.

A primary motivation for the development of this model was to provide a tool for understanding the implications of the frequency glide for stimulus coding. The glide is level independent, but because it interacts with the compressive nonlinearity, its influence on the rate and timing of AN responses is complex. Using techniques that have been developed in recent studies (e.g., Heinz *et al.*, 2001a,b, 2002; Tan, 2003), this computational AN model can be used in future studies to explore the implications of the frequency glide for predictions based on AN responses of psychophysical performance on basic tasks. This model can also be applied to studies of AN responses to more complex stimuli, such as speech sounds (e.g., Bruce, 2003). Tan (2003) has recently applied this AN model to a study of the coding of vowel-like stimuli, including predicting the limits of psychophysical performance in formant-frequency discrimination in quiet and in noise based on population AN responses.

ACKNOWLEDGMENTS

We acknowledge the helpful comments and suggestions of Michael Heinz, Susan Early, Xuedong Zhang, Ian Bruce, Satish Iyengar and Jayant Datta. This work is supported by Grant No. IBN-9983567 from the National Science Foundation and Grant No. DC01641 from the National Institutes of Health. Dr. Christopher Plack kindly provided the computer code for the DRNL model. Dr. Don Johnson generously provided the synchronization coefficient data for AN fibers. Dr. Roger Miller generously provided the Q_{10} data for AN fibers.

APPENDIX A: EXAMPLES OF CF-DEPENDENT PARAMETER VALUES FOR THREE CFS

CF (Hz)	500	1000	2200	Description
σ (rad/s)	954	1258.9	1725.7	Real part of the pole closest to the imaginary axis
P_ω (rad/s)	436.7	979.4	2281.9	Imaginary part of the pole closest to the imaginary axis
P_a (rad/s)	835.1	1697.1	3801.9	Relative locations of poles in signal path, real part
P_b(rad/s)	−496	357.1	3187.8	Relative locations of poles in signal path, imaginary part
X_{zero} (rad/s)	1407.5	3981	12990	The distance from the origin to the zeroes along the negative real axis
G_{control}	650.1	1445.6	3023.8	Gain in control path

APPENDIX B: CALCULATION OF INSTANTANEOUS FREQUENCY

Two methods can be used to calculate the instantaneous frequency of a model or data revcor function, $s(t)$. The first method was based on the Hilbert transform. A complex signal is made by taking $s(t)$ as the real part and the Hilbert transform of $s(t)$ as the imaginary part. The time-domain derivative of the phase of this complex signal is the instantaneous frequency of the revcor function s(t). The second method was based on identifying the zero-crossing points in the revcor functions. The instantaneous frequency of the impulse response was estimated as the reciprocal of the instantaneous period, which is equal to two times the distance between sequential zero crossings.

Before estimating the instantaneous frequency with either method, a three-point average of the original revcor function was performed to smooth the revcor function. Also, the instantaneous frequency profile was only estimated over a time interval for which the envelope amplitude was at least 25% of the maximum value (de Boer and Nuttall, 1997), which avoided noisy fluctuations where the revcor function amplitude was too small to make accurate estimates. In general, these two methods gave similar results; however, the method based on the Hilbert transform was more sensitive to noise in the revcor function and therefore showed more oscillation in the results. The results presented in Figs. 6 and 14 are therefore based on the zero-crossing method.

[1]Best frequency (BF) is the frequency at which the fiber response is strongest at a certain SPL (peak frequency of the response area curve, see Fig. 10). The characteristic frequency (CF) is the frequency at which the threshold of the fiber is lowest.

[2]The C code used to implement this model is available at http://earlab.bu.edu

[3]To get real responses in the time domain after the inverse Laplace transform, poles were arranged in conjugate pairs in the Laplace domain.

Anderson, D. J., Rose, J. E., Hind, J. E., and Brugge, J. F. (**1971**). ''Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: Frequency and intensity effects,'' J. Acoust. Soc. Am. **49**, 1131–1139.

Arthur, M. A., Pfeiffer, R. R., and Suga, N. (**1970**). ''Properties of 'two-tone inhibition' in primary auditory neurons,'' J. Physiol. **212**, 593–1139.

Bruce, I. C., Sachs, M. B., and Young, E. D. (**2003**). ''An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses,'' J. Acoust. Soc. Am. **113**, 369–388.

Carney, L. H. (**1993**). ''A model for the responses of low-frequency auditory nerve fibers in cat,'' J. Acoust. Soc. Am. **93**, 401–417.

Carney, L. H. (**1994**). ''Spatiotemporal encoding of sound level: Models for normal encoding and recruitment of loudness,'' Hear. Res. **76**, 31–44.

Carney, L. H. (**1999**). ''Temporal response properties of neurons in the auditory pathway,'' Curr. Opin. Neurobiol. **9**, 442–446.

Carney, L. H., and Yin, T. C. T. (**1988**). ''Temporal coding of resonances by low-frequency auditory nerve fibers: Single fiber responses and a population model,'' J. Neurophysiol. **60**, 1653–1677.

Carney, L. H., McDuffy, M. J., and Shekhter, I. (**1999**). ''Frequency glides in the impulse responses of auditory-nerve fibers,'' J. Acoust. Soc. Am. **105**, 2384–2391.
Cheatham, M. A., and Dallos, P. (**1999**). ''Response phase: A view from the inner hair cell,'' J. Acoust. Soc. Am. **105**, 799–810.
Cheatham, M. A., and Dallos, P. (**2001**). ''Inner hair cell response patterns: implications for low-frequency hearing,'' J. Acoust. Soc. Am. **110**, 2034–2044.
Colburn, H. S., Carney, L. H., and Heinz, M. G. (**2003, in press**). ''Quantifying the information in auditory-nerve responses for level discrimination,'' JARO.
de Boer, E. (**1997**). ''Connecting frequency selectivity and nonlinearity for models of the cochlea,'' Aud. Neurosci. **3**, 377–388.
de Boer, E., and de Jongh, H. R. (**1978**). ''On cochlear encoding: Potentialities and limitations of the reverse-correlation technique,'' J. Acoust. Soc. Am. **63**, 115–135.
de Boer, E., and Nuttall, A. L. (**1997**). ''The mechanical waveform of the basilar membrane. I. Frequency modulations (glides) in impulse responses and cross-correlation functions,'' J. Acoust. Soc. Am. **101**, 3583–3592.
de Boer, E., and Nuttall, A. L. (**2000**). ''The mechanical waveform of the basilar membrane. III. Intensity effects,'' J. Acoust. Soc. Am. **107**, 1497–1507.
Delgutte, B. (**1990**). ''Two-tone rate suppression in auditory-nerve fibers: Dependence on suppressor frequency and level,'' Hear. Res. **49**, 225–246.
Evans, E. F. (**1977**). ''Frequency selectivity at high signal levels of single units in cochlear nerve and nucleus,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 185–192.
Geisler, C. D., and Sinex, D. G. (**1980**). ''Responses of primary auditory fibers to combined noise and tonal stimuli,'' Hear. Res. **3**, 317–334.
Goldstein, J. L. (**1990**). ''Modeling rapid wave form compression on the basilar membrane as multiple-band-pass-nonlinearity filtering,'' Hear. Res. **49**, 39–60.
Goldstein, J. L. (**1995**). ''Relations among compression, suppression, and combination tones in mechanical responses of the basilar membrane: data and MBPNL model,'' Hear. Res. **89**, 52–68.
Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001a**). ''Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve,'' Neural Comput. **13**, 2273–2316.
Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001b**). ''Rate and timing cues associated with the cochlear amplifier: Level discrimination based on Monaural cross-frequency concidence detection,'' J. Acoust. Soc. Am. **110**, 2065–2084.
Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2002**). ''Quantifying the implications of nonlinear cochlear tuning for auditory-filter estimates,'' J. Acoust. Soc. Am. **111**, 996–1011.
Heinz, M. G., Zhang, X., Bruce, I. C., and Carney, L. H. (**2001c**). ''Auditory-nerve model for predicting performance limits of normal and impaired listeners,'' ARLO **2**, 91–96.
Irino, T., and Patterson, R. D. (**1997**). ''A time-domain, level-dependent auditory filter: the gammachirp,'' J. Acoust. Soc. Am. **101**, 412–419.
Irino, T., and Patterson, R. D. (**2001**). ''A compressive gammachirp auditory filter for both physiological and psychophysical data,'' J. Acoust. Soc. Am. **109**, 2008–2022.
Johnson, D. H. (**1980**). ''The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones,'' J. Acoust. Soc. Am. **68**, 1115–1122.
Kiang, N. Y. S., and Moxon, E. C. (**1974**). ''Tails of tuning curves of auditory-nerve fibers,'' J. Acoust. Soc. Am. **55**, 620–630.
Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). ''Discharge patterns of single fibers in the cat's auditory nerve,'' MIT Research Monograph No. 35 (MIT, Cambridge, MA).
Liberman, M. C. (**1978**). ''Auditory-nerve responses from cats raised in a low-noise chamber,'' J. Acoust. Soc. Am. **63**, 442–455.
Lin, T., and Guinan, J. J. (**2000**). ''Auditory-nerve-fiber responses to high-level clicks: Interference patterns indicate that excitation is due to the combination of multiple drives,'' J. Acoust. Soc. Am. **107**, 2615–2630.
Marquardt, D. W. (**1963**). ''An Algorithm for Least-Squares Estimation of Nonlinear Parameters,'' J. Soc. Ind. Appl. Math. **11**, 431–441.
Meddis, R., O'Mard, L. P., and Lopez-Poveda, E. A. (**2001**). ''A computational algorithm for computing nonlinear auditory frequency selectivity,'' J. Acoust. Soc. Am. **109**, 2852–2861.
Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D. (**1997**). ''Effects of acoustic trauma on the representation of the vowel /ɛ/ in cat auditory nerve fibers,'' J. Acoust. Soc. Am. **101**, 3602–3616.
Møller, A. R. (**1977**). ''Frequency selectivity of single auditory-nerve fibers in response to broadband noise stimuli,'' J. Acoust. Soc. Am. **62**, 135–142.
Mountain, D. C., and Hubbard, A. E. (**1996**). ''Computational analysis of hair cell and auditory nerve processes,'' in *Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 121–156.
Nomoto, M., Suga, N., and Katsuki, Y. (**1964**). ''Discharge pattern and inhibition of primary auditory nerve fibers in the monkey,'' J. Neurophysiol. **27**, 768–787.
Patuzzi, R. (**1996**). ''Cochlear micromechanics and macromechanics,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 186–257.
Pfeiffer, R. R. (**1970**). ''A model for two-tone inhibition of single cochlear-nerve fibers,'' J. Acoust. Soc. Am. **48**, 1373.
Recio, A., Narayan, S. S., and Ruggero, M. A. (**1997**). ''Wiener-kernel analysis of basilar membrane response to noise,'' in *Diversity in Auditory Mechanics*, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. Hecht-Poinar (World Scientific, Singapore), pp. 325–331.
Recio, A., Narayan, S. S., and Ruggero, M. A. (**1998**). ''Basilar-membrane responses to clicks at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **103**, 1972–1989.
Rhode, W. S. (**1971**). ''Observations of the vibration of the basilar membrane in squirrel monkeys using the Mossbauer technique,'' J. Acoust. Soc. Am. **49**, 1218–1231.
Rhode, W. S., and Cooper, N. P. (**1996**). ''Nonlinear mechanics in the apical turn of the chinchilla,'' Aud. Neurosci. **3**, 101–120.
Robert, A., and Eriksson, J. L. (**1999**). ''A composite model of the auditory periphery for simulating responses to complex sounds,'' J. Acoust. Soc. Am. **106**, 1852–1864.
Robles, L., Rhode, W. S., and Geisler, C. D. (**1976**). ''Transient response of the basilar membrane measured in squirrel monkeys using the Mossbauer effect,'' J. Acoust. Soc. Am. **59**, 926–939.
Rosowski, J. J. (**1996**). ''Models of External- and Middle-Ear Function,'' in *Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 15–61.
Ruggero, M. A., and Rich, N. C. (**1991**). ''Furosemide alters organ of corti mechanics: evidence for feedback of outer hair cells upon the basilar membrane,'' J. Neurosci. **11**, 1057–1067.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Sachs, M. B., and Kiang, N. Y. S. (**1968**). ''Two-tone inhibition in auditory-nerve fibers,'' J. Acoust. Soc. Am. **43**, 1120–1128.
Shekhter, I., and Carney, L. H. (**1997**). ''A nonlinear auditory nerve model for CF-dependent shifts in tuning with sound level,'' Assoc. Res. Otolaryngol. **20**, 617.
Shera, C. A. (**2001a**). ''Frequency glides in click responses of the basilar membrane and auditory nerve: Their scaling behavior and origin in traveling-wave dispersion,'' J. Acoust. Soc. Am. **109**, 2023–2034.
Shera, C. A. (**2001b**). ''Intensity-invariance of fine time structure in basilar-membrane click responses: Implications for cochlear mechanics,'' J. Acoust. Soc. Am. **110**, 332–348.
Smith, R. L. (**1988**). ''Encoding of sound intensity by auditory neurons,'' in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 243–274.
Tan, Q. (**2003**). ''Computational and statistical analysis of auditory peripheral processing for vowel-like signals,'' Ph.D. dissertation, Boston University.
Tan, Q., and Carney, L. H. (**1999**). ''A phenomenological model for auditory nerve responses: Including the frequency glide in the impulse response,'' Proc. IEEE 25th Annual Northeast Bioengineering Conference, pp. 23–24.
Westerman, L. A., and Smith, R. L. (**1988**). ''A diffusion model of the transient response of the cochlear inner hair cell synapse,'' J. Acoust. Soc. Am. **83**, 2266–2276.
Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (**2001**). ''A phenomenological model for the responses of auditory-nerve fibers. I. Nonlinear tuning with compression and suppression,'' J. Acoust. Soc. Am. **109**, 648–670.

Time-frequency analyses of transient-evoked stimulus-frequency and distortion-product otoacoustic emissions: Testing cochlear model predictions[a)]

Dawn Konrad-Martin
VA RR&D National Center For Rehabilitative Auditory Research, Portland VA Medical Center, 3710 SW US Veterans Hospital Road, Portland, Oregon 97239and Rush University, Chicago, Illinois[b),c)]

Douglas H. Keefe
Boys Town National Research Hospital, Omaha, Nebraska

(Received 24 January 2003; revised 29 May 2003; accepted 2 June 2003)

Time-frequency representations (TFRs) of otoacoustic emissions (OAEs) provide information simultaneously in time and frequency that may be obscured in waveform or spectral analyses. TFRs were applied to transient-evoked stimulus-frequency (SF) and distortion-product (DP) OAEs to test cochlear model predictions. SFOAEs and DPOAEs were elicited in 18 normal-hearing subjects using gated tones and tone pips. Synchronous spontaneous (SS) OAEs were measured to assess their contributions to SFOAEs and DPOAEs. A common form of TFR of measured OAEs was a collection of frequency-specific components often aligned with SSOAE sites, with each component characterized by one or more brief segments or a single long-duration segment. The spectral envelope of evoked OAEs differed from that of the evoking stimulus. Strong emission regions or cochlear ''hot spots'' were detected, and sometimes accounted for OAE energy observed outside the stimulus bandwidth. Contributions of hot spots and multiple internal reflections to the OAE, and differences between measured and predicted OAE spectra, increased as stimulus level decreased, consistent with level-dependent changes in the estimated cochlear reflectance. Suppression and frequency-pulling effects between components were observed. A recursive formulation was described for the linear coherent reflection emission theory [Zweig and Shera, J. Acoust. Soc. Am. **98**, 2018–2047 (1995)] that is well suited for time-domain calculations. [DOI: 10.1121/1.1596170]

PACS numbers: 43.64.Jb, 43.64.Kc [BLM]

I. INTRODUCTION

Otoacoustic emissions (OAEs) have been explained as arising from a combination of two cochlear mechanisms: coherent, linear reflection and nonlinear distortion. Reflections are described as scattering from multiple randomly spaced discontinuities along the basilar membrane. Only those reflections that sum constructively with the incident traveling wave (i.e., those that display phase coherence) and that arise for an incident traveling wave that is both broad and tall will have sufficient amplitude to contribute to the recorded OAE (Zweig and Shera, 1995). This type of traveling wave response involves nonlinear mechanical amplification supplied by the cochlear outer hair cells that has its highest gain at the lowest stimulus levels. The reflection process itself, however, is thought to be linear. In contrast, the nonlinear distortion mechanism for OAE generation is described as a byproduct of intermodulation distortion in the basilar membrane traveling-wave response. Both coherent linear reflection and nonlinear distortion are thought to contribute to all evoked OAEs, with the level of the stimulus determining, in part, which mechanism is dominant (Shera and Guinan, 1999; Talmadge *et al.*, 2000).

Stimulus-frequency otoacoustic emissions (SFOAEs) and click-evoked otoacoustic emissions (CEOAEs) at low to moderate levels are thought to be predominantly generated by linear reflection near the tonotopic region associated with the spectral content of the stimulus. The SFOAE may have a quasi-regular fluctuation in amplitude (or level) and phase with small changes in the stimulus frequency. This spectral fine structure has a local frequency separation between maxima that is inversely related to the round-trip time delay between the ear-canal microphone and the apical (characteristic frequency, or CF) reflection site in the cochlea (Shera and Zweig, 1993). Interference from multiple reflections between the oval window and the apical reflection site may produce a temporal fine structure in the response envelope, and also may contribute to the spectral fine structure.

In contrast to SFOAEs and CEOAEs, distortion product otoacoustic emissions (DPOAEs) are thought to be initiated by nonlinear distortion in the basilar membrane's response to the two-tone stimulus (with frequencies f_1, f_2, where $f_1 < f_2$). For the cubic distortion product $(2f_1 - f_2)$ and other DPs tuned more apically than the f_2 place a secondary component arises as a coherent linear reflection near the place associated with the DP frequency (e.g., Talmadge *et al.*,

[a)]Portions of this paper were presented at the 2002 June Meeting of the Acoustical Society of America.

[b)]First author's current and previous affiliations, respectively. Work was completed at both institutions listed. Electronic mail: dawn.martin@med.va.gov

[c)]Portions of the data were collected when the first author was affiliated with BTNRH in Omaha, NE.

1999; Knight and Kemp, 2000; Kalluri and Shera, 2001). DPOAE fine structure has been explained as originating from interference between these two ''sources'' (Kim, 1980; Kemp and Brown, 1983) and from multiple reflections of these two components between the oval window and the DP tonotopic place (Dhar *et al.*, 2002). Another proposed source of the spectral fine structure in evoked OAEs is related to spatial variations in the magnitude of the reflected wave, which may be due to variations in the effective reflectance with position along the basilar membrane (Zweig and Shera, 1995; Shera and Guinan, 1999).

OAEs may also be elicited using short-duration tones that are compact in both the time and frequency domains. Such stimuli have the advantage of allowing latencies of zeroth-order (i.e., representing propagation to the site of OAE generation and back to the ear canal) and higher-order reflections to be measured directly in the time domain. Such responses are useful in examining the underlying cochlear mechanisms (Wit and Ritsma, 1980; Talmadge *et al.*, 1998; Tubis *et al.*, 2000a). The presence and number of higher order reflections depends on the basal reflectance at the oval window and the apical reflectance, which is thought to arise from linear coherent reflection near the forward traveling wave peak, but herein is described after its reverse propagation to the basal end of the cochlea. These transient OAE responses may provide a means of analyzing contributions to the OAE fine structure, and may allow estimation of the magnitude of the round-trip reflectance, which is the product of the magnitudes of the apical and basal reflectances. The basal reflectance is fixed by the middle-ear and ear-canal properties while the apical reflectance is hypothesized to decrease in level with increasing stimulus level (Shera and Zweig, 1993). The level dependence of the apical reflectance, which is associated with the compressive basilar-membrane mechanics, is predicted to play a critical role in determining the level of OAEs that arise as coherent reflections as well as the amount of fine structure in evoked OAE recordings.

The coherent reflection mechanism proposed for SFOAEs and CEOAEs has also been proposed for the production of spontaneous otoacoustic emissions (SOAEs) (e.g., Zweig and Shera, 1995). SOAEs are narrow-band emissions measurable in the ear canal in the absence of acoustic stimulation, which are generated within the cochlea by stable limit-cycle oscillations (Tubis and Talmadge, 1998). Based on responses measured in the human ear, SOAEs are associated with the energy reflectance in the ear canal sometimes exceeding unity so that the round-trip losses are balanced by an external source of energy (Burns *et al.*, 1998). For a given ear, the frequency spacing of SOAEs resembles the fine structure of swept-tone SFOAEs and DPOAEs, as well as the fine structure in behavioral measures of hearing threshold (Talmadge *et al.*, 1998). The fine-structure spacing approximates 0.4 bark or $\frac{1}{15}$ oct (i.e., the frequency spacing scales with frequency) and is attributed to the round-trip latency of the OAE. Interference in the ear canal pressure is thought to occur between the applied stimulus and the evoked OAE (Zwicker and Peisl, 1990; He and Schmiedt, 1993; Talmadge *et al.*, 1998, 2000). These experimental results suggest that evoked OAEs contain components in common with SOAE sites, the locations of which are unique to an individual ear.

Closely related to SOAEs and easier to measure is the synchronous spontaneous otoacoustic emission (SSOAE), elicited using a low-level click that synchronizes the OAE response to the measurement duration (Wilson, 1980; Ruggero *et al.*, 1983). The SSOAE is typically measured by discarding the initial 20–25 ms of the response following the click presentation. The next 60–70 ms of the response is sufficiently delayed with respect to the click that the CEOAE response present in the initial 20–25 ms has decayed, and the SSOAE is calculated using a time-averaged spectrum of the windowed, late-duration response. The SSOAE spectrum tends to have some components in common with the SOAE but also contains additional components that are intermittently present in some ears in so-called dynamically linked SOAEs. These are SOAEs in which frequency components are present at one time, but not at another, indicative of a plurality of structurally stable SOAEs in the same ear (Burns *et al.*, 1998). To study the effect of SOAEs on other types of evoked OAEs, the SSOAE is the appropriate response because it shares with evoked OAEs the property that it is elicited synchronously with the stimulus presentation. The tonotopic locations at CFs with strong SSOAEs are termed ''hot spots'' in this report, indicative of strong sites of cochlear reflections. The prevalence of SOAEs is similar in neonatal and young adult ears and is on the order of 62%–64% (Burns *et al.*, 1992). It is assumed that the prevalence of SSOAEs in ears of young adults (and younger subjects) is similar to the prevalence of SOAEs; hence, SSOAEs should be observed in many of the ears included in this study.

OAEs elicited using transient stimuli vary in both time and frequency domains. Such measurements provide direct (OAE waveform) or indirect (OAE spectral phase) information on OAE latencies, but these are interpreted with respect to the delay of a single component whereas there may be multiple frequency and/or latency components within the same OAE signal. Time-frequency representations (TFR) of multiple-component OAE signals provide information simultaneously in both time and frequency that may be useful for their interpretation, but which may be obscured in waveform or spectral analyses. An introduction to time-frequency analysis is given by Cohen (1995). Time-frequency analyses have been reported for both click-evoked and tone-burst-evoked OAEs using the Wigner distribution (Ozdamar *et al.*, 1997), and for CEOAEs using the pseudo-Wigner distribution (Cheng, 1995), wavelet analysis (Wit *et al.*, 1994; Tognola *et al.*, 1997), and Gabor and adaptive spectrograms (Hatzopoulos *et al.*, 2000). The Cohen kernel is a useful class of bilinear time-frequency representations (TFR), for which constraints on the choice of kernel function lead to control of artifacts in the TFR. One commonly used kernel leads to the Choi–Williams TFR, which has been used to analyze click-evoked and tone burst-evoked OAEs (Ozdamar *et al.*, 1997). Using a ''cone kernel'' TFR (Zhao *et al.*, 1990) from the Cohen class, TFRs of spontaneous OAEs (Burns and Pitton, 1993) have been measured as have been evoked OAEs for click responses up to 10 kHz and for tone-burst-evoked SFOAE and DPOAE responses (Keefe *et al.*, 1999).

The aim of the present research was to use time-frequency analyses of OAEs to test cochlear model predictions for a variety of transient-stimulus classes. A double-evoked (2E) measurement technique (Keefe, 1998; Keefe and Ling, 1998), which measures an OAE as a nonlinear residual, provides a unified framework in which to compare OAEs elicited by continuous tones to those obtained using short-duration stimuli. The use of a pair of pips (band-limited impulses), a pair of gated tones (tonal stimuli with well-defined onset, steady state and decay), or presentation of a gated tone with a continuous tone at different frequencies and levels provides comparative measurements of the spectral envelopes of the stimulus and OAE responses. Time-frequency representations of such responses may provide insight into nonlinear cochlear dynamics through the analysis of OAE responses at multiple frequencies, each with its own temporal envelope. A particular goal of this study was to determine whether the magnitude of the product of the basal and apical reflectances could be estimated from a TFR of a transient-evoked OAE.

II. METHODS

A. Subjects

OAEs were measured in 18 subjects aged 12–48 years (mean=19.8 years, standard deviation (SD)=10.5 years) with normal hearing, which was defined as 20 dB HL or better pure-tone thresholds at octave frequencies from 250 Hz through 8.0 kHz, and normal middle-ear function as assessed by a normal 226-Hz tympanogram at the time of testing. The subjects included 15 females and 3 males. During OAE testing, subjects were seated in a comfortable chair within a single-walled sound-attenuated booth.

B. OAE recording technique

Stimuli were digitally generated and recorded at a 24-kHz sample rate using a signal processing board (CAC model DSP32) in a personal computer using custom-designed software. The probe microphone system (Etymotic Research ER-10C) included two loudspeakers to deliver stimuli and a microphone to record the pressure response. The probe was modified by Etymotic to provide 20 dB additional gain in the signal driving each loudspeaker. A 2E OAE waveform p_d was calculated using three waveform responses, p_1, p_2, and p_{12}, stored in separate buffers each of length 1024 samples (42.7 ms), by (Keefe, 1998)

$$p_d = p_1 + p_2 - p_{12}, \tag{1}$$

in which p_1 is the microphone response to a stimulus s_1 presented through loudspeaker 1, p_2 is the response to a stimulus s_2 presented through loudspeaker 2, and p_{12} is the response to the simultaneous presentation of s_1 through loudspeaker 1 and s_2 through loudspeaker 2 (this is called the s_1+s_2 condition). Transient SFOAEs were obtained using s_1 and s_2 tones that were equal in frequency, and transient DPOAEs were recorded using different stimulus frequencies.

For both types of recordings, the 2E subtraction in Eq. (1) removed the linear stimulus response and any linear-growth component of the OAE, eliminated single-channel harmonic distortion, and generated the output distortion signal p_d, which is interpreted as a nonlinear OAE response. To test the extent to which intermodulation distortion produced by the measurement system was negligible, coupler measurements were performed using an artificial human ear simulator (Bruel & Kjaer type 4157) that approximated the impedance of an average adult ear. Interleaving sets of stimulus waveforms (s_1, s_2, and s_1+s_2) and use of artifact rejection resulted in a 2E OAE response that was highly repeatable within and across test sessions. The total stimulus buffer was created using five presentations of s_1, five presentations of s_2, and five presentations of s_1+s_2. To control for response discontinuities when the stimulus changes, the response to each of the first presentations of s_1, s_2, and s_1+s_2 was discarded, so that a single total response buffer comprised four independent responses for each of p_1, p_2, p_{12}, and p_d, the latter calculated using Eq. (1). A total response buffer was valid if there was no artifact, as determined using an artifact rejection test (Keefe and Ling, 1998). An average total response buffer was stored as the time average of 32 valid total response buffers.

The discrete Fourier transform (DFT) was calculated for each of the p_1, p_2, p_{12}, and p_d responses in the average total response buffer. The noise signal corresponding to each of the p_1, p_2, p_{12}, and p_d spectra was calculated in terms of the standard error of the mean of the four independent responses within the average total response buffer, based on the variance of the spectral estimates in each frequency bin. This is described in more detail by Schairer *et al.* (2003).

C. Stimulus classes

DPOAEs were measured for primary frequencies $f_1=3.3$ kHz and $f_2=4$ kHz, which were presented at unequal levels $L_1=L_2+10$ dB. This placed the $2f_1-f_2$ DPOAE at 2.7 kHz. The primary stimuli were a continuous f_2 paired with a gated f_1, or a gated f_2 paired with a gated f_1. The former stimulus paradigm is similar to that introduced by Whitehead *et al.* (1996) and Talmadge *et al.* (1999), except they used a continuous f_1 paired with a gated f_2. The latter stimulus paradigm is related to an OAE waveform measurement procedure introduced by Martin *et al.* (1998) in which gated tones were used to elicit DPOAEs. The continuous f_2/gated f_1 paradigm allows examination of the influence of the gated f_1 tone on the production of the DPOAE and on its suppression of the SFOAE at f_2, which are both thought to be produced near the f_2 tonotopic place. The rationale for the use of the continuous f_2/gated f_1 is to compare whether features of cochlear mechanics are more easily observed with such a paradigm compared to the gated-tone paradigms; in particular, the continuous f_2 produces less suppression of the DP than the continuous f_1 so that multiple internal reflections may be more easily observed with the continuous f_2/gated f_1. It would have been of similar interest to use a gated f_2/continuous f_1 paradigm, inasmuch as the responses measured using either paradigm could be used to test existing OAE models.

SFOAEs were measured using gated tones and tone pips presented at center frequencies of 4 kHz (equivalent to the f_2 used for DPOAE measurements) and 2.7 kHz (equivalent to

the f_{DP} in DPOAE measurements). The present study focused on SFOAEs at 2.7 kHz. The s_1 and s_2 were equal in level ($L_1 = L_2$), so that the level of the $s_1 + s_2$ condition was $L_{12} = L_1 + 6$ dB. SFOAE primary levels based on total energy in the 1024 sample buffer varied from 20 to 60 dB SPL in 5-dB steps.

To produce the gated tone, which had a nominal steady-state duration of 512 samples (21.3 ms), one-half of a Hamming window was used as an onset and offset window, with each half-window duration equal to the period of the tone being gated. For the tone pips, the sinusoidal signal was gated with a complete Hamming window of duration equal to 10.5 times the period of the tone being gated.

Unless otherwise noted, the SPLs reported were the total SPL for the 1024-sample buffer. However, the portion of the response buffer containing the stimulus varied depending on whether a continuous tone, gated tone, or tone pip was used. For the tone pips, the pip SPL at the center frequency was less than the total SPL by the following levels as listed by center frequency: −11.7 dB at 2.7 kHz, −12.7 dB at 3.3 kHz, and −13.5 dB at 4 kHz. These varied with frequency because the pip window duration varied with frequency. Another way to characterize the tone pip spectra is by the SPL of the continuous tone prior to application of the pip window. The differences between the total SPL of the pip and the total SPL of the equivalent continuous tone were −14.4 dB at 2.7 kHz, −15.4 dB at 3.3 kHz, and −16.2 dB at 4.0 kHz. For the gated tones, for which each signal occupied half the buffer with the other half of the buffer zeroed, the steady-state SPL during the gated tone was approximately 3 dB higher than the total SPL. Thus, the gated-tone spectra can be characterized by the SPL of an equivalent continuous tone by adding 3 dB to the total SPL of the gated tone.

In order to characterize spontaneous sources of narrowband OAEs in each ear that might interact with the transient SFOAE and DPOAE measurements, SSOAEs were elicited in response to a 60-dB SPL click. The SSOAE measurement utilized a modified recording technique with a longer buffer (2048 samples or 85.3-ms duration). The initial 22 ms of the response was zeroed, the next 2 ms ramped on the response with a 2-ms cosine-squared window, and the remaining part of the response was unmodified. The SSOAE was calculated from the final 1024 samples of the unmodified response.

Table I lists distortion levels measured in the coupler at the SFOAE or DPOAE frequency (2.7 kHz) corresponding to each stimulus condition, in which the term distortion level includes the joint effects of measurement-system distortion plus noise. Distortion at this frequency was negligible; the mean across the 25 stimulus conditions was −29.2 dB SPL with a SD of 2.3 dB. Moreover, the maximum distortion in the coupler never exceeded −18.4 dB SPL at any frequency between 0.8 and 8.0 kHz. The mean maximum distortion measured across stimulus conditions was −20.7 dB SPL (SD = 0.96), indicating that measurement artifact in these recordings was unlikely to affect the interpretation of OAE responses recorded in the ears of the subjects.

TABLE I. Distortion levels for each stimulus condition tested measured at the nominal SFOAE and DPOAE frequency (2.672 kHz) in an artificial human ear simulator (Bruel & Kjaer type 4157).

L1,L2 (dB SPL)	Tone-pip-evoked SFOAE Distortion level (dB SPL)	Gated-tone-evoked SFOAE Distortion level (dB SPL)
60,60	−23.8	−23.3
55,55	−28.3	−29.0
50,50	−29.0	−27.7
45,45	−30.6	−29.2
40,40	−29.0	−28.5
35,35	−31.4	−30.0
30,30	−30.8	−26.0
	Gated-f_1/ Continuous-f_2-evoked DPOAE Distortion level (dB SPL)	Gated-f_1/ Gated-f_2-evoked DPOAE Distortion level (dB SPL)
65,55	−31.3	−30.3
55,45	−29.6	−30.5
45,35		−31.7
35,25		−32.9
52,45	−30.7	
42,35	−27.3	
32,25	−28.5	
22,15	−30.2	
11,5	−30.7	

D. TFR analysis

Time-frequency representations (TFRs) analyze the 2E signal simultaneously along the dimensions of time and frequency. The discrete cone-kernel method of calculating these TFRs is described in the Appendix. The TFR plots include three-dimensional (3-D) contour plots and two-dimensional (2-D) TFR ''slices'' (or trajectories). These plot the temporal response at a given frequency bin, and express the OAE signal as a level in decibels (dB)—akin to $10 \log_{10} C(t, \omega)$ in Eq. (A2) of the Appendix. The spectral responses below 0.8 kHz were omitted in all TFR contour plots, as these components were dominated by noise while the signal frequencies of interest were well above 0.8 kHz.

The calibration reference for the TFR level varied with the length of the TFR low-pass filter (see the Appendix), as listed in Table II. For example, the calibration level was 17 dB for a TFR low-pass filter length of 90 samples. This means that if a continuous sine tone at 60 dB SPL was recorded and its stimulus response analyzed as a TFR, the TFR trajectory in the frequency bin containing the sine tone frequency had a continuous level that was 17 dB lower. Thus, the table provides the level difference to add to the TFR level in order to convert it to an equivalent continuous SPL at that

TABLE II. Effect of TFR window length on TFR level calibration.

Length (samples)	Level (dB)
90	17
120	18
150	19
180	21

frequency. This calibration reference is specific to the cone-kernel TFR.

III. SFOAE THEORY

The linear coherent reflection emission theory (Zweig and Shera, 1995) describes SFOAE generation in terms of a cochlear apical reflectance $R_a(f)$, which is defined as a function of frequency f to be the ratio of the reverse to the forward (cochlear) pressure traveling wave, with each referenced to the base of the cochlea. In a swept-sine measurement of a SFOAE, the total ear-canal pressure $P_{ec}(f;R_a)$ is related to the reflectance by (Zweig and Shera, 1995)

$$P_{ec}(f;R_a)=\left(1+\frac{T_-R_aT_+}{(1+R_+)[1-R_-R_a]}\right)P_{ec}(f;0). \quad (2)$$

$P_{ec}(f;0)$ is the total ear-canal stimulus pressure in the absence of SFOAEs (for $R_a=0$), $T_+(f)$ is the forward middle-ear pressure transmittance from the ear canal to the basal end of the cochlea, $R_a(f)$ is the apical cochlear pressure reflectance (called ''apical reflectance'' in the Introduction), $T_-(f)$ is the reverse middle-ear pressure transmittance from the basal end of the cochlea to the ear canal, $R_+(f)$ is the forward middle-ear pressure reflectance that is measured in ear-canal reflectance measurements, and $R_-(f)$ is the reverse middle-ear pressure reflectance at the basal end of the cochlea (called ''basal reflectance'' in the Introduction). The middle-ear scattering parameters (T_+,T_-,R_+,R_-) (Shera and Zweig, 1992) and the magnitude of the apical reflectance $|R_a|$ are thought to vary slowly with frequency, the former because middle-ear mechanics varies more slowly with frequency than the cochlear mechanics underlying OAE generation, and the latter because of approximate scaling symmetry on the cochlea and the assumption that the irregularities in cochlear stiffness vary slowly. The term $(1-R_-R_a)$ in the denominator of Eq. (2), which describes multiple internal reflections within the cochlea, fluctuates rapidly with frequency due to the rapid variation of the phase of R_a.

It is convenient to denote the ear-canal pressure including OAEs as $P=P_{ec}(f;R_a)$, and the ear-canal (stimulus) pressure in the absence of OAEs as $Q=P_{ec}(f;0)$. The above equation is rewritten after multiplying through by the term in the denominator and rearranging terms as

$$P_{oae}=T_-R_aT_+Q-R_+P_{oae}+R_-R_a[1+R_+]P_{oae}, \quad (3)$$

in which the total OAE pressure spectrum $P_{oae}(f)$ is defined as the difference between the ear-canal pressures in the presence and absence of OAEs:

$$P_{oae}(f)=P-Q. \quad (4)$$

Its complement in the time domain is obtained via the inverse Fourier transform (FT) as

$$p_{oae}(t)=t_-*r_a*t_+*q-r_+*p_{oae}+r_-*r_a*[\delta+r_+]*p_{oae}, \quad (5)$$

in which the operator $*$ denotes the convolution between each pair of waveforms. Throughout this report, upper-case names are used for frequency-domain variables (except for the use of an upper-case variable in the Appendix to represent the TFR), and lower-case names are used for time-domain variables. The explicit argument of frequency f or time t is often omitted. For example, the forward middle-ear transmittance is $t_+(t)$ in the time domain and $T_+(f)$ in the frequency domain. The Dirac delta function is $\delta(t)$. The interpretation of each term on the right-hand side of Eq. (5) has a clear physical significance. The first term, which is causally interpreted from right to left, includes the effects on the ear-canal stimulus waveform $q(t)$ due to forward transmission through the middle ear, forward-nonlinear cochlear transmission to apical place, linear coherent reflection from the apical region, reverse-nonlinear cochlear transmission back to the cochlear base, and reverse transmission through the middle ear into the ear canal as the total OAE waveform $p_{oae}(t)$. The second term $-r_+*p_{oae}$ represents the effect of internal reflection of the OAE within the ear canal. The third term, which is proportional to the convolution between the total-internal-reflection filter (r_-*r_a) and the resulting output OAE (p_{oae}), represents the transmission back into the ear canal of the multiple internal reflections. This term is present only when the basal reflectance r_- is nonzero. The second and third terms are recursive because the output variable $p_{oae}(t)$ occurs within the convolution.

Each recursive expression in Eq. (5) is causal because each of the middle-ear scattering parameters and the apical reflectance is causal. For example, suppose that the ear-canal pressure $p(t)$ equals zero prior to turning on the stimulus $q(t)$ at $t=0$. A causal time-domain reflectance has the property that it is zero for all times less than or equal to zero (McIntyre *et al.*, 1983), from which it follows that the convolution of r_a, r_+ or r_- with either of q or p is zero for all $t\leqslant 0$. Thus, the solution to the above equation at $t=0$ is $p_{oae}(0)=0$, which implies from the inverse FT of Eq. (4) that $p(0)=q(0)$. The equation may be evaluated at later times t, with the recursive term only involving the values of $p_{oae}(t')$ for $t'<t$, so that $p_{oae}(t)$ on the left-hand side of Eq. (5) is a function of $p_{oae}(t')$ at earlier times t'. OAEs may be simulated in the time domain in response to an arbitrary stimulus waveform q by specifying models for middle-ear and cochlear transmission.

In the absence of multiple internal reflections (i.e., $r_-=0$), Eq. (5) for the total OAE simplifies to an expression that is equivalent to, but distinct from, that given by Shera and Zweig (1995):

$$p_{oae}(t)=t_-*r_a*t_+*q-r_+*p_{oae}, \quad (6)$$

For a given stimulus waveform $q(t)$, Eq. (6) predicts that the resulting SFOAE is delayed relative to the stimulus largely due to the convolution with the time-domain apical reflectance $r_a(t)$, which includes the round-trip travel time between the cochlear base and the tonotopic place. It also predicts that the SFOAE is rescaled in magnitude by the composite effects of the middle-ear scattering parameters and the apical reflectance. Because the magnitude of the apical reflectance is thought to vary slowly with position along the BM, it follows that the predicted SFOAE response to a tone pip stimulus is a time-delayed and attenuated tone pip, the predicted SFOAE response to a gated tone stimulus is a

time-delayed and attenuated gated tone, and so forth. This prediction is tested in the measurements reported below. One significant exception is that the presence of SSOAEs in the spectral bandwidth of the stimulus evoking the SFOAE will produce discrepancies between the SFOAE and stimulus responses shapes, but an ear without SSOAEs is predicted to have SFOAEs of this form.

In some experiments, it is necessary to generate model predictions for observables that are nonlinear residuals of the total OAE. In the present experiments, such an observable is the distortion pressure defined in Eq. (1). The distortion pressure predicted by the above model is constructed in this section for the stimulus condition in which the s_1 and s_2 are presented at equal frequencies and levels so that $s_{12}=2s$. This stimulus condition was used in the present experiments to measure all transient SFOAE responses. The corresponding stimulus pressures arising in the FT of Eq. (1) in the absence of OAEs are $P_1=P_2=Q$, so that $P_{12}=2Q$. Each of these scales the level dependence of its corresponding total OAE. The level-dependent apical reflectance measured at the input stimulus level of either P_1 or P_2 is notated as $R_a(Q)$, which should not be confused with the fact that apical reflectance is also a function of frequency, and that measured at the level of P_{12} is notated as $R_a(2Q)$. It is reasonable to assume that the middle-ear scattering parameters are approximately independent of level.

Therefore, the pressures P_1 and P_{12} each satisfy an expression similar to that in Eq. (2):

$$P_1=\left(1+\frac{T_-R_a(Q)T_+}{(1+R_+)[1-R_-R_a(Q)]}\right)Q,$$
$$P_{12}=\left(1+\frac{T_-R_a(2Q)T_+}{(1+R_+)[1-R_-R_a(2Q)]}\right)2Q, \tag{7}$$

and in the equal-level 2E paradigm, the distortion spectrum is calculated from the FT of Eqs. (1) and (7) as

$$\begin{aligned}P_d(f)&=P_1+P_2-P_{12}\\&=2P_1-P_{12}\\&=\frac{T_-}{(1+R_+)}\left(\frac{R_a(Q)}{[1-R_-R_a(Q)]}\right.\\&\quad\left.-\frac{R_a(2Q)}{[1-R_-R_a(2Q)]}\right)T_+2Q.\end{aligned} \tag{8}$$

The latter equation is multiplied through by the denominator terms and simplified:

$$\begin{aligned}P_d&=2T_-[R_a(Q)-R_a(2Q)]T_+Q-R_+P_d\\&\quad+R_-[R_a(Q)+R_a(2Q)-R_-R_a(Q)R_a(2Q)]\\&\quad\times(1+R_+)P_d.\end{aligned} \tag{9}$$

The inverse FT of this equation gives the time-domain solution:

$$\begin{aligned}p_d(t)&=2t_-*[r_a(q)-r_a(2q)]*t_+*q-r_+*p_d\\&\quad+r_-*[r_a(q)+r_a(2q)-r_-*r_a(q)*r_a(2q)]\\&\quad*(\delta+r_+)*p_d.\end{aligned} \tag{10}$$

The inverse FT of $R_a(Q)$ is notated as $r_a(q)$. This relation for the observable $p_d(t)$ is analogous to that for the total OAE in Eq. (5). The first term on the right-hand side is directly proportional to the stimulus $q(t)$ but also dependent on the nonlinearity in the apical reflectance. The remaining terms on the right-hand side depend on the output variable $p_d(t')$ at earlier times $t'<t$. When $q(t)$ is first turned on at $t=0$, the initial value satisfies $p_d(0)=0$. A limitation of the above is that Eqs. (5) and (10) have been derived from of the frequency-domain equations using the inverse FT, and this operator only exists for a linear system. These equations might equally well have been derived directly in the time domain, thus removing the requirement that the inverse FT exist. Such an approach is preferable, because the nonlinear forms of Eqs. (5) and (10) exist as long as the nonlinear function $r_a(q)$ exists, even when the inverse FT is undefined. An advantage of the present derivation is that it proceeds from the more familiar frequency-domain form of the model.

In the absence of multiple internal reflections, i.e., with $r_-=0$, the OAE residual simplifies to

$$\begin{aligned}p_d(t)&=2t_-*[r_a(q)-r_a(2q)]*t_+*q-r_+*p_d\\&\approx 2\left(\frac{T_-T_+}{1+R_+}\right)[r_a(q)-r_a(2q)]*q.\end{aligned} \tag{11}$$

The latter approximate result is accurate if the middle-ear function properties are assumed to be slowly varying, i.e., $t_-(t)\approx T_-\delta(t)$, $t_+(t)\approx T_+\delta(t)$, and $r_+(t)\approx R_+\delta(t-\varepsilon)$. The time offset $\varepsilon>0$ is introduced in the last relation to satisfy the causal reflectance property $r_+(t)=0$ for $t\leqslant 0$, but ε is otherwise extremely small. In practical ear-canal reflectance measurements, there is a small time delay ε corresponding to the acoustic round-trip travel time between the microphone and the eardrum. The frequency-domain complement to the top expression in Eq. (11) is

$$P_d=2\left(\frac{T_-T_+}{1+R_+}\right)[R_a(Q)-R_a(2Q)]Q, \tag{12}$$

in which the middle-ear parameters are arbitrary functions of frequency.

Such a recursive formulation as that of Eqs. (5) and (10) has proven useful for time-domain oscillatory models of sound generation in the vocal tract (Sondhi and Resnick, 1983), and sound generation by musical instruments (McIntyre *et al.*, 1983). This formulation may be useful for studying nonlinear models of OAE generation in the time domain.

With a nonzero basal reflectance, the output OAE waveform in Eq. (5) has a term proportional to a convolution of itself with the round-trip reflectance within the cochlea. Because convolution with the apical reflectance produces time delay (i.e., the round-trip latency) and a change in magnitude, the output represents the first-order apical reflection with an infinite cascade of internal reflections. The direct formulation of the time-domain model as an infinite cascade, which arises from series expansions of the terms in the denominator of Eq. (2) followed by application of the inverse FT, is

$$p_{oae}(t)=t_{-}*r_a*t_{+}*\left(\sum_{n=0}^{\infty}(-1)^n r_{+}^{(n)}\right)*q$$
$$*\left(\delta+\sum_{m=1}^{\infty}(r_{-}*r_a)^{(m)}\right). \quad (13)$$

In the above, the variable $r^{(n)}$ for arbitrary r denotes the convolution product of n instances of r with itself (e.g., $r^{(3)}=r*r*r$), and with $r^{(0)}=\delta$ for $n=0$. The sum over the index n represents multiple reflections from the eardrum, while the mth term in the sum over the index m corresponds to the mth multiple internal reflection. Equations (5) and (13) are formally equivalent. Practical calculations with Eq. (13) involve more computation and a truncation of the power series, but Eq. (13) makes analytically explicit the contribution to the recorded OAE of the individual multiple internal reflections. With the additional assumption that the middle-ear parameters are constant over the bandwidth of the stimulus, Eq. (13) takes the simpler form (Zweig and Shera, 1995):

$$p_{oae}(t)=\left(\frac{T_{-}T_{+}}{1+R_{+}}\right)r_a*q*\left(\delta+\sum_{m=1}^{\infty}R_{-}^{m}r_a^{(m)}\right). \quad (14)$$

The nature of this assumption restricts the use of Eq. (14) to relatively narrow-band OAEs, whereas Eq. (5) remains accurate (within the framework of the linear coherent reflection emission theory).

By a similar power series expansion of the denominators in the terms of Eq. (8) followed by application of the inverse FT, the distortion waveform may be directly calculated as

$$p_d(t)=2t_{-}*\left(\sum_{n=0}^{\infty}(-1)^n r_{+}^{(n)}\right)*\left([r_a(q)-r_a(2q)]\right.$$
$$\left.+\sum_{m=1}^{\infty}r_{-}^{(m)}*[r_a^{(m+1)}(q)-r_a^{(m+1)}(2q)]\right)*t_{+}*q. \quad (15)$$

This expression for the distortion waveform is equivalent to Eq. (10). There is a tradeoff between the recursive and direct expressions for p_d that is similar to the tradeoff between the analogous expressions for p_{oae}. This tradeoff is between ease of calculation in the recursive model and ease of interpretation in the direct model of the individual multiple internal reflections.

A two-source model for DPOAE generation including multiple internal reflections is mathematically expressed as follows for the case (as in the present experiments) that the tonotopic DP place is apical to the f_2 place (Shera and Zweig, 1992; Talmadge *et al.*, 1998, 1999):

$$P_{DP}(f)=\frac{P_l+R_aP_r}{1-R_{-}R_a}, \quad (16)$$

in which $P_{DP}(f)$ is the total DPOAE pressure spectrum in the ear canal at the DP frequency f, P_l is the DPOAE generated near the f_2 place that propagates (to the ''left'') to the basal end of the cochlea and into the ear canal, and P_r is the DPOAE generated near the f_2 place that propagates (to the ''right'') to the tonotopic DP place where it is reflected back to the ear canal via the linear coherent reflection emission mechanism in terms of the apical reflectance R_a introduced earlier. The functions P_l and P_r are functions of the middle-ear scattering primaries at the primary frequencies and of R_{+} at the DP frequency. Relative to the DPOAE component directly generated at f_2 place, Talmadge *et al.* (1998) predict that the secondary DPOAE component that is reflected from its tonotopic place can be either smaller or larger in level ($|R_aP_r|<|P_l|$ or $|R_aP_r|\geqslant|P_l|$), and experimental results confirming this prediction have been reported (Talmadge *et al.*, 1999; Dhar *et al.*, 2002). After multiplication through by the term in the denominator and using the inverse FT, the recursive time-domain form of Eq. (16) is

$$p_{DP}(t)=p_l+r_a*p_r+(r_{-}*r_a)*p_{DP}. \quad (17)$$

The direct time-domain form is (Talmadge *et al.*, 1998)

$$p_{DP}(t)=[p_l+r_a*p_r]*\left[\delta+\sum_{m=1}^{\infty}(r_{-}*r_a)^{(m)}\right]. \quad (18)$$

The functional form of the multiple internal reflections in DPOAE generation is the same as for the total SFOAE [Eq. (13)] but differs from that of the SFOAE observable [Eq. (15)]. The recursive formulation may be useful in models to calculate DPOAEs that include a nonzero basal reflection, but which otherwise truncate the infinite sum of the direct formulation. For example, this summation has sometimes been truncated in analyses retaining only the $m=1$ term, e.g., in Talmadge *et al.* (1998, 2000), although these authors have been careful to restrict the generalizability of their results to the limit that the total-internal-reflection filter magnitude is sufficiently small [$|R_{-}(f)R_a(f)|\ll 1$]. The recursive form is well suited to time-domain calculations including nonlinearity in R_a without such a limitation on the magnitudes of the basal and apical reflectance. Subsequent frequency-domain results for all time-domain models are obtained using the FT of the solution after it has attained a steady state.

The time delay Δt between adjacent reflections is predicted by the above relations to be related to the fine structure spacing by (Zweig and Shera, 1995)

$$\Delta t=1/\Delta f. \quad (19)$$

For a given stimulus waveform, the SFOAE output in the time domain is predicted to be a sequence of delayed and attenuated replicas of the stimulus waveform, except that each ''replica'' becomes more widely dispersed in time with increasing reflection order through processes similar to those described for DPOAEs by Tubis *et al.* (2000a). The SFOAE output in the frequency domain is predicted to be the spectrum of the stimulus with additional fine structure spacing produced by the presence of the sequence of delayed components.

The ratio ρ_{oae} of the total SFOAE or DPOAE output amplitudes of successive multiple internal reflections is calculated from the FT of Eq. (13) or (18), respectively, as

$$\rho_{oae}=|R_{-}||R_a|. \quad (20)$$

This amplitude ratio may thus be used to estimate the round trip cochlear reflectance. Inspection of the same equation

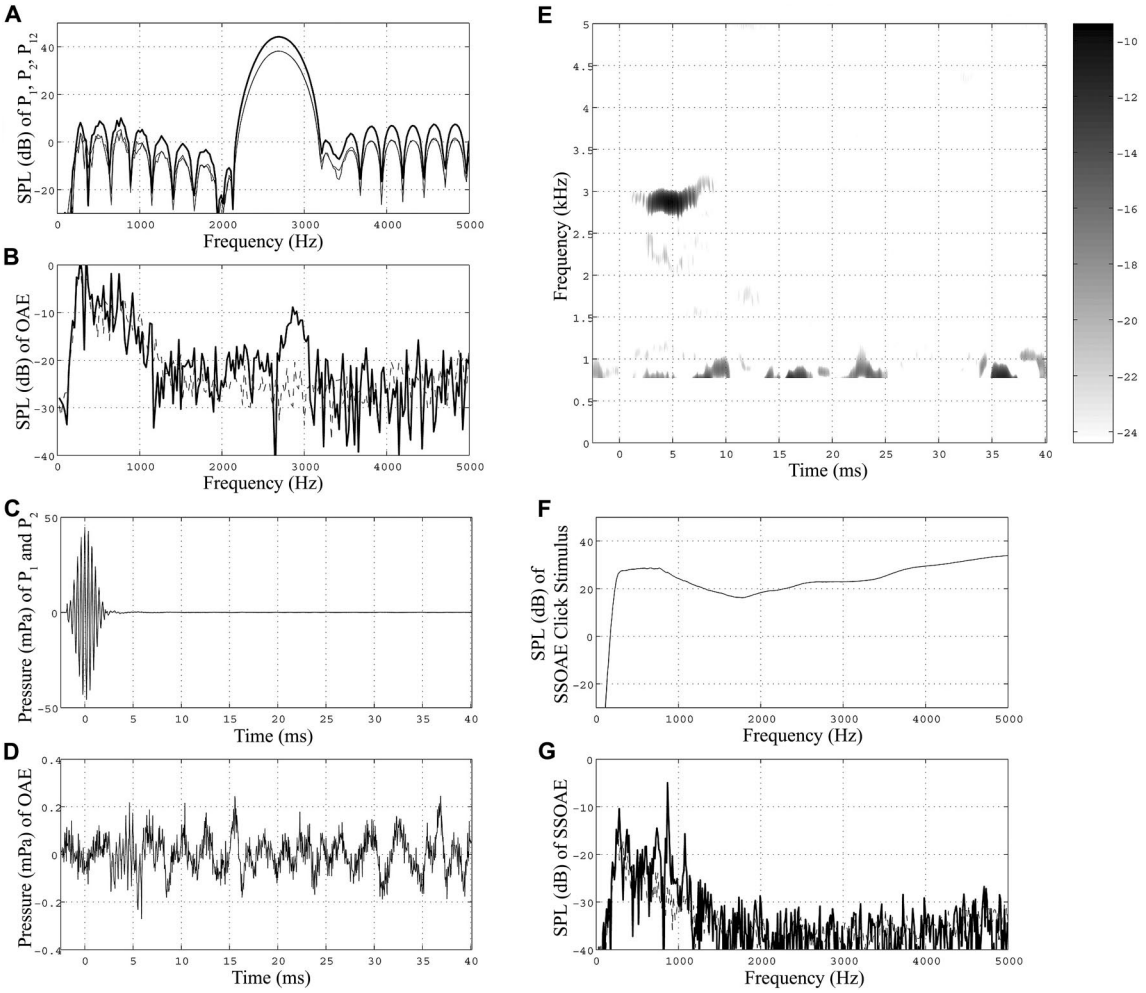


FIG. 1. Pip-evoked SFOAE in an ear with no SSOAEs within the stimulus pass-band. $f_1, f_2 = 2.7$ kHz. $L_1, L_2 = 50$ dB SPL. (a) The total pressure spectra for p_1, p_2, and p_{12}. (b) The SFOAE spectrum. (c) The corresponding waveforms for p_1, p_2, and p_{12}. (d) The 2E SFOAE waveform. (e) The TFR of the 2E waveform (results plotted from 0.8 to 5.0 kHz). (f) Click stimulus spectrum used to elicit SSOAEs. (g) The SSOAE spectrum.

shows that ρ_{oae} is also equal to the ratio of the amplitude of the first internal reflection to that of the directly generated OAE.

The distortion waveform in Eq. (15), which is estimated in the experimental results via the TFR, has a qualitatively similar form, except that the amplitude ratio ρ_m of the term corresponding to m internal reflections to that with $m-1$ internal reflections is

$$\rho_m = |R_-| \left| \frac{R_a^{m+1}(Q) - R_a^{m+1}(2Q)}{R_a^m(Q) - R_a^m(2Q)} \right|. \tag{21}$$

This amplitude ratio varies with the number of internal reflections, unlike that of Eq. (20). The amplitude ratio ρ_1 of the first internal reflection to the directly propagated OAE is

$$\begin{aligned} \rho_1 &= |R_-||R_a(Q) + R_a(2Q)| \\ &\approx |R_-|(|R_a(Q)| + |R_a(2Q)|), \end{aligned} \tag{22}$$

with the latter expression incorporating the approximation that the nonlinear shift in apical reflectance phase is small. The amplitude ratio ρ_2 of the second to the first internal reflection is

$$\rho_2 = |R_-| \left| \frac{R_a(Q)^3 - R_a(2Q)^3}{R_a(Q)^2 - R_a(2Q)^2} \right| \tag{23}$$

It is obscured in these amplitude ratios, but apparent in Eq. (15), that for later reflections (after turning off the stimulus) such that the apical reflectance has returned to its linear-growth limit, the apical reflectance should satisfy $R_a(Q) = R_a(2Q)$. In this case, no further internal reflections should be observed in the SFOAE residual.

The amplitude ratios measured for transient SFOAE observable responses are predicted to differ from those measured for transient DPOAEs. The correspondence between measured amplitude ratios and round-trip cochlear reflectance may be more easily interpreted from the transient DPOAE than from transient SFOAE responses because the number of internal reflections within the measurement window duration is predicted to be less for transient SFOAE observable responses compared to transient DPOAE observable responses. However, the transient SFOAE responses may provide an estimate of the point at which the apical reflectance returns to its low-level, linear-growth regime.

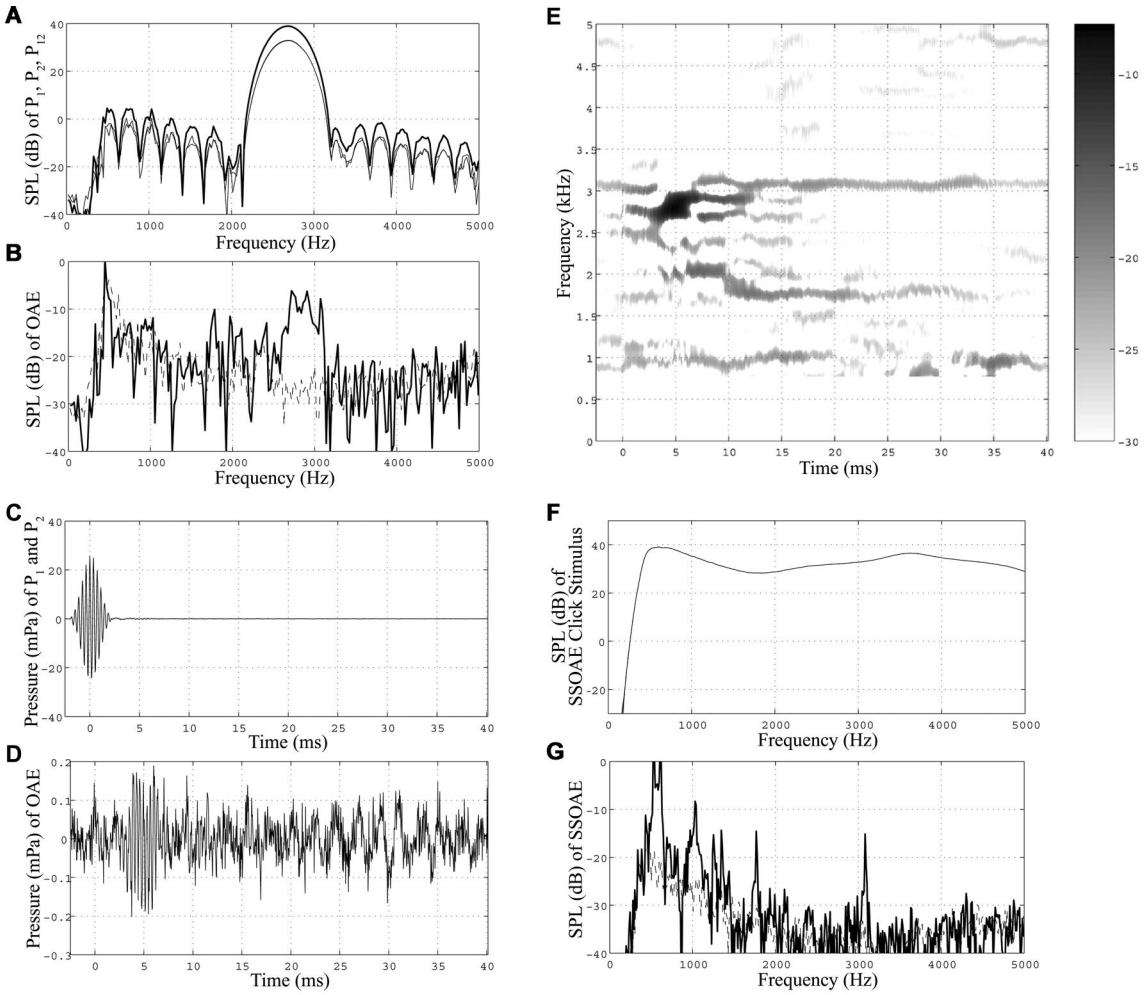


FIG. 2. Pip-evoked SFOAE in an ear with SSOAEs within the pip-stimulus pass-band, and multiple internal reflections. $f_1, f_2 = 2.7$ kHz. $L_1, L_2 = 45$ dB SPL. Format is the same as for Fig. 1.

IV. RESULTS

The experimental measurements are described in terms of the stimulus class evoking each response.

A. Pip-evoked SFOAEs

Figure 1 shows a representative pip-evoked SFOAE obtained in an ear with no SSOAEs within the stimulus pass-band. [Contrast the SSOAE spectrum in Fig. 1(g) with the stimulus pass band in Fig. 1(a)]. The SFOAE spectrum [Fig. 1(b)] somewhat resembled the predicted attenuated and delayed 2.7-kHz pip, except that the lower frequencies in the stimulus spectrum were reduced or absent in the SFOAE [compare responses at 2.1–2.6 kHz in Figs. 1(b) and (a)]. This reduced OAE level in the lower frequencies of the stimulus bandwidth was a common feature of tone-pip-evoked SFOAEs in this study. The total ear canal pressure waveforms (p_1 and p_2) are shown in Fig. 1(c) with maximal levels centered at 0 ms. The corresponding 2E waveform is illustrated in Fig. 1(d), but is difficult to interpret. The TFR of this waveform in Fig. 1(e) shows a SFOAE that is centered at 2.85 kHz in the frequency dimension and approximately 5 ms in the time dimension. This latency is similar to the mean SFOAE group delay (5.6 ms for 2.9 kHz) measured using a continuous-stimulus technique (Shera *et al.*, 2002). The spectrum of the click stimulus that was used to evoke the SSOAE is illustrated in Fig. 1(f).

Figure 2 shows a pip-evoked SFOAE in another subject. In this case, the ear has SSOAEs within the pip-stimulus pass-band and fine structure is present in the SFOAE spectrum as a ripple in the spectral peak region from 2.6 to 3.0 kHz. Such a ripple in the spectral envelope was much less apparent in Fig. 1(b). The SFOAE spectrum contains three strong maxima within the main band of the response (2.7, 2.9, and 3.1 kHz), and other narrow-band components (at 1.7, 1.9, 2.1, and 2.4 kHz), some of which are, however, lower in frequency than the lower edge of the stimulus pass-band. At least two peaks in the SFOAE spectrum (1.7 and 3.1 kHz) correspond to strong narrow-band signals in the SSOAE spectrum and are interpreted as cochlear ''hot spots.'' The absence of strong SSOAEs at other peak frequencies within the SFOAE spectrum suggests that such peaks are due to some other factor.

In order to reveal low-level OAE components and better separate closely spaced frequency components, the lowest level depicted by the TFR was set to −30 dB SPL and the

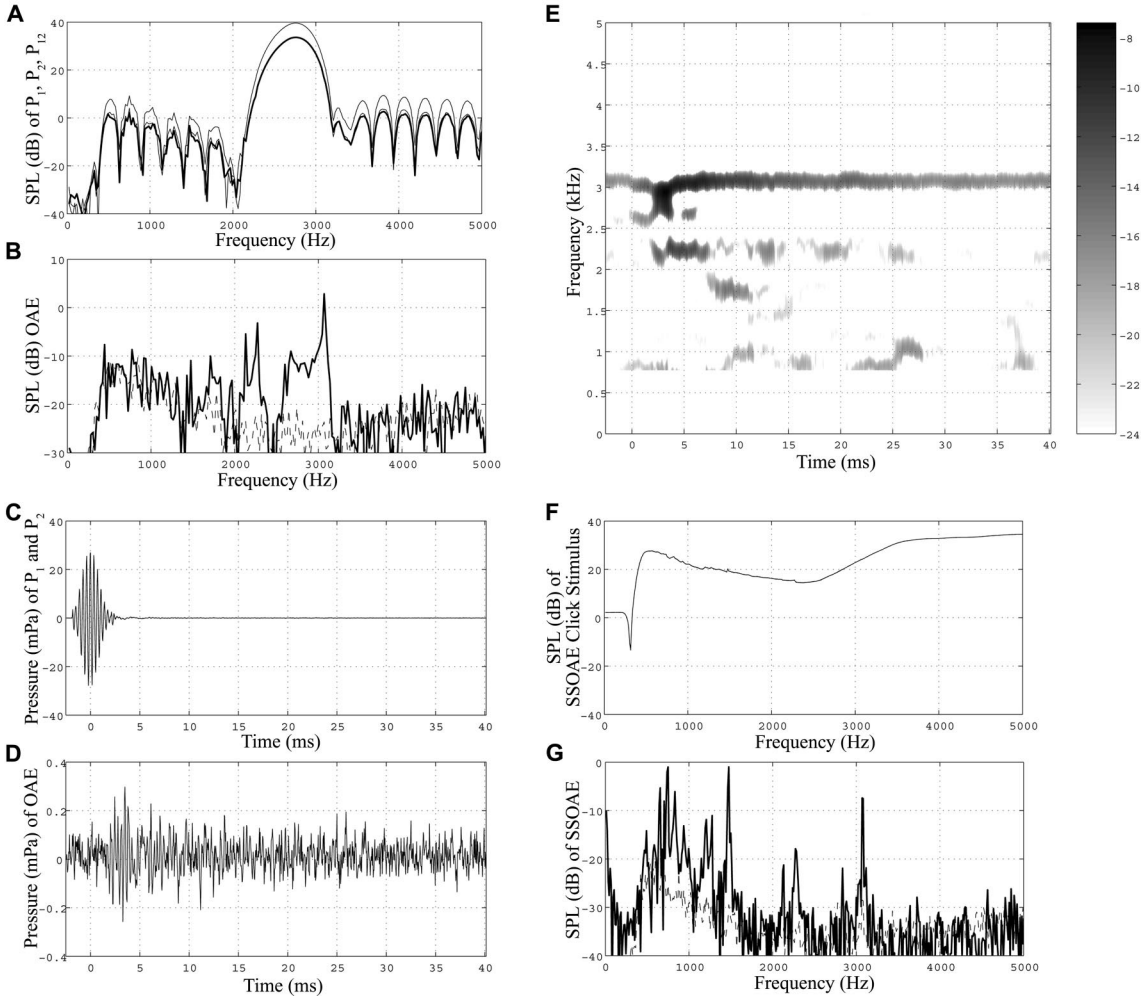


FIG. 3. Pip-evoked SFOAE in an ear with multiple SSOAEs. $f_1, f_2 = 2.7$ kHz. $L_1, L_2 = 45$ dB SPL. Format is the same as for Figs. 1 and 2.

TFR filter length was increased from 120 to 180 ms (see the Appendix for discussion of the effects of filter length on the TFR). This representation reveals multiple short OAE bursts present in the TFR in separate frequency bins. At approximately 2.7 kHz, for example, an OAE burst began near the stimulus onset, a second segment began at about 6–7 ms, and a third at 11–13 ms. The time interval separating the center of energy of these OAE bursts was approximately 6 ms, and $\Delta f = 167\,\text{Hz} = 1/(6\,\text{ms})$. Thus, the 170-Hz ripple in the SFOAE spectrum in Fig. 2(b) is accurately predicted by Eq. (19) as a spectral effect of multiple internal reflections on the OAE fine structure.

Figure 3 shows a 2.7-kHz pip-evoked SFOAE in a third subject's ear. This ear produced multiple SSOAEs within the stimulus pass-band. The SFOAE spectrum [Fig. 3(b)] is illustrative of responses obtained in ears with strong SSOAEs in that it did not resemble the spectral shape of the evoking pips. Broadband spectral components were present from 2.6 to 3.0 kHz and narrow-band components were present at frequencies within (at 1.7, 2.1, 2.3, and 3.1 kHz) or outside of the stimulus spectrum. Narrow-band components above 2 kHz were aligned with SSOAEs [compare Figs. 3(b) and (g)], and are concluded to have a common origin. Thus, the SFOAE fine structure in this ear is partially explained by the hot spots associated with multiple SSOAE components.

The broadband component of the SFOAE spectrum is characterized by a ''single'' high-level response near 3.5 ms in the TFR [Fig. 3(e)]. As expected based on the high-level SSOAE near 3.1 kHz, the 3.1-kHz component is nearly continuous in the TFR but with variations in level. There is a frequency pulling of the 3.1-kHz SSOAE towards the 2.7-kHz pip response near 3.5 ms. The 2.2-kHz component had four brief segments; the last three were spaced 6.3 ms apart and are interpreted as higher-order reflections. This pattern was observed to be robust with respect to the TFR calculations: these multiple reflections were also present when the TFR filter length was changed in order to better separate closely spaced frequency components (not shown).

B. Gated SFOAEs and DPOAEs

This section describes results obtained using gated sine tones of 22-ms duration combined with either a second gated sine tone or a continuous tone. Figure 4 shows a gated-tone-evoked SFOAE in the same subject's ear introduced in Fig. 3. The SFOAE spectrum is a series of narrow-band components and strongly differs from the spectrum of the eliciting stimulus. The major spectral peaks in the 2.1- to 3.1-kHz range line up with the fine structure observed in the pip-evoked SFOAE for this ear [compare Fig. 4(b) to Fig. 3(b)].

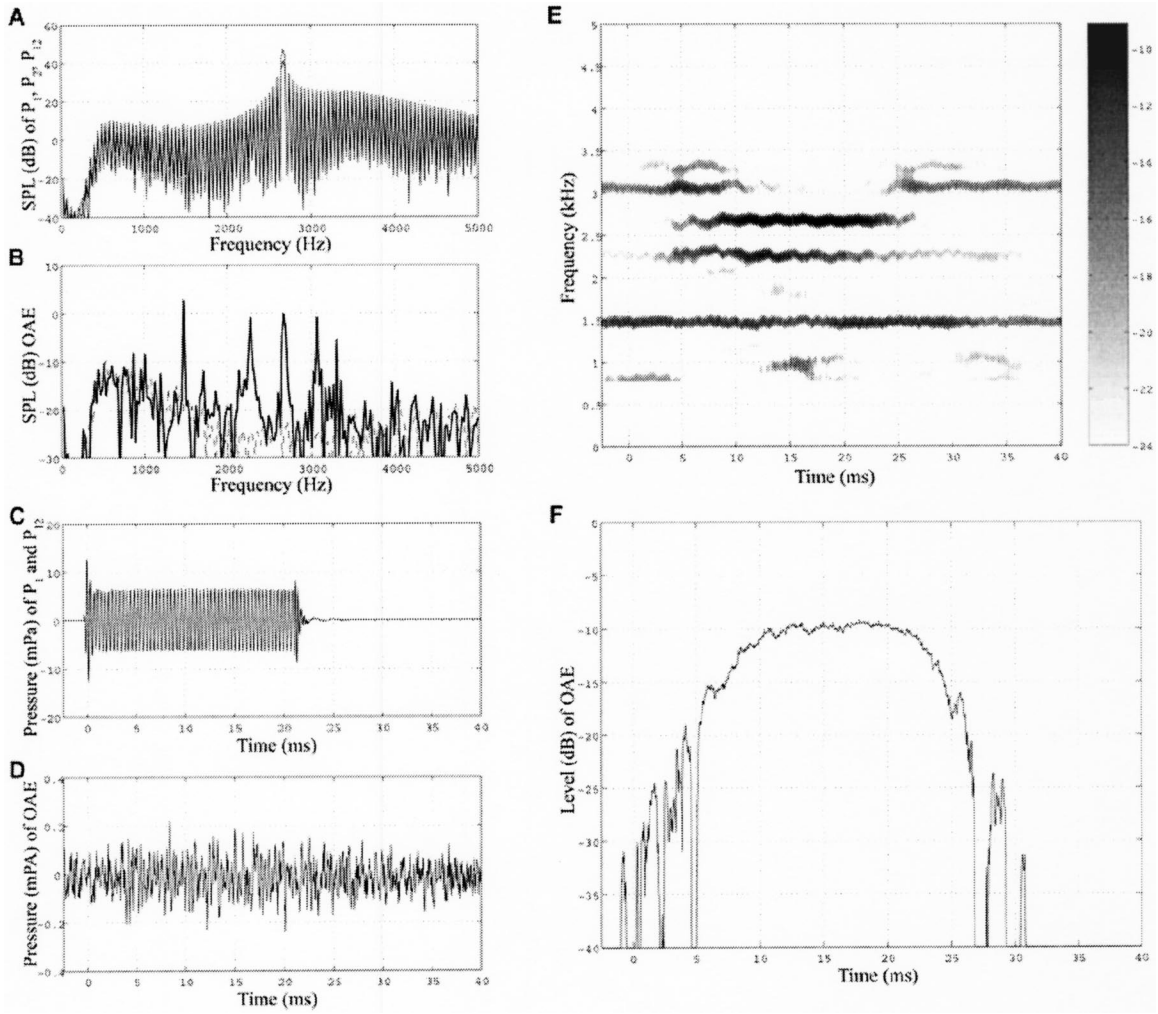


FIG. 4. Gated-tone-evoked SFOAE in an ear with numerous SSOAEs (same ear as Fig. 3). f_1,,$f_2=2.7$ kHz. L_1,$L_2=45$ dB SPL. (a) The total pressure spectra for p_1, p_2, and p_{12}. (b) The SFOAE spectrum. (c) The corresponding waveforms for p_1, p_2, and p_{12}. (d) The 2E SFOAE waveform. (e) The TFR of the 2E waveform. (f) Slice from the TFR at 2.7 kHz.

As predicted, the bandwidths of the stimulus-related peaks at 2.7 kHz are similar for the stimulus and SFOAE spectra [Figs. 4(a) and (b)]. Except for the SFOAE peak at 3.4 kHz, the other major peaks in the gated SFOAE response correspond to SSOAE sites [compare Fig. 4(b) to Fig. 3(g)].

Interpreting SFOAE temporal characteristics using the 2E SFOAE waveform in Fig. 4(d) is difficult and is aided by the TFR analysis, although this and other 2E waveforms are mainly deterministic with small influence from noise. The TFR in Fig. 4(e) shows components in common with the SSOAEs at 1.5, 2.25, and 3.1 kHz, and with the elicited SFOAE at 2.7 kHz. When the 2.7-kHz SFOAE is high in level, the SSOAE level at 3.1 kHz is reduced, and that at 2.2 kHz is increased. Thus, the SFOAE suppressed the nearby higher-frequency SSOAE component, and enhanced the nearby lower-frequency SSOAE component. Figure 4(f) shows a slice through the TFR taken at 2.7 kHz. For the gated responses, the OAE latency is defined as the latency of the increasing TFR envelope response that is 3 dB down from the amplitude at the center of the response envelope, and is measured with respect to the corresponding −3-dB latency of the stimulus. OAE onset latencies were calculated with respect to stimulus onset, and OAE offset latencies were independently calculated with respect to stimulus offset. The SFOAE onset latency was 6.2 ms, and its duration was similar to that of the stimulus. The spectral peak at 3.4 kHz is exhibited in the TFR by responses at the onset and offset of the evoking gated stimulus, and is likely due to transient wideband splatter associated with the onset and offset ramps.

The next set of results (Fig. 5) illustrates the OAEs produced by a combination of a continuous tone ($f_2=4$ kHz) and a gated tone ($f_1=3.3$ kHz), i.e., responses measured in the continuous f_2/gated f_1 DPOAE paradigm. Results were measured for the same ear as for the responses shown in Figs. 3 and 4. Both tones were presented at a moderate level ($L_2=45$, $L_1=L_2+10$ dB). The $2f_1-f_2$ DPOAE at 2.7 kHz is similar in level to the 2.7-kHz gated SFOAE shown previously for this ear [compare response maxima near −10 dB in Figs. 5(f) and 4(f)]. In addition, the overall spectrum of this DPOAE response resembles the spectrum of the gated SFOAE for this ear [compare Fig. 5(b) to Fig. 4(b)]. Both contain strong narrow-band components at frequencies that correspond to SSOAE sites [compare Figs. 5(b) and 4(b) with the SSOAE spectrum in Fig. 3(g)].

The TFR shown in Fig. 5(e) has two main frequency components. One corresponds to the SFOAE at the f_1 fre-

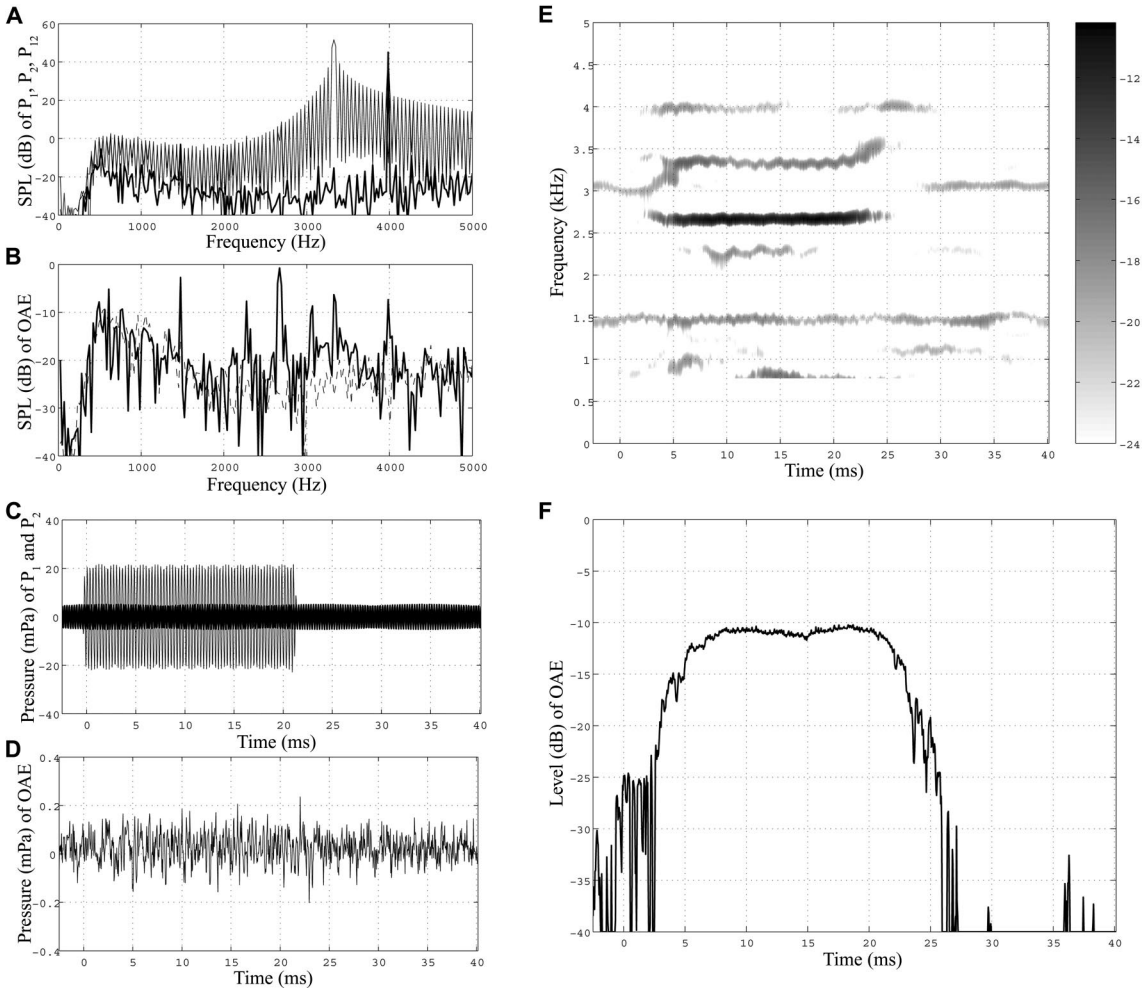


FIG. 5. Continuous f_2/gated f_1 DPOAE with $L_2=45$ and $L_1=L_2+10$ dB. $f_1=3.34$ kHz, $f_2=4$ kHz. Data are from the same ear shown in Figs. 3 and 4. Format is the same as for Fig. 4.

quency (3.3 kHz) and the other to the DPOAE frequency (2.7 kHz). SSOAEs are present at 1.5 and 3.1 kHz, and for a restricted duration at 2.25 kHz. SFOAE energy is present at the f_2 of 4 kHz only near the stimulus onset and offset. The 3.1-kHz SSOAE was suppressed during the gated f_1 tone, and its SFOAE trajectory is evident at 3.3 kHz. Frequency glides at f_1 onset and offset are present, although perhaps limited in resolution by the TFR properties. These results demonstrate that ''hot spots'' associated with SSOAE sites contributed to the DPOAE and SFOAE fine structure in this ear, and there were complex spectral-temporal interactions between SSOAE and stimulus-evoked response components.

Because the f_2 tone at 4 kHz was on continuously, intermodulation distortion was initiated when the onset response to the gated-f_1 stimulus reached the f_2 tonotopic place (or slightly basal to it). Thus, the DPOAE onset in the 2-D TFR trajectory is earlier than the onset of the gated 2.7-kHz SFOAE with a similar response level [compare Fig. 5(f) to Fig. 4(f)]. The DPOAE onset latency (3 dB down from the amplitude at the center of the response envelope) taken relative to the stimulus waveform is 3.9 ms. As described in Sec. V, this latency is consistent with model predictions and measurements.

Figure 6 shows gated tone-evoked SFOAE results for the subject whose tone pip-evoked SFOAE responses were shown in Fig. 2; this ear produced numerous SSOAEs [Fig. 2(g)]. The SFOAE spectrum resembles the pip-evoked SFOAE in that the frequencies corresponding to spectral peaks were similar in both recordings [compare Fig. 6(b) to Fig. 2(b)]. For the gated tone, however, the level of the OAE spectral peak near 2.7 kHz was at least 15 dB greater than the levels of other spectral components.

The TFR in Fig. 6(e) shows the SFOAE at 2.7 kHz and a set of OAEs above and below 2.7 kHz clustered at the SFOAE onset and offset; the excitation of this set of OAEs is attributed to the gating of the stimulus. The TFR latencies decreased with increasing frequency, and were similar to average SFOAE group-delays at corresponding frequencies (e.g., 5.5-ms latency at 3 kHz) reported by Shera *et al.* (2002). An OAE was present near 2.0 kHz, as in the pip-evoked response for this ear [compare Fig. 6(e) to Fig. 2(e)], which was not present in the SSOAE [Fig. 2(g)]. Its latency was similar to the mean group delay (7.1 ms) obtained for 2.0-kHz SFOAEs (Shera *et al.*, 2002).

Figure 6(f) shows a slice through the TFR taken at 2.7 kHz, and shows a greater dynamic range than that in the TFR contour plot. The 2.7-kHz SFOAE increased more slowly at onset than its decay rate at offset. This resulted in an onset

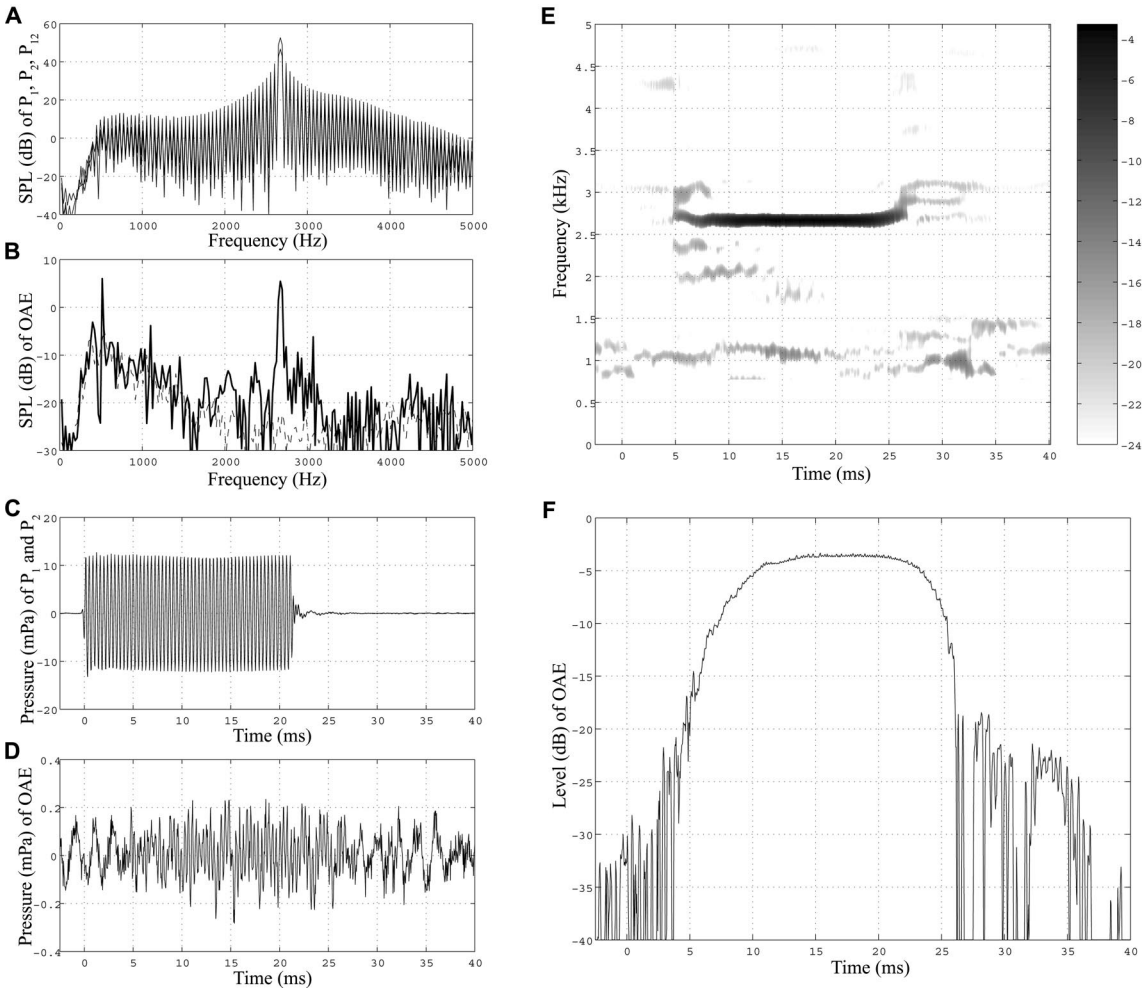


FIG. 6. Gated-tone-evoked SFOAE in an ear with numerous SSOAEs (same ear as Fig. 2). $f_1, f_2 = 2.7$ kHz. $L_1, L_2 = 50$ dB SPL. Format is the same as for Fig. 4.

latency (7.8 ms) that was longer than its offset latency (4.5 ms). There were two delayed ''humps'' in the TFR slice following the stimulus offset, one at 27 ms, and the other at about 33 ms. These were interpreted as multiple internal reflections, as they appeared following the offset of the gated tone, but were not present before the corresponding onset. Their round-trip delays (~6 ms) are consistent with other measures of round-trip latency; e.g., their values are intermediate between the onset and offset latencies of the direct gated response of the OAE.

Figure 7 shows results for a continuous f_2/gated f_1 DPOAE recorded in the same ear as for the results shown in Figs. 2 and 6. The DPOAE stimulus conditions used were $f_1 = 3.3$ kHz, $f_2 = 4$ kHz, $L_2 = 55$ dB SPL, and $L_1 = L_2 + 10$ dB. The DPOAE spectrum and TFR for this ear have two main components, one at 2.7 kHz (the $2f_1 - f_2$ DP) and another at 4 kHz (the f_2 primary's SFOAE response). The 2E waveform shows the presence of a gated OAE response, but not its decomposition into multiple spectral components. Its TFR shows that some energy was present near the onset and offset of the 2.7-kHz DPOAE, presumably due to gating of the f_1 stimulus. The duration of the SFOAE response at $f_2 = 4$ kHz was similar to that of the gated f_1 stimulus, even though the f_2 tone was continuously presented. This is evidence for some SFOAE suppression at the f_2 frequency due to the presence of a ''suppressor'' at the f_1 frequency. Another brief component was present at the f_1 frequency (3.3 kHz) with a delay of about 4 ms. Its amplitude was reduced to the noise floor by about 7 ms, possibly due to a transient suppression effect. This is consistent with the observation that higher frequency tones do suppress lower frequency SFOAEs (Kemp *et al.*, 1990). The spectral energy observed between 2.8 and 3.2 kHz in Fig. 7(b) is observed in the TFR to be predominantly an OAE response near f_1 onset. The DPOAE response in Fig. 7(f) has a notch in the temporal envelope at 23 ms and a set of multiple internal reflections that are further described in Sec. V.

Figures 6(f) and 7(f) show SFOAE and DPOAE responses at a similar maximum TFR level and frequency in the same ear. The two responses were presumably generated by disparate sources/mechanisms. Comparing the temporal characteristics of such responses may provide a means to test models of OAE generation. To that end, a similar but more extensive comparison of SFOAE and DPOAE responses across different stimulus levels is presented in Fig. 8 for the same ear. Each panel in Fig. 8 depicts 2-D slices taken from TFRs obtained for a SFOAE (thin line type) and DPOAE (heavy line type) at 2.7 kHz. For both responses within a

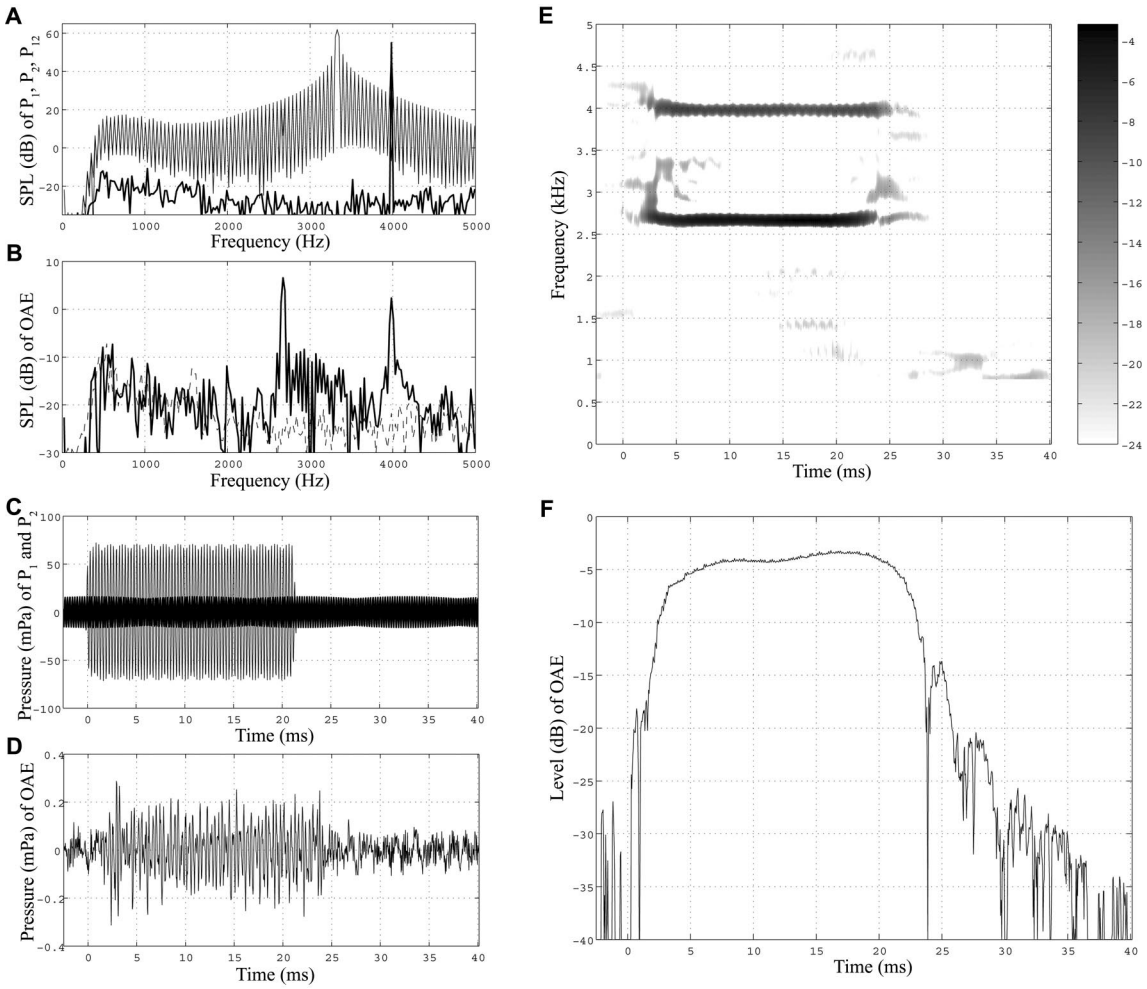


FIG. 7. Continuous f_2/gated f_1 DPOAE with $f_1=3.34$ kHz, $f_2=4$ kHz, $L_2=55$, and $L_1=L_2+10$ dB (same ear as in Figs. 2 and 6). Format is the same as for Fig. 4.

panel, the level of the main lobe of the response envelope is similar. The SFOAEs in Fig. 8 were measured using a pair of gated tones as shown previously [i.e., the SFOAE in Fig. 8(a) for L_1, $L_2=50$ dB SPL is re-plotted from Fig. 6(f)]. The DPOAEs in Fig. 8 were measured using the gated f_1/gated f_2 paradigm. The DPOAE in Fig. 8(a) is comparable in many respects to the DPOAE measured using the gated f_1/continuous f_2 DPOAE paradigm shown in Fig. 7(f), although the respective onset latencies differ. Across subjects, the DPOAEs measured in the same ear using these two paradigms were similar with respect to the presence and total number of prominent notches and higher order reflections at comparable stimulus levels.

Comparing SFOAE plots across panels in Fig. 8 illustrates the effect of stimulus level on SFOAEs for this ear. As expected based on early transient OAE studies (e.g., Kemp, 1978; Neely *et al.*, 1988), SFOAE latencies increased slightly with decreasing stimulus intensity. At low stimulus levels, there was a notch at 8 ms in the SFOAE response [see Figs. 8(c) and (d)]. This notch may involve wave cancellation caused by interaction between the directly propagated stimulus and the first multiple internal reflection. Such notches in the time-domain envelope of OAE responses may be analogous to an OAE input/output function that happens to pass through a minimum of fine structure at a particular level. Across ears, SFOAEs were more likely to show evidence for higher-order reflections at lower stimulus levels [Figs. 8(c) and (d)] than moderate levels [Figs. 8(a) and (b)]. These trends are consistent with the magnitude of apical reflectance increasing with decreasing stimulus level. To the extent that round trip reflections produce OAE fine structure, these data represent one mechanism whereby fine structure may be increased with decreasing stimulus level.

Comparing DPOAE plots across panels in Fig. 8 shows a slight increase in DPOAE latency with decreasing stimulus level for this ear. At the highest stimulus level [Fig. 8(a)], the DPOAE onsets were more rapid than the SFOAE onsets. A notch was present at 23 ms near the offset of the DPOAE response shown in Fig. 8(a), and was even more prominent in Fig. 7(f). The DPOAE notch location shifted from near the offset for moderate-level stimuli to a less deep notch just after the onset [4 ms in Fig. 8(b)] for lower-level stimuli [Figs. 8(b) and (c)]. The relation of such DPOAE notch properties is further described in Sec. V.

Figure 9 shows changes in the SFOAE spectrum as a function of stimulus level for the same ear as those data shown in Figs. 3–5. For comparison, responses to tone pips are plotted in the left column [Figs. 9(a)–(d)] and responses

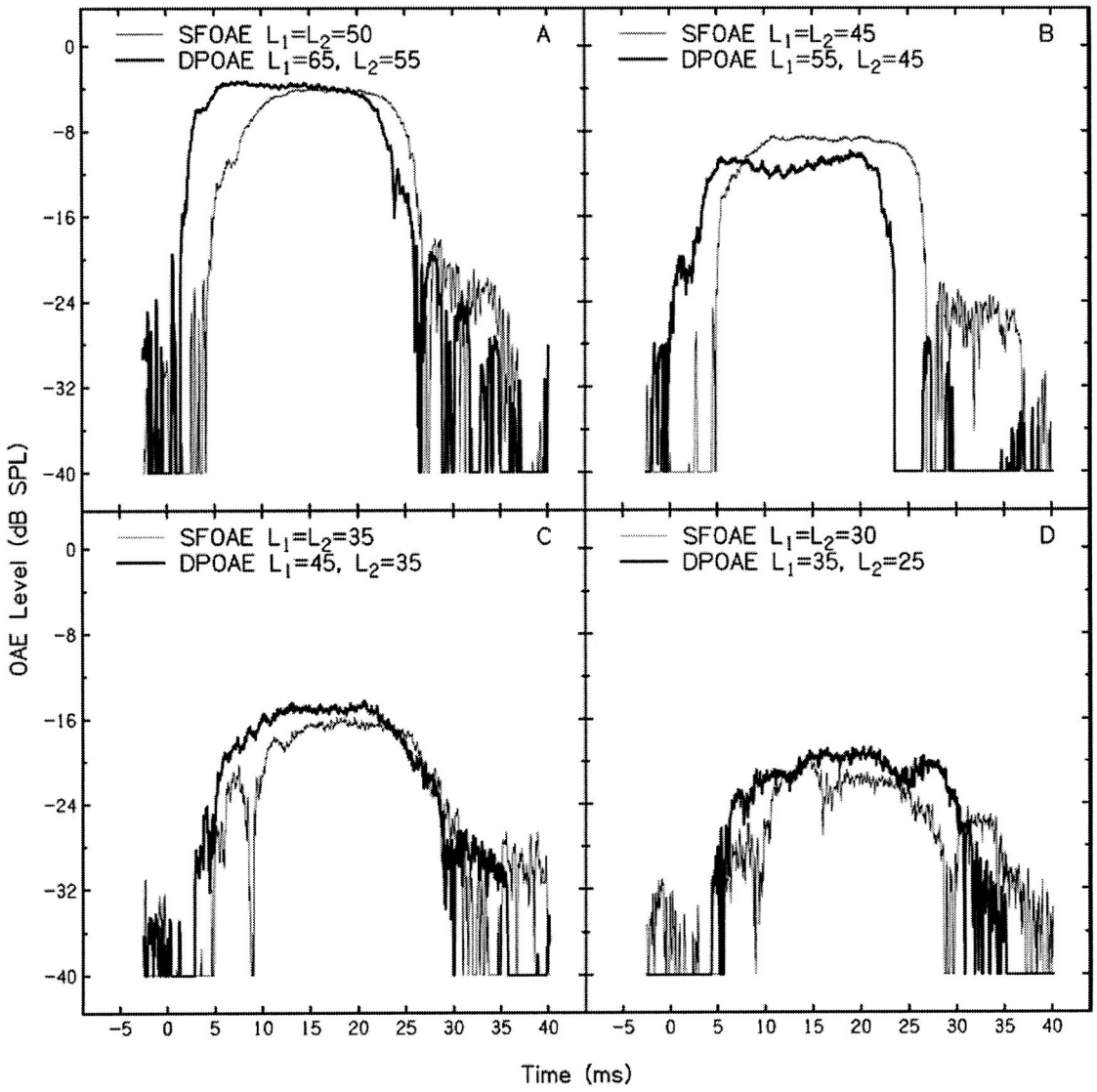


FIG. 8. Series of TFR slices at 2.7 kHz taken from measurements of DPOAEs and SFOAEs obtained in the same ear for a range of stimulus sound pressure levels, using stimuli that descend in level from panels (a)–(d) (as indicated in the legend of each panel). The gated f_2/gated f_1 DPOAE trajectory is shown in a thick line, and the gated SFOAE responses are shown in a thin line on each panel.

to gated tones are shown on the right [Figs. 9(e)–(h)]. The SSOAE spectrum for this ear is presented in Fig. 3(g). The progression down each column shows that for both tone pips and gated tones, the OAE spectrum is less likely to resemble the spectrum of the eliciting stimulus at low stimulus levels. In addition, the relative level of the response component at the nominal stimulus frequency (2.7 kHz) is low at low stimulus levels compared to other narrow-band components within the same recording. For example, compare the levels of the 2.7-kHz component in the responses shown in Figs. 9(d) and (h) with the levels of components at 2.3 and 3.1 kHz, which are SSOAE sites in this ear. This is one mechanism in which the contribution to the ear canal recorded OAE by fine structure within the stimulus pass-band increases with decreasing stimulus level, as has been noted by others (He and Schmiedt, 1993; Talmadge *et al.*, 1998). This suggests that the stimulus level of the pip was too low at the lower stimulus levels used in these measurements to evoke a nonlinear residual via the coherent reflection mechanism, and the resulting OAE spectrum was dominated by the SSOAE components.

At high levels, the gated tone evoked extensive intermodulation produced by interactions between the stimulus and possibly the OAE components [Fig. 9(e)]. This is similar in some respects to the intermodulation reported in guinea pigs for OAEs recorded in response to filtered clicks (Yates and Withnell, 1999), although the general tendency is for much stronger intermodulation in guinea pig than in human ears and the guinea pig data may not have required the presence of SSOAEs. Most transient SFOAEs recorded in human ears in the present study showed no significant intermodulation.

V. DISCUSSION

A. Summary

The purpose of the present study was to test cochlear model predictions for stimulus-frequency otoacoustic emis-

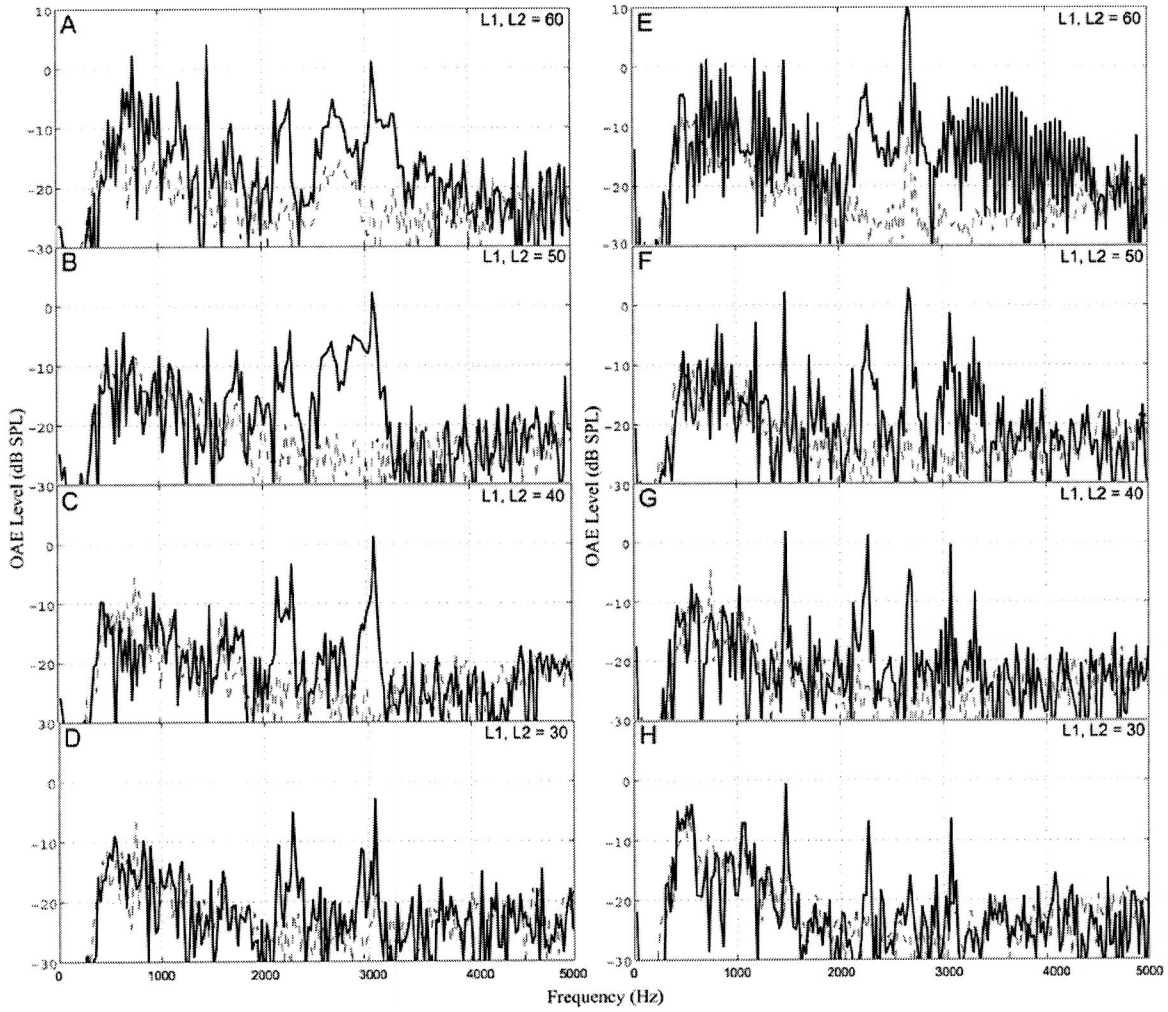


FIG. 9. SFOAE spectra elicited using pairs of 2.7-kHz tone pips (left) or gated tones (right) obtained for a range of stimulus intensity levels. The L_1, L_2 levels (dB SPL) in each panel decrease from the top to bottom panels. Data are for the same ear as the data in Figs. 3–5.

sions (SFOAEs) and distortion-product otoacoustic emissions (DPOAEs) obtained using a variety of transient stimulus classes. This was accomplished by applying time-frequency representations (TFRs) to the ear-canal recorded responses. The theory of coherent reflection emissions extrapolated to nonsinusoidal stimuli at moderate levels accounts for the spectro-temporal relationships between stimulus and OAE responses only for ears with weak SSOAEs, and, even then, there were deviations from model predictions. As expected from measurements of SOAE prevalence and the close connection between SOAEs and SSOAEs, SSOAEs were observed in most responses analyzed in this study. For tone-pip-evoked SFOAEs, narrow-band components corresponding to the higher-frequency portion of the stimulus pass-band tended to dominate recordings. DPOAEs had additional energy at SSOAE locations within the frequency band corresponding to $2f_1-f_2$. A common form of transient SFOAEs and DPOAEs generated from tone-pips and gated-tones was a collection of frequency-specific components, with each component characterized in the TFR by one or more brief segments or a single long-duration segment. When multiple internal reflections of OAEs were present, the time-differences between successive higher-order reflections obtained for 2.7-kHz DPOAEs and SFOAEs in the same ear were consistent with the two-source model of DPOAE generation, and the measures of round-trip latency were consistent with other SFOAE delay measurements.

B. Spectral composition of transient SFOAEs and DPOAEs

Models incorporating cochlear reflectance predict that an OAE response would resemble the eliciting stimulus, as expressed by Eq. (6). For example, the model predicts that the SFOAE spectral energy should lie within the pass-band of the stimulus energy, and that the SFOAE spectrum, aside from fine structure, should be similar to the stimulus spectrum within the pass-band. The results of this study show that other factors are involved. For tone-pip-evoked SFOAEs, the OAE spectrum is narrow compared to the eliciting stimulus, and response components correspond primarily to the higher-frequency portion of the stimulus pass-band. This is apparent in comparing Figs. 1(b) and (a), 2(b) and (a), and 3(b) and (a). The fact that the higher-frequency region is always higher in level than the lower-frequency region suggests that the explanation is not due to spatial variations in apical reflectance magnitude along the basilar

membrane. If that were the case, one would expect to find individual ears in which the higher frequency region was lower in level, but such effects were not observed in these data. An explanation in accord with these data is that the basilar-membrane mechanisms responsible for two-tone SFOAE suppression, for which the onset of suppression is best for a suppressor slightly higher in frequency than the reference tone (Kemp *et al.*, 1990; Brass and Kemp, 1993), may also influence coherent cochlear reflections from transient stimuli. The most sensitive frequency for SFOAE suppression of continuous tones is typically for a continuous suppressor tone that is 20% higher in frequency than that of the probe tone (Keefe *et al.*, 2003); this frequency bandwidth lies within the bandwidth of the tone pips [e.g., Fig. 1(a)]. Thus, the higher-frequency part of the tone-pip stimulus spectrum may have been more effective at suppressing the lower-frequency part of the tone-pip SFOAE than was the lower-frequency part of the tone-pip stimulus spectrum at suppressing the higher-frequency part of the tone-pip SFOAE. The resulting SFOAE spectrum would have the asymmetry observed in these data.

This suggests that the coherent reflection theory of OAE generation is incomplete for describing responses at levels (at approximately 20 dB SPL and higher) for which the basilar membrane response becomes compressive due to the saturation effects of outer-hair cell functioning. The difficulty is not with the linear coherent reflection mechanism itself, but rather with its extrapolated application to any stimulus other than a single, low-level sine tone. The processes of two-tone suppression appear to be involved in accounting for the OAE responses to such stimuli as tone pips, which are essential spectral-temporal building block towards the construction of more realistic sounds common to speech and music. Because the onset of distortion-production generation is thought to be inextricably linked with the mechanism of two-tone suppression (Lukashkin and Russell, 1998), there is also a need to include two-tone suppression effects in theories of DPOAE generation. Inclusion of two-tone suppression effects may be useful in accounting for transient OAE responses. A related finding of Dhar *et al.* (2002) is that $|R_a|$, which is implicit in the two-source DPOAE model of Eq. (16), is likely to be larger at wider f_2/f_1 ratios due to less suppression of the DP traveling wave by the more distant primaries.

Previous studies have shown that SOAEs influence the spectra of toneburst-evoked OAEs, CEOAEs (e.g., Wilson, 1980; Probst *et al.*, 1986), and DPOAEs (e.g., Moulin *et al.*, 1993). The spectral composition of evoked OAEs in the present study was strongly influenced by cochlear regions producing strong SSOAEs. In many cases, the transient (tone pip or gated tone) stimulus appeared to synchronize SOAEs located within the stimulus pass-band, such that the SFOAE or DPOAE included components of the ear's SSOAE structure. Increasing the level of the stimulus resulted in a wider frequency bandwidth of SSOAE-related components. For cases in which OAE energy was present outside the SFOAE stimulus pass-band, that energy also tended to correspond to SSOAE locations. However, at relatively high stimulus levels, an additional mechanism for fine structure may be intermodulation distortion.

To account for the presence of OAE energy outside the stimulus pass-bands in results in guinea pig observed using CEOAEs, intermodulation distortion has been proposed as a source component of CEOAEs (Withnell and Yates, 1998; Yates and Withnell, 1999). In the present study in human ears, evidence for distortion outside the stimulus bands (and two-tone-stimulus linked $2f_1-f_2$ pass-bands) was lacking for most SFOAEs and DPOAEs elicited by low-to-moderate level transient stimuli, although one contrasting response with extensive intermodulation distortion was shown in Fig. 9(e). Frequency pulling, suppression, and enhancement effects between components were often observed.

The interpretation of the sources of OAE spectral components was enhanced by the use of time-frequency analysis. For example, the SFOAE spectra obtained for a moderate-level 2.7-kHz tone pip [Fig. 9(b)] and gated tone [Fig. 9(f)] contained a component near 1.75 kHz. Only low-level components were evident in the SSOAE spectrum for this ear [see Fig. 3(g)] between about 1.5 and 2.2 kHz, suggesting that the 1.75-kHz component was not an SSOAE. The TFR for both recordings [see Fig. 3(e) for the tone pip-evoked response] revealed the 1.75-kHz component as a burst in the TFR at about 9 ms. This component may be an intermodulation component between the 2.7-kHz SFOAE and 2.2-kHz SSOAE, and the 9-ms delay may be a clue to modeling the underlying temporal dynamics.

There were cases in which SFOAE fine structure appeared in the spectrum of a single time-domain response (i.e., there was evidence of multiple internal reflections). For example, peaks in the SFOAE spectrum shown in Fig. 2(b) occurred with the approximately 170-Hz spacing predicted by Eq. (19) and the measured time delay between successive internal reflections.

C. Frequency-specific time domain measurements of OAE latency

Round trip latencies were usually consistent with model predictions (Zweig and Shera, 1995; Talmadge *et al.*, 1998). Level-dependent shifts in SFOAE latency were observed, as previously reported (Neely *et al.*, 1988). SFOAE onset latencies in Fig. 8 increased nonmonotonically as the stimulus level was decreased: stimulus levels of 50, 45, 35, and 30 dB SPL had corresponding onset latencies of 7.8, 6.3, 9.4, and 9.5 ms. SFOAE offset latencies in Fig. 8 increased monotonically with decreasing level: stimulus levels of 50, 45, 35, and 30 dB SPL had corresponding offset latencies of 4.5, 4.7, 6.5, and 7.5 ms. At some stimulus levels, the observed SFOAE latencies were shorter than the onset latencies for tone burst-evoked SFOAE reported by Neely *et al.* (1988). Methodological differences between studies may account for these differences, as the definition of level was not the same and the latencies measured by Neely *et al.* were for the entire analyzed waveform rather than the stimulus frequency-specific latency obtained from the TFR. Shera *et al.* (2002) reported SFOAE latencies using a moderate-level swept frequency stimulus that were similar to those shown here.

In each of the panels in Fig. 8, the SFOAE onsets were longer than the corresponding offsets. This asymmetry is related to the presence of a notch near the onset, which was increasingly pronounced as stimulus level decreased. In Fig. 8(c) the deepened ''notch'' created a hump near 6.5 ms. In Fig. 8(d), two humps are present at 7 and 13 ms near the onset of the main response lobe. The presence of the notch influenced the onset delays reported above, and such nonmonotonic effects may have influenced other published reports of level-dependent latencies. The SFOAE response in Fig. 8(d) was interpreted earlier as having a 9-ms latency due to the notch at approximately 8 ms, but a lower level response was present at a 5-ms latency. The notch at 8 ms may be interpreted as a temporal cancellation between the level-dependent apical reflectances of the zeroth and first-order SFOAE components. The absence of subsequent cancellations at later times may possibly involve the relative strengths of the direct and re-reflected contributors to the SFOAE, with the direct transmission dominating the response at later times. Thus, the interpretation of OAE latencies measured at low stimulus levels may be complicated by temporal interactions observable in TFRs. Higher order reflections may only play a role at low levels [as in Figs. 8(d) and (c)] for which the level difference between the SFOAE and the stimulus is less. An additional factor involved in the production of the early notch in the SFOAE temporal envelope may be that the spatial region on the cochlea over which the coherent reflected signal is generated is itself increasing with time at the onset of the tone.

D. The dominant DPOAE source changes with stimulus level

Time-differences between successive higher-order reflections obtained for 2.7-kHz DPOAEs and SFOAEs in the same ear are consistent with the two-source model of DPOAE generation expressed in Eq. (16). As a result of disparate generation locations and mechanisms, the latencies of these two DPOAE components would differ (e.g., Kemp and Brown, 1983; Shera and Zweig, 1992; Talmadge *et al.*, 1998; Knight and Kemp 2000), and, depending on their relative phase, could combine to produce notches or peaks in the transient DPOAE envelope. Such notches in the DPOAE envelope near the stimulus onset and offset have been observed (Whitehead *et al.*, 1996; Talmadge *et al.*, 1999; Tubis *et al.*, 2000a).

Ears with a strong f_2-place source are predicted to have a notch in the transient DPOAE envelope following the stimulus offset; a notch in the DPOAE response near its onset is predicted for ears with a strong DP-place source (Talmadge *et al.*, 1999; Tubis *et al.*, 2000a). The data in Fig. 8 reveal a change in the location of a prominent notch in the DPOAE envelope with decreasing stimulus level, and a corresponding shift in the DPOAE onset latency. DPOAE latencies at moderate stimulus levels were shorter compared to SFOAE latencies obtained in the same ear using gated tones presented at the DP frequency (2.7 kHz) [Figs. 8(a) and (b)]. In contrast, at low stimulus levels, DPOAE latencies were similar to SFOAE latencies obtained using 2.7-kHz gated tones [Fig. 8(d)]. The notch position and relative DPOAE-SFOAE latencies at the highest level tested [Fig. 8(a)] are consistent with model predictions for a DPOAE dominated by nonlinear distortion arising near f_2. The notch position and relative DPOAE-SFOAE latencies at lower levels [Figs. 8(b)–(d)] are consistent with model predictions for a DPOAE dominated by the DP-place source.

It is possible to compare the relative SFOAE and DPOAE latencies using the WKB theory of cochlear mechanics and the assumption of scaling invariance, using relations described in Tubis *et al.* (2000a, b). The model is expressed in terms of three empirically derived constants with values appropriate to a human cochlea taken from Talmadge *et al.* (1998): the wave number at activity pattern maximum is $\hat{k}=75\,\mathrm{cm}^{-1}$, the frequency map exponential constant is $k_\omega=1.382\,\mathrm{cm}^{-1}$, and the geometrical wave number of the cochlea is $k_0=31\,\mathrm{cm}^{-1}$. The SFOAE and the DPOAE responses were compared at the same frequency 2.7 kHz, which is denoted here as f_{DP}. The DPOAE was generated using an f_2 of 4 kHz. The latency τ_{SF} of the SFOAE is the round-trip time at frequency f_{DP}, and is calculated in the notation of the Tubis *et al.* model by

$$\tau_{\mathrm{SF}}=\frac{2\hat{k}}{k_\omega 2\pi f_{\mathrm{DP}}}=6.5\ \ \mathrm{ms}. \tag{24}$$

This latency is slightly longer than the 6-ms latency measured for SFOAEs by Shera *et al.* (2002), but is sufficiently close for present purposes.

In the gated f_1/gated f_2 stimulus condition, the latency of the DPOAE is the sum of the latency of the f_2 component to its tonotopic place and the latency of the generated f_{DP} back to the base of the cochlea. This is because the gated f_1 tone arrives at the f_2 place before the gated f_2 tone does (about 10% earlier according to a simple WKB theory), so that it is the forward latency of the onset of the f_2 tone that initiates the transient distortion-product generation. The DPOAE latency τ_{DP} is

$$\tau_{\mathrm{DP}}=\frac{\hat{k}}{k_\omega 2\pi f_2}+\frac{k_0}{k_\omega 2\pi\sqrt{f_2^2-f_{\mathrm{DP}}^2}}=3.4\ \ \mathrm{ms}. \tag{25}$$

Thus, the DPOAE latency is predicted to be shorter than the SFOAE latency, and this is in agreement with the measured latencies at all levels. As discussed in Sec. IV and outside the scope of the above theory, the latency differences decreased as the stimulus level decreased and the complicated SFOAE onset in Fig. 8(d) made it difficult to define a single onset latency. This appears to limit the applicability of this simplified theoretical approach, which does not take stimulus level into account. Figures 4(f), 6(f) and 8 provide evidence that the temporal onset and decay of SFOAE responses is more complicated than predicted. The temporal amplitude envelope of the SFOAE observable is different from that of the stimuli eliciting the SFOAE. While multiple internal reflections may play a role in the growth of the low-level SFOAE responses in Fig. 8 during the time in which the stimulus is present, it is likely that the time evolution of the distributed nonlinearity of basilar-membrane mechanics may also be involved.

In human ears, a level-dependent shift in the dominant DPOAE generator mechanism was demonstrated in previous studies (Stover *et al.*, 1996; Konrad-Martin *et al.*, 2001) examining DPOAEs measured in the frequency domain using a fixed-f_2 stimulus paradigm and transformed into an equivalent time representation using inverse fast Fourier transform (IFFT) techniques. Peak latencies of the IFFT output were relatively constant for a given subject with changes in the stimulus level, however, the relative amplitude of peaks varied, with the earliest peak becoming more dominant as the stimulus level increased. Results obtained using a suppressor tone presented near the DP frequency provide empirical evidence suggesting that even though the IFFT analytic signal has no direct physical correspondence, it provides useful information about DPOAE sources (Konrad-Martin *et al.*, 2001). The suppressor reduced or eliminated later energy (as measured along the IFFT delay axis), but left the earliest IFFT peak unaffected. The most parsimonious explanation for this result is that the secondary source (at the DP place) was suppressed and the primary source (near f_2) was not. The proportion of the IFFT waveform energy eliminated by the suppressor (and presumably originating near the DP tonotopic place) increased as the stimulus decreased from a moderate ($L_2=45$ dB SPL) to a low level ($L_2=25$ dB SPL). The average latency for the earliest peak in the IFFT waveform was 2.9 ms (s.d.=53, $N=45$) for $f_2=4$ kHz and L_1 levels ranging from 25 to 45 dB SPL. This delay is similar to the onset latencies shown here for DPOAEs obtained using moderate-level stimuli. Models of DPOAE generation for a fixed-f_2 swept-stimulus paradigm with $f_2=4$ kHz predicted the latency of the zeroth-order f_2-place component to be (Tubis *et al.*, 2000b)

$$\tau_{\text{fixed-}f_2}=\frac{k_0}{k_\omega 2\pi\sqrt{f_2^2-f_1^2}}+\frac{k_0}{k_\omega 2\pi\sqrt{f_2^2-f_{\text{DP}}^2}}=2.8\ \text{ms} \tag{26}$$

compared to the 3.4 ms predicted above for the gated f_1/gated f_2 paradigm. These DPOAE predictions and measurements from studies using different methodologies appear consistent with the present results at moderate levels [Figs. 8(a) and (b)].

E. Estimates of round-trip cochlear reflectance magnitude

It was hoped that measurements of the onset response of the gated DPOAE might differentiate between the relative time delays between the onsets of the two sources [see Eq. (18)], and the offset response would enable measurement of the round-trip cochlear reflectance magnitude [see Eq. (20)]. The goal of examining the onsets was based on the assumption that the onsets of the direct generated DPOAE source and of the secondary source reflected at the DP place would each be abrupt. However, the measurements of gated SFOAEs [as in Fig. 6(f) and Fig. 8] showed a slow-onset behavior or a more complex behavior at lower stimulus levels, which would tend according to the DPOAE model to blur the onset transient. Except for the observed shifts in the DPOAE envelope notch from onset to offset as described in the previous section, it was not possible to separate the two sources in terms of their onset behavior.

Regarding the offset responses, the theory predicts that paired DPOAE zeroth-order components (arising near the f_2 and DP places, respectively) serve as ''stimuli'' for an additional pair of (first-order) apical reflections, which serve as stimuli for second-order reflections and so on, with all higher-order reflections arising at the DP place in the form of SFOAEs. Consequently, the latency between every other transient DP component reaching the microphone should approximate the round trip latency to the $2f_1-f_2$ place [for a more detailed discussion, see Konrad-Martin *et al.* (2001)].

In the DPOAE response in Fig. 7(f) (for the $2f_1-f_2$ DPOAE at 2.7 kHz), following a sharp notch at 24 ms, there were humps following the stimulus offset and centered at about 25, 27.5, 31, and 33.5 ms. The notch at 24 ms was approximately equal to the sum of the DPOAE onset delay (approximately 3 ms) and the duration of the gated-f_1 stimulus (approximately 21.3 ms). The latency of the notch confirms the notch behavior predicted by Talmadge *et al.* for a DPOAE dominated by the generator source at f_2, as described in the previous section for the DPOAE responses in Fig. 8. A notch would be produced if the two sources were similar in level and opposite in phase—they would be similar in level at the particular time close to offset as the dominant generator source starts to decay and just before the secondary sources initiates its decay. The sequence of later humps were separated in latency by the round-trip cochlear travel delay, which was 6 ms in Fig. 7(f). This measurement matched the round trip delay for SFOAEs elicited in the same ear using moderate-level 2.7-kHz pips [Fig. 2(e)], and thus is evidence in support of the two-source model of DPOAE generation.

The relative levels of successive humps represent the attenuation level in each round trip within the cochlea. The DPOAE response trajectory in Fig. 7(f) decayed approximately 18 dB from the main lobe of the zeroth-order emission (−3 dB SPL) to the hump at 27.5 ms (−21 dB SPL), which was interpreted as the first-order apical reflection of this f_2-place component. An 18-dB attenuation in level corresponds to a round-trip amplitude reduction of $|R\_R_a|=0.125$ according to the prediction in Eq. (20). If the magnitude of the cochlear reflectance varies inversely with the traveling-wave amplitude as predicted by Talmadge *et al.* (2000), the round-trip level differences would decrease between successive pairs of higher order reflections. Confirming this prediction, the level difference between the pair of humps at 25 and 31 ms was approximately 15 dB, and the level difference between the pair of humps at 27.5 and 33.5 ms was approximately 9 dB. Thus, the level of the next round-trip reflectance was as much as $18-9=9$ dB larger than that of the first round-trip reflectance, although this is an upper bound as the later reflection at 33.5 ms was close to the noise floor. This measured increase is consistent with the predicted relative increase in the apical reflectance with increasing numbers of multiple internal reflections.

The main lobe of the DPOAE in Fig. 7(f) and the main lobe of the SFOAE shown in Fig. 6(f) for the same ear were comparable in ''steady-state'' level. In both cases, a 2.7-kHz

OAE with a level measured in the ear canal of about −3 dB served as input for higher order reflections at the 2.7-kHz place. Equations (20) and (21) predict that the round-trip level differences should differ between the transient DPOAEs and SFOAEs, and such differences were observed. The difference between the main lobe of the SFOAE (−3 dB SPL) and the hump at 27.5 ms (approximately −19 dB SPL) was 16 dB; SFOAE amplitude decreased by a factor of 0.15, which is similar to the DPOAE result. However, the next round trip resulted in a 3.5-dB decrease (factor of 0.67 amplitude change), which is less than the corresponding DPOAE result. These results suggest a relative difference in the extent of the increase in the apical reflectance magnitude as stimulus levels decrease. There are other potential sources of difference as the DPOAE had suppressors of the 2.7-kHz tone at f_1 and f_2, whereas the SFOAE was dependent only on the signal at 2.7 kHz. The fact that a second-order internal reflection was observed in the SFOAE approximately 12 ms after turning off the stimulus demonstrates that the apical reflectance (in this particular ear under this particular stimulus condition) had not yet returned to its assumed final low-level, linear-growth state after 12 ms. A longer period of silence following the end of the gated tones would be useful in further interpreting such responses.

Talmadge *et al.* (2000) have described a cochlear model incorporating both randomly spaced inhomogeneities and nonlinearity. Model predictions indicated that nonlinearity would have important implications for the recorded OAE, including a level-dependent cochlear reflectance. He and Schmiedt (1993) used a similar explanation to account for observed increases in DPOAE fine structure with decreasing stimulus intensity. The round-trip reflectance magnitude estimates, which are presented herein, are compatible with these viewpoints.

F. TFR filter length

The cone-kernel TFR (Pitton and Atlas, 1995) included a low-pass filter of variable length. While some filtering was always needed to achieve satisfactory results, the filter length altered responses to some extent. A shorter length provided better resolution of temporal events, whereas a longer length provided better separation of nearby frequencies but introduced errors in the extracted envelope amplitudes over long times. Examining responses at a variety of filter lengths assisted interpretation of potential multiple internal reflections. TFRs shown here utilized filter lengths from 120 to 180 samples (3.8–7.5 ms). However, longer filter lengths were used to verify that multiple internal reflections were not due to artifact in the TFR caused by incomplete separation of closely spaced frequency components.

VI. CONCLUSIONS

Many normal-hearing ears have significant SSOAEs, suggesting that actual OAE responses may be different from theoretical predictions, at least when transient stimuli are used. In these ears, multiple SSOAE sources and the presence of multiple internal reflections in a subset of the ears influence the fine structure of transient-evoked OAEs elicited by low to moderate-level stimuli. In addition, for pip-evoked SFOAEs, the bandwidth of the evoked SFOAE tends to be smaller than the bandwidth of the stimulus, which is consistent with a within-band suppression of the lower-frequency OAE components by the slightly higher-frequency components. This suggests the importance of two-tone suppression processes to the interpretation of OAE responses elicited by any stimuli more complicated than a single sine tone. The relative contributions by SSOAEs and multiple internal reflections to the total OAE response increased with decreasing stimulus level. This is reasonable because the limit of vanishing small stimulus level would correspond to a SSOAE, after its early-latency CEOAE is subtracted out. This means that the evoked OAE spectrum, as measured by a nonlinear residual technique, is less likely to resemble that of the eliciting stimulus (and thus model predictions) as the stimulus level decreases. The round-trip magnitude of the cochlear reflectance was measured and found to decrease with increasing effective input level, which means that realistic models of OAE generation should include a nonlinearity in the apical cochlear reflectance. A recursive formulation of the coherent reflection theory was presented that may be useful in time-domain simulations. Predictions of DPOAE and SFOAE latencies by cochlear models were in general accord although the onset of the SFOAE and the lack of level dependence in the simplified cochlear models presents difficulties. An additional effect at high stimulus levels in some ears is the observed dynamical linking between SSOAE and stimulus-evoked OAE components, which produces intermodulation at frequencies not present in the original SSOAE or in the pass-band of the stimulus.

ACKNOWLEDGMENTS

This research was supported by the NIH (Grant Nos. RO1 DC003784 and T32 DC00013). We thank Jeffrey Simmons for performing data collection and audiometric measurements, and Skip Kennedy for help in generating the figures. We thank Dr. Robert Withnell and an anonymous reviewer for valuable comments on an earlier draft of this manuscript.

APPENDIX: CALCULATION OF TIME-FREQUENCY REPRESENTATIONS

The TFR $C(t,\omega)$ is a function of time t and radian frequency ω that is bilinear in the waveform $s(t)$ of which it is the TFR. The waveform, as in the case of the waveforms analyzed in this study, is assumed to be real. The Cohen class of TFRs used in the present research differs according to the particular choice of kernel function $\Phi(\theta,\tau)$, which is a function of internal radian frequency θ and time-delay τ variables. The Cohen class of TFRs for continuous time and frequency variables is defined by (Cohen, 1995)

$$C(t,\omega)=\frac{1}{2\pi}\int\int\int s(u-\tau/2)s(u+\tau/2) \times\Phi(\theta,\tau)e^{-j\theta t}e^{-j\omega\tau}\,du\,d\tau\,d\theta. \quad \text{(A1)}$$

Particular choices of the kernel function lead to desirable properties including the attenuation of so-called cross terms in the TFR that otherwise severely complicate its interpretation of signal properties of the input waveform $s(t)$. The trivial choice of kernel $\Phi(\vartheta,\tau)=1$ leads to the Wigner distribution TFR. Other choices of kernel lead to the Choi–Williams and reduced interference distribution TFRs. The present investigation focuses on the cone kernel TFR, which may be expressed as (Zhao *et al.*, 1990)

$$C(t,\omega)=\frac{1}{2\pi}\int w(\tau)e^{-j\omega\tau}\int_{t-|\tau|/2}^{t+|\tau|/2}s(u-\tau/2)\times s(u+\tau/2)du\,d\tau. \quad \text{(A2)}$$

The cone kernel TFR includes a general low-pass filter $w(\tau)$ on the temporal delay variable τ. The bilinear signal terms include the product, or time-interference, of signals at times $u-\tau/2$ and $u+\tau/2$, so that the duration of interference is extended over the range of τ. By low-pass filtering along the τ dimension, for example, by the use of a tapered window centered at $\tau=0$ (this is the vertex of the ''cone'' of the kernel), the time-interference of signal components is limited. This limits the generation of cross-terms in the TFR, which are artifacts evidenced by spurious components on the time-frequency plane (Loughlin *et al.*, 1993).

The input signal (OAE) waveform is the averaged 2E waveform of Eq. (1). This waveform eliminates all stimulus response so that the problem of extracting a small-amplitude OAE from a large-amplitude stimulus response is avoided. Except for Keefe *et al.* (1999), which also used the 2E waveform as the input signal, previous research on time-frequency analyses of transient-evoked OAEs have used click-evoked and tone-burst-evoked signals in which the dynamic range problem of the stimulus and OAE responses is present. One common solution has been to remove the short-latency part of the OAE, as for example in standard techniques to differentially measure CEOAEs, but this also discards the high-frequency portion of the OAE response. Use of the 2E waveform allows measurement of high-frequency CEOAEs, at least up to 10 kHz (Keefe *et al.*, 1999).

TFRs of 2E OAE waveforms have been compared using the Wigner, Choi–Williams, reduced interference, and cone kernel distributions (Keefe *et al.*, 1999). The results show the Wigner distribution to be the worst choice, because it is highly contaminated with cross-terms. The reduced interference distribution, and, to a lesser extent, the Choi–Williams, have more significant cross-terms than the cone kernel (as deduced based on analyses of various OAE signals), and this is the main reason for the choice of the cone kernel in the present research. A secondary reason is that the Choi–Williams distribution blurs the temporal onsets and offsets of TFR components compared to the cone kernel. The cone kernel is preferable for identifying discontinuities of the signal in time. The cone kernel is certainly not an ideal choice of distribution in the Cohen class, as, indeed, the definition of ''ideal'' is problematical: significant theoretical progress has been achieved by using other forms of TFRs that are not bilinear in the signal [see Cohen (1995) for an introductory review]. Nevertheless, the cone kernel augmented by a reasonable choice of low-pass filter provides acceptable results for analyzing OAE TFRs. One reason has to do with the particular nature of OAEs, which are signals with well-defined latencies and decays that tend to have relatively invariant spectral features. This is due to the fact that particular choices of stimulus frequencies lead to particular frequency components in the response. The cone kernel also has a lateral inhibition effect that enhances spectral peaks in the TFR (Zhao *et al.*, 1990; Loughlin *et al.*, 1993).

A discrete-time implementation of the cone kernel TFR is used in the present research based on Pitton and Atlas (1995), in which the distribution is sampled at discrete frequencies using the DFT, and computations are performed at discrete times based on the measurement sample rate (24 kHz). In previous research on OAEs using older discrete-time implementations of the cone kernel TFR, it was necessary to artificially double the signal sample rate in the calculation of the TFR (Keefe *et al.*, 1999). This significantly lengthened the calculation time. The only modification from the algorithm of Pitton and Atlas is the additional property that the TFR should have circular rotational invariance in time. This is the property such that if the waveform is circular rotated in its buffer by N samples, the TFR should be circular rotated along its temporal dimension by N samples with no effect along the frequency dimension. This eliminates artifacts in the TFR that otherwise occur at time samples close to the beginning or end of the buffer. To build in this property, the signal buffer is increased by $N/4$ samples at both the beginning and end. The beginning of the buffer is filled by the final $N/4$ samples from the end of the original signal buffer, and the end of the buffer is filled by the initial $N/4$ samples from the beginning of the original signal buffer. Any rotation of the signal includes rotation of these extra samples. The TFR of the resulting augmented buffer has the desired circular-invariance property.

The particular low-pass filter used in the TFR in the present research is a Kaiser window with shape parameter 7.04 with user-selectable length. Such a window has similar sideband rejection to a Blackman window of the same length, but a central lobe that is narrower in frequency by a factor of 0.77 (Oppenheim and Schafer, 1989). A narrowed central lobe enhances the frequency selectivity of the low-pass filter. Filter lengths used in these analyses ranged from 90 to 180 samples, or 3.8 to 7.5 ms. A shorter length provided better resolution of temporal events, whereas a longer length provided better separation of nearby frequencies but introduced errors in the extracted envelope amplitudes over long times (see Sec. V for more detail).

Brass, D., and Kemp, D. T. (**1993**). ''Suppression of stimulus frequency otoacoustic emissions,'' J. Acoust. Soc. Am. **93**, 920–939.

Burns, E. M., and Pitton, J. W. (**1993**). ''Time-frequency analyses of coherent frequency fluctuations among spontaneous otoacoustic emissions,'' J. Acoust. Soc. Am. **93**, 2314(A).

Burns, E. M., Arehart, K. H., and Campbell, S. L. (**1992**). ''Prevalence of spontaneous otoacoustic emissions in neonates,'' J. Acoust. Soc. Am. **91**, 1571–1575.

Burns, E. M., Keefe, D. H., and Ling, R. (**1998**). ''Energy reflectance in the ear canal can exceed unity near spontaneous otoacoustic emission frequencies,'' J. Acoust. Soc. Am. **103**, 462–474.

Cheng, J. (**1995**). ''Time-frequency analysis of transient evoked otoacoustic emissions via smoothed pseudo Wigner distribution,'' Scand. Audiol. **24**, 91–96.
Cohen, L. (**1995**). *Time-Frequency Analysis* (Prentice–Hall, New York).
Dhar, S., Talmadge, C. L., Long, G. R., and Tubis, A. (**2002**). ''Multiple internal reflections in the cochlea and their effect on DPOAE fine structure,'' J. Acoust. Soc. Am. **112**, 2882–2897.
Hatzopoulos, S., Cheng, J., Grzanka, A., Morlet, T., and Martini, A. (**2000**). ''Optimization of TEOAE recording protocols: a linear protocol derived from parameters of a time-frequency analysis: a pilot study on neonatal subjects,'' Scand. Audiol. **29**, 21–27.
He, N. J., and Schmiedt, R. A. (**1993**). ''Fine structure of the 2f1-f2 acoustic distortion product: changes with primary level,'' J. Acoust. Soc. Am. **94**, 2659–2669.
Jeong, J., and Williams, W. J. (**1992**). ''Kernel design for reduced interference distributions,'' IEEE Trans. Signal Process. **40**, 402–412.
Kalluri, R., and Shera, C. A. (**2001**). ''Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation,'' J. Acoust. Soc. Am. **109**, 622–637.
Keefe, D. H. (**1998**). ''Double-evoked otoacoustic emissions. I. Measurement theory and nonlinear coherence,'' J. Acoust. Soc. Am. **103**, 3489–3498.
Keefe, D. H., and Ling, R. (**1998**). ''Double-evoked otoacoustic emissions: II. Intermittent noise rejection, calibration and ear canal measurements,'' J. Acoust. Soc. Am. **103**, 3499–3508.
Keefe, D. H., Piskorski, P., and Gorga, M. P. (**1999**). ''Double-evoked otoacoustic emissions: Time-frequency representations and high-frequency transient-response measurements,'' Assoc. Res. Otolaryngol. Abs. **22**, 395 (abstract).
Keefe, D. H., Ellison, J., Fitzpatrick, D., and Gorga, M. P. (**2003**). ''Two-tone suppression of stimulus frequency otoacoustic emissions,'' Assoc. Res. Otolaryngol. Abs. **26**, 681 (abstract).
Kemp, D. T. (**1986**). ''Otoacoustic emissions, traveling waves and cochlear mechanisms,'' Hear. Res. **22**, 95–104.
Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from the human auditory systems,'' J. Acoust. Soc. Am. **64**, 1386–1391.
Kemp, D. T. (**1997**). ''Otoacoustic emissions in perspective,'' in *Otoacoustic Emissions: Clinical Applications*, edited by M. S. Robinette and T. J. Glattke (Thieme, New York), pp. 1–21.
Kemp, D. T., and Brown, A. M. (**1983**). ''An integrated view of cochlear mechanical nonlinearities observable from the ear canal,'' in *Cochlear Mechanics*, edited by E. de Boer and M. A. Viergever (Delft U.P., Delft), pp. 75–82.
Kemp, D. T., Brass, D., and Souter, M. (**1990**). ''Observations on simultaneous SFOAE and DPOAE generation and suppression,'' in *Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer-Verlag, New York), pp. 202–209.
Knight, R. D., and Kemp, D. T. (**2000**). ''Indications of different distortion product otoacoustic emission mechanisms from a detailed f1, f2 area study,'' J. Acoust. Soc. Am. **107**, 457–473.
Kim, D. O. (**1980**). ''Cochlear mechanics: Implications of electrophysiological and acoustical observations,'' Hear. Res. **2**, 297–317.
Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (**2001**). ''Sources of DPOAEs revealed by suppression experiments and IFFTs in normal ears,'' J. Acoust. Soc. Am. **109**, 2862–2879.
Loughlin, P. J., Pitton, J. W., and Atlas, L. E. (**1993**). ''Bilinear time-frequency representations: New insights and properties,'' IEEE Trans. Signal Process. **41**, 750–767.
Lukashkin, A. N., and Russell, I. J. (**1998**). ''A descriptive model of the receptor potential nonlinearities generated by the hair cell mechanoelectrical transducer,'' J. Acoust. Soc. Am. **103**, 973–980.
Martin, G. K., Jassir, D., Stagner, B. B., Whitehead, M. L., and Lonsbury-Martin, B. L. (**1998**). ''Locus of generation for the 2f1-f2 vs 2f2-f1 distortion-product otoacoustic emissions in normal-hearing humans revealed by suppression tuning, onset latencies, and amplitude correlations,'' J. Acoust. Soc. Am. **103**, 1957–1971.
McIntyre, M. E., Schumacher, R. T., and Woodhouse, J. (**1983**). ''On the oscillations of musical instruments,'' J. Acoust. Soc. Am. **74**, 1325–1345.
Moulin, A., Collet, L., Veuillet, E., and Morgon, A. (**1993**). ''Interrelations between transiently evoked otoacoustic emissions, spontaneous otoacoustic emissions and acoustic distortion products in normally hearing subjects,'' Hear. Res. **65**, 216–233.
Neely, S. T., Norton, S. J., Gorga, M. P., and Jesteadt, W. (**1988**). ''Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli,'' J. Acoust. Soc. Am. **83**, 652–656.
Oppenheim, A. V., and Schafer, R. W. (**1989**). *Discrete-time Signal Processing* (Prentice–Hall, Englewood Cliffs, NJ).
Ozdamar, O., Zhang, J., Kalayci, T., and Ulgen, Y. (**1997**). ''Time-frequency distribution of evoked otoacoustic emissions,'' Br. J. Audiol. **31**, 461–471.
Pitton, J., and Atlas, L. E. (**1995**). ''Discrete-time implementation of the cone-kernel time-frequency representation,'' IEEE Trans. Signal Process. **43**, 1996–1998.
Probst, R., Coats, A. C., Martin, G. K., and Lonsbury-Martin, B. L. (**1986**). ''Spontaneous, click-, and toneburst-evoked otoacoustic emissions from normal ears,'' Hear. Res. **21**, 261–275.
Ruggero, M. A., Rich, N. C., and Freyman, R. (**1983**). ''Spontaneous and impulsively evoked otoacoustic emissions: Indicators of cochlear pathology?'' Hear. Res. **10**, 283–300.
Schairer, K. S., Fitzpatrick, D., and Keefe, D. H. (**2003**). ''Input-output functions for stimulus-frequency otoacoustic emissions in normal-hearing adult ears,'' J. Acoust. Soc. Am. (to be published)
Shera, C. A., and Guinan, J. J. (**1999**). ''Evoked otoacoustic emissions arise by two fundamentally different mechanisms: a taxonomy for mammalian OAEs,'' J. Acoust. Soc. Am. **105**, 782–798.
Shera, C., and Zweig, G. (**1992**). ''Analyzing reverse middle-ear transmission: Noninvasive Gedanken experiments,'' J. Acoust. Soc. Am. **92**, 1371–1381.
Shera, C. A., and Zweig, G. (**1993**). ''Noninvasive measurement of the cochlear traveling-wave ratio,'' J. Acoust. Soc. Am. **93**, 3333–3352.
Shera, C. A., Guinan, J. J., and Oxenham, A. J. (**2002**). ''Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements,'' Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.
Sondhi, M., and Resnick, R. (**1983**). ''The inverse problem for the vocal tract: Numerical methods, acoustical experiments and speech synthesis,'' J. Acoust. Soc. Am. **73**, 985–1002.
Stover, L. J., Neely, S. T., and Gorga, M. P. (**1996**). ''Latency and multiple sources of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **99**, 1016–1024.
Talmadge, C. L., Long, G. R., Rubis, A., and Dhar, S. (**1999**). ''Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **105**, 275–292.
Talmadge, C. L., Tubis, A., Long, G. R., and Pikorski, P. (**1998**). ''Modeling otoacoustic emission and hearing threshold fine structures,'' J. Acoust. Soc. Am. **104**, 1517–1543.
Talmadge, C. L., Tubis, A., Long, G. R., and Tong, C. (**2000**). ''Modeling the combined effects of basilar membrane nonlinearity and roughness on stimulus frequency otoacoustic emission fine structure,'' J. Acoust. Soc. Am. **108**, 2911–2932.
Tognola, G., Grandori, F., and Ravazzani, P. (**1997**). ''Time-frequency distributions of click-evoked otoacoustic emissions,'' Hear. Res. **106**, 112–122.
Tubis, A., and Talmadge, C. L. (**1998**). ''Ear canal reflectance in the presence of spontaneous otoacoustic emissions. I. Limit-cycle oscillator model,'' J. Acoust. Soc. Am. **103**, 454–461.
Tubis, A., Talmadge, C. L., and Tong, C. (**2000a**). ''Modeling the temporal behavior of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **107**, 2112–2127.
Tubis, A., Talmadge, C. L., Tong, C., and Dhar, S. (**2000b**). ''On the relationships between the fixed-f1, fixed-f2, and fixed-ratio phase derivatives of the 2f1-f2 distortion product otoacoustic emission,'' J. Acoust. Soc. Am. **108**, 1772–1785.
Whitehead, M. L., Stagner, B. B., Martin, G. K., and Lonsbury-Martin, B. L. (**1996**). ''Visualization of the onset of distortion-product otoacoustic emissions, and measurement of their latency,'' J. Acoust. Soc. Am. **100**, 1663–1679.
Wilson, J. P. (**1980**). ''Evidence for cochlear origin for acoustic re-emissions, threshold fine structure, and tonal tinnitus,'' Hear. Res. **2**, 233–252.
Wit, H. P., and Ritsma, R. J. (**1980**). ''Evoked acoustical responses from the human ear: some experimental results,'' Hear. Res. **2**, 253–261.
Wit, H. P., van Dijk, P., and Avan, P. (**1994**). ''Wavelet analysis of real ear and synthesized click evoked otoacoustic emissions,'' Hear. Res. **73**, 141–147.

Withnell, R. H., and Yates, G. K. (**1998**). ''Enhancement of the transient-evoked otoacoustic emission produced by the addition of a pure tone in the guinea pig,'' J. Acoust. Soc. Am. **104**, 344–349.
Yates, G. K., and Withnell, R. H. (**1999**). ''The role of intermodulation distortion in transient-evoked otoacoustic emissions,'' Hear. Res. **136**, 49–64.
Zhao, Y., Atlas, L. E., and Marks, R. J. (**1990**). ''The use of cone-shaped kernels for generalized time-frequency representations of nonstationary signals,'' IEEE Trans. Acoust., Speech, Signal Process. **38**, 1084–1091.
Zweig, G., and Shera, C. A. (**1995**). ''The origin of periodicity in the spectrum of evoked otoacoustic emissions,'' J. Acoust. Soc. Am. **98**, 2018–2047.
Zwicker, E., and Peisl, W. (**1990**). ''Cochlear preprocessing in analog models, in digital models and in human inner ear,'' Hear. Res. **44**, 209–216.

Physiological vulnerability of distortion product otoacoustic emissions from the amphibian ear

Pim van Dijk[a)]
Department of Otorhinolaryngology and Head & Neck Surgery, University Hospital Maastricht, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands

Peter M. Narins and Matthew J. Mason[b)]
Department of Physiological Science, University of California at Los Angeles, Los Angeles, California 90095-1606

(Received 22 April 2003; revised 4 July 2003; accepted 16 July 2003)

The physiological vulnerability of distortion product otoacoustic emissions (DPOAEs) was investigated in the leopard frog, *Rana pipiens pipiens*. For each frog, DPOAEs were recorded from the amphibian and the basilar papillae. Measurements were taken before and after either the arrest of oxygen supply due to cardioectomy, or the destruction of the central nervous system (CNS). DPOAEs in response to high-level stimuli (>75 dB SPL) were rather robust to these insults during the first two hours post surgery. In contrast, DPOAE amplitudes in response to low-level stimuli (<75 dB SPL) decreased significantly. On average, low-level emissions from the amphibian papilla disappeared within 6 min for cardioectomy, and after 13 min for CNS destruction. In the basilar papilla, low-level DPOAEs disappeared more slowly: on average after 34 min following cardioectomy, and after 58 min for CNS destruction. The difference in physiological vulnerability between low- and high-level emissions is similar to that in mammals and a lizard. The difference between the DPOAE decay rate of the frog's amphibian and basilar papillae suggests important differences between the hearing mechanisms of the papillae. [DOI: 10.1121/1.1608957]

PACS numbers: 43.64.Jb, 43.64.Ri [BLM]

I. INTRODUCTION

Distortion product otoacoustic emissions (DPOAEs) in mammals and lizards are vulnerable to acoustic overstimulation, anoxia, and pharmacological agents (Zurek *et al.*, 1982; Kemp and Brown, 1984; Lonsbury-Martin *et al.*, 1987; Whitehead *et al.*, 1992; Manley *et al.*, 1993; Mills *et al.*, 1993; Rebillard *et al.*, 1993; Frolenkov *et al.*, 1998). While DPOAEs generated by low-level stimuli (below about 60 dB SPL) are particularly vulnerable, high-level DPOAEs (above 60 dB SPL stimulus level) are more robust. This observation has led to the conclusion that low- and high-level DPOAEs are generated by different inner ear mechanisms. In particular, mammalian low-level DPOAEs are assumed to be related to motility of outer hair cells. In contrast, high-level DPOAEs may originate from ''passive'' nonlinear responses of intracochlear structures.

Somatic outer hair cell motility (Brownell *et al.*, 1985) has been proposed as a mechanism for otoacoustic emission generation. Somatic motility relies on the presence of the protein prestin in the hair cell basolateral membrane (Zheng *et al.*, 2000; Liberman *et al.*, 2003). Prestin-driven somatic motility has been confirmed in mammalian outer hair cells (Zheng *et al.*, 2000) and its effect on otoacoustic emissions shows that it is a necessary and important component of the mammalian cochlear amplifier (Liberman *et al.*, 2003). In amphibians and other nonmammalian species there is no evidence of somatic motility (He *et al.*, 2003; Manley, 2001). An alternative mechanism for emission generation is the active movements of hair bundles. Although in the frog such bundle movements have only been shown in saccular hair cells at low frequencies (Assad *et al.*, 1989; Assad and Corey, 1992; Benser *et al.*, 1996; Martin and Hudspeth, 1999), it is conceivable that similar motility exists at audio frequencies and in other vertebrate species. Active hair bundle movements have also been shown for hair cells from the turtle basilar papilla (Crawford and Fettiplace, 1985) and the chicken basilar papilla (Hudspeth *et al.*, 2000). Active bundle movements could be a generation mechanism common across vertebrate species, perhaps acting in concert with somatic cell motility in the case of mammals.

Here, we investigated the vulnerability of DPOAEs from the amphibian ear. If DPOAEs in amphibians show similar physiological vulnerability to that in other vertebrates, the hypothesis that similarities exist between DPOAE generation mechanisms in all vertebrates is supported. The common generation mechanism across species may be hair bundle motility, while somatic motility is an additional function that may have become predominant in mammals.

The amphibian inner ear includes the amphibian papilla (AP) and the basilar papilla (BP), both of which are auditory end organs. Based on a comparison between DPOAE frequencies and auditory nerve fiber characteristic frequencies for both papillae, Van Dijk and Manley (2001) concluded that in *Hyla cinerea* the AP generates both spontaneous otoacoustic emissions (SOAEs) and DPOAEs, while the basilar

a)Electronic mail: pvd@kno.unimaas.nl
b)Current address: University of Cambridge, Department of Zoology, Downing Street, Cambridge CB2 3EJ, United Kingdom.

papilla generates DPOAEs only. This suggests that important differences between the papillae may exist. We will describe differences between the vulnerability of DPOAEs from the AP and the BP.

II. MATERIALS AND METHODS

Four male and four female leopard frogs, *Rana pipiens pipiens*, were used in these experiments, all obtained from commercial suppliers. DPOAE measurements were performed using a sensitive microphone probe (Etymotic Research, model ER10-C). This probe contains two miniature speakers and a microphone. It is typically used in human subjects with a special foam tip, which fits the human auditory meatus. We used the probe with such a foam tip, but we fit a piece of rubber tubing (length 21 mm, inner diameter 9.5 mm) over the tip, thus extending the probe assembly by about 10 mm. Then, the microphone probe was acoustically coupled to the frog's ear, by placing the open end of the tubing on the skin around the tympanic membrane. The connection between the tubing and the skin was sealed with silicon grease.

The two speakers were used to present two stimulus tones simultaneously, with frequencies f_1 and f_2, and levels L_1 and L_2. Stimulus tones were generated using a DA-converter (Tucker Davis Technologies, Gainesville, FL, model RP2).

The microphone signal was analyzed with an FFT network analyzer (Stanford Research, model SR770) set for an analysis range of 780 Hz and frequency resolution of 1.95 Hz. The average amplitude spectrum of 16 successive 50%-overlapping time windows was obtained for each set of stimulus parameters. The peak amplitudes of the distortion products at $2f_1-f_2$ and $2f_2-f_1$ were recorded from the average spectrum.

DPOAE input–output functions were measured for two pairs of stimulus tones at $f_1=1011$ Hz, $f_2=1112.1$ Hz, $f_2/f_1=1.1$, and at $f_1=2011$ Hz, $f_2=2212.1$ Hz, $f_2/f_1=1.1$. These stimulus frequencies were chosen to be within the sensitivity ranges of the amphibian papilla (AP) and the basilar papilla (BP), respectively (Ronken, 1990, 1991; Christensen-Dalsgaard and Narins, 1993). The levels of the stimulus tones were equal ($L_1=L_2$), and were varied from 40 to 90 dB SPL in 5-dB steps. In addition to input–output functions, DP-grams were recorded for frequencies from $f_1=211$ to 3011 Hz, in 200-Hz frequency steps, with $L_1=L_2=85$ dB SPL and $f_2/f_1=1.1$.

Emission measurements were performed both on live animals and *post mortem*. Initially, animals ($n=8$) were anesthetized with an intramuscular injection of pentobarbital sodium solution (Nembutal, Abbott Laboratories, 50 mg/ml: 1.0–1.3 μl/g body mass), after which baseline DPOAE measurements were performed. Then, animals were killed by cardioectomy ($n=3$ animals) or by destruction of the central nervous system (CNS, $n=4$). In one animal, both cardioectomy and CNS destruction were applied. In cardioectomy, an incision was made in the animal's abdomen. The heart was located, and removed by cutting its connections with a pair of scissors. CNS destruction was by means of double pithing, a standard technique for frog euthanasia. A sharp needle was inserted cranially between the skull and atlas, penetrating through the foramen magnum and, by rotation, mechanically destroying the brain. The needle was then withdrawn, and reinserted caudally to destroy the spinal cord. Following cardioectomy or CNS destruction, DPOAE measurements were taken continuously for 2 h. In three frogs, additional DPOAE measurements were performed beyond the 2 h time window. These animals were followed for 45, 51 and 72 h, respectively. Body temperature of the animals was not actively controlled; room temperature was about 20 °C.

The ''Principles of Animal Care'' (NIH publication 85-23, revised 1985) and USA regulations were followed throughout this study, and protocols were approved by the University of California Animal Research Committee.

III. RESULTS

Prior to euthanasia, DPOAEs were present in all animals investigated. In response to the high-level stimulus, with $L_1=L_2=85$ dB SPL, the distortion product at $2f_1-f_2$ was typically detectable for stimulus frequencies from $f_1=411$ Hz to $f_1=2611$ Hz. Within this range, emission levels were highest for $f_1=1811$ or 2011 Hz, with peak levels ranging from 20 to 48 dB SPL with an average peak level of 32 dB SPL. The DPOAE at $2f_2-f_1$ was detectable for $f_1=411$ Hz to $f_1=2211$ Hz, and peaked at 1611 or 1811 Hz, with peak levels ranging from 22 to 51 dB SPL with an average of 35 dB SPL.

Input–output functions consisted of two segments, separated by a knee point near 75 dB SPL input level (see Fig. 1). We will refer to the corresponding emissions as low-level ($<$75 dB SPL input level) and high-level ($>$75 dB SPL input level) DPOAEs, respectively.

Low-level DPOAEs disappeared during the first hour *post mortem*. That is, during this period, the knee point disappeared, and the input–output function essentially became a straight line (see Fig. 1). With CNS destruction ($n=4$ animals), low-level amphibian papilla DPOAEs (i.e., $f_1=1011$ Hz) disappeared on average in 13 minutes (range across animals 5–25 min; see Table I). In contrast, in cardioectomized animals ($n=4$, including the animal which underwent both cardioectomy and CNS destruction), low-level DPOAEs in the amphibian papilla always disappeared before the first *post-mortem* recording, i.e., within 6 min. In the basilar papilla, low-level DPOAEs always disappeared slower than those from the amphibian papilla. After CNS destruction the average time to full DPOAE extinction was 58 min on average (range 53–62 min), while after cardioectomy time to extinction was faster: 34 min on average (range 15–52 min). These values are presented in Table I. In summary, the DPOAE decay rate was slower in animals with CNS destruction than in cardioectomized animals. Also, the rate of DPOAE decay was slower in the BP than in the AP.

The high-level DPOAEs either increased or decreased during the first 2 h *post mortem*. Following CNS destruction, amphibian papilla DPOAEs changed on average by +9.3 dB (range −18.4 to +26.9 dB), and basilar papilla DPOAEs changed on average by −1.3 dB (range −5.4 to +3.5 dB). Following cardioectomy, amphibian papilla DPOAEs

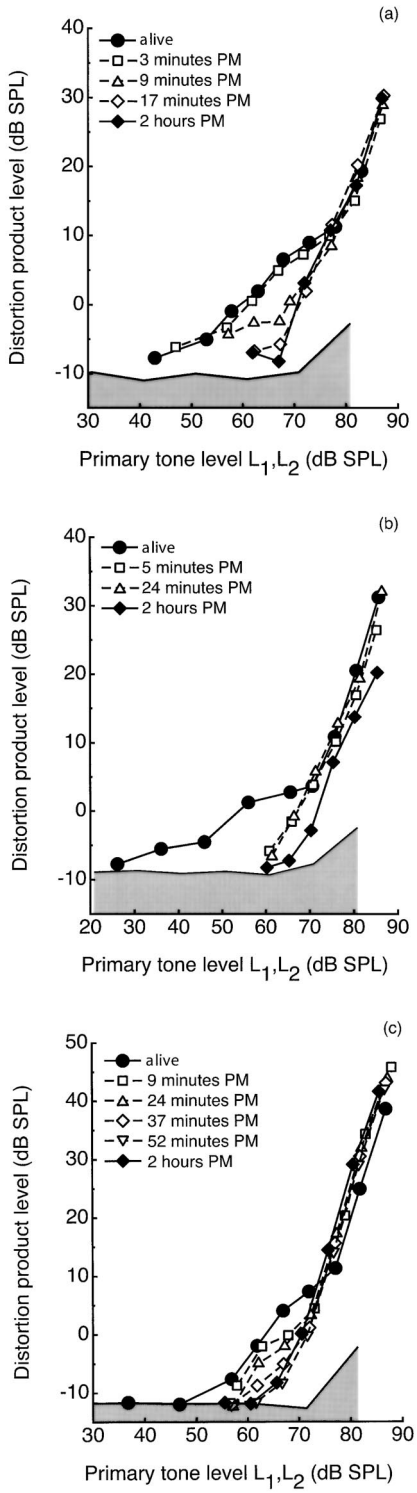


FIG. 1. Input–output curves for the DPOAE at $2f_1-f_2$ for various times *post mortem* (PM). Equal primary levels, $L_1=L_2$. Ratio $f_2/f_1=1.1$. (a) Amphibian papilla, $f_1=1011$ Hz, before and after CNS destruction. (b) Amphibian papilla, $f_1=1011$ Hz, before and after cardioectomy. (c) Basilar papilla, $f_2=2011$ Hz, same subject as panel (b). In live animals, input–output curves consisted of two parts separated by a knee-point near 75 dB SPL input level. The corresponding emissions are referred to as low-level (below 75 dB SPL) and high-level (above 75 dB SPL) DPOAEs. *Post mortem*, the low-level DPOAEs disappeared, while high-level DPOAEs are relatively unaffected. Panels (a) and (b) illustrate that the rate of decay of low-level DPOAE was slower after CNS destruction than after cardioectomy. Panels (b) and (c) show that the decay was more rapid in the amphibian papilla than in the basilar papilla. The gray shaded areas indicate system noise (below 70 dB SPL input level) and system distortion (above 70 dB SPL input level).

TABLE I. Decay times of low-level DPOAEs for CNS destruction versus cardioectomy, and amphibian papilla versus basilar papilla. Time to full disappearance is indicated in minutes. The mean across subjects is in bold; the range of values is in brackets.

	CNS destruction ($n=4$)	Cardioectomy ($n=4$)
Amphibian papilla ($f_1=1011$ Hz)	**13** (5–25)	<6
Basilar papilla ($f_1=2011$ Hz)	**58** (53–62)	**34** (15–52)

changed by −9.2 dB (range −17.0 to −0.2 dB), and basilar papilla DPOAEs changed by −3.4 dB (range −9.4 to +0.8 dB).

In three frogs, DPOAEs were monitored beyond 2 h post mortem ($n=1$ cardioectomy, $n=1$ CNS destruction, $n=1$ cardioectomy and CNS destruction). During this time interval, only high-level DPOAEs were detectable. Initially, DPOAEs would decrease by 20–25 dB to a minimum at 150–400 min *post mortem*. After this initial phase of decrease, DPOAEs reemerged, following a time-course (Fig. 2) which was inconsistent across subjects. Within the same animal, the time-course of the AP- and BP-emissions were also inconsistent. This is related to the observation that after the reemerging of DPOAEs, the DP-grams no longer reflected the frequency ranges of the amphibian and basilar papillae (Fig. 3). In the three frogs, clear acoustic distortion products were still observed at 45, 51, and 72 h after surgery, respectively. At these times the experiment was terminated.

IV. DISCUSSION

The results are qualitatively similar to those in mammals and lizards (Zurek *et al.*, 1982; Kemp and Brown, 1984; Lonsbury-Martin *et al.*, 1987; Whitehead *et al.*, 1992; Manley *et al.*, 1993; Mills *et al.*, 1993; Rebillard *et al.*, 1993; Frolenkov *et al.*, 1998): DPOAEs in response to low-level primary tones (below about 75 dB SPL in frogs) are very sensitive to disruption of inner ear physiology. In contrast, DPOAEs generated by high-level stimuli are more robust.

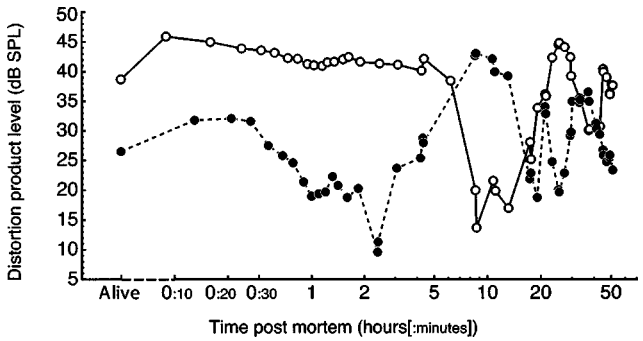


FIG. 2. Example of the long-term time course of high-level DPOAEs at $2f_1-f_2$ in one animal. Method of euthanasia: cardioectomy. $L_1=L_2=85$ dB SPL, $f_2/f_1=1.1$. Closed symbols: $f_1=1011$ Hz, amphibian papilla. Open symbols: $f_1=2011$ Hz, basilar papilla. The DPOAE from the amphibian papilla decays to a minimum just after 2 h. Then, it reappears and follows a complex relationship with time. The DPOAE from the basilar papilla follows a similar time-course, but on a different time scale. The initial minimum amplitude for this emission was observed after about 8 h *post mortem*.

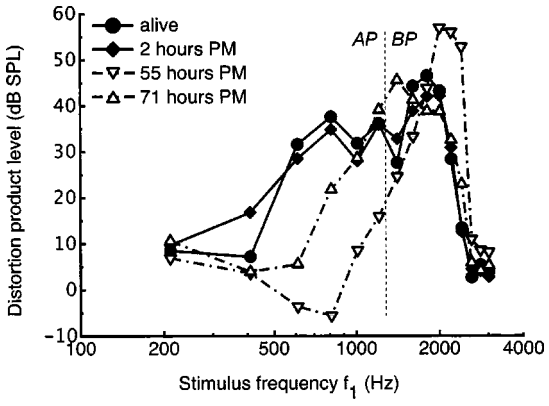


FIG. 3. Example of the long-term time course of the high-level DPOAE at $2f_1-f_2$ following cardioectomy. $L_1=L_2=85$ dB, $f_2/f_1=1.1$. The thin vertical line approximately separates the range of characteristic frequencies of auditory neurons from the amphibian papilla (AP) and the basilar papilla (BP) (Ronken, 1991). In the live animal (filled circles) and 2 h *post mortem* (filled diamonds), both the amphibian and basilar papillae emit DPOAEs. After an initial decay of these components (see Fig. 2), DPOAEs reappeared (open symbols) but emission frequencies no longer corresponded to the papillar frequency ranges.

In the experiments described here, no artificial respiration was used. In most frog species, oxygen is absorbed into the animal's body by diffusion through the moist skin, and is distributed through the body via blood circulation. One result of cardioectomy is the arrest of blood circulation. Consequently, the main effect of cardioectomy on hearing is presumably the near elimination of the oxygen supply to the inner ear.

In contrast, with CNS destruction, the heart typically continues to be functional for several hours (our observation), and so oxygen supply to the inner ear is presumably maintained. Rather, CNS destruction disrupts efferent input to the inner ear. In addition, mechanical destruction of the CNS as applied here (''double-pith'' procedure), probably also mechanically destroys the endolymphatic and perilymphatic sacs, which are in the brain case. The disturbance of the efferent system will only affect the amphibian papilla, since the basilar papilla in ranid frogs lacks efferent innervation (Frishkopf and Flock, 1974; Robbins *et al.*, 1967).

We showed that low-level DPOAEs decay rapidly following cardioectomy, but slower after CNS destruction. We infer that cardioectomy immediately stops oxygen supply, which in turn results in the rapid decay of DPOAEs. In contrast, the physiological effect of CNS destruction on DPOAEs takes place in several minutes to an hour.

The difference in DPOAE decay time between the amphibian papilla and the basilar papilla, even within the same animal [compare Figs. 1(b) and (c)], is of interest. Apparently, DPOAEs from the AP are more vulnerable to physiological insult than are those from the BP. This suggests that different mechanisms may underly these emissions. The vulnerability of DPOAEs in the AP correlates with the ability of the AP to generate spontaneous otoacoustic emissions (Van Dijk and Manley, 2001). Possibly, hearing in the AP is supported by an active mechanism *sensu* Gold (1948), i.e., via a feedback mechanism which involves cycle-by-cycle amplification. Such a mechanism could account for the presence of spontaneous emissions in the AP and is expected to be highly vulnerable to oxygen supply disruption. In contrast, DPOAEs from the basilar papilla, which does not generate spontaneous emissions, could reflect a passive nonlinearity, for example due to the nonlinear stiffness of hair bundles (Howard *et al.*, 1988; Howard and Hudspeth, 1988). Note that ''passive'' does not imply independence of oxygen supply. In fact, the integrity of the inner ear, and in particular the hair cell tip-links, is essential for generation of such a passive nonlinear response, and is expected to be related in the long term to the inner ear oxygen supply. Rather, ''passive'' implies the absence of a cycle-by-cycle amplifier, which could generate spontaneous emissions.

A further difference between the AP and the BP relates to their efferent innervation. While in *Xenopus laevis* both the AP and BP receive efferent innervation (Hellmann and Fritzsch, 1996), in *Rana catesbeiana* it was found to be present in the AP, but absent in the BP (Frishkopf and Flock, 1974; Robbins *et al.*, 1967). This latter finding can probably be extrapolated to *Rana pipiens*, the congeneric species investigated here. In humans (Veuillet *et al.*, 1991; Collet *et al.*, 1992) and a bird (Manley *et al.*, 1999), stimulation of the efferent system affects otoacoustic emissions. Presumably, the efferent system plays a role in controlling the active mechanism in the inner ear. According to the argument presented above, the presence of efferent innervation in the AP would correlate with the presence of an active mechanism, while in the BP of ranid frogs both the active mechanism and efferent innervation would be absent. Regardless of this possible relationship between efferent innervation and active hearing, the difference in innervation of the AP and BP may contribute to the difference of DPOAE decay rate between the papillae, in particular in the case of CNS destruction.

The robust DPOAEs in response to high-level stimuli reflect the frequency sensitivity ranges of the AP and BP, respectively, at least during the first 2 h *post mortem*. Consequently, an explicit conclusion can be drawn: high-level DPOAEs must originate from structures related to the amphibian and basilar papillae. Our results are therefore consistent with the conclusion of Whitehead *et al.* (1992), who showed that high-level *post mortem* DPOAEs in the rabbit are sensitive to acoustic trauma, and therefore must originate from the hearing epithelium. In guinea pigs, high-level DPOAEs may be present despite hair cell damage (Brown *et al.*, 1989). Apparently, components other that hair cell contribute to these emissions. We speculate that nonlinear (mechanical) properties of auxiliary structures may generate the high-level distortions. Possible candidates in the frog ear are passive nonlinear responses of supporting cells, the tectorial membranes, or the contact membranes, which separate the endolymphatic and perilymphatic fluids in the AP and the BP.

Beyond 2 h *post mortem*, DPOAEs were still present, but emission frequencies were no longer related to the frequency ranges of the AP and the BP. It is unclear how these distortion products are generated, but the changes in frequency range presumably reflect *post mortem* changes in the middle and inner ears. Since these distortions do not seem to be related to ''normal'' inner ear physiology, they are of no

further interest here. However, their presence shows that the detectability of DPOAEs does *not* imply that the inner ear is functioning normally.

Our main observation is that DPOAE behavior during the first 2 h *post mortem* in the frog is similar to that in the other vertebrates investigated: DPOAEs in response to low-level stimuli are physiologically vulnerable, while DPOAEs for higher stimulus levels are more robust. Nevertheless, our results suggest that important differences exist between the emission characteristics of the two inner ear papillae: the vulnerability of the low-level DPOAEs from the AP is consistent with the presence of an active process, while DPOAEs from the BP need not involve such an active process.

ACKNOWLEDGMENTS

This work was supported by a grants to PvD from the Netherlands Organization for Scientific Research (NWO) and the Heinsius Houbolt Foundation, and by a grant to PMN from the National Institutes of Health, No. DC-00222.

Assad, J. A., and Corey, D. P. (**1992**). ''An active motor model for adaptation by vertebrate hair cells,'' J. Neurosci. **12**, 3291–3309.

Assad, J. A., Hacohen, N., and Corey, D. P. (**1989**). ''Voltage dependence of adaptation and active bundle movement in bullfrog saccular hair cells,'' Proc. Natl. Acad. Sci. U.S.A. **86**, 2918–2922.

Benser, M. E., Marquis, R. E., and Hudspeth, A. J. (**1996**). ''Rapid, active hair bundle movements in hair cells from the bullfrog's sacculus,'' J. Neurosci. **16**, 5629–5643.

Brown, A., McDowell, B., and Forge, A. (**1989**). ''Effects of chronic gentamycin treatment on hair cells can be monitored using acoustic distortion products,'' Hear. Res. **42**, 143–156.

Brownell, W. E., Bader, C. R., Bertrand, D., and De Ribaupierre, Y. (**1985**). ''Evoked mechanical responses of isolated cochlear outer hair cells,'' Science **227**, 194–196.

Christensen-Dalsgaard, J., and Narins, P. M. (**1993**). ''Sound and vibration sensitivity of VIIIth nerve fibers in the frogs *Leptodactylus albilabris* and *Rana pipiens pipiens*,'' J. Comp. Physiol., A **172**, 653–662.

Collet, L., Veuillet, E., Bene, J., and Morgon, A. (**1992**). ''Effects of contralateral white noise on click-evoked emissions in normal and sensorineural ears: towards an exploration of the medial olivocochlear system,'' Audiology **31**, 1–7.

Crawford, A. C., and Fettiplace, R. (**1985**). ''The mechanical properties of ciliary bundles of turtle cochlear hair cells,'' J. Physiol. (London) **364**, 359–379.

Frishkopf, L. S., and Flock, A. (**1974**). ''Ultrastructure of the basilar papilla, and auditory organ in the bullfrog,'' Acta Oto-laryngol. **77**, 176–184.

Frolenkov, G. I., Belyantseva, I. A., Kurc, M., Mastroianni, M. A., and Kachar, B. (**1998**). ''Cochlear outer hair cell electromotility can provide force for both low and high intensity distortion product otoacoustic emissions,'' Hear. Res. **126**, 67–74.

Gold, T. (**1948**). ''Hearing II. the physical basis of the action of the cochlea,'' Proc. R. Soc. E. **B135**, 492–498.

He, D. Z. Z., Beisel, K. W., Chen, L., Ding, D.-L., Jia, S., Fritsch, B., and Salvi, R. (**2003**). ''Chick hair cells do not exhibit voltage-dependent somatic motility,'' J. Physiol. (London) **546**, 511–520.

Hellmann, B., and Fritzsch, B. (**1996**). ''Neuroanatomical and histochemical evidence for the presence of common lateral line and inner ear efferents and of efferents to the basilar papilla in a frog, *Xenopus laevis*,'' Brain Behav. Evol. **47**, 185–194.

Howard, J., and Hudspeth, H. J. (**1988**). ''Compliance of the hair bundle associated with gating of the mechanoelectrical transduction channels in the bullfrog's saccular hair cell,'' Neuron **1**, 189–199.

Howard, J., Roberts, W. M., and Hudspeth, H. J. (**1988**). ''Mechanoelectrical transduction by hair cells,'' Annu. Rev. Biophys. Biophys. Chem. **17**, 99–124.

Hudspeth, A. J., Choe, Y., Mehta, A. D., and Martin, P. (**2000**). ''Putting ion channels to work: mechanoelectrical transduction, adaptation, and amplification by hair cells,'' Proc. Natl. Acad. Sci. U.S.A. **97**, 11765–11772.

Kemp, D. T., and Brown, A. M. (**1984**). ''Ear canal acoustic and round window electrical correlates of $2f_1-f_2$ distortion generated in the cochlea,'' Hear. Res. **13**, 39–46.

Liberman, M. C., Gao, J., He, D. Z. Z., Wu, X., Jia, S., and Zuo, J. (**2003**). ''Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier,'' Nature (London) **419**, 300–304.

Lonsbury-Martin, B. L., Martin, G. K., Probst, R., and Coats, A. C. (**1987**). ''Acoustic distortion products in rabbit ear canal. I. Basic features and physiological vulnerability,'' Hear. Res. **28**, 173–189.

Manley, G. A. (**2001**). ''Evidence for an active process and a cochlear amplifier in nonmammals,'' J. Neurophysiol. **86**, 541–549.

Manley, G. A., Köppl, C., and Johnstone, B. M. (**1993**). ''Distortion-product otoacoustic emissions in the bobtail lizard. I. General characteristics,'' J. Acoust. Soc. Am. **93**, 2820–2833.

Manley, G. A., Taschenberger, G., and Oeckinghaus, H. (**1999**). ''Influence of contralateral acoustic stimulation on distortion-product and spontaneous otoacoustic emissions in the barn owl,'' Hear. Res. **138**, 1–12.

Martin, P., and Hudspeth, A. (**1999**). ''Active hair-bundle movements can amplify a hair cell's response to oscillatory mechanical stimuli,'' Proc. Natl. Acad. Sci. U.S.A. **7**, 14306–14311.

Mills, D. M., Norton, S. J., and Rubel, E. W. (**1993**). ''Vulnerability and adaptation of distortion product otoacoustic emissions to endocochlear potential variation,'' J. Acoust. Soc. Am. **94**, 2108–2122.

Rebillard, G., Klis, J. F. L., Lavigne-Rebillard, M., Devaux, P., Puel, J. L., and Pujol, R. (**1993**). ''Changes in $2f_1-f_2$ distortion product otoacoustic emissions following alteration of cochlear metabolism,'' Br. J. Audiol. **27**, 117–121.

Robbins, R. G., Bauknight, B. S., and Honrubia, M. D. (**1967**). ''Anatomical distribution of the efferent fibers in the VIIIth cranial nerve of the bullfrog (*Rana catesbeiana*),'' Acta Otolaryngol. **64**, 436–448.

Ronken, D. A. (**1990**). ''Basic properties of auditory-nerve responses from a ''simple'' ear: The basilar papilla of the frog,'' Hear. Res. **47**, 63–82.

Ronken, D. A. (**1991**). ''Spike discharge properties that are related to the characteristic frequency of single units in the frog auditory nerve,'' J. Acoust. Soc. Am. **90**, 2428–2440.

Van Dijk, P., and Manley, G. A. (**2001**). ''Distortion product otoacoustic emissions in the tree frog *Hyla cinerea*,'' Hear. Res. **153**, 14–22.

Veuillet, E., Collet, L., and Duclaux, R. (**1991**). ''Effect of contralateral acoustic stimulation on active cochlear micromechanical properties in human subjects: dependence on stimulus variables,'' J. Neurophysiol. **65**, 724–735.

Whitehead, M. L., Lonsbury-Martin, B. L., and Martin, G. K. (**1992**). ''Evidence for two discrete sources of $2f_1-f_2$ distortion-product otoacoustic emission in rabbit II. Differential physiological vulnerability,'' J. Acoust. Soc. Am. **92**, 2662–2682.

Zheng, J., Shen, W., He, D. Z. Z., Long, K. B., Madison, L. D., and Dallos, P. (**2000**). ''Prestin is the motor protein of cochlear outer hair cells,'' Nature (London) **405**, 149–155.

Zurek, P. M., Clark, W. W., and Kim, D. O. (**1982**). ''The behavior of acoustic distortion products in the ear canals of chinchillas with normal or damaged ears,'' J. Acoust. Soc. Am. **72**, 774–780.

Electrical field interactions in different cochlear implant systems

Colette Boëx,[a] Chloé de Balthasar, Maria-Izabel Kós, and Marco Pelizzone
Centre Romand d'Implants Cochléaires, University Hospital, CH-1211 Geneva 14, Switzerland

(Received 28 January 2003; revised 12 July 2003; accepted 26 July 2003)

The goal of this study was to evaluate electrical field interactions produced by the stimulation of different types of intracochlear electrodes in 12 adult subjects (three Ineraid™, four Clarion™ S-Series, three S-Series with the electrode positioning system-EPS and two Clarion™ HiFocus-I with the EPS). Psychophysical measurements were conducted with biphasic stimuli (813 pulse per second, 153.8 μs/phase). ''Perturbation'' signals (300 ms) were applied to one electrode chosen at the middle of the array and their effects on detection thresholds of ''probe'' signals (30 ms) were measured on the neighbor basal electrode. Perturbation levels were set below the detection threshold of the perturbation electrode (-2 dB *re* threshold). Measurements were first conducted for simultaneous stimulation of the probe and of the perturbation electrodes, for monopolar for all subjects and for bipolar stimulus configurations for both Clarion™ HiFocus-I subjects. The tested Clarion™ electrodes did not present lower monopolar interactions than the Ineraid™ electrodes. Nevertheless, considering the shorter distance between electrodes for the Clarion™ than for the Ineraid™, the tested Clarion™ electrodes might be more selective than the Ineraid™. We did not find any significant monopolar electrical field-interaction differences between subjects who received the S-Series array with and without the EPS. We did not find lower interactions for both subjects who received the HiFocus-I array than for subjects who received the S-Series. Electrical field interactions were lower for bipolar than for monopolar configurations for both HiFocus-I subjects. A second set of measurements was conducted for nonsimultaneous stimulation similar to the one used in continuous interleaved sampling sound strategy. These measurements showed that interactions evaluated for simultaneous biphasic stimuli were larger than for nonsimultaneous stimulation. [DOI: 10.1121/1.1610451]

PACS numbers: 43.64.Me, 43.66.Dc [BLM]

I. INTRODUCTION

Most modern cochlear implant systems use sound coding strategies related to the continuous interleaved sampling strategy (CIS, Wilson *et al.*, 1991). In these strategies the auditory nerve fibers are excited by interleaved stimuli. These stimuli, usually biphasic pulses, carry envelopes of sounds of a limited number of frequency bands (8 to 22 per cycle in present systems). Stimuli are delivered on tonotopically ordered intracochlear electrodes according to the frequency bands they code. An increase in the number of frequency band signals delivered on respective intracochlear electrodes should improve the information transmitted on spectral structure of sounds. This benefit has been already shown in speech understanding in noisy environments (Friesen *et al.*, 2001). A different way of improving speech coding strategies would be to transmit in an efficient way the fine temporal structure of sounds which is important for speech perception and which has also been shown to be important for the perception of music and localization of sounds (Smith *et al.*, 2002). To transmit the fine temporal structure of sounds, the simultaneous stimulation of different intracochlear electrodes is usually required. This has been the method adopted in some sound coding strategies (Eddington, 1980; Hochmair-Desoyer *et al.*, 1983; Merzenich *et al.*, 1984). But there is evidence that simultaneous stimulation increases electrical field interactions across electrodes (Eddington *et al.*, 1978; Shannon, 1983b; White *et al.*, 1984; Hartmann and Klinke, 1989; Favre and Pelizzone, 1993; de Balthasar *et al.*, in press). The reason for using interleaved stimulation in CIS strategies was to minimize electrical field interactions across the intracochlear electrodes.

The efficiency of simultaneous stimulation of intracochlear electrodes depends on their selectivity and the latter is modulated by stimulus configuration, electrode positions, or electrode design. Many studies were conducted on the effects of stimulus configuration on stimulation selectivity. In monopolar stimulus configuration, the electrode of excitation is one of the intracochlear electrodes and the electrical return path can be an external electrode placed under the temporal muscle, as with Ineraid subjects (Eddington, 1980) or the base of the implanted receiver-stimulator case, as with Clarion subjects (Advanced Bionics Corporation, Sylmar, CA). In bipolar configuration, excitation and return electrodes are both chosen among the intracochlear electrodes. Many studies were devoted to analyze the selectivity of a bipolar configuration in comparison to a monopolar configuration. They were conducted physiologically in animals (Merzenich and White, 1977; Black and Clark, 1980; van den Honert and Stypulkowski, 1987; Ryan *et al.*, 1990; Kral *et al.*, 1998; Rebscher *et al.*, 2001; Bierer and Middlebrooks, 2002), in humans through psychophysical measures (Shannon, 1983b; White *et al.*, 1984; Lim *et al.*, 1989), and through measures of evoked auditory potentials (Brown

[a]Electronic mail: colette.boex@hcuge.ch

TABLE I. Month and year of birth and implantation, etiology of deafness, and type of electrode array implanted for each subject.

Subject	Birth	Implantation	Etiology of deafness	Electrode array
I03	Jan. 1928	Sep. 1988	Trauma+Mondini	Ineraid
I09	Jan. 1953	Apr. 1988	Streptomycin+Genetic?	Ineraid
I34	Nov. 1940	May 1997	Viral Labyrinthitis+Otitis	Ineraid
C05	Sep. 1939	Nov. 1997	Ménière's Syndrome	S-Series
C06	Mar. 1962	Dec. 1997	Otosclerosis	S-Series
C24	Jan. 1956	July 1999	Unknown (progressive)	S-Series
C30	Jun. 1937	May 2000	Unknown	S-Series
Cp08	Nov. 1953	Jun. 1998	Streptomycin ototoxicity	S-Series with EPS
Cp14	Jun. 1943	Nov. 1998	Multiple Otitis	S-Series with EPS
Cp18	Sep. 1949	May 1999	Traumatic (+progressive)	S-Series with EPS
H26	Jan. 1947	Nov. 1999	Multiple Otitis	HiFocus-I with EPS
H29	Jul. 1958	Apr. 2000	Congenital	HiFocus-I with EPS

et al., 1996). In most studies, the monopolar configuration is found to excite a larger population of neurons than the bipolar configuration for the same current level while the monopolar configuration would produce a louder sensation for the same current level.

Another aspect worth considering is the proximity of the electrode to the modiolus. Kawano *et al.* (1998) have shown that the electrical detection thresholds decreased with the electrode distance to the Rosenthal's canal in humans. Shepherd *et al.* (1993) also measured lower evoked auditory potential thresholds in implanted cats when electrodes were closer to the modiolus. These observations are in agreement with models of the stimulated implanted human cochlea (Frinjs *et al.*, 2001; Hanekom, 2001). These issues lead to the development of new designs of electrode arrays. For instance, the precurved Nucleus™ Contour electrode (Cohen *et al.*, 2001) is designed to come closer to the modiolus after insertion. Advanced Bionics (Sylmar, CA) also developed the electrode positioning system (EPS; Kuzma and Balkany, 1999; Fayad *et al.*, 2000) that consists of an independent silicone part (a ''positioner'') inserted inside the cochlea, laterally to the array, in an attempt to bring it close to the modiolus. Also, the Clarion™ HiFocus-I electrode array (Kuzma and Balkany, 1999) affords to produce simultaneous or nonsimultaneous stimulation of 16 intracochlear electrodes, in monopolar or in bipolar configurations.

We conducted psychophysical measurements in order to characterize the spread of electrical fields and the selectivity of electrodes of different cochlear implant arrays. We evaluated these interactions by applying a ''perturbation'' signal to a selected electrode to measure its effect on the threshold of a ''probe'' signal applied to an adjacent electrode. Our experiments were based on the hypothesis that the current from the perturbation electrode spreads itself close to the fibers excited by a probe current and modifies the probe detection threshold. The modifications in the perturbed probe threshold would describe the effects of the interactions of the perturbation electrode on the probe electrode.

We evaluated interactions obtained for different types of electrode arrays worn by 12 subjects in monopolar configuration (Ineraid™, Clarion™ S-Series alone or implanted in conjunction with the EPS and the Clarion™ HiFocus-I in conjunction with the EPS) and in bipolar configuration for two Clarion™ HiFocus-I+EPS subjects. We also evaluated interactions with nonsimultaneous sequential stimulation of adjacent electrodes in monopolar configuration to compare the interactions resulting from simultaneous stimulation to the interactions resulting from nonsimultaneous stimulation, like in CIS strategies.

II. METHODS

A. Subjects

Twelve subjects participated in this study. Three subjects received the Ineraid™ implant (I03, I09 and I34), four subjects (C05, C06, C24 and C30) received the standard Clarion™ S-Series electrode array, three subjects (Cp08, Cp14 and Cp18) received the Clarion™ S-Series electrode array with EPS, and two subjects (H26 and H29) received the HiFocus-I electrode with EPS. At the time of this study, the Ineraid subjects have been using a CIS strategy for at least 4 years (Pelizzone *et al.*, 1995, 1999). The Clarion subjects who received the S-Series electrode array have been using a CIS strategy. The Clarion subjects who received the HiFocus-I electrode array have been using a SAS strategy (simultaneous analogue strategy, Battmer *et al.*, 1999; Osberger and Fisher, 1999). All Clarion subjects have been using their implant for at least 1 year.

Their causes of deafness were Mondini malformation, Ménière's disease, progressive, traumatic, streptomycin, or unknown (Table I). At the time of the study, the youngest subject was 39 years old and the oldest was 73 years old (mean: 53 years and 10 months). All subjects were volunteers. They were selected because of their willingness to participate in this study. This study followed the ethical guidelines of the Declaration of Helsinki.

B. Experimental design

The interactions between electrodes were evaluated by applying a ''perturbation'' signal to a selected electrode at the middle of the array in order to measure its effect on the threshold of a ''probe'' signal applied to an adjacent electrode. Psychophysical detection thresholds were measured

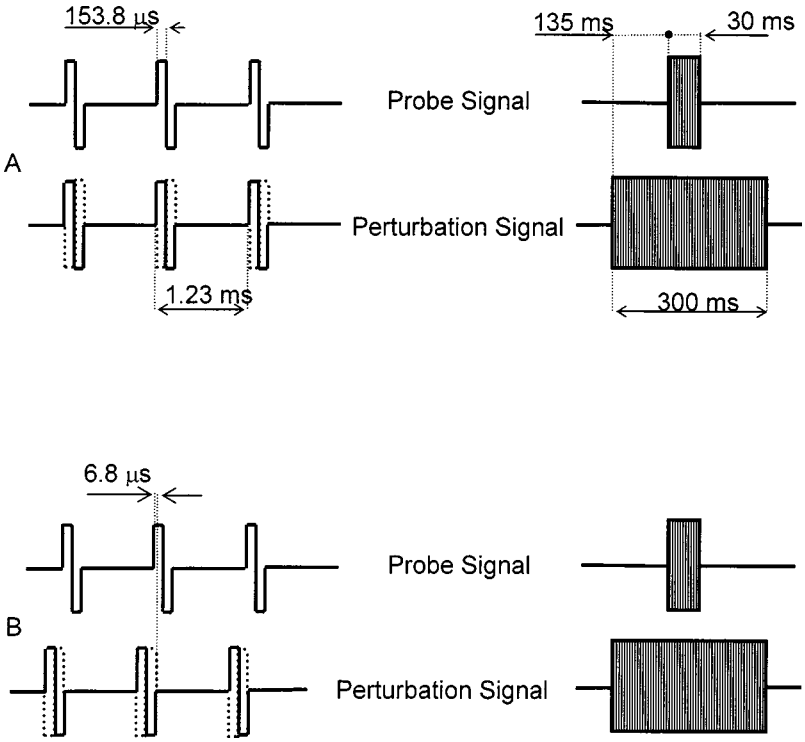


FIG. 1. Stimuli were all biphasic (813 pulses per second, 153.8 μs/phase). Perturbation signals lasted 300 ms (zero rise-decay times). Probe signals lasted 30-ms duration (zero rise-decay times). The probe signals overlapped the perturbation signals and occurred 135 ms after the onset of the perturbation signal. Perturbation signals were either positive or negative first biphasic pulses (dashed lines). (A) Perturbation and probe signals were presented simultaneously. (B) Perturbation and probe signals were presented sequentially.

using a three-alternative-forced-choice (3AFC) ''one up/two down'' adaptive procedure (one single incorrect response results in an increase in the probe signal by one step, two successive correct responses result in a decrease in the probe signal by one step). This procedure converges to a level where approximately 70.7% of presentations are detected (Levitt, 1971). Each observation interval was indicated by successive flashes in three alternative boxes displayed on a computer screen. The subjects were asked to indicate in which interval they could perceive a sound. Feedback was provided. Each measurement started with a clearly detectable signal. One detection threshold measurement was obtained after gathering six reversals (descending and ascending segments). Signals were decreased initially by steps set to about 15% of the initial amplitude. For the second and third reversals the steps were halved. One threshold was specified by the last four descending segments of the staircase procedure. The detection thresholds indicated in the present study were calculated as the mean of at least two threshold measurements. Using this psychophysical procedure, the detection thresholds of the probe signal alone and of the perturbation signal alone were first measured. Then the detection thresholds of the probe signal together with the perturbation signal were measured. The probe signal occurred in addition to the perturbation signal randomly in one of the three intervals.

Interaction measurements were conducted for biphasic stimuli (813 pulse per second, 153.8 μs/ph). The perturbation signals were set 2 dB below the detection threshold of the perturbation electrode. First [Fig. 1(a)] the perturbation pulses were presented simultaneously to the probe pulses, in phase or out of phase with the probe signal (plain or dashed lines). Second [Fig. 1(b)] the perturbation pulses were presented sequentially (short 6.8 μs overlap) to the probe pulses and could be either positive or negative first (plain or dashed lines).

All psychophysical measurements were realized using the Clarion™ Research Interface (CRI, Wygonski *et al.*, 1999) which allows control of all stimulus parameters. The exact current amplitudes delivered on intracochlear electrodes were not perfectly proportional to the clinical units over the entire electrical output range. They were slightly dependent on electrode impedance. These limitations can lead to measurement error in psychophysical experiments. Therefore we devoted special attention to improve the control of the symmetry and the amplitude of current biphasic pulses. For this, we used special lookup tables based on calibrations. One table was built measuring, for each clinical unit, the effective current delivered by the current source of the research implantable cochlear stimulator (RICS). We defined from these tables the clinical units required to provide the desired current amplitudes. These tables were defined for each electrode impedance reported for all subjects in the Ap-

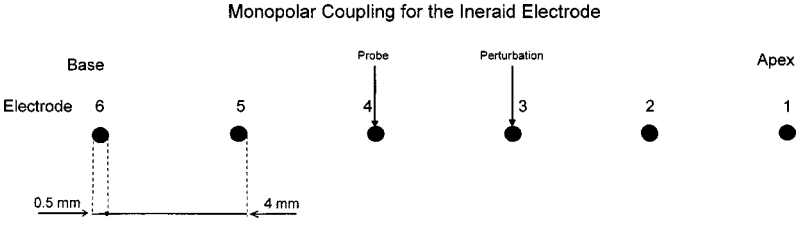


FIG. 2. Scheme of the Ineraid™ electrode array. Each electrode is a platinum 0.5-mm-diam ball. Electrodes were numbered from the most apical electrode to the most basal pair from 1 to 6. The distance between electrodes is about 4 mm. The probe and perturbation electrodes are indicated.

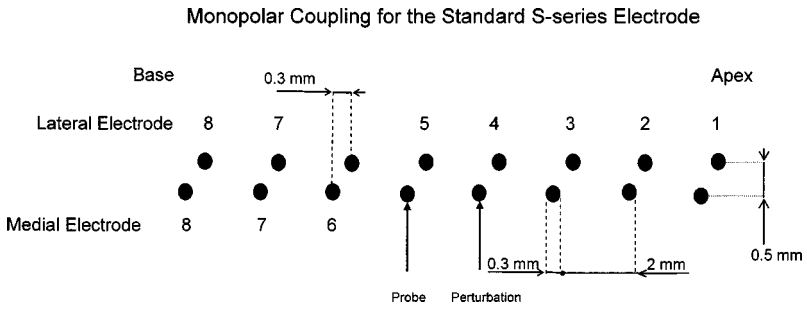


FIG. 3. Scheme of the standard Clarion™ S-Series electrode array. Each electrode is iridium-platinum (90:10) 0.3-mm-diam ball. Electrode pairs (one medial, one lateral) are numbered from the most apical electrode pair to the most basal pair from 1 to 8. The distance between pairs of electrodes is 2 mm. The probe and perturbation signals are indicated.

pendix. Monopolar impedances were evaluated from the clinical setup impedance measurements[1] (Clarion™ SCLIN) using sine waves (1000 Hz) at low stimulation levels. We also evaluated monopolar impedances of Ineraid™ electrodes using sine waves (1000 Hz) at low stimulation levels. In monopolar configuration, maximum impedances and currents used in the present study were low (respectively 27 kΩ and 60 μA). In these limits current sources behaved almost as perfect current sources. As an example, if impedance decreased from 27 to 20 kΩ (26% lower), stimulation amplitude would increase from 60 to 66 μA, or from 35 to 38 μA.

In bipolar configuration, we did not report measurements conducted with the Clarion™ S-Series subjects because the lateral electrodes were not stimulated in the CIS speech strategy they used. Actually nonstimulated electrodes had unusual high impedances that quickly decreased if electrodes were stimulated. This impedance drops consequent to short stimulation periods had been observed previously (Loeb *et al.*, 1983; Charlet de Sauvage *et al.*, 1997). Because the HiFocus-I subjects were SAS speech strategy users and because that strategy stimulates all electrodes, monopolar impedance of each electrode used could be considered and summed to define the bipolar impedances for these subjects. In bipolar configuration, maximum impedances and maximum stimulation currents found were 150 μA for 25 kΩ or 200 μA for 15 kΩ. If impedance decreased from 15 to 11 kΩ (26% lower), stimulation amplitude would increase from 200 to 219 μA. If impedance decreased from 25 to 18 kΩ (26% lower), stimulation amplitude would increase from 150 to 166 μA. We connected directly the channel output of the RICS to the Ineraid™ electrodes via their percutaneous plug to conduct these measurements in Ineraid subjects.

C. Electrode types

Three different implanted systems have been tested. The **Ineraid**™ electrode array has six intracochlear platinum electrodes. The distance between the centers of electrodes is approximately 4 mm (Fig. 2). One extra cochlear electrode is placed under the temporalis muscle and is used as a far-field ground electrode. In the monopolar configuration, the stimuli were applied between one intracochlear electrode and the ground electrode. The probe signals were applied to electrode 4 and the perturbation signals were applied to electrode 3 at the middle of the array. The distance between probe and perturbation electrodes was approximately 4 mm.

The standard **Clarion™ S-Series** electrode array (Kessler, 1999) was implanted either alone or with the **Electrode Positioning System** (EPS). The Clarion™ S-Series electrode has eight pairs of electrodes. The distance between electrodes is 2 mm (Fig. 3). Each pair consists of one medial and one lateral electrode positioned radially, 0.6 mm distant from each other. In monopolar configuration, the stimuli were applied between a medial electrode and the base of the implanted receiver-stimulator case used as a far field ground in the Clarion™ S-series system. The probe signals were applied to medial electrode 5 and the perturbation signals were applied to medial electrode 4 at the middle of the array.

The **Clarion™ HiFocus-I** electrode array has 16 electrodes. The distance between the centers of electrodes is 1 mm (Fig. 4). All subjects who received the HiFocus-I Electrode array also received the EPS. In monopolar configuration, the stimuli were applied between one electrode and the base of the implanted receiver-stimulator case used as a far-field ground in the Clarion™ HiFocus-I system. The probe signals were applied to electrode 10 and the perturbation signals were applied to electrode 8 at the middle of the array. The distance between the probe and the perturbation electrodes was about 2 mm. In the bipolar configuration, the stimuli were applied between two adjacent electrodes with the basal electrode used as a ground electrode. Thus, the distance between the source and the ground electrodes in bipolar configuration was 1 mm. The probe signals were applied between electrodes 10 and 11 and the perturbation signals were applied between electrodes 8 and 9. The nominal distance between the probe and the perturbation-electrode pairs was 2 mm.

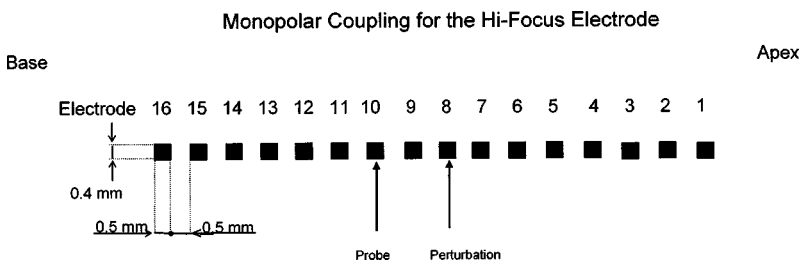


FIG. 4. Scheme of the Clarion™ HiFocus-I electrode array. Each contact is rectangular in shape (0.4×0.5 mm) and made from pure platinum. Electrodes are numbered from the most apical electrode to the most basal from 1 to 16. Distance between electrodes is 1 mm. The probe and perturbation electrodes are indicated for monopolar and bipolar configuration.

TABLE II. Amplitude of perturbation signals used in monopolar and in bipolar configurations. They were set 2 dB under the detection threshold of the perturbation electrode alone. Interaction measurements were not conducted in bipolar configuration for Ineraid and Clarion™ S-Series subjects.

Subject	Perturbation signal monopolar configuration (μA)	Perturbation signal bipolar configuration (μA)
I03	19	···
I09	19	···
I34	20	···
C05	13	···
C06	18	···
C24	18	···
C30	18	···
Cp08	13	···
Cp14	20	···
Cp18	10	···
H26	19	79
H29	14	91

III. RESULTS

A. Simultaneous and nonsimultaneous monopolar interactions

Simultaneous and nonsimultaneous interaction experiments were conducted presenting subthreshold signals on the perturbation electrodes. These levels were set 2 dB under the detection threshold of the perturbation electrodes. The perturbation levels adopted for each subject are listed in Table II.

Figure 5 describes changes of the probe-detection thresholds due to the simultaneous stimulation of the perturbation electrodes. All thresholds were expressed in μA peak and were obtained in monopolar configuration. The black squares in this figure are the detection thresholds of the probe alone electrode. The probe thresholds obtained by Ineraid subjects were higher than those of Clarion subjects ($t=3.5$, $p<0.001$). We did not find any difference in probe threshold between the S-Series subjects with and without an EPS ($t=0.82$, $p=0.43$). The probe thresholds decreased significantly when in phase perturbation signals (triangles pointing upwards) were presented with the probe signals ($t=-15.7$, $p<0.001$). The probe thresholds increased significantly when out of phase perturbation signals (triangles pointing downwards) were presented ($t=12.8$, $p<0.001$). Most subjects showed important interaction effects which could be observed through the increases and the decreases of the probe thresholds symmetrical to the probe alone thresholds.

Figure 6 describes changes in the probe-detection thresholds due to nonsimultaneous stimulation of the perturbation electrodes. When the perturbation and the probe signals were presented with positive first phases (triangles pointing upwards), the probe thresholds slightly increased ($t=2.3$, $p<0.03$). On average, the probe thresholds decreased slightly when the probe signal was presented with a positive first phase and the perturbation signal was presented with a negative first phase (triangles pointing downwards; $t=-2.5$, $p<0.02$). Threshold changes were much more significant for simultaneous stimuli than for nonsimultaneous stimuli.

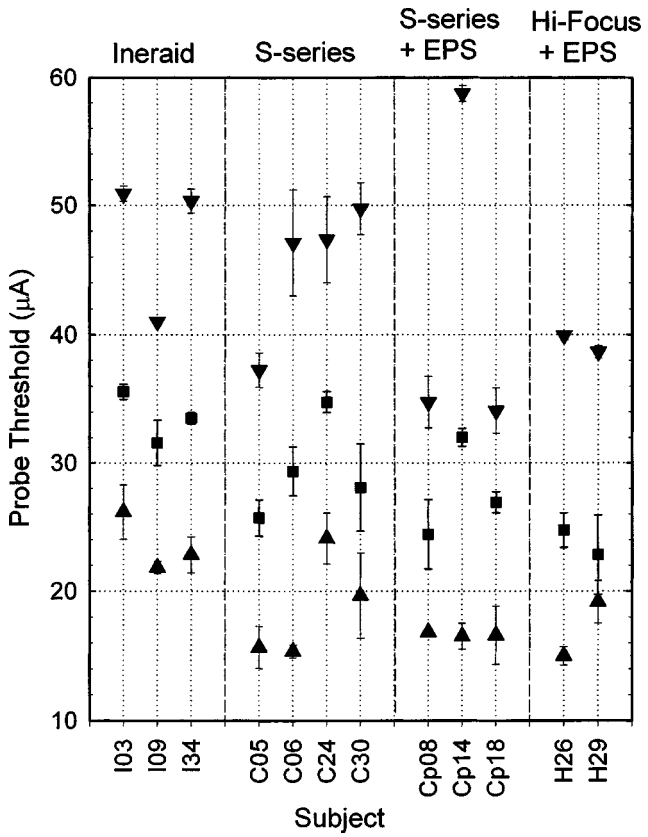


FIG. 5. Interaction measurements conducted for simultaneous stimulation. Black squares represent probe alone thresholds (same as in Fig. 6). When the perturbation signal is presented in phase with the probe signal, the probe threshold is represented by a triangle pointing upwards. When the perturbation signal is presented out of phase with the probe signal, the probe threshold is represented by a triangle pointing downwards.

B. Simultaneous bipolar interactions

Figure 7 shows probe thresholds (white squares) obtained in bipolar configurations for both HiFocus-I subjects. As expected probe thresholds obtained in monopolar configuration were much lower (about four to seven folds) than thresholds obtained in bipolar configuration for both subjects. In bipolar configuration, simultaneous interactions caused also a significant decrease in probe thresholds when perturbation signals were presented in phase. When perturbation and probe signals were out of phase (triangles pointing downwards) an increase in probe thresholds was observed in both subjects.

IV. DISCUSSION

The symmetry between in phase and out of phase stimuli shown for most of the simultaneous subthreshold measurements (Figs. 5 and 7) indicated in this case that direct electrical field summation was responsible for the observed interactions in this condition. It is of interest to note that this phenomenon is similar in the four types of electrode arrays tested. Such interactions had been shown previously for bi-

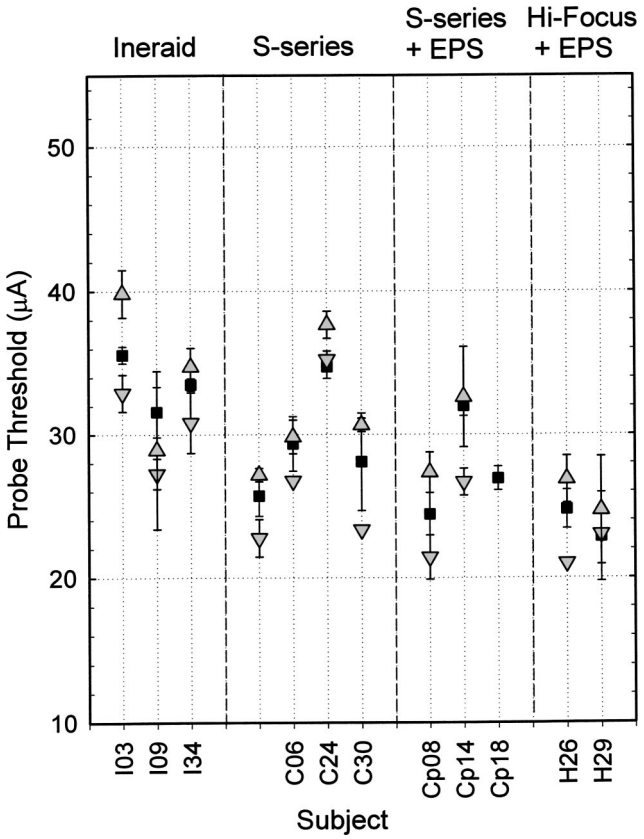


FIG. 6. Interaction measurements conducted for nonsimultaneous stimulation. Black squares represent probe alone thresholds. When perturbation and probe signals were presented with a positive first phase, the symbols are triangles pointing upwards. When probe signals were presented with a positive first phase and the perturbation signals were presented with a negative first phase, symbols are triangles pointing downwards.

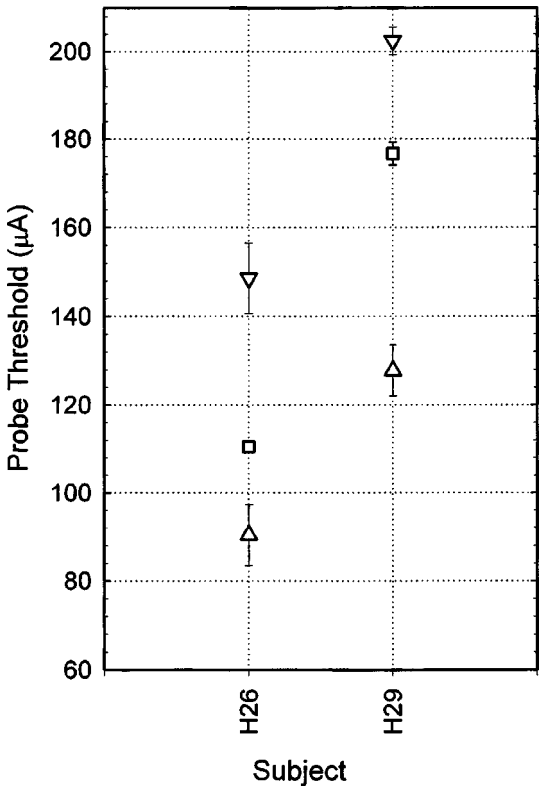


FIG. 7. Simultaneous interactions conducted in bipolar configuration for HiFocus-I subjects. White squares represent probe alone thresholds. Symbols for perturbation and probe signals in phase are triangles pointing upwards. Symbols for perturbation and probe signals out of phase are triangles pointing downwards.

phasic stimuli (Eddington, 1978; White *et al.*, 1984; Favre and Pelizzone, 1993) and in a recent paper (de Balthasar *et al.*, 2003) where they were discussed in detail.

To compare interactions across subjects, we computed a simultaneous interaction index (Eddington *et al.*, 1991) which normalizes the probe threshold change $(THR_{\text{Probe}}(\text{Pert}-) - THR_{\text{Probe}}(\text{Pert}+), \mu\text{A})$ relative to the level of perturbation signal $(I_{\text{Pert}}, \mu\text{A})$ applied. We computed the following ratio $THR_{\text{Probe}}(\text{Pert}-) - THR_{\text{Probe}}(\text{Pert}+)/(2*I_{\text{Pert}})$ which describes the threshold change of the probe electrode induced by the presence of the perturbation signal presented to the neighbor basal electrode. The smaller this ratio is (min=0), the smaller interactions are and the higher this ratio is (max=1), the higher interactions are. Figure 8 shows this index for all subjects and for both configurations for both HiFocus-I subjects. Black symbols indicated the simultaneous interaction indexes obtained in monopolar configuration. The monopolar indexes were not significantly lower for the Ineraid subjects than for the other subjects ($t = -2.11$, $p = 0.06$) while the distance between probe and perturbation electrodes for the Ineraid subjects (4 mm) was twice as large as those of the other subjects (2 mm). This observation indicates that the Clarion™ electrodes were not less selective than the Ineraid™. It suggests even that they may be more selective than the Ineraid™ considering that the distances between electrodes is shorter for the Clarion™ electrode arrays. In addition we found higher detection thresholds for Ineraid subjects in comparison to Clarion subjects. The location of the electrodes in the cochlea which partly determines the distance between electrodes and fibers to be excited could be of importance in determining the electrical excitation thresholds (Shepherd *et al.*, 1993; Kawano *et al.*, 1998; Frinjs *et al.*, 2001; Hanekom, 2001) as well as the shape of the electrodes influencing the electrical field in the cochlea.

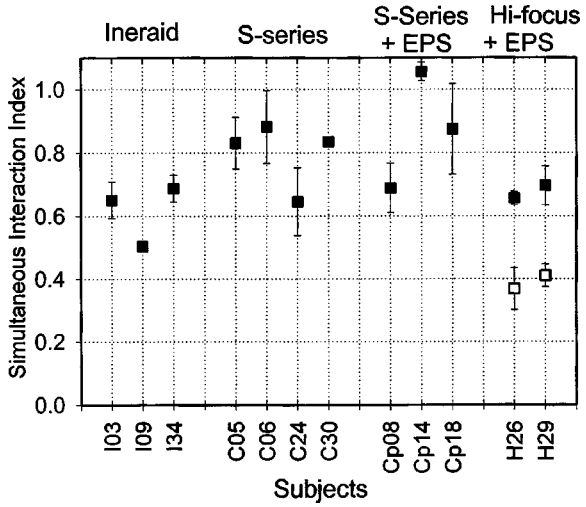


FIG. 8. Simultaneous interactions index: $THR_{\text{Pr}\,obe}(Pert-) - THR_{\text{Pr}\,obe}(Pert+)/(2*I_{Pert})$. Black symbols describe the simultaneous interaction indexes obtained in monopolar configuration. Open symbols indicate the interaction indexes obtained in bipolar configuration. Error bars are the standard deviations.

We did not find differences in probe-detection thresholds between the S-Series subjects with or without the EPS. Young and Grohne's (2001) or Donaldson *et al.*'s (2001) studies indicated lower detection thresholds and most comfortable loudness levels for children receiving the EPS. Although our observations contradict these studies, we can not verify that the EPS specifically placed the electrodes 4 and 5 closer to the modiolus in the three subjects who received the S-Series with EPS. Interaction indexes (Fig. 8) also show that simultaneous interactions were not different between subjects implanted with EPS and without EPS. In Fig. 8 open symbols indicate the indexes obtained in bipolar configuration for both HiFocus-I subjects. They were lower in bipolar configuration than in monopolar configuration for both subjects. A similar observation has been made by other investigators with other implant devices in humans. White *et al.* (1984) conducted simultaneous interaction measurements in one subject by loudness summation with biphasic stimuli (100 μs/ph, 100 pps, 300 ms). They obtained lower interactions for a near radial bipolar configuration than for a monopolar configuration. More recently and for similar S-Series with or without EPS and for HiFocus-I electrode arrays, Stickney *et al.* (2001) obtained also lower simultaneous interactions in bipolar configuration.

In addition, we noticed higher detection thresholds in bipolar configuration than in monopolar configuration. The same observation has been already made (Shannon, 1983a; Hartmann and Klinke, 1989; Busby *et al.*, 1994; Pfingst *et al.*, 1995).

The comparison of probe threshold changes that we obtained for simultaneous (Fig. 5) and for nonsimultaneous interactions (Fig. 6) indicated that nonsimultaneous stimulation presents much fewer interactions across electrodes for the four types of electrode arrays tested than simultaneous stimulation. But we also observed that when a positive last phase-perturbation signal occurred just before a positive first probe signal (downwards triangles), the probe thresholds were slightly lower. On the contrary, when a negative last phase-perturbation signal occurred just before a positive first probe signal (upwards triangles), the probe thresholds increased slightly. This suggests that the short (6.8 μs) overlap of the perturbation signals on the probe signals improved or worsened the probe detection according to its polarity. This observation could also be due to residual polarization. de Balthasar *et al.* observed the same results with subthreshold tests conduced with strictly sequential biphasic stimuli in Ineraid subjects. In addition, in this study, this effect was studied and discussed in regard to the level of the perturbation signal and in regard to the delay between the perturbation and the probe signals.

V. CONCLUSION

Interactions evaluated for simultaneous biphasic stimuli were larger than for nonsimultaneous stimuli across four different types of electrode arrays. Monopolar electric field interactions in the Clarion subjects were similar to those in Ineraid subjects. Hence, the tested Clarion™ electrodes were as selective as the Ineraid™ for shorter distances between electrodes. We did not find differences between S-Series subjects with or without the EPS. We did find lower detection thresholds for both subjects who received the HiFocus-I array than for subjects who received the S-Series. Nevertheless, monopolar subthreshold interactions were similar. Interactions were lower in bipolar configuration than in monopolar configurations for both HiFocus-I subjects.

These subthreshold measurements do not provide information on the spread of excitation at higher levels of stimulation, such as those used in speech coding strategies. To complete the present data we conducted a second study (Boëx *et al.*, 2003) designed to evaluate the neural interactions in the same group of subjects.

ACKNOWLEDGMENTS

This work has been supported by the ''Fonds National pour la Recherche Scientifique,'' Contract No. 315388.98. Most of all we are thankful to the subjects who participated in this study for the time and the effort they devoted to our study. We would like to thank R. V. Shannon and one anonymous reviewer for their constructive comments on the manuscript. We would like to thank Patrick Boyle for his help in programming the Clarion™ Research Interface. Finally, we thank all members of the ''Center Romand d'Implants Cochléaires.''

APPENDIX: ELECTRODE IMPEDANCES IN MONOPOLAR CONFIGURATION

TABLE III. Electrode impedances (kΩ) measured in monopolar configuration of all Ineraid subjects who participated in the study.

Ineraid subjects	Electrode 1 impedance (kΩ)	Electrode 2 impedance (kΩ)	Electrode 3 impedance (kΩ)	Electrode 4 impedance (kΩ)
I03	9.1	9.3	10.3	8.5
I09	5.8	5.6	7.5	7.8
I34	7.2	6.6	6.6	6.6

TABLE IV. Electrode impedances (kΩ) measured in monopolar configuration of all S-Series subjects who participated in the study.

S-Series subjects	Electrode 1 impedance (kΩ)	Electrode 2 impedance (kΩ)	Electrode 4 impedance (kΩ)	Electrode 5 impedance (kΩ)	Electrode 6 impedance (kΩ)	Electrode 8 impedance (kΩ)
C05	21	19	20	15	24	29
C06	30	20	23	27	27	23
C30	13	13	17	17	17	14
C24	20	17	19	23	22	25

TABLE V. Electrode impedances (kΩ) measured in monopolar configuration of all S-Series+EPS subjects who participated in the study.

S-Series +EPS subjects	Electrode 1 impedance (kΩ)	Electrode 2 impedance (kΩ)	Electrode 4 impedance (kΩ)	Electrode 5 impedance (kΩ)	Electrode 6 impedance (kΩ)	Electrode 8 impedance (kΩ)
Cp08	13	17	20	18	18	26
Cp14	18	24	19	18	20	19
Cp18	18	11	16	16	18	24

TABLE VI. Electrode impedances (kΩ) measured in monopolar configuration of all HiFocus-I subjects who participated in the study.

HiFocus-I subjects	El. 2 Imp. (kΩ)	El. 3 Imp. (kΩ)	El. 4 Imp. (kΩ)	El. 5 Imp. (kΩ)	El. 8 Imp. (kΩ)	El. 9 Imp. (kΩ)	El. 10 Imp. (kΩ)	El. 11 Imp. (kΩ)	El. 12 Imp. (kΩ)	El. 13 Imp. (kΩ)	El. 14 Imp. (kΩ)
H26	26	21	18	15	13	10	10	15	18	18	14
H29	12	15	13	11	8	12	8	7	8	8	9

[1]Tables were built for our RICS; they could not be built for the implanted stimulator of each subject as we could not measure current amplitudes delivered on implanted electrodes as well as impedances with our present research interface. We were not aware of any reports on description of impedance changes with amplitude of stimulation in Clarion™ S-Series or HiFocus-I electrodes.

Battmer, R. D., Zilberman, Y., Haake, P., and Lenarz, T. (**1999**). ''Simultaneous Analog Stimulation (SAS)-Continuous Interleaved Sampler (CIS) pilot comparison study in Europe,'' Ann. Otol. Rhinol. Laryngol. **108**, 69–73.

Black, R. C., and Clark, G. M. (**1980**). ''Differential electrical excitation of the auditory nerve,'' J. Acoust. Soc. Am. **67**, 868–874.

Boëx, C., Kós, M.-I., and Pelizzone, M. (**2003**). ''Forward masking in different cochlear implant systems,'' J. Acoust. Soc. Am. **114**, 2058–2065.

Bierer, J. A., and Middlebrooks, J. C. (**2002**). ''Auditory cortical images of cochlear-implant stimuli: dependence on electrode configuration,'' J. Neurophysiol. **87**(1), 478–492.

Brown, C. J., Abbas, P. J., Borland, J., and Bertschy, M. R. (**1996**). ''Electrically evoked whole nerve action potentials in Ineraid cochlear implant users: Responses to different stimulating electrode configurations and comparison to psychophysical responses,'' J. Speech Hear. Res. **39**, 453–467.

Busby, P. A., Whitford, L. A., Blamey, P. J., Richardson, L. M., and Clark, G. M. (**1994**). ''Pitch perception for different modes of stimulation using the Cochlear multiple-electrode prosthesis,'' J. Acoust. Soc. Am. **95**, 2658–2669.

Charlet de Sauvage, R., Lima da Costa, D., Erre, J.-P., and Aran, J. M. (**1997**). ''Electrical and physiological changes during short-term and chronic electrical stimulation of the normal cochlea,'' Hear. Res. **110**, 119–134.

Cohen, L. T., Saunders, E., and Clark, G. M. (**2001**). ''Psychophysics of a prototype peri-modiolar cochlear implant electrode array,'' Hear. Res. **155**, 63–81.

de Balthasar, C., Boëx, C., and Pelizzone, M. (**2003**). ''Channel interactions with high-rate biphasic electric stimulation in cochlear implant subjects,'' Hear. Res. **182**, 77–87.

Donaldson, G. S., Peters, M. D., Ellis, M. R., Friedman, B. J., Levine, S. C., and Rimell, F. L. (**2001**). ''Effects of the Clarion Electrode Positioning System on auditory thresholds and comfortable loudness levels in pediatric patients with cochlear implants,'' Arch. Otolaryngol. Head Neck Surg. **127**, 956–960.

Eddington, D. K. (**1980**). ''Speech discrimination in deaf subjects with cochlear implants,'' J. Acoust. Soc. Am. **68**, 885–891.

Eddington, D. K., Girzon, G., and Cuneo, P. (**1991**). ''An electroanatomical model of intracochlear electrical stimulation II: Tests of model predictions for monopolar electrodes,'' personal communication.

Eddington, D. K., Dobelle, W. H., Brackmann, D. E., Mladejovsky, M. G., and Parkin, J. L. (**1978**). ''Auditory prostheses research with multiple channel intracochlear stimulation in man,'' Arch. Oto-Rhino-Laryngol. **687**, Suppl. 53, Part 2, 1–39.

Favre, E., and Pelizzone, M. (**1993**). ''Channel interactions in patients using the Ineraid multichannel cochlear implant,'' Hear. Res. **66**, 150–156.

Fayad, J. N., Luxford, W., and Linthicum, F. H. (**2000**). ''The Clarion Electrode Positioner: Temporal Bone Studies,'' Am. J. Otol. **21**, 226–229.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). ''Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants,'' J. Acoust. Soc. Am. **110**, 1150–1163.

Frinjs, J. H. M., Briaire, J. J., and Grote, J. J. (**2001**). ''The importance of human cochlear anatomy for the results of the modiolus-hugging multichannel cochlear implants,'' Otol. Neurotol. **22**, 340–349.

Hanekom, T. (**2001**). ''Three-dimensional spiraling finite element model of the electrically stimulated cochlea,'' Ear Hear. **22**(4), 300–315.

Hartmann, R., and Klinke, R. (**1989**). ''Response characteristics of nerve fibers to patterned electrical stimulation,'' in *Cochlear Implants, Models of the Electrically Stimulated Ear*, edited by J. M. Miller and F. A. Spelman (Springer-Verlag, New York), pp. 135–160.

Hochmair-Desoyer, I. J., Hochmair, E. S., Burian, K., and Stiglbrunner, H. K. (**1983**). ''Percepts from the Vienna cochlear prosthesis,'' Ann. N.Y. Acad. Sci. **405**, 295–306.

Kawano, A., Seldon, H. L., Clark, G. M., Ramsden, R. T., and Raine, C. H. (**1998**). ''Intracochlear factors contributing to psychophysical percepts following cochlear implantation,'' Acta Otolaryngol. (Stockh, **118**, 313–326.

Kessler, D. K. (**1999**). ''The Clarion® Multi-Strategy™ Cochlear Implant,'' Ann. Otol. Rhinol. Laryngol. **108**, 8–16.

Kral, A., Hartmann, R., Mortazavi, D., and Klinke, R. (**1998**). ''Spatial resolution of cochlear implants: the electrical field and excitation of auditory afferents,'' Hear. Res. **121**, 11–28.

Kuzma, J. A., and Balkany, T. J. (**1999**). ''New generation Clarion electrodes for highly focused stimulation,'' 1999 Conference on Implantable Auditory Prostheses, Pacific Grove, CA.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467.

Lim, H. H., Tong, Y. C., and Clark, G. M. (**1989**). ''Forward masking patterns produced by intracochlear electrical stimulation of one and two electrode pairs in the human cochlea,'' J. Acoust. Soc. Am. **86**, 971–980.

Loeb, G., Byers, C., Rebscher, S., Casey, D., Fong, M., Schindler, R., Gray, R., and Merzenich, M. (**1983**). ''Design and fabrication of an experimental cochlear prosthesis,'' Med. Biol. Eng. Comput. **21**, 241–254.

Merzenich, M. M., and White, M. M. (**1977**). ''Cochlear implant: The interface problem,'' Biomed. Eng. Instrum. Ser. **3**, 321–340.

Merzenich, M. M., Rebscher, S. J., Loeb, G. E., Byers, C. L., and Schindler, R. A. (**1984**). ''The UCSF cochlear implant project, State of development,'' Adv. Audiol. **2**, 119–144.

Osberger, M. J., and Fisher, L. (**1999**). ''SAS-CIS preference study in postlingually deafened adults implanted with the Clarion® cochlear implant,'' Ann. Otol. Rhinol. Laryngol. **108**, 74–79.

Pelizzone, M., Cosendai, G., and Tinembart, J. (**1999**). ''Within-Patient Longitudinal Speech Reception Measures with Continuous Interleaved Sampling Processors for Ineraid Implanted Subjects,'' Ear Hear. **20**, 228–237.

Pelizzone, M., Boëx-Spano, C., Sigrist, A., François, J., Tinembart, J., De-

give, C., and Montandon, P. (**1995**). ''First field trials with a portable CIS processor for the Ineraid multichannel cochlear implant,'' Acta Otolaryngol. (Stockh) **115**, 622–628.
Pfingst, B. E., Morris, D. J., and Miller, A. M. (**1995**). ''Effects of electrode configuration on threshold functions for electrical stimulation of the cochlea,'' Hear. Res. **85**, 76–84.
Rebscher, S. J., Snyder, R. L., and Leake, P. A. (**2001**). ''The effect of electrode configuration and duration of deafness on threshold and selectivity of responses to intracochlear electrical stimulation,'' J. Acoust. Soc. Am. **109**, 2035–2048.
Ryan, A. F., Miller, J. M., Wang, Z.-X., and Woolf, N. K. (**1990**). ''Spatial distribution of neural activity evoked by electrical stimulation of the cochlea,'' Hear. Res. **50**, 57–70.
Shannon, R. R. (**1983a**). ''Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics,'' Hear. Res. **11**, 157–189.
Shannon, R. R. (**1983b**). ''Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction,'' Hear. Res. **12**, 1–16.
Shepherd, R. K., Hatsushika, S., and Clark, G. M. (**1993**). ''Electrical stimulation of the auditory nerve: The effect of electrode position on neural excitation,'' Hear. Res. **66**, 108–120.
Smith, Z. M., Delgutte, B., and Oxenham, A. (**2002**). ''Chimaeric sounds reveal dichotomies in auditory perception,'' Nature (London) **416**, 87–90.
Stickney, G. S., Loizou, P. C., Assmann, P. F., Shannon, R. V., and Opie, J. M. (**2001**). ''Electrode interaction and speech intelligibility in multichannel cochlear implants,'' 2001 Conference on Implantable Auditory Prostheses, Pacific Grove, CA.
van den Honert, C., and Stypulkowski, P. H. (**1987**). ''Temporal response patterns of single auditory nerve fibers elicited by periodic electrical stimuli,'' Hear. Res. **29**, 207–222.
Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). ''Better speech recognition with cochlear implants,'' Nature (London) **352**, 236–238.
White, M. M., Merzenich, M. M., and Gardi, J. N. (**1984**). ''Multichannel cochlear implants, Channel interactions and processor design,'' Arch. Otolaryngol. **110**, 493–501.
Wygonski, J. J., Lee, J., Faltys, M., Shannon, R., and Robert, M. (**1999**). ''Configurable speech strategy implementation using the Clarion research interface,'' 1999 Conference on Implantable Auditory Prostheses, Pacific Grove, CA.
Young, N. M., and Grohne, K. M. (**2001**). ''Comparison of pediatric Clarion recipients with and without the electrode postioner,'' Otol. Neurotol. **22**, 195–199.

Forward masking in different cochlear implant systems

Colette Boëx,[a] Maria-Izabel Kós, and Marco Pelizzone
"Centre Romand d'Implants Cochléaires," University Hospital, Geneva, Switzerland

(Received 28 January 2003; revised 12 July 2003; accepted 26 July 2003)

The goal of this study was to evaluate, from a psychophysical standpoint, the neural spread of excitation produced by the stimulation of different types of intracochlear electrode arrays: the Ineraid™, the Clarion™ S-Series on its own or with the Electrode Positioning System (EPS), and the Clarion™ HiFocus-I with the EPS. The EPS is an independent silicone part designed to bring the electrode array close to the modiolus. Forward masking was evaluated in 12 adult subjects (3 Ineraid™, 4 Clarion™ S-Series, 3 Clarion™ S-Series+EPS, 3 HiFocus-I+EPS) by psychophysical experiments conducted using trains of biphasic stimuli (813 pulses per second, 307.6 μs/phase). Masker signals (+8 dB *re*: threshold, 300 ms) were applied to the most apical electrode. Probe signals (30 ms, 10-ms postmasker) were delivered to more basal electrodes. Masked and unmasked detection thresholds of probe signals were measured. For both Clarion™ HiFocus-I subjects, measurements were conducted in both monopolar and bipolar stimulus configurations. No major differences were found in forward masking between the different intracochlear electrode arrays tested in the monopolar configuration at suprathreshold levels equivalent to those used in speech-coding strategies, but significant differences were found between subjects. A significant negative correlation also was found between the level of forward masking and the consonant identification performance. These measurements showed that the neural spread of excitation was more restricted in the bipolar configuration than in the monopolar configuration for HiFocus-I subjects. It was found that CIS strategies implemented without using apical electrodes, which showed high levels of masking, could improve consonant identification. [DOI: 10.1121/1.1610452]

PACS numbers: 43.64.Me, 43.66.Dc [BLM]

I. INTRODUCTION

Existing multichannel cochlear implant systems allow adult patients that became deaf after the acquisition of language to recover oral speech communication. Cochlear implants also allow congenitally as well as prelingually deaf children to adopt oral language as a main mode of communication (Svirsky *et al.*, 2000). Some of the efficiency of multichannel cochlear implants is thought to be due to the stimulation selectivity of each of the intracochlear electrodes. This stimulation selectivity contributes to the frequency selectivity of the acoustic sounds that cochlear implant systems transmit to the central auditory system. We expect that if each electrode excites a limited population of auditory-nerve fibers, the stimulation of each electrode should provide spectral information. At present, this selectivity and its effect on speech reception is poorly understood.

The excitation spread produced by electrical stimulation can be measured with forward-masking experiments (Shannon, 1983a, 1983b; Tong and Clark, 1986; Lim *et al.*, 1989; Cohen *et al.*, 1996; Chatterjee and Shannon, 1998; Throckmorton and Collins, 1999). Shannon (1983a) proposed that the level of masking obtained in forward masking was due in part to the overlap of the population of fibers excited by the masker with the population of fibers excited by the probe. Indeed, the nature of forward masking is the result of different neural masking phenomena. Neural masking can be due to refractory effects of the auditory-nerve fibers or to central effects. The refractory effects are due to the fact that (1) fibers cannot produce a spike in response to an excitation while not having partially recovered from the generation of a previous spike or that (2) fibers need stronger excitation to produce a spike in response to an excitation while not having completely recovered from the generation of a previous spike see (Miller *et al.*, 2001 for a review). If the delay between the masker and the probe is long enough (about 7 ms), the refractory effect should not influence the results of forward-masking experiments. Central effects remain poorly understood but they certainly play an important role in forward-masking results obtained with electric hearing (Shannon, 1990; Shannon and Otto, 1990; Nelson and Donaldson, 2001, 2002).

In order to analyze the stimulation selectivity (the spread of excitation) of different intracochlear electrode arrays, we measured forward masking at levels similar to those used in sound coding strategies. We measured forward masking at comfortable hearing levels (masker: 8 dB above threshold) in a group of 12 subjects. We conducted forward masking for subjects who received the Ineraid™ electrode array, the Clarion™ S-Series electrode array with or without the Electrode Positioning System (EPS), and the Clarion™ HiFocus-I electrode array with the EPS. The EPS is an independent silicone part designed to place the electrode array close to the modiolus. Forward masking was measured successively in different electrodes for a masker applied on the most apical electrode available for stimulation. For both us-

[a]Electronic mail: colette.boex@hcuge.ch

ers of the Clarion™ HiFocus-I electrode arrays (for whom the masker level could be reached within the stimulator-compliance limits), we also measured forward masking for bipolar stimulus configuration. Forward masking was not measured in bipolar configuration for Ineraid subjects as this implant was not designed to support this stimulus configuration. In this case, the Ineraid implant would present only 5 channels and very large bipolar pairs (electrodes are 4-mm distant). They were also not conducted in Clarion™ S-Series because of compliance limits in these cases. It was not possible to reach suprathreshold levels (+8 dB) in bipolar configuration. In order to analyze the effect of forward masking on speech reception performances, all subjects were also subjected to consonant identification tests. In addition, the information gathered through these experiments was used to adapt CIS strategies (Wilson *et al.*, 1991) for subjects who showed apical electrodes with high levels of masking.

II. METHODS

A. Subjects

All 12 subjects who participated in this study also participated in previous electrical interaction experiments [Boëx *et al.*, (2003)], Three subjects received the Ineraid™ implant (I03, I09, and I34), four subjects (C05, C06, C24, and C30) the standard Clarion™ S-Series electrode array, three subjects (Cp08, Cp14, and Cp18) the S-Series electrode with EPS, and two subjects (H26 and H29) received the HiFocus-I electrode with EPS. All subjects except subjects Cp08 and Cp18 were bilaterally, profoundly, and postlingually deaf. Subjects Cp08 and Cp18 suffered from severe deafness in the contralateral ear. Within this group different etiologies of deafness were found (Mondini, Ménière's syndrome, traumatic, streptomycin, multiple otitis, viral labyrinthitis, otosclerosis, congenital, or unknown). At the time of the study, subjects were between the ages of 39 and 73 years old (mean: 53 years and 10 months). These experiments were conducted within the ethical guidelines of the Declaration of Helsinki.

B. Experiment design

Detection thresholds were measured for the probe alone, the masker alone, and the probe in presence of the masker. This was carried out using a 3-alternative-forced-choice (3AFC) ''one up/two down'' adaptive procedure converging to a level where approximately 70.7% of stimuli are detected (Levitt, 1971). The subjects were asked to indicate in which interval they perceived the probe signal in addition to the masker. Feedback was provided. Each measurement started with a clearly detectable signal. One detection threshold measurement was obtained after gathering six reversals (descending and ascending segments). Signals were decreased initially by steps of about 15% of the initial amplitude. For the second and third reversals the steps were halved. One threshold was computed as the mean of the last four descending segments of the staircase procedure. The detection thresholds indicated in the present study were calculated as the mean of at least two threshold measurements. Forward masking was measured for biphasic stimuli (813 pulses per second, 307.6 μs/phase, Fig. 1). The masker (300 ms) was set at 8 dB above the detection threshold of the masker electrode. It was clearly perceived and comfortable for all subjects. This level of masker was expected to be high enough to allow the observation of the masking effect on a probe electrode distant more than 6 mm away from the masker. Masker was applied to the most apical electrode, and its effects on the detection thresholds of the probe signals on other electrodes were measured. The probe signal (30 ms) was presented 10 ms after the end of the masker. This delay was applied to limit refractory effects and residual polarization effects on the forward-masking measurements.

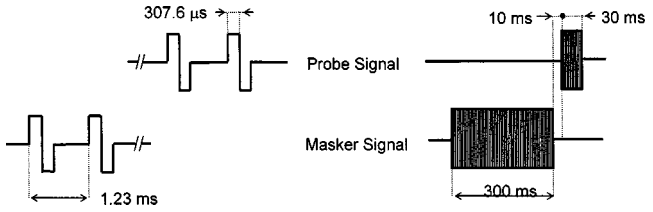


FIG. 1. Scheme of the masker and probe signals used in the forward-masking measurements. Stimuli were all biphasic (813 pulses per second, 307.6 μs/phase). Masker signals were 300-ms duration (zero rise–decay times). Probe signals were 30-ms duration (zero rise–decay times). Probe and masker signals were always first positive biphasic pulses. The probe signal was presented 10 ms after the setting off of the masker signal.

The electric stimuli were generated using the Clarion™ Research Interface (CRI, Wygonski *et al.*, 1999), which allows custom control of all stimulation parameters. The current amplitudes delivered through intracochlear electrodes were not exactly proportional to the requested currents over the entire electrical output range. They depended slightly on electrode impedance. These limitations can lead to measurement error in psychophysical experiments. Therefore, we devoted special attention to improve the control of the symmetry and the amplitude of current biphasic pulses. For this, we used special look-up tables based on calibrations. One table was built measuring, for each clinical unit, the effective current delivered by the current source of the Research Implantable Cochlear Stimulator (RICS). We defined from these tables the clinical units required to provide the desired current amplitudes. These tables were defined for each electrode impedance reported for all subjects in Boëx *et al.* (2003). Impedances were evaluated from the clinical setup impedance measurements[1] (SCLIN) using sine waves (1000 Hz) at very low levels. We also evaluated impedances of Ineraid™ electrodes using sine waves (1000 Hz) at very low levels. In monopolar configuration, maximum impedances and currents used in the present study were low (respectively, 18 kΩ and 60 μA). In these limits current sources behaved almost as perfect current sources. As an example, if impedance decreased from 18 to 13 kΩ (27% lower), stimulation amplitude would increase from 60 to 62 μA. In bipolar configuration, because the HiFocus-I subjects were SAS (Simultaneous Analog Strategy, Battmer *et al.*, 1999; Osberger and Fisher, 1999) speech strategy users and, because that strategy stimulates all electrodes, monopolar impedance of each electrode used could be considered and summed to define the bipolar impedances. For these subjects, maximum bipolar impedances and currents used in the present study

were low (24 kΩ and 130 μA or 33 kΩ and 100 μA). In these limits current sources behaved almost as perfect current sources. As an example, if impedance decreased from 24 to 18 kΩ (27% lower) or from 33 kΩ to 24 kΩ (27% lower), stimulation amplitude would increase from 130 to 139 μA or from 100 to 109 μA, respectively.

To conduct these measurements in Ineraid subjects, we connected directly the channel output of the RICS provided with the CRI to the Ineraid™ electrodes via their percutaneous plug.

C. Electrode types

Three different types of electrode arrays have been tested.

The Ineraid™ electrode array has six intracochlear platinum electrodes. Each electrode is a platinum 0.5-mm-diameter ball. The distance between the centers of electrodes is about 4 mm. In monopolar configuration, the stimuli were applied between one intracochlear electrode and an external electrode used as a far-field ground electrode, placed under the temporalis muscle. Electrodes were numbered from the most apical electrode to the most basal pair from 1 to 6. The masker signals were applied to electrode 1 (most apical) and the probe signals were successively applied to electrodes 1, 2, 3, and 4. Thus, the nominal distance between the probe and the masker electrodes was 0, 4, 8, and 12 mm, respectively.

The preformed Clarion™ S-Series electrode array (Kessler, 1999) was implanted either alone or with the Electrode Positioning System (EPS). The Clarion™ S-Series electrode array has eight pairs of electrodes (one medial, one lateral). Each electrode is an iridium–platinum (90:10) 0.3-mm-diameter ball. The distance between each pair is 2 mm. Each pair consists of one medial and one lateral electrode positioned radially, 0.6 mm distant from each other. Electrode pairs are numbered from the most apical electrode pair to the most basal pair from 1 to 8. In monopolar configuration, the stimuli were applied between a medial electrode and the base of the implanted receiver–stimulator case used as a far-field ground in the Clarion™ S-series system. The masker signals were applied to medial apical electrode 1 and the probe signals were applied successively to medial electrodes 1, 2, 4, 6, and 8. Thus, the nominal distance between the probe and the masker electrodes was 0, 2, 6, 10, and 14 mm, respectively.

The Clarion™ HiFocus-I electrode array has 16 electrodes. Each contact is rectangular in shape (0.4×0.5 mm) and made from pure platinum. The distance between the centers of electrodes is 1 mm. This array was implanted with the EPS. Electrodes are numbered from the most apical electrode to the most basal from 1 to 16. In the monopolar configuration, stimuli were applied between one electrode and the base of the implanted receiver–stimulator case used as a far-field ground in the Clarion™ HiFocus system. The masker signals were applied to electrode 2 and the probe signals were applied successively to electrodes 2, 4, 8, 12, and 14. Thus, the distances between the probe and the masker electrodes were 0, 2, 6, 10, and 12 mm, respectively. In the bipolar configuration, stimuli were applied between two adjacent electrodes, with the more basal electrode of the pair used as the ground electrode. The masker signals were applied between electrodes 2 and 3. The probe signals were applied successively between electrodes 4 and 5, 8 and 9, 12 and 13. Thus, the nominal distance between the probe and the masker electrode pairs was 2, 6, and 10 mm, respectively.

D. Consonant identification tests

Speech reception was evaluated through closed-set medial consonant tests (Pelizzone *et al.*, 1993) and initial consonant tests. Each test consisted in the presentation of 56 tokens using the 14 French consonants /b, d, f, g, k, l, m, n, p, r, s, t, v, and z/, in the form ''aBa,'' ''aDa,'' etc. Tokens were presented in a random order and feedback was not provided. Medial consonant tokens were spoken by one male speaker. Initial consonant tokens were spoken by two female and two male talkers.

Speech tests were conducted without visual cues. Subjects were seated 1 m from the loudspeaker (Fostex™ UP203 S) in a sound-proof chamber (IAC 1201A). Tests were played from the Turtle Beach™ Pinnacle™ Pro Series sound card. Sound levels were adjusted with an EMB™ P 300 amplifier. The overall level of the tokens was about 75 dB peak SPL A. Scores were expressed in percent of correctly identified consonants.

E. Speech processor

The Ineraid subjects had been using a CIS strategy for at least 4 years (Geneva Wearable Processors, Pelizzone *et al.*, 1995, 1999). The Clarion subjects who received the S-Series electrode array (with or without the EPS) had been using the standard Clarion™ CIS strategy. The Clarion subjects who received the HiFocus-I electrode array had been using a SAS strategy. The Clarion subjects had been using their implant for at least 1 year and for less than 4 years (mean:2 years).

Subjects C05, Cp08, Cp14, Cp18, and C30 tested different CIS strategies, implemented with reduced numbers of stimulation channels. All strategies were implemented using the clinical Clarion™ (SCLIN) platform. Each time a channel was switched off, the same input frequency range was shared across the remaining electrodes and the rate of stimulation maximized.

III. RESULTS

A. Forward-masking experiments

The masker signals were presented on the most apical electrode that could be stimulated. Masker levels were set at 8 dB above the masker threshold. The masker amplitudes used for each subject are reported in Table I.

In Fig. 2, the thresholds for the probe alone and the masked probe conditions are plotted as a function of the distance between the masker and the probe electrodes for monopolar stimulation and in Fig. 3 for bipolar stimulation. The nominal distances (abscissa) between the probe electrode and the masker electrode were determined from the electrode designs. The probe thresholds (triangles pointing upwards) were generally lower than the masked probe

TABLE I. Amplitudes of masker signals set 8 dB above the threshold of the masker electrode in the monopolar and bipolar stimulus configurations. Suprathreshold levels (8 dB) could not be reached (because of compliance limits) in the bipolar configuration for S-Series subjects. Neural masking measurements were not conducted in bipolar configuration for Ineraid subjects.

Subjects	Monopolar masker signal (μA)	Bipolar masker signal (μA)
I03	55	⋯
I09	31	⋯
I34	68	⋯
C05	26	⋯
C06	25	⋯
C24	36	⋯
C30	32	⋯
Cp08	22	⋯
Cp14	19	⋯
Cp18	60	⋯
H26	21	79
H29	23	209

thresholds (triangles pointing downwards). In some cases, the thresholds for the probe alone and masked probe conditions were similar (e.g., subject H29 at 12 mm, subject H26 at 10 mm in monopolar configuration), indicating a negligible masking at those distances. In bipolar configuration, subject H26 did not show any differences between his probe thresholds and his masked probe thresholds, indicating an absence of forward masking.

Thresholds and masked thresholds could be very different across subjects, even for the same type of electrode array (e.g., subject Cp18 in comparison to subjects Cp08 or Cp14). Masking was computed as the ratio of masked probe thresholds to probe alone thresholds (masked probe threshold/probe threshold alone) in the monopolar (Fig. 4) and in the bipolar configurations (Fig. 5).

In the monopolar configuration, the largest masking was observed when the probe and the masker electrodes were both on the same most apical electrode (0 mm). The masking at 0 mm could be very different across subjects implanted

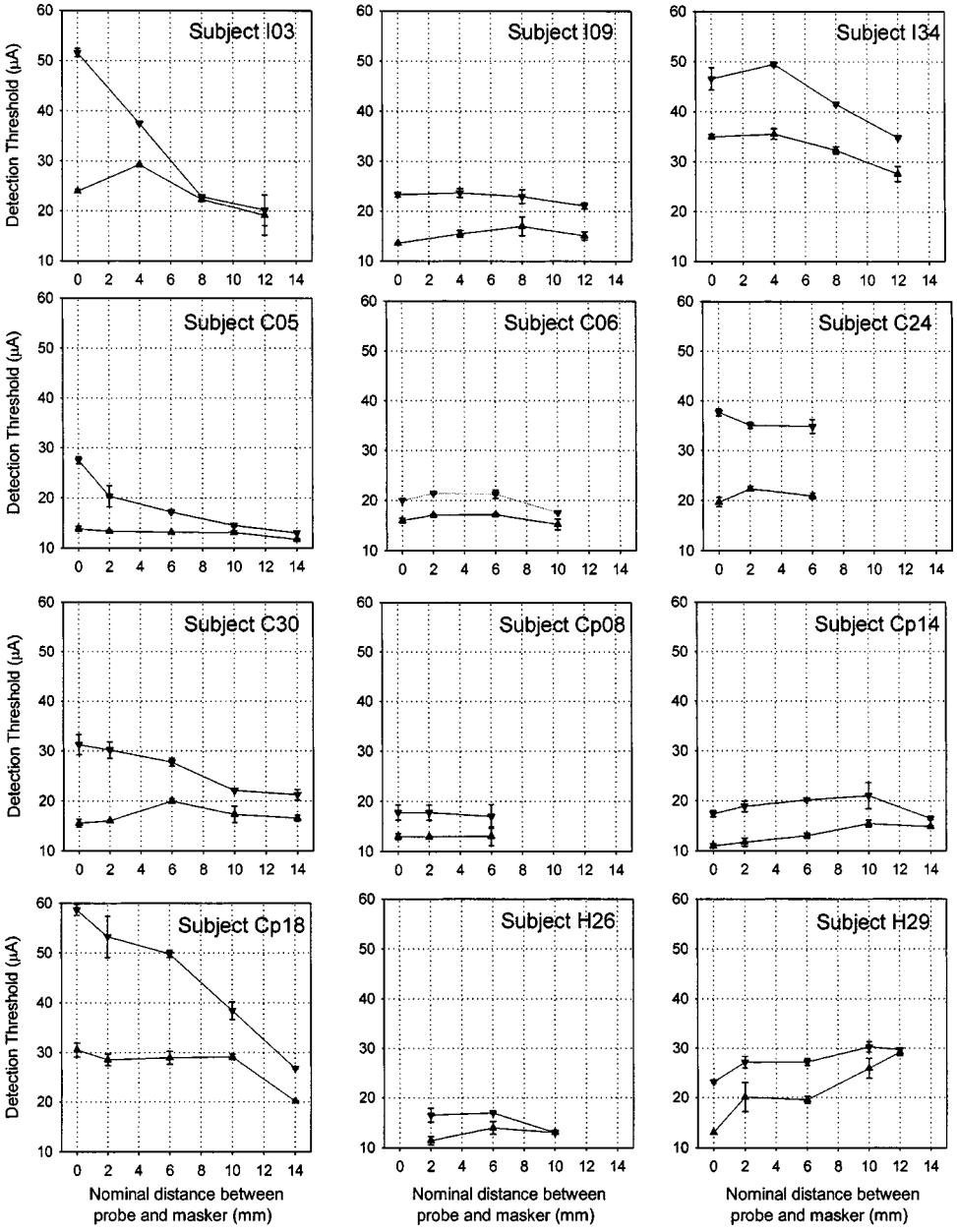


FIG. 2. Probe thresholds measured with and without masker in monopolar configuration. In all graphs, subject probe thresholds (μA) are identified by triangles pointing upwards. Masked probe thresholds (μA) are identified by the triangles pointing downwards. The nominal distance between the masker and the probe electrodes tested is indicated in mm on the abscissa.

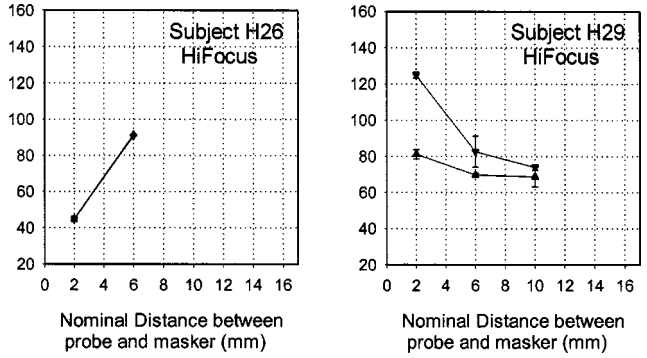


FIG. 3. Probe thresholds measured with and without masker in bipolar configuration. Subject probe thresholds (μA) are identified by triangles pointing upwards and masked probe thresholds (μA) are identified by triangles pointing downwards. The nominal distance between the masker and the probe electrodes tested is indicated in mm on the abscissa.

with the same electrode array (e.g., Cp08, Cp14, and Cp18). Overall, masking decreased as the distance increased between probe and masker electrodes (Table II), but it did not decrease identically for all subjects. For some subjects, masking did not decrease up to a 6-mm distance between probe and masker electrodes (Cp08,Cp14,C06,I34). With other subjects (C05,C24,H29), masking decreased significantly at 2 mm ($P<0.001$) in the monopolar configuration.

Overall masking computed from all electrode masking

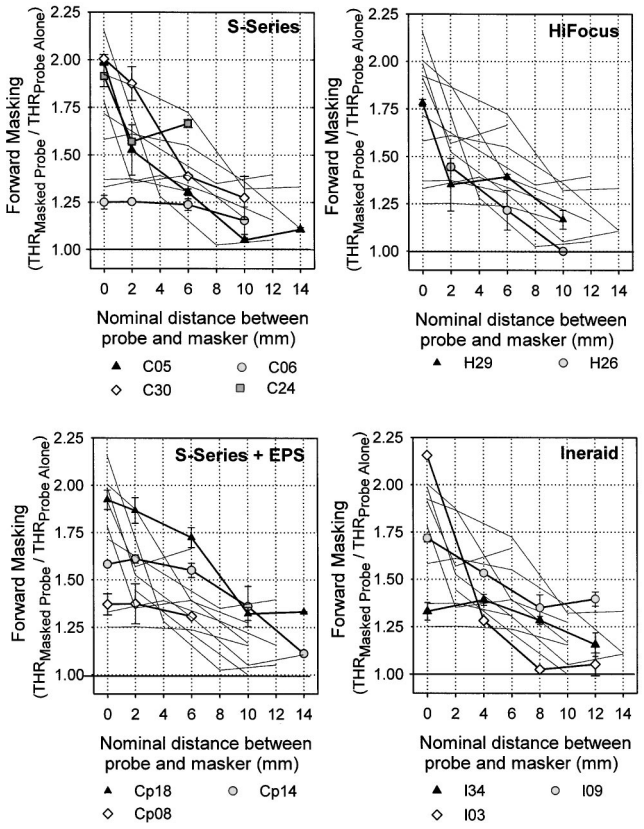


FIG. 4. Forward-masking measurements (ratio of masked probe thresholds to probe alone thresholds) obtained in the monopolar configuration. In all graphs the thin lines without symbols represent the curves obtained for all 12 subjects. In each graph, the forward masking obtained by subjects implanted with the specified electrode array is indicated by specified symbols and solid lines. Error bars are the standard errors of the mean.

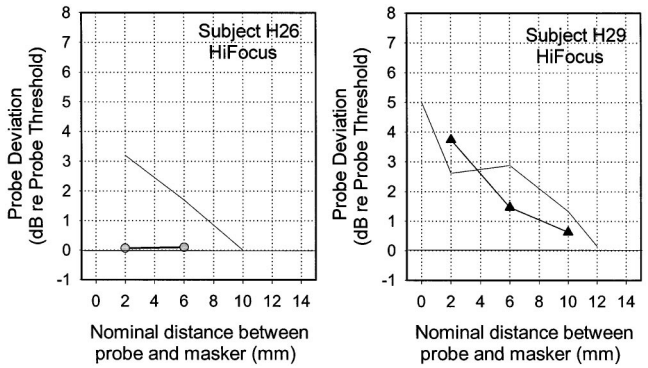


FIG. 5. Forward-masking measurements (ratio of masked probe thresholds to probe alone thresholds) obtained in the bipolar configuration for both HiFocus-I subjects. Masked threshold could not be reached (because of compliance limits) at 0-mm masker–probe separation (not measured for subject H26). The thin lines without symbols represent the curves obtained in the monopolar configuration by the same subject. Error bars are the standard errors of the mean.

was not significantly different for subjects implanted with the Ineraid than for subjects implanted with the S-Series with or without the EPS. Overall masking was not significantly different for subjects implanted with the S-Series alone or with the EPS.

Subject H26 (HiFocus-I) did not show masking in the bipolar configuration. In subject H29, masking in bipolar configuration was slightly higher at 2 mm and was slightly lower than in monopolar configuration for more distant probes.

B. Consonant identification scores

We analyzed the correlation between medial consonant identification scores and masking data. Consonant identification scores were expressed in percent of correct responses and were computed for at least six consonant tests obtained within two sessions. All subjects used the CIS sound coding strategy, except H26 and H29 who used a SAS strategy (seven channels in both subjects). All Clarion CIS-user scores were obtained with the standard eight channel Clarion™ CIS strategy. We used masking data obtained at 4 mm to consider masking for the same distance[2] between probe and masker electrodes for all different types of electrode arrays. To calculate masking at a distance of 4 mm between probe and masker electrodes we made a linear interpolation based on data obtained at 2 and 6 mm in Clarion subjects. The consonant identification scores with masking are shown in Fig. 6. We obtained a regression coefficient of -0.60 (R, Pearson product moment; $p<0.04$) which indicates that there was a statistically significant negative correlation between masking and consonant identification. The latter accounted, however, for only 36% of the variance.

C. Electrode selection for the CIS strategy

Subjects Cp18 and C30 showed high levels of forward masking at 2 mm (electrode 2) and for subject Cp18 interpolated forward masking was also high at 4 mm. We concluded that for these subjects, electrodes 1 and 2 stimulated partly the same population of fibers. Hence, we proposed to test a

TABLE II. Average monopolar masking (masked probe threshold/probe threshold alone) for masker to probe distances ranging from 0 to 14 mm. These average changes were computed from all available electrode measurements. Standard errors (SE) are indicated.

Nominal distance between masker and probe electrodes	0 mm	2 mm	4 mm	6 mm	8 mm	10 mm	12 mm	14 mm
Number of subjects	11	9	3	9	3	7	3	3
Mean changes	1.73	1.54	1.40	1.42	1.22	1.19	1.20	1.18
(SE)	(0.09)	(0.07)	(0.07)	(0.06)	(0.10)	(0.05)	(0.10)	(0.07)

seven-channel CIS strategy that eliminated the use of electrode 1. We suggested further to subject Cp18 to test a six-channel CIS strategy that eliminated electrode 2 in addition to electrode 1. Subject Cp18 tested each strategy for 2 weeks. His initial consonant identification scores increased (nonsignificantly) from 50.86% (5.52) to 53.5% (5.2) with the seven-channel processor and significantly increased to 58.25%($4.27; P<0.05$) with the six-channel processor (Fig. 7). He chose then to adopt the six-channel CIS processor permanently. Subject C30 tested the seven-channel strategy for 2 weeks. Her initial consonant identification scores increased (nonsignificantly) from 40.4% (6.8) to 46.4% (3.21) with the seven-channel strategy. When she tested a CIS strategy implemented with six channels, the initial consonant scores were not significantly different, varying from 40.4% (6.8) to 45.8% (3.19). She chose then to adopt the seven-channel strategy permanently.

We also tested CIS strategies implemented with seven channels for subjects C05 and Cp08. The latter, differently from the former two subjects, presented lower masking at 2 mm. Their initial consonant identification scores decreased significantly ($P<0.05$), even after 2 weeks of daily use. They reverted to their initial standard eight-channel strategy.

IV. DISCUSSION

We did not find differences in monopolar masking between the groups of subjects using the three types of electrode arrays under study (Ineraid™, Clarion™ S-Series, and HiFocus-I). We noted, however, significant forward-masking differences across subjects (e.g., Cp18 and C08). Important similar variations across subjects were also observed by Chatterjee and Shannon (1998) and by Cohen *et al.* (in bipolar configuration, 1996). For example, subject C06 presented a low level of forward masking, indicating that the tested electrodes did not stimulate significantly overlapping neural populations. Some subjects like subject Cp18 showed important masking at 0 mm which did not decrease significantly when the probe was switched from electrode 1 to electrode 2. In his case, if forward masking was due in part to the extent of the population of fibers excited, this would suggest that the population of fibers excited by electrode 1 was similar to the population of fibers excited by electrode 2. In agreement with this hypothesis, consonant tests showed that subject Cp18 performed better when the CIS-speech processor was fitted without using the two most apical electrodes, while subject Cp08 did not. In addition, we also observed that subject Cp18 assigned about the same 100-Hz tone pitch [acoustic to electric pitch-comparison experiments in Boëx *et al.*,

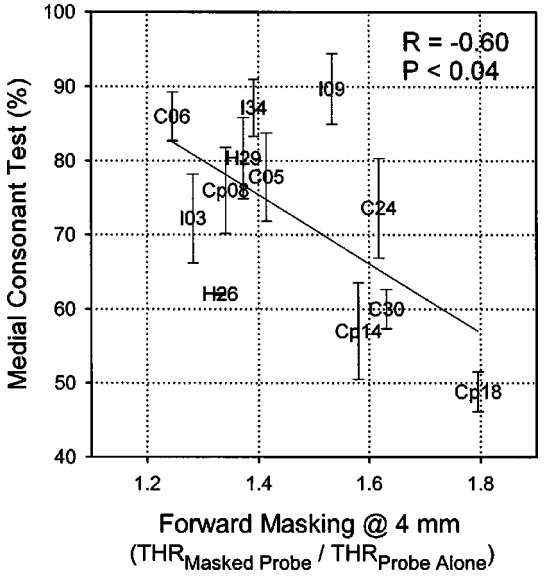


FIG. 6. Medial consonant identification scores are expressed in percent of correct responses versus forward masking measured at 4 mm for Ineraid subjects and calculated at 4 mm from 2- and 6-mm data for Clarion subjects. Scores are the mean percentage of correct responses computed on six consonant tests (6×56 tokens), collected in two different test sessions. Error bars are standard deviations.

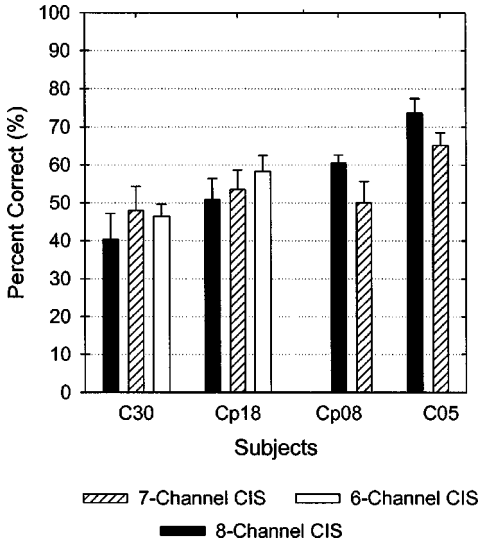


FIG. 7. Initial consonant identification scores are expressed in percent of correct responses for S-Series subjects who tested CIS strategies implemented with different numbers of channels. Each processor was used daily for at least 2 weeks. Scores are the mean percentage of correct responses computed on three consonant tests (3×56 tokens). Error bars are standard deviations.

(submitted)] to electrodes 1 and 2. In contrast, we observed that subject Cp08, who showed low masking levels between electrodes 1 and 2, assigned to electrode 1 a 383-Hz tone pitch and to electrode 2 a 505-Hz tone pitch. This observation is in agreement with the hypothesis that in the case of subject Cp08, electrode 1 and 2 stimulated different populations of fibers as indicated by the low masking levels observed. When we reduced the number of channels subject Cp18 showed improvements in his performance. Zwolan *et al.* (1997) made a similar observation when they reported improvements in speech perception for some subjects who tested speech-coding strategies that did not use nondiscriminable electrodes. The improvements we obtained discarding the most apical electrodes could also be in agreement with the proposition of Shannon *et al.* (2001) that signals should be delivered to pitch-matched electrodes to allow for the best speech reception, particularly for apical electrodes. The negative correlation shown in our data between forward masking and consonant reception suggests that the overlap of the populations of fibers excited may limit speech reception. Throckmorton and Collins (1999) also obtained significant correlations between average forward masking and some speech recognition test scores.

We compared forward masking between monopolar and bipolar configurations in both HiFocus-I subjects. In the bipolar configuration the forward masking was lower for both subjects, suggesting that the use of a bipolar configuration would significantly improve the selectivity of HiFocus-I stimulation. Improvement of selectivity offered by a bipolar configuration was observed previously for different intracochlear electrodes. Tong and Clark (1986) and Tong *et al.* (1987) also observed a decrease in masking with increases in distances between masker and probe electrodes in users of the Nucleus electrode array with bipolar configuration. Shannon (1983b) conducted forward-masking measurements for sinusoid stimuli in one subject, using different bipolar and monopolar electrode pairs placed at different positions in the cochlea and for different stimulus levels. He reported better spatial selectivity in the bipolar configuration than in the monopolar configuration. Lim *et al.* (1989) also conducted experiments showing that larger spatial extents of bipolar pairs resulted in spatially larger patterns of masking.

V. CONCLUSION

We did not find group differences in the monopolar selectivity of stimulation among the subjects wearing the Ineraid™, the Clarion™ S-Series, and the HiFocus-I electrode arrays. But, we found important differences between subjects implanted with the same types of electrode array. We found a statistically significant negative correlation between the level of forward masking and the initial consonant identification performance in our group of 12 users, implanted with different cochlear implant systems. One subject performed better when the most apical electrodes presenting high levels of masking were not used in CIS processors. Forward-masking experiments can help determine the selectivity of intracochlear electrode stimulation and select electrodes to be used in sound-coding strategies for each subject. This is even more relevant with present cochlear implants, which have more intracochlear electrodes.

In addition, we found low levels of masking in bipolar stimulation for both HiFocus-I subjects, indicating a better selectivity of stimulation in that configuration. Interestingly, both subjects had adopted the bipolar simultaneous analog strategy (SAS) since the beginning of their implant use. The use of more selective electrical stimulation channels should increase the number of effective spectral channels and should allow better implementation of simultaneous strategies.

ACKNOWLEDGMENTS

This work has been supported by the ''Fonds National pour la Recherche Scientifique,'' Contract No. 31-053881.98. Most of all, we are thankful to the subjects for the time and the effort they devoted to the study. We would like to thank D. K. Eddington for his constructive comments on an earlier version of the manuscript as well as R. V. Shannon and one anonymous reviewer. We would like to thank Chloé de Balthasar for the development of the MATLAB interface used in the measurements. We would like to thank Patrick Boyle for his help in programming the Clarion™ Research Interface. Finally, we thank all members of the ''Center Romand d'Implants Cochléaires.''

[1]Tables were built for our RICS; they could not be built for each subject-implanted stimulator as we could not measure current amplitudes as well as impedances on implanted electrodes with our present research interface. We were not aware of any reports on description of impedance changes with amplitude of stimulation in Clarion™ S-Series or HiFocus-I electrodes.

[2]This choice was made because masking could not be measured at 2 and 6 mm in Ineraid subjects and because we did not want to use the 0-mm data.

Battmer, R. D., Zilberman, Y., Haake, P., and Lenarz, T. (**1999**). ''Simultaneous analog stimulation (SAS)—Continuous interleaved sampler (CIS) pilot comparison study in europe,'' Ann. Otol. Rhinol. Laryngol. **108**, 69–73.

Boëx, C., de Balthasar, Cc., Kós, M.-I., and Pelizzone, M. (**2003**). ''Electrical field interactions in different cochlear implant systems,'' J. Acoust. Soc. Am. **114**, 2049–2057.

Boëx, C., Cosendai, G., Kós, M.-I., and Pelizzone, M. (**submitted**). ''Pitch matching with very deep electrode insertion,'' Ear Hear.

Chatterjee, M., and Shannon, R. V. (**1998**). ''Forward masked excitation patterns in multielectrode electrical stimulation,'' J. Acoust. Soc. Am. **103**(5), 2565–2572.

Cohen, L. T., Busby, P. A., and Clark, G. M. (**1996**). ''Cochlear implant place psychophysics. I. Pitch estimation with deeply inserted electrodes,'' Audiol. Neuro-Otol. **1**, 265–277.

Dynes, S. B. C. (**1996**). ''Discharge characteristics of auditory nerve fibers for pulsatile electrical stimuli,'' Doctoral thesis, Massachusetts Institute of Technology.

Kessler, D. K. (**1999**). ''The Clarion® Multi-Strategy™ Cochlear Implant,'' Ann. Otol. Rhinol. Laryngol. **108**, 8–16.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**(2), 467.

Lim, H. H., Tong, Y. C., and Clark, G. M. (**1989**). ''Forward masking patterns produced by intracochlear electrical stimulation of one and two electrode pairs in the human cochlea,'' J. Acoust. Soc. Am. **86**, 971–980.

Miller, C. A., Abbas, P. J., and Robinson, B. (**2001**). ''Response properties of the refractory auditory nerve fiber,'' J.A.R.O. **2**, 216–232.

Nelson, D. A., and Donaldson, G. S. (**2001**). ''Psychophysical recovery from single-pulse forward masking in electric hearing,'' J. Acoust. Soc. Am. **109**(6), 2921–2933.

Nelson, D. A., and Donaldson, G. S. (**2002**). ''Psychophysical recovery from pulse-train forward masking in electric hearing,'' J. Acoust. Soc. Am. **112**(6), 2932–2947.

Osberger, M. J., and Fisher, L. (**1999**). ''SAS-CIS preference study in postlingually deafened adults implanted with the Clarion® cochlear implant,'' Ann. Otol. Rhinol. Laryngol. **108**, 74–79.
Pelizzone, M., Boëx, C., and Montandon, P. (**1993**). ''Vowel and consonant identification tests can be used to compare performances in a multilingual group of cochlear implant patients,'' ORL J. Otorhinolaryngol. Relat. Spec. **55**, 341–346.
Pelizzone, M., Boëx-Spano, C., Sigrist, A., François, J., Tinembart, J., Degive, C., and Montandon, P. (**1995**). ''First field trials with a portable CIS processor for the Ineraid multichannel cochlear implant,'' Acta Otolaryngol. (Stockh) **115**, 622–628.
Pelizzone, M., Cosendai, G., and Tinembart, J. (**1999**). ''Within-patient longitudinal speech reception measures with continuous interleaved sampling processors for Ineraid implanted subjects,'' Ear Hear. **20**, 228–237.
Shannon, R. V., Fu, Q.-J., Baskent, D., and Galvin, J. (**2001**). ''Frequency-place alignment is critical for speech recognition,'' Conference on Implantable Auditory Prostheses, Pacific Grove, CA.
Shannon, R. V. (**1990**). ''Forward masking in patients with cochlear implants,'' J. Acoust. Soc. Am. **88**, 741–744.
Shannon, R. V., and Otto, S. R. (**1990**). ''Psychophysical measures from electrical stimulation of the human cochlear nucleus,'' Hear. Res. **47**(1–2), 159–168.
Shannon, R. V. (**1983a**). ''Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics,'' Hear. Res. **11**, 157–189.
Shannon, R. V. (**1983b**). ''Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction,'' Hear. Res. **12**, 1–16.
Svirsky, M. A., Robbins, A. M., Kirk, K. I., Pisoni, D. B., and Miyamoto, R. T. (**2000**). ''Language development in profoundly deaf children with cochlear implants,'' Psychol. Sci. **11**, 153–158.
Throckmorton, C. S., and Collins, M. (**1999**). ''Investigation of the effects of temporal and spatial interactions on speech-recognition skills in cochlear-implant subjects,'' J. Acoust. Soc. Am. **105**(2), 861–873.
Tong, Y. C., and Clark, M. (**1986**). ''Loudness summation, masking, and temporal interactions for sensations produced by electric stimulation of two sites in the human cochlea,'' J. Acoust. Soc. Am. **79**(6), 1958–1966.
Tong, Y. C., Clark, M., and Lim, H. H. (**1987**). ''Estimation of the effective spread of neural excitation produced by a bipolar pair of scala tympani electrodes,'' Ann. Otol. Rhinol. Laryngol. Suppl. **128**, 37–38.
Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). ''Better speech recognition with cochlear implants,'' Nature (London) **352**, 236–238.
Wygonski, J. J., Lee, J., Faltys, M., Shannon, R., and Robert, M. (**1999**). ''Configurable speech strategy implementation using the Clarion research interface,'' 1999 Conference on Implantable Auditory Prostheses, Pacific Grove, CA.
Zwolan, T. A., Collins, L. M., and Wakefield, G. H. (**1997**). ''Electrode discrimination and speech recognition inpostlingually deafened adult cochlear implant subjects,'' J. Acoust. Soc. Am. **102**(6), 3673–3685.

Desynchronization of electrically evoked auditory-nerve activity by high-frequency pulse trains of long duration[a)]

Leonid M. Litvak[b)] and Zachary M. Smith
Eaton-Peabody Laboratory and Cochlear Implant Research Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Speech and Hearing Bioscience and Technology Program, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

Bertrand Delgutte
Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

Donald K. Eddington
Cochlear Implant Research Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Research Laboratory of Electronics and Neural Prosthesis Research Center, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

(Received 13 March 2003; accepted for publication 4 August 2004)

Rubinstein *et al.* [Hear. Res. **127**, 108–118 (1999)] suggested that the neural representation of the waveforms of electric stimuli might be improved by introducing an ongoing, high-rate, desynchronizing pulse train (DPT). A DPT may desynchronize neural responses to electric stimulation in a manner similar to spontaneous activity in a healthy ear. To test this hypothesis, responses of auditory-nerve fibers (ANFs) to 10-min-long electric pulse trains (5 kpps) were recorded from acutely deafened, anesthetized cats. Stimuli were delivered via an intracochlear electrode, and their amplitude was chosen to elicit a response in most ANFs. Responses to pulse trains showed pronounced adaptation during the first 1–2 min, followed by either a sustained response or cessation of spike discharges for the remainder of the stimulus. The adapted discharge rates showed a broad distribution across the ANF population like spontaneous activity. However, a higher proportion of fibers (46%) responded to the DPT at rates below 5 spikes/s than for spontaneous activity, and 12% of the fibers responded at higher rates than any spontaneously active fiber. Interspike interval histograms of sustained responses for some fibers had Poisson-like (exponential) shapes, resembling spontaneous activity, while others exhibited preferred intervals and, occasionally, bursting. Simultaneous recordings from pairs of fibers revealed no evidence of correlated activity, suggesting that the DPT does desynchronize the auditory nerve activity. Overall, these results suggest that responses to an ongoing DPT resemble spontaneous activity in a normal ear for a substantial fraction of the ANFs.
[DOI: 10.1121/1.1612492]

PACS numbers: 43.64.Me, 43.64.Pg [WPS]

I. INTRODUCTION

Cochlear implant patients rely heavily on temporal information for speech reception (Van Tasell *et al.*, 1987, 1992; Rosen, 1992). For example, in continuous interleaved sampling (CIS) sound-processing strategies, temporal information about sounds is encoded in the amplitude modulations of pulse trains (Wilson *et al.*, 1991). Proper representation of stimulus waveforms in the temporal discharge patterns of the auditory nerve is therefore an important goal in these and other strategies.

Despite the importance of temporal information in cochlear implants, the temporal discharge patterns of auditory-nerve fibers to electric stimuli are typically very different from acoustic responses in a normally functioning ear [see Javel (1990) for review]. In response to sinusoidal electric stimuli, spikes occur only during a small fraction of the stimulus cycle, resulting in a poor representation of the stimulus waveform (Hartmann *et al.*, 1984; van den Honert and Stypulkowski, 1987; Parkins, 1989; Litvak *et al.*, 2001). For low frequency (<800 Hz) stimuli, spikes tend to occur exactly once per stimulus cycle (Moxon, 1967; van den Honert and Stypulkowski, 1987; Javel and Shepherd, 2000; Litvak *et al.*, 2001). Such entrainment does not occur for acoustic stimulation. On the other hand, because electric stimulation can produce much higher discharge rates than acoustic stimulation, neural refractoriness causes severe distortions in the temporal discharge patterns with electric sinusoids, particularly for frequencies above 800 Hz (Parkins, 1989; Javel, 1990; Javel and Shepherd, 2000; Wilson *et al.*, 1997; Litvak *et al.*, 2001). With complex waveforms such as vowels, the narrow dynamic range to electric stimulation (only a few decibels) allows only the largest waveform peaks to be represented in the temporal discharge patterns (van den

[a)]Portions of this work were presented as a poster at the ARO Midwinter Meeting in St. Petersburg Beach, FL, 2001 and the CIAP Conference in Monterey, CA, 2001.
[b)]Electronic mail: leonid.litvak@advancedbionics.com

Honert and Stypulkowski, 1987; Knauth *et al.*, 1994). In short, the combination of a limited dynamic range, neural refractoriness, abnormally high discharge rates and exaggerated synchrony make the temporal discharge patterns to electric stimuli highly unnatural compared to normal acoustic stimulation. The goal of the present paper and its companions (Litvak *et al.*, 2003a, b) is to test the physiological validity of one proposed scheme (Rubinstein *et al.*, 1999) for restoring more natural temporal cues with cochlear implants.

Rubinstein *et al.* (1999) proposed that the temporal coding of electric stimuli might be improved by introducing a sustained, high-frequency, ''desynchronizing'' pulse train (DPT) in addition to the basic processor output. Their rationale is that noise in the neural membrane, as well as across-fiber differences in sensitivity, refractoriness, adaptation, and other properties, will result in the responses across fibers being desynchronized shortly after the onset of DPT stimulation. Such desynchronization would lead to an improved representation of the stimulus waveform in temporal discharge patterns. Rubinstein *et al.* (1999) further argued that responses to the DPT might mimic the spontaneous activity in a healthy ear. Because fibers with higher spontaneous discharge rates have lower thresholds (Liberman, 1978), spontaneous activity is thought to enhance detection of weak signals. The mechanism for enhanced detection may be related to the phenomenon of stochastic resonance (Wiesenfeld and Moss, 1995). If spontaneous activity could be restored by the DPT, then similar benefits could be provided to cochlear implantees. Finally, Kiang *et al.* (1970) suggested that some forms of tinnitus, a common complaint among the hearing impaired, are related to the absence of spontaneous activity in a portion of the cochlea. By restoring a form of spontaneous activity under electric stimulation, it may be possible to reduce tinnitus in these patients.

Only a few studies have investigated the responses of auditory-nerve fibers (ANFs) to the high-rate, electric pulse trains proposed by Rubinstein *et al.* (1999). Moxon (1967) studied responses of ANFs in cats to 900-pps electric pulse trains of relatively long duration (200 s). He found that discharge rates were large at stimulus onset, but decreased markedly over the course of the stimulus. Following adaptation, the discharges were stochastic in that they skipped one, two, or more stimulus pulses in an irregular sequence, and thus their timing was similar to that of normal spontaneous activity (Kiang *et al.*, 1965). However, Moxon reported that sustained discharge rates could still exceed 500 spikes/s, which is much higher than the rates of spontaneous activity in a healthy ear (Liberman, 1978). On the other hand, the higher pulse rates (5 kpps) proposed by Rubinstein *et al.* (1999) might produce greater adaptation of the neural response and, therefore, discharge rates more consistent with spontaneous activity.

In a previous report (Litvak *et al.*, 2001), we investigated responses to short (200 ms), high-rate (1–5 kpps) DPTs. We found that ANF responses to the DPT were stochastic and resembled spontaneous activity for low to moderate firing rates ($<$150 spikes/s), but that such moderate rates could only be achieved over a narrow (2–3 dB) range of stimulus levels. Since neural thresholds differ by as much as 10 dB across fibers, this finding implies that only a small fraction of the fibers can respond to any given DPT at rates appropriate for spontaneous activity. Here, we hypothesize that neural adaptation over the course of a long-duration DPT might bring the discharge rates in response to the DPT within the range of normal spontaneous activity for a substantial fraction of the auditory-nerve fiber population. In addition to using short DPTs, our previous study did not address the safety of sustained DPT stimulation. Several studies have reported changes in excitability of the auditory nerve following sustained, high-frequency electric stimulation (Duckert and Miller, 1982; Shepherd, 1986; Killian, 1994; Tykocinski *et al.*, 1995b, 1997), suggesting that such stimulation might cause long-term nerve damage.

In the first part of this paper, we report the results of single-unit experiments designed to test whether responses of ANFs to a long-duration DPT resemble spontaneous activity in a healthy ear. Specifically, (1) the sustained discharge rates evoked by the DPT are compared to the distribution of spontaneous discharge rates (Liberman, 1978, 1990); (2) interval histograms computed from DPT responses are compared to histograms for spontaneous activity (Kiang *et al.*, 1965); and (3) cross-fiber correlations computed from pairs of simultaneously recorded DPT responses are compared to correlations for spontaneous activity (Johnson and Kiang, 1976).

In the second part of this paper we describe measurements of electrically evoked compound action potentials (ECAPs) for a single-pulse probe following DPT stimulation of long duration. While a truly continuous DPT may be unsafe, a DPT that is turned on only when needed may be less traumatic to the auditory nerve. To test this hypothesis, we recorded ECAPs following a 10-min DPT, and compared these responses to ECAPs recorded after 1 h of DPT stimulation.

II. METHODS

A. Animal preparation

The animal preparation was similar to that described by Litvak *et al.* (2001). Cats were anesthetized using dial in urethane (75 mg/kg), and the bulla opened to expose the round window. An intracochlear stimulating electrode was inserted about 8 mm into the cochlea through the round window. The electrode was a Pt/Ir ball, 400 μm in diameter. A similar ball electrode was inserted just inside the round window and used as the return electrode. The round window was then sealed with connective tissue. The resulting wide bipolar configuration was found to reduce stimulus artifact compared to the monopolar configuration (using the ear bar as a return electrode) used in our previous work.

Co-administration of kanamycin (subcutaneous, 300 mg/kg) followed by slow intravenous injection of ethacrinic acid (1 mg/min) was used to deafen the animals (Xu *et al.*, 1993). A 100-μs condensation click was used to measure compound action potentials (CAPs), which were continuously monitored in the non-implanted ear while ethacrinic acid was administered. This measurement provides an assessment of the effect of the drug protocol on hair cell function that is not confounded by the uncontrolled additional trauma caused by

TABLE I. Residual hearing status, DPT level, type of DPT modulation, and number of data records for the ten animals included in this study. Hearing status is assessed by the CAP threshold for a 100-μs click in the non-implanted ear. A click at 0 dB *re*: 1 V approximately corresponds to a 100 dB SPL peak sound pressure. When two thresholds are given, the first one was measured just after deafening, and the other one at the end of the experiment. DPT modulation is coded as follows: U is unmodulated, S is sinusoidal modulation, and V is vowel modulation.

Cat	CAP threshold (dB *re*: 1 V)	DPT level (dB *re*: 1 mA)	DPT modulation	No. of records
lml11	−35/2	6	U	9
lml13	0	6	U/S	26
lml15	−35/−17	8	U/S	7
lml16	−30/>10	0	U/S	30
lml17	>10	6	S	33
lml18	−22/−4	−1	S	25
lml19	−12	4	V	4
lml20	−7	8	V	5
lml21	−40	8	V	21
zms2	−30/−5	4	V	13

electrode insertion. CAP threshold, which is defined as the lowest click level evoking a visually detectable CAP response, exhibited a rapid increase within 1 h of ethacrinic acid injection. However, consistent with previous reports which used a similar protocol (Shepherd and Javel, 1997), we failed to achieve complete deafening in six out of ten animals, even with the maximum safe dose of 40 mg/kg. Table I gives the residual CAP threshold in the non-implanted ear for the ten animals used in this study. Residual CAP thresholds ranged from 80 to >110 dB peak SPL, and often dropped slowly during the course of the experiment (which lasted for 2–3 days). Although CAP thresholds in the implanted ear were not measured, based on our previous work (Litvak *et al.*, 2001), they would be expected to exceed 110 dB SPL, reflecting the additional trauma caused by electrode insertion. Thus, our measure of residual hearing is highly conservative.

B. Stimuli

In the four initial experiments, we recorded responses of single ANFs to unmodulated, 10-min, 5-kpps ''desynchronizing'' electric pulse trains (DPTs). Pulse trains were composed of biphasic (cathodic/anodic) pulses (25 μs per phase). In later experiments, these unmodulated DPTs were replaced by pulse trains that were modulated by either a low-frequency sinusoid or a filtered vowel for 200–400 ms every second (Table I). Details of the modulation waveforms are described in the companion papers (Litvak *et al.*, 2003a, b). For the purpose of this paper, only responses recorded during the unmodulated portions were analyzed. These responses were only used for analysis of average discharge rates; temporal discharge patterns were not analyzed for intermittently modulated stimuli because this analysis requires longer continuous data records than the 600–800-ms unmodulated segments.

For the most part, a single level of the DPT was used for each animal so as to simulate a DPT-enhanced stimulation strategy where an ongoing DPT would be superimposed upon the output of the sound processor, as suggested by Rubinstein *et al.* (1999). Our experimental design is therefore that of a population study, where responses to the same DPT stimulus are obtained for all the sampled fibers. To maximize data yield, the standard DPT level was selected to stimulate a high percentage of the fibers. For this purpose, electrically evoked compound action potentials (ECAPs) were used to assess the dynamic range of the auditory nerve to electric stimulation for each animal prior to the single-unit recordings. The ECAP recording electrode (a 400-μm-diam Pt/Ir ball) was placed on the auditory nerve near the internal auditory meatus. The stimulus used to evoke the ECAP was a biphasic pulse (25 μs per phase) presented at a 15/s repetition rate. The pulse polarity alternated with each presentation and 30 single-pulse responses were averaged. Because the stimulus artifact is largely reversed by a polarity change, while the neural response retains the same polarity, most of the artifact is cancelled by averaging the responses to opposite polarities.

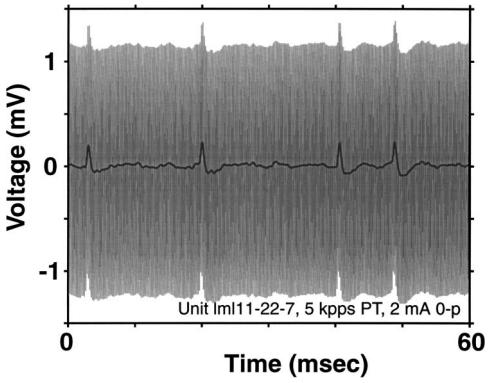


FIG. 1. Artifact cancellation technique. The gray trace shows the raw signal recorded from the micropipette, containing spikes as well as the 5-kpps stimulus artifact. The black line is the signal after processing through a 0.2-ms moving-average filter. The artifact has been entirely removed, while spikes are preserved.

All stimuli were digitally synthesized using a 40-kHz, 16-bit D/A converter, and presented through a custom-built, high-bandwidth stimulus isolator. All stimulus levels are expressed as peak currents.

C. Single-unit recordings

Standard techniques were used to expose the auditory nerve via a dorsal approach (Kiang *et al.*, 1965). We recorded from single fibers in the auditory nerve using glass micropipettes filled with 3M KCl. A 25-ms, 5-kpps burst of electric pulses presented at the standard DPT level with a repetition rate of 10/s was used as the search stimulus. Neural signals were recorded with a 20-kHz sampling rate. A digital signal processor (DSP) was used to separate spike discharges from the stimulus artifact during the unmodulated stimulus segments in real time. Specifically, a digital moving average filter with a width equal to the 0.2-ms period of the pulse train removed most of the stimulus artifact while preserving the spike shape. Figure 1 shows an example of artifact cancellation for a signal recorded during stimulation with a 5-kpps pulse train. Recorded waveforms were also streamed to disk for off-line analysis.

The threshold of each fiber to a short electric pulse train (25 ms, 5 kpps) was measured prior to DPT stimulation using a tracking procedure (Taylor and Creelman, 1967). Threshold was defined as the stimulus level that evoked a single spike during the last 10 ms of the stimulus for 50% of the presentations.

In general, the limited contact time with a fiber allowed us to record responses to only one 10-min DPT at the standard level. For a few fibers, however, we were able to record responses to both unmodulated and partially modulated DPTs, or repeat the same stimulus twice. In a few cases, we also presented DPT stimuli 1–4 dB above the standard level in order to examine the effects of larger discharge rates on temporal discharge patterns.

Stimulation with a sustained DPT may have long-term effects on ANF responsiveness. To minimize these long-term effects on our single unit recordings, we waited for at least 10 min between recordings of responses to the DPT (which was at most 10 min long). Whenever possible, we also recorded single-unit thresholds after DPT stimulation, and verified that these returned to prestimulation levels prior to collecting further data. Typically, thresholds were elevated by no more than 5 dB immediately after DPT stimulation, and recovered to prestimulation levels within 2 to 3 min. During a 2–3-day single-unit experiment, the total DPT exposure was less than 3 h.

Possible hair-cell mediated activity (''electrophonic hearing'') may also complicate the interpretation of ANF responses to electric stimulation. Hair cell function may not have been completely eliminated by the combined effects of the acute deafening protocol and the mechanical trauma caused by electrode insertion into the cochlea. To mitigate the effects of remaining hair cells, only units lacking sustained spontaneous activity are included in this report. Depending on the animal, between 0% and 30% of the units were rejected based on this criterion. In one animal with particularly large residual acoustic responses, an additional five units were eliminated because they responded to acoustic clicks delivered to the implanted ear.

D. Identification of vestibular responses

Both the auditory nerve and the vestibular nerve pass through the internal auditory meatus. The border between the vestibular and the auditory nerve can usually be visualized from our dorsal view (Kiang *et al.*, 1965). Stereotaxic and neurophysiological criteria have been developed in our laboratory to identify vestibular nerve fibers (Liberman and Kiang, 1978) (see below). These criteria were only strictly adopted in later experiments. Consequently, some of the recordings from the earlier experiments may be from vestibular fibers. This issue is discussed in Sec. III E.

To quantitatively assess the possible influence of vestibular recordings on results of the earlier experiments, we deliberately recorded from both vestibular and auditory-nerve fibers in some experiments. The recording microelectrode was oriented 45° from the horizontal plane and 10°–15° medio-lateral from the sagittal plane (Liberman and Kiang, 1978). We labeled a fiber as vestibular if three conditions were met: (1) the point of entry of the electrode was anterior to the border between the vestibular nerve and the auditory nerve; (2) the depth of the electrode was less than 1 mm from the nerve surface; and (3) a unit with regularly firing spontaneous activity was found deeper in the nerve. Since only vestibular units have regular spontaneous activity, detection of regular spontaneous activity is an unambiguous indication that the electrode is in the vestibular nerve, and not in the auditory nerve beneath (Walsh *et al.*, 1972). A fiber was labeled as acoustic if either the point of entry of the electrode was posterior to the auditory/vestibular nerve border, or the fiber was at least 1 mm from the nerve surface and no regular vestibular unit was found deeper (we always looked at least 0.3 mm deeper). Neurons that could not be classified as either vestibular or acoustic were excluded from analysis.

E. Data analysis

Both average discharge rates and, for unmodulated DPTs, temporal discharge patterns were analyzed. Peristimulus time (PST) histograms were constructed by computing the average discharge rate in successive 1-s windows as a function of time following DPT onset. For unmodulated DPTs, temporal discharge patterns were analyzed using interspike interval histograms and tests of serial dependence between intervals. Because the tails of interspike interval distributions and measures of serial dependence are highly sensitive to deviations from stationarity, selecting a response window over which the discharges behave in a manner consistent with stationarity is critical for these analyses (Kelly *et al.*, 1996). In practice, we used an *ad hoc* statistical criterion for identifying a window over which the discharge rate is nearly constant. Specifically, a response window was accepted as stationary if the slope of the regression line to that portion of the PST histogram was ''well within'' the distribution of slopes computed from random spike sequences having the same length and same interspike interval distribution as the data. By ''well within,'' we mean that the slope had to be smaller in absolute value than 30% of the slopes expected for a random sequence of spikes. This 30% criterion leads to rejection of 70% of the records from the random sequence, and is therefore very stringent.

For each recorded response to the DPT, the stationary time window was determined by beginning with the entire data record, and iteratively decreasing the window size by 10 s on the left, and 2 s on the right, until the window passed the stationarity test. In practice, small windows were selected for responses that slowly adapted throughout DPT stimulation, while larger windows were selected for units that discharged at a nearly constant rate after initial adaptation. The window size was 150 s on average, and ranged from 5 to 500 s across fibers.

III. RESULTS

A. DPT level selection

To assess the sensitivity of the auditory nerve to electric stimulation, ECAPs were measured by averaging responses to 30 biphasic pulses (25 μs/phase) which alternated in po-

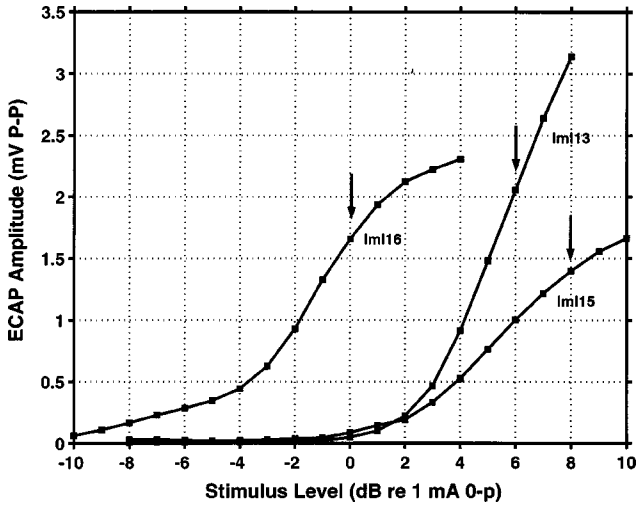


FIG. 2. ECAP growth curves and DPT levels for three cats. Each point represents the peak-to-peak ECAP amplitude evoked by an alternating, biphasic pulse presented at a rate of 15/s. In most animals, the DPT level (arrow) was chosen to be roughly 1 dB below the ''knee,'' where the ECAP growth curve begins to saturate. In animals where a knee was not apparent (e.g., lml13), DPT level was set at the highest stimulus level that did not evoke a facial twitch.

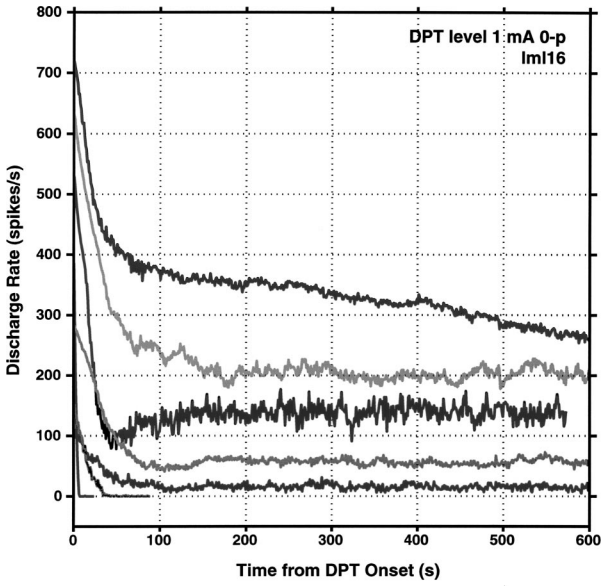


FIG. 3. Discharge rate as a function of time from the onset of the DPT stimulus for 7 units from one animal (lml16). The stimulus was a 10-min, 5-kpps electric train of biphasic pulses presented at 1 mA 0-p. Rate was estimated over successive 1-s windows.

larity. Figure 2 shows the peak-to-peak ECAP amplitude as a function of pulse level for three animals. These curves were chosen to illustrate interanimal variability. All animals showed a segment of rapid ECAP growth at low stimulus levels. As the level increased further, the rate of growth slowed in two of the three animals, suggesting that fewer fibers were recruited by similar increments in pulse level. Assuming that the distribution of fiber thresholds is bell-shaped, levels where the ECAP nearly saturates probably stimulate a large percentage of the fibers in the auditory nerve. Whenever possible, these levels were chosen for the DPT (arrows). In some animals (such as lml13), the stimulus produced facial twitches before the ECAP saturated. In these cases, a DPT level just below the twitch threshold was selected. The DPT levels used in the ten experiments ranged from −1 to 8 dB *re*: 1 mA, with a median of 6 dB (Table I).

B. Time course and distribution of discharge rate evoked by the DPT

Our results are based on 170 data records obtained from 138 auditory-nerve fibers in ten cats. None of these fibers had spontaneous activity, and, if tested, none responded to click sounds. The data set includes 46 responses to unmodulated DPTs and 124 responses to partially modulated DPTs (for which only the unmodulated segments were analyzed in the present paper). In most cases (143 records), responses to the DPT were recorded for at least 2.5 min; in the 27 remaining cases, the fiber stopped responding after less than 2.5 min of DPT stimulation, and the record was not used further except to note that the response completely adapted.

Figure 3 shows the discharge rate as a function of time from the onset of an unmodulated, 1 mA peak DPT for seven auditory-nerve fibers from one cat. While initially the discharge rates could be very large (exceeding 700 spikes/s for some fibers), they decreased considerably over the course of DPT stimulation. Five units showed rapid adaptation over the first 100 s of stimulation, followed by either slower adaptation or a steady response. Two other units completely stopped discharging after 1–2 min of DPT stimulation.

Figure 4 shows the discharge rate distribution after 100 s of DPT stimulation for all fibers that initially responded to the DPT at the standard level. In this figure, responses to unmodulated pulse trains are pooled together with responses during the unmodulated parts of the partially modulated pulse trains. The solid line is the cumulative distribution of discharge rates evoked by the DPT. The dashed line is the cumulative distribution of spontaneous discharge rates in

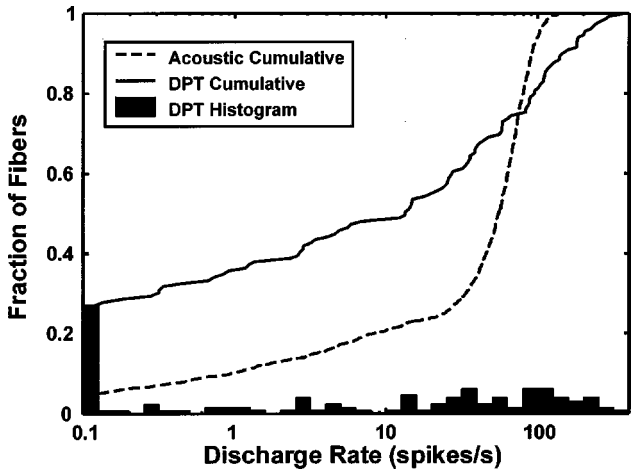


FIG. 4. Discharge rate distributions for spontaneous activity and DPT-evoked responses. The black vertical bars represent a histogram (bin width 2.5 spikes/s) of the discharge rates in response to the DPT. For each unit, the discharge rate was averaged from 100 to 130 s after stimulus onset. Only responses to the DPT presented at the standard level in each animal (Table I) are included, and each ANF is included only once, even if multiple data records are available. The black line shows the cumulative distribution of discharge rates to the DPT. The dashed line shows the cumulative distribution of spontaneous discharge rates from four healthy ears (Liberman, 1990).

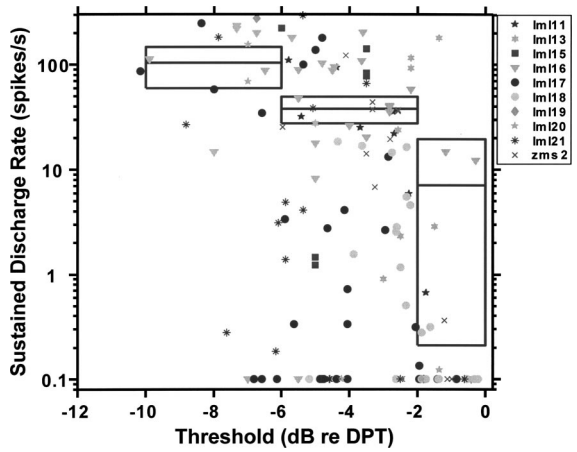


FIG. 5. Average discharge rate 100–130 s after DPT onset as a function of fiber threshold. To allow comparisons across experiments, threshold is expressed relative to DPT level. Data from ten cats are coded by different symbols. The horizontal lines show the mean discharge rate for fibers that have thresholds in the range represented by the horizontal extent of each line. Boxes around each line represent the 99% confidence interval for the mean discharge rate estimate. The confidence intervals were computed using the bootstrap method (Efron and Tibshirani, 1993).

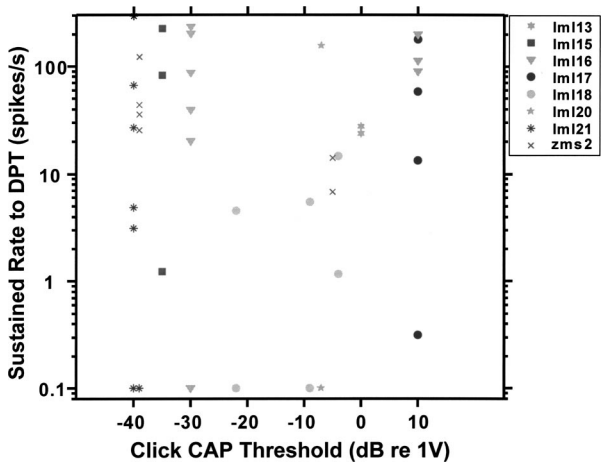


FIG. 6. Sustained discharge rate of individual fibers in response to the DPT as a function of the residual hearing status of the animal. Hearing status was assessed by the click-evoked CAP threshold in the non-implanted ear. To be included in this figure, a record had to meet three criteria: (1) the DPT was presented at the standard level for each animal (Table I), (2) single-pulse threshold was at least 2 dB below DPT level, and (3) a CAP threshold was measured within 5 h of the single-recordings. Because the CAP threshold changed slowly over the course of an experiment, data from one cat may be represented at several horizontal positions in this figure. Animals with CAP thresholds above 10 dB *re*: 1 V had no detectable residual hearing.

four healthy ears (Liberman, 1990). Discharge rates in response to the DPT vary over a wide range, as do spontaneous rates. However, the rate distribution for the DPT differs from that for spontaneous activity in that (1) a larger proportion (46% vs. 17%) of units respond to the DPT with discharge rates below 5 spikes/s, including the 26% that stopped discharging after 100 s of stimulation, and (2) about 13% of the units respond to the DPT at higher discharge rates ($>$122 spikes/s) than any spontaneously responding unit. Overall, given the large differences in thresholds and the narrow dynamic ranges for responses to short electric stimuli (van den Honert and Stypulkowski, 1984; Javel *et al.*, 1987; Miller *et al.*, 1999; Litvak *et al.*, 2001), the discharge rate distribution for the DPT is surprisingly similar to that for spontaneous activity.

Figure 5 shows that nearly all units with thresholds within 2 dB of the standard DPT level either stopped responding to the DPT, or responded at a very low discharge rate ($<$20 spikes/s). In contrast, the discharge rates of units with thresholds more than 2 dB below the DPT level varied widely from 0 to over 300 spikes/s. Although the mean discharge rate correlates negatively with threshold [$R = -0.41$, $p < 0.001$, permutation test for correlation (Good, 2000, p. 46)], there is considerable variability in discharge rates among units that have a given threshold. For example, fibers with thresholds 4 dB below DPT level had sustained discharge rates ranging from 0 to over 200 spikes/s. This wide range of discharge rates among units having similar thresholds is a consistent finding across animals.

The cross-fiber differences in sustained discharge rates might reflect differences in the status of residual hair cells. The degree to which hair cells were affected by the drug protocol was assessed by measuring click-evoked CAP thresholds in the non-implanted ear. Figure 6 shows the sustained discharge rate to the DPT against the CAP threshold in the non-implanted ear. To be included in the plot, the CAP threshold had to be measured within 5 h of the single-unit recordings. This 5-h criterion was chosen based on the slow rate of increase in CAP threshold over the course of the experiment, which had a time constant of 12–24 h. Because CAP thresholds were measured sporadically, the number of units in this figure is smaller than in Figs. 4 and 5. Regression analysis revealed no significant correlation between CAP threshold and discharge rate to the DPT ($p = 0.94$, permutation test for correlation). Moreover, there was a large variability in discharge rates even among fibers from nearly deaf animals (CAP thresholds $>$10 dB re: 1 V, about 110 dB peak SPL). Thus, the results suggest that hearing status alone cannot account for the differences in DPT-evoked discharge rates across fibers.

C. Temporal discharge patterns

The temporal discharge patterns in response to unmodulated DPTs were analyzed quantitatively and compared to temporal patterns of spontaneous activity in healthy ears. This analysis was performed for a time window over which the response was determined to be approximately stationary using a statistical algorithm described in Sec. II D. To ensure that histogram estimates would be reliable, the stationary window had to contain at least 2000 spikes for a record to be used in this analysis. Twenty-nine records met this criterion.

1. Interspike interval histograms

First-order interspike interval histograms (IH) for spontaneous activity in a healthy ear are nearly exponential for intervals exceeding the refractory period (Kiang *et al.*, 1965), as would be expected for a Poisson process. In contrast, responses to the DPT during the stationary window exhibited a variety of IH shapes illustrated in Fig. 7. The top row shows the IH for intervals between 0 and 20 ms, while the middle row shows the same histogram on a logarithmic

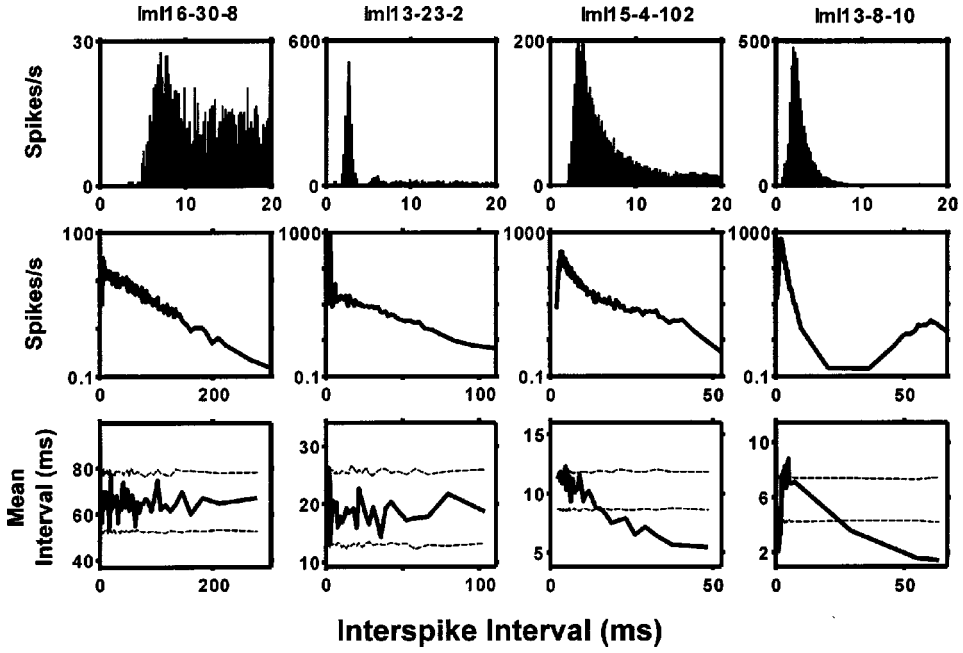


FIG. 7. Characterization of temporal discharge patterns in response to the DPT for four auditory-nerve fibers. The top row shows the first 20 ms of the interspike interval histograms (IH) computed with a fixed 200-μs bin width. The second row shows the entire interval histogram on a logarithmic vertical scale. To accommodate different average discharge rates, the bin width is adjusted so that each bin contains exactly 50 intervals. The histograms in the top two rows are normalized to represent the interval probability density function in spikes/s. The bottom row shows the conditional mean interval length as a function of the preceding interval (solid line) using the same bins as in the second row. The plot also shows 99% confidence intervals for the conditional mean interval assuming that the intervals show no serial dependence.

vertical scale where an exponential histogram would appear as a straight line. The time scale is adjusted in each panel to include all the intervals for that record.

Seven of the 29 data records (24%) exhibited a nearly exponential IH, like most spontaneously responding fibers in a healthy ear (left column). Another eight records (28%) exhibited one or more interval modes below 10 ms, followed by a nearly exponential tail (second column from left). Both patterns have been reported previously for brief (250 ms) stimulation with high-rate pulse trains (Litvak *et al.*, 2001). A third type of response found in 14 records (48%) consisted of one or several modes below 10 ms followed by a nonexponential tail beyond 20 ms (two right-most columns). The rightmost interval histogram corresponds to a ''bursting'' response. In this case, bursting was regular, with a repetition period near 60 ms reflected as a mode of the IH.

A formal definition of interval histogram type was developed based on a metric called the interval-histogram exponential shape factor (IH-ExpSF) (Litvak *et al.*, 2001). The IH-ExpSF test was separately applied to both the initial histogram segment (extending from the largest mode to 20 ms) and the histogram ''tail'' extending from 20 ms to the largest interval.[1] For each of these two segments, the IH-ExpSF was computed by first fitting the data with both a single exponential and, piecewise, with three exponentials. For the piecewise fit, the data were partitioned so that each piece contained an equal number of intervals in order to make the fit equally reliable for each piece. We computed the root mean squared (rms) error of each fit to the data, and defined the IH-ExpSF as the ratio of the rms error for the piecewise fit to that for the single exponential fit. The IH-ExpSF for samples from a Poisson process is approximately 1. A histogram segment was labeled as nonexponential if its IH-ExpSF was smaller than 99% of the IH-ExpSFs computed from simulated samples of a Poisson process having both the same number of spikes and the same average discharge rate as the data.

A Type IA histogram is one that was classified as exponential by the IH-ExpSF test over both the initial segment and the tail (e.g., Fig. 7, left). Type IB histograms are exponential in the tail, but not in the initial segment (e.g., Fig. 7, second from left). Type II histogram are not exponential in the tail (e.g., Fig. 7, two rightmost fibers).

Responses with histograms of Type IA tend to have lower discharge rates than Type IB and II responses. The difference in the mean discharge rates between IA histograms and the other responses is highly significant [$p = 0.0095$, permutation test (Good, 2000, p. 37)]. The differences between the mean discharge rates of Type IB and Type II responses are not significant. The dependence of IH type on average discharge rate suggests the type may also correlate with stimulus level and sensitivity.

Both Type IB and II responders occurred in deaf preparations (CAP threshold >0 dB) as well as in those with residual hearing. Although exponential (Type IA) interval histograms were seen exclusively in deaf preparations, in only two units were measurements of DPT responses that showed IA interval histograms made sufficiently close in time to a click-evoked CAP to make the comparison useful. Thus, residual hearing does not seem to account for the nonexponential response types.

2. Serial dependence between intervals

Spontaneously responding ANFs in a healthy ear show little dependence between successive interspike intervals (Kiang *et al.*, 1965; Lowen and Teich, 1992); consequently, they can be approximately described as a renewal process (Cox, 1967). A formal test of serial dependence is the conditional mean test (Rodieck *et al.*, 1962; Johnson, 1996). First, the mean interval is computed as a function of the previous interval. Next, it is determined whether this mean is significantly dependent on the previous interval.

The bottom row of Fig. 7 illustrates the conditional

TABLE II. Conditional mean test results as a function of interval histogram type.

IH type	Conditional mean test	
	No. passed	No. failed
IA	6	1
IB	5	3
II	0	14

mean test for four units. For each unit, the solid line shows the mean interval as a function of the previous interval. The dashed lines show 99% confidence intervals for the mean interval assuming that there is no serial dependence between intervals. These confidence intervals were obtained by repeatedly shuffling the intervals and recomputing the conditional mean interval. A record would fail the test ($p<0.01$) if the mean interval exceeded the confidence intervals more than three times for each 100 points. With this criterion, the two left records pass the test, while the two right records fail.

About 60% (18/29) of the DPT responses showed a marked dependence of the mean interval on the previous interval, with a tendency for shorter intervals to follow longer ones (e.g., Fig. 7, bottom row, two right-most plots). The results of the conditional mean test were strongly correlated with the interval histogram type (Table II). Of the six records that had Type IA interval histograms, only one failed the conditional mean test, while all the Type II responders failed the test. The results were nearly split for Type IB responders.

D. Simultaneous recordings from pairs of fibers

Simultaneous recordings of spontaneous activity from pairs of auditory nerve fibers show no evidence for cross-fiber correlation in spike times (Johnson and Kiang, 1976). To assess the degree of cross-fiber correlation with DPT stimulation, we simultaneously recorded responses to a DPT from pairs of fibers, and computed the cross-correlation histograms (Perkel *et al.*, 1967). This analysis was performed over a segment in which both responses were judged to be stationary. We examined lags up to ± 40 ms at a resolution of 200 μs so that any correlation related to the 200-μs period of the stimulus would average out. An example cross-correlation histogram is shown in Fig. 8. The histogram appears flat, suggesting no significant correlation between the two fibers.

A statistical test was devised to determine whether fluctuations in the cross-correlation histograms are within limits of statistical variability. For each fiber pair, a 99% confidence interval was established by randomly shuffling the interspike intervals in each record 100 times and computing the cross-correlation histogram for each shuffle. Because shuffling is performed independently for each fiber in the pair, the resulting spike times are guaranteed to be uncorrelated. Under the null hypothesis, we expect the cross-correlation histogram to cross the confidence intervals on average four times. More than eight crossings would cause the uncorrelated response hypothesis to be rejected ($p<0.01$). The pair in Fig. 8 showed only one crossing, and therefore passed the correlation test.

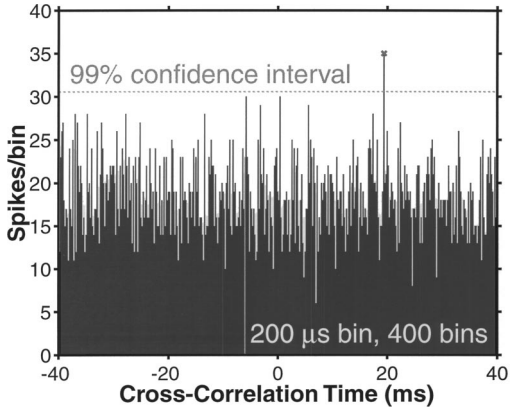


FIG. 8. Cross-correlation histogram (bin width 200 μs) in response to the DPT for a pair of fibers recorded simultaneously. The dashed line marks the 99% confidence interval based on the null hypothesis that the two spike trains are uncorrelated. The X indicates a crossing of the confidence interval. Because there are 400 bins, about four crossings are expected on average under the null hypothesis.

We failed to detect any significant correlation in all four pairs from which we recorded (Table III). This result, which resembles that found by Johnson and Kiang (1976) for spontaneous activity, is consistent with responses to the DPT being statistically independent from fiber to fiber.

E. Vestibular responses

Some of the recordings described in the previous sections may have come from vestibular fibers lacking spontaneous activity (Walsh *et al.*, 1972). Vestibular responses to electric stimulation of the cochlea have been reported (Hartmann *et al.*, 1984; Shepherd and Javel, 1997). To address

TABLE III. Number of times that the cross-correlation histogram between spike trains simultaneously recorded from two fibers crosses the 99% confidence interval.[a]

Data record	Record duration (s)	No. of spikes	No. of crossings
lml13-87-2a	10	3069	5
lml13-87-2b		1799	
lml13-88-3a	10	2721	7
lml13-88-3b		2953	
lml13-89-2a	60	5013	3
lml13-89-2b		1118	
zms2-3-2a	200	40 158	4.2[b]
zms2-3-2b		85 724	

[a]To avoid effects of slow adaptation in discharge rate that occurred for both fibers in Record zms2-3-2, the test was applied separately to ten consecutive 20-s data windows. This segmentation was predicated by applying the test to simulated spike trains from two independent, slowly adapting, Poisson processes. Both the average rate and the rate of adaptation of the simulated spike trains were matched to those in zms2-3-2. Although these spike trains were statistically independent by construction, the cross-correlation test failed for windows longer than 20 s. Because simulated spike trains did pass the test when there was no adaptation, we assume that adaptation is responsible for this failure.

[b]Average number of crossings for ten consecutive 20-s windows of data.

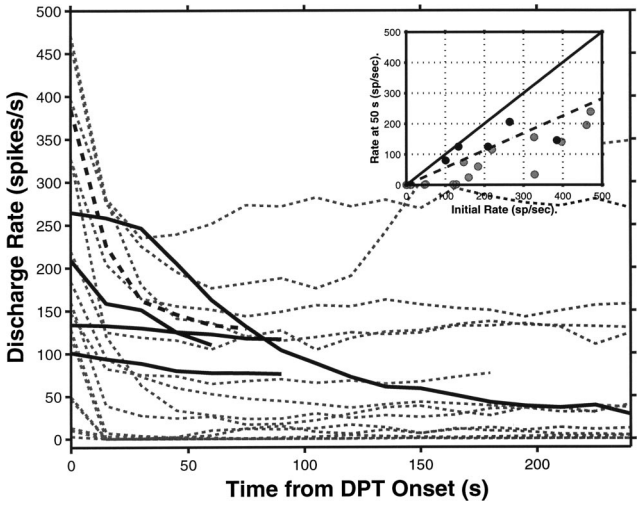


FIG. 9. Average discharge rate as a function of time from DPT onset for both auditory fibers (gray, dotted) and non-spontaneously-active vestibular fibers (black). The inset shows the initial discharge rate (from 0 to 15 s after DPT onset) versus the adapted discharge rate (from 50 to 75 s) for both auditory (gray) and vestibular (black) fibers. For four out of five vestibular fibers, the adapted discharge rate is more than 0.56 times larger than the initial discharge rate (dashed line). A sixth vestibular fiber had a nearly zero rate. In contrast, all the auditory fiber data lie below the dashed line. The solid black line in the inset shows where the rate after 50 s equals the initial discharge rate. Units whose responses fall on this line would show no adaptation.

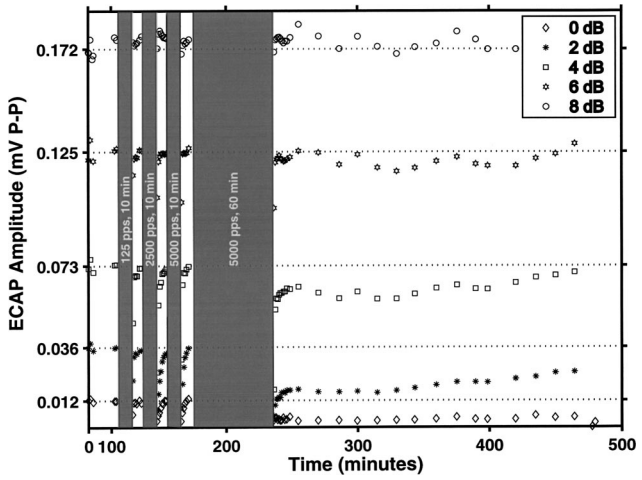


FIG. 10. ECAP amplitude for a probe pulse stimulus as a function of time for one experiment (zms2) in which various conditioning stimuli were applied. Symbols indicate probe levels in dB *re*: 1 mA 0-p. The gray areas mark times during which a 6-dB *re*: 1 mA conditioning pulse train was presented and no ECAP measurements were made. The pulse rate and duration of the conditioning stimulus are indicated inside the gray areas.

this issue, we deliberately recorded from both vestibular and auditory fibers in some experiments to determine how the two differ in their responses to the DPT.

Fibers identified as vestibular based on stereotaxic and physiological criteria often responded spontaneously. Discharge patterns for these units were similar to those encountered in a healthy ear, and included both regular and irregularly discharging responses (Walsh *et al.*, 1972). In addition, some vestibular fibers did not respond spontaneously and were detected only because they responded to the search stimulus.

Vestibular fibers with spontaneous activity had high thresholds to biphasic electric pulses. For 70% of the vestibular fibers that we contacted, threshold was higher than the standard DPT level for that animal. However, some vestibular fibers were easily excited by electric stimulation. One fiber's threshold was 7.5 dB below the DPT level. These results are consistent with previous reports of vestibular fibers with low thresholds to electric stimulation of the cochlea (Hartmann *et al.*, 1984; Shepherd and Javel, 1997).

Figure 9 shows that the time course of adaptation of responses to the DPT in vestibular fibers with zero spontaneous rate usually differs from that in auditory fibers. Of the five vestibular fibers, four (solid black lines) adapted more slowly during the first 50 s of DPT stimulation than did any auditory fiber. For one vestibular unit, however, the adaptation was unremarkable when compared to that of auditory-nerve fibers (dashed black line).

The inset plots the rate after 50 s of DPT stimulation against the initial discharge rate for both vestibular (black) and auditory (gray, dotted) units. For five of the six vestibular fibers, the discharge rate at 50 s was greater than that in any auditory fiber when normalized by the initial rate. The dashed straight line representing a 0.56 ratio of the discharge rate at 50 s to the initial rate almost completely separates auditory from vestibular units. This ratio criterion was applied to data from earlier experiments to determine which of these units might be vestibular. Only 15% of the records were labeled as vestibular by this criterion. None of the conclusions of the earlier analyses would be changed if these records were excluded from analysis, and this is also true even if 30% of the unclassified units with the lowest adaptation rates are excluded.

F. Recovery of evoked responses from DPT stimulation

To reliably evoke a sustained discharge rate in an ANF, the DPT level must exceed the single-pulse threshold by at least 2 dB (Fig. 5). Consequently, only high-level DPTs can evoke sustained responses in a large number of fibers. The influence of such intense DPTs on subsequent evoked responses was investigated methodically in two cats after completion of the single-unit recordings. Because the results were similar in both cats, we focus on results from one experiment. This cat had little residual hearing (click-evoked CAP threshold of 2 dB *re*: 1 V, about 102 dB peak SPL) in the non-implanted ear at the time that the conditioning experiment was undertaken.

Figure 10 shows the peak-to-peak ECAP amplitude as a function of time for five different probe levels (0, 2, 4, 6, and 8 dB *re*: 1 mA 0-p). In between these ECAP measurements, the auditory nerve was stimulated with a conditioning pulse train presented at 6 dB *re*: 1 mA. The rate and the duration of the conditioning pulse train were systematically varied.

As a control, ECAPs were first measured over a course of 2 h without a conditioning stimulus, and found to be stable within ± 0.25 dB. Next, a 125-pps conditioning pulse train was applied for 10 min. This low-rate pulse train is representative of carriers used in some CI processors. Evoked responses were measured immediately and repeat-

edly after the end of the conditioning stimulus. For all probe levels, evoked responses recovered rapidly from this low-frequency stimulation, with the recovery being nearly complete 2 min after the end of the conditioning stimulus.

Next, 10-min conditioning stimuli of 2.5 and 5 kpps were applied. These higher pulse rates are appropriate for a DPT-enhanced strategy. Evoked responses recorded immediately after these stimuli were strongly suppressed, particularly for low-level probes. The response to a 0 dB *re*: 1 mA probe was completely suppressed. However, the response to an 8-dB probe was not affected by a 6-dB conditioner of either frequency. For all probe levels, responses recovered to near prestimulation level within 5 to 10 min.

Finally, after 1-h stimulation with a 5-kpps conditioner, responses to low-level probes were strongly suppressed. In contrast to the effects of 10-min conditioners, the suppression was long-lasting, with responses recovering only partially 4 h after the end of stimulation. However, remarkably, the response to the 8-dB probe was still not affected by the 1-h DPT.

These results suggest that high-rate electric stimulation lasting an hour or longer may cause long-term shifts in neural excitability. On the other hand, two or three applications of high-frequency pulse trains of 10 min or less separated by recovery periods of 10 min or more seem to cause only temporary shifts in neural thresholds.

IV. DISCUSSION

The goals of introducing a DPT are twofold. The primary goal is to improve the representation of the waveforms of electric stimuli in the temporal discharge patterns of auditory-nerve fibers. We postpone a discussion of stimulus coding until the companion papers (Litvak *et al.*, 2003a, b). The second goal is to elicit a pseudo-spontaneous response in the auditory nerve. Such pseudo-spontaneous activity may be of clinical significance, because lack of spontaneous activity has been suggested as a cause of tinnitus (Kiang *et al.*, 1970). Restoring a form of spontaneous activity on the auditory nerve may reduce or eliminate tinnitus in some patients.

A. Discharge rate distribution and adaptation

A major finding of this paper is that the distribution of discharge rates after 100 s of stimulation by a 5-kpps sustained DPT broadly resembles the distribution of spontaneous discharge rates in a healthy ear. This result is particularly striking given the small dynamic ranges and large differences in thresholds reported for the ANF responses to brief electric stimuli (van den Honert and Stypulkowski, 1984; Litvak *et al.*, 2001). This result is testimony to the profound effect of adaptation on responses to a sustained DPT.

To produce a sustained discharge rate in a fiber over the course of 10 min, the DPT needs to be at least 2 dB above that fiber's threshold. On average, the sustained discharge rate is larger for fibers with low thresholds relative to DPT level. However, fibers with thresholds at least 2 dB below the DPT level show large variability in discharge rates even if they have similar thresholds.

The result that threshold poorly predicts the adapted discharge rate to the DPT is consistent with earlier studies of responses to short, high-frequency electric stimuli. These studies have found that units differ widely in the degree of adaptation to supra-threshold stimuli (Moxon, 1967; van den Honert and Stypulkowski, 1987; Dynes and Delgutte, 1992; Killian, 1994; Litvak *et al.*, 2001). This is true even if the stimulation level is set at a fixed value relative to the threshold of individual fibers (Litvak *et al.*, 2001).

Fiber-to-fiber variability in both adaptation and sustained discharge rate may be related to a difference in the ability of fibers to maintain homeostasis in the face of increasing metabolic demands. One factor that may account for cross-fiber variability in metabolic resources is fiber diameter. Larger fibers may have more metabolic resources per membrane area, and therefore adapt slower to a taxing stimulus such as the DPT. Consistent with this hypothesis, when differences in threshold are controlled for, vestibular fibers, which are larger, adapt more slowly than do auditory fibers. Peripherally, ANF diameters differ by a factor of 2 (Liberman and Oliver, 1984). Interestingly, in a healthy ear, fibers with low spontaneous discharge rates have smaller diameters than high spontaneous fibers. Because smaller fiber diameter may be correlated with both faster adaptation and higher thresholds to electrical stimulation (Rubinstein, 1991), it is possible that many fibers with low spontaneous discharge rates would only respond transiently to a DPT.

B. Temporal discharge patterns

Using a biophysical model of the neural membrane, Rubinstein *et al.* (1999) predicted that responses to a 5-kpps pulse train would exhibit exponential interval histograms, resembling those of spontaneously active fibers in an intact ear. In our data, only about 25% of the interval histograms computed from DPT responses were strictly exponential (Type IA). Most of the records with exponential interval histograms had discharge rates below 50 spikes/s, and showed no serial dependence from one interval to the next, as revealed by the conditional mean test.

Interval histograms were generally nonexponential for fibers whose discharge rates exceeded 100 spikes/s. We distinguished between two types of nonexponential interval histograms. One (Type IB) was nonexponential in only the first 20 ms, but had an exponential tail. About half of responses associated with this type of histogram showed a serial dependence between successive intervals. Type II histograms were nonexponential in the tail as well as the initial segment. These late deviations in the interval histograms were highly correlated with serial dependence. Bursting was noted in some of these units. The interburst intervals were frequently regular (e.g., rightmost unit in Fig. 7, showing an interburst interval of 60 ms).

Both exponential and nonexponential histograms were previously noted by Litvak *et al.* (2001) in responses of ANFs to short ($<$500 ms) electric pulse trains at 5 kpps. These recordings were generally not long enough to distinguish the exponential tails of Type IB responses from the nonexponential tails of Type II responses. Litvak *et al.* also noted that responses to short, 1-kpps pulse trains were mostly exponential; it is unclear, however, whether exponen-

tial interval histograms would also be observed in responses to long-term stimulation at 1 kpps.

Mechanisms that generate the complex interval histograms are not well understood. Random sequences of long and short intervals have been observed in responses to dc stimulation in a squid giant axon (Guttman and Barnhill, 1970; Guttman *et al.*, 1980). These responses can be understood in terms of an interaction between the nonlinear dynamics of the Hodgkin-Huxley model and noise in membrane channels (Schneidman *et al.*, 1998). Although the high-frequency pulses in a DPT stimulus differ from the dc currents used in studies of the squid axon, some parallels may exist between the responses to the two stimuli. Specifically, rectification and low-pass filtering, both of which are known to occur in neural membranes, may transform the extracellular pulse train into an intracellular stimulus with a significant dc component. A model that correctly captures the dynamics of the ion channels in auditory-nerve fibers might accurately predict the Type IB responses of ANFs to high-frequency electric pulse trains.

The complex, Type II interval histograms may be related to processes other than those of ion channel dynamics. One example of such a process is the dynamic maintenance of sodium, potassium, and calcium concentrations in the intracellular and extracellular spaces. Slow oscillations in the concentration of these substances inside the neuron may lead to excitability changes on the scale of several tens of milliseconds, and produce bursting responses. Consistent with this hypothesis, neurons that exhibit complex histograms have higher discharge rates, which may cause greater load on homeostasis, and therefore greater fluctuations in ion concentrations. Intracellular recordings near the stimulating site during DPT stimulation may be required to establish a precise basis for the complex interval histograms.

C. Safe DPT stimulation

Although we only investigated the safety of DPT stimulation in two animals at a single stimulus level, our results do suggest that hour-long stimulation with a suprathreshold DPT may cause long-term threshold shifts in the most sensitive auditory-nerve fibers. This finding is consistent with earlier findings that sustained electric high-frequency stimulation cause long-term threshold shifts in auditory nerve responses (Tykocinski *et al.*, 1995b). However, the threshold shifts that we observed are distinct from those reported for dc stimulation (Tykocinski *et al.*, 1995a), because a dc blocking capacitor was used to filter out any residual dc component of the stimuli. While the threshold shift is not necessarily deleterious in itself, it may indicate permanent damage to the cell that may eventually lead to cell death. Alternatively, over a period of days, the cell metabolism might change and even reverse part of the original shift.

Our data suggests that 10-min stimulation followed by a rest period of 10 min or more might be safe. Therefore, short-term laboratory tests in human cochlear implant subjects to evaluate the effectiveness of the DPT concept may be appropriate. It is interesting that near-threshold, continuous high-frequency stimulation is already used routinely in some cochlear implant devices (Eddington, personal communication). In these devices, the carrier pulse train is at a subthreshold level even when there is no sound. This stimulation, while safe, is probably at a level well below that required to evoke sustained responses from many auditory-nerve fibers.

In the present study, a strongly supra-threshold DPT was presented on a single intracochlear electrode. The level of the DPT was chosen to stimulate a large fraction of the ANFs. Consequently, the DPT exceeded the threshold of the most sensitive fibers (presumably those that are near the stimulating electrode) by as much as 8 dB. Alternatively, a cochlear implant processor might simultaneously present a DPT on several intra-cochlear electrodes. Each electrode would stimulate only a local population of fibers. This approach may be safer because the DPT level could be lower relative to threshold for most fibers. Alternatively, a cochlear implant processor could temporarily evoke pseudo-spontaneous activity in auditory-nerve fibers by utilizing a DPT that is turned on and off as needed over the course of several minutes. More work is necessary to establish the safety and effectiveness of such fluctuating DPT stimulation.

D. Do responses to a DPT mimic spontaneous activity?

One of the goals of introducing a DPT is to evoke activity on the auditory nerve that mimics spontaneous activity in a healthy ear. This activity may improve the neural representation of electric stimuli superimposed upon the DPT, as well as alleviate some forms of tinnitus.

We found that a majority of the fibers respond in a sustained manner to a DPT presented at a level more than 2 dB above threshold. After 2 min of DPT stimulation, the electrically evoked activity in these fibers exhibited broadly distributed interval histograms, like spontaneous activity. These broad interval histograms suggest that responses are random, and that they may be uncorrelated from one fiber to the next. We directly tested this hypothesis by simultaneously recording from pairs of auditory-nerve fibers during stimulation with a DPT. We failed to detect significant correlations between the responses of any of the four pairs tested.

Despite these broad similarities, DPT-evoked responses differ from spontaneous responses in a healthy ear in several respects. First, a greater proportion of fibers have sustained discharge rates to the DPT below 5 spikes/s compared to the proportion of the fibers that have low spontaneous discharge rates in a healthy ear. In addition, a significant fraction of the fibers respond to the DPT with higher rates than reported for any spontaneously active fiber in a normal ear. Interval histograms also differ in shape from those of spontaneous activity for a majority of sustained responses to the DPT. While spontaneous responses generally have nearly exponential histograms (Kiang *et al.*, 1965), about 75% of the interval histograms for DPT responses were at least partly nonexponential.

While these differences between DPT-evoked responses and spontaneous activity are interesting from a modeling perspective, their functional significance, if any, is unclear. Studies in normal-hearing cats have identified a very small population of spontaneously responding fibers with nonexpo-

nential interval histograms that resemble the Type IB histograms in this study (Teich and Khanna, 1985; Gaumond, 1980). Interestingly, these fibers also tend to have high spontaneous rates (Miller and Wang, 1993). Nearly 50% of responses to the DPT have serial correlations between interspike interval that exceed the correlations reported for spontaneously active fibers by a factor of 10. Because, at least on a time scale of 40 ms, the DPT responses appear to be uncorrelated from fiber to fiber, these serial correlations may be of little functional significance for the response of central neurons if these neurons integrate activity across many ANF inputs.

While our experiments suggest that a high-level DPT may introduce pseudo-spontaneous activity on the auditory nerve, psychophysical experiments will be necessary to determine whether implanted subjects can learn to interpret DPT responses as spontaneous activity (i.e., silence). Recent experiments suggest that this is indeed the case (Rubinstein, personal communication). Initial experiments with DPT stimulation also show promise as a tinnitus treatment (Rubinstein *et al.*, 2003).

In summary, stimulation of the auditory nerve with a high-rate pulse train of long duration appears to reproduce many of the essential features of spontaneous activity in about half of the ANFs. In the companion papers (Litvak *et al.*, 2003a, b), we examine whether this pseudo-spontaneous activity improves the representation of electric stimuli superimposed upon a DPT in the temporal discharge patterns of auditory-nerve fibers.

ACKNOWLEDGMENTS

The authors would like to acknowledge the tireless efforts of Leslie Liberman without whose surgical skills this work would be impossible, and the assistance of K. M. Brinsko for figure preparation. This work was supported by the W. M. Keck Foundation and NIH Grant Nos. DC00361, DC02258, and N01-DC-6-2100.

[1]Because fibers with very low discharge rates (<50 spikes/s) yielded too few intervals shorter than the 20-ms cutoff, the cutoff was set to the mean interval for these units.

Cox, D. R. (**1967**). *Renewal Theory* (Methuen, London).

Duckert, L. G., and Miller, J. M. (**1982**). ''Acute morphological changes in guinea pig cochlea following electrical stimulation. A preliminary scanning electron microscope study,'' Ann. Otol. Rhinol. Laryngol. **91**, 33–40.

Dynes, S. B. C., and Delgutte, B. (**1992**). ''Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea,'' Hear. Res. **58**, 79–90.

Efron, B., and Tibshirani, R. (**1993**). *An introduction to the bootstrap* (Chapman and Hall, New York).

Gaumond, R. P. (**1980**). ''Studies of the stimulus and recovery dependence of cat cochlear nerve fiber discharge probability,'' Ph.D. thesis, Washington University, St. Louis, MO.

Good, P. I. (**2000**). *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses* (Springer-Verlag, New York).

Guttman, R., and Barnhill, J. (**1970**). ''Oscillation and repetitive firing in squid axons. Comparison of experiments with computations,'' J. Gen. Physiol. **55**, 104–118.

Guttman, R., Lewis, S., and Rinzel, J. (**1980**). ''Control of repetitive firing in squid axon membrane as a model for a neuroneoscillator,'' J. Physiol. (London) **305**, 377–395.

Hartmann, R., Topp, G., and Klinke, R. (**1984**). ''Discharge patterns of cat primary auditory fibers with electrical stimulation of the cochlea,'' Hear. Res. **13**, 47–62.

Javel, E. (**1990**). ''Acoustic and Electrical Encoding of Temporal Information,'' in *Cochlear Implants. Models of the Electrically Stimulated Ear*, edited by J. M. Miller and F. A. Spelman (Springer-Verlag, New York), pp. 247–295.

Javel, E., and Shepherd, R. K. (**2000**). ''Electrical stimulation of the auditory nerve. III. Response initiation sites and temporal fine structure,'' Hear. Res. **140**, 45–76.

Javel, E., Tong, Y. C., Shepherd, R. K., and Clark, G. M. (**1987**). ''Responses of cat auditory nerve fibers to biphasic electrical current pulses,'' Ann. Otol. Rhinol. Laryngol. Suppl. **128**, 26–30.

Johnson, D. H. (**1996**). ''Point process models of single-neuron discharges,'' J. Comput. Neurosci. **3**, 275–299.

Johnson, D. H., and Kiang, N. Y. S. (**1976**). ''Analysis of discharges recorded simultaneously from pairs of auditory nerve fibers,'' Biophys. J. **16**, 719–734.

Kelly, O. E., Johnson, D. H., Delgutte, B., and Cariani, P. (**1996**). ''Fractal noise strength in auditory-nerve fiber recordings,'' J. Acoust. Soc. Am. **99**, 2210–2220.

Kiang, N. Y. S., Moxon, E., and Levine, R. (**1970**). ''Auditory-nerve activity in cats with normal and abnormal cochleas,'' in *Sensorineural Hearing Loss*, CIBA Found. Symp., pp. 241–273.

Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, T. F. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT, Cambridge, MA).

Killian, M. J. P. (**1994**). ''Excitability of the Electrically Stimulated Auditory Nerve,'' Ph.D. thesis, University of Utrecht.

Knauth, M., Hartmann, R., and Klinke, R. (**1994**). ''Discharge pattern in the auditory nerve evoked by vowel stimuli: a comparison between acoustical and electrical stimulation,'' Hear. Res. **74**, 247–258.

Liberman, M. C. (**1978**). ''Auditory-nerve response from cats raised in a low-noise chamber,'' J. Acoust. Soc. Am. **63**, 442–455.

Liberman, M. C. (**1990**). ''Effects of chronic cochlear de-efferentation on auditory-nerve response,'' Hear. Res. **49**, 209–224.

Liberman, M. C., and Kiang, N. Y. S. (**1978**). ''Acoustic trauma in cats. Cochlear pathology and auditory-nerve activity,'' Acta Oto-Laryngol., Suppl. **358**, 1–63.

Liberman, M. C., and Oliver, M. E. (**1984**). ''Morphometry of intracellularly labeled neurons of the auditory nerve: Correlations with functional properties,'' J. Comp. Neurol. **223**, 163–176.

Litvak, L. M., Delgutte, B., and Eddington, D. K. (**2001**). ''Auditory nerve fiber responses to electric stimulation: modulated and unmodulated pulse trains,'' J. Acoust. Soc. Am. **110**, 368–379.

Litvak, L. M., Delgutte, B., and Eddington, D. K. (**2003a**). ''Improved temporal coding of sinusoids in electric stimulation of the auditory nerve using desynchronizing pulse trains,'' J. Acoust. Soc. Am. **114**, 2079–2098.

Litvak, L. M., Delgutte, B., and Eddington, D. K. (**2003b**). ''Improved neural representation of vowels in electric stimulation using desynchronizing pulse trains,'' J. Acoust. Soc. Am. **114**, 2099–2111.

Lowen, S. B., and Teich, M. C. (**1992**). ''Auditory-nerve action potentials form a non renewal point process over short as well as long time scales,'' J. Acoust. Soc. Am. **92**, 803–806.

Miller, C. A., Abbas, P. J., Robinson, B. K., Rubinstein, J. T., and Matsuoka, A. J. (**1999**). ''Electrically evoked single-fiber action potentials from cat: responses to monopolar, monophasic stimulation,'' Hear. Res. **130**, 197–218.

Miller, M. I., and Wang, J. (**1993**). ''A new stochastic model for auditory-nerve discharge,'' J. Acoust. Soc. Am. **94**, 2093–2107.

Moxon, E. C. (**1967**). ''Electric stimulation of the cat's cochlea: a study of discharge rates in single auditory nerve fibers,'' M.S. thesis. MIT, Cambridge, MA.

Parkins, C. W. (**1989**). ''Temporal response patterns of auditory nerve fibers to electrical stimulation in deafened squirrel monkeys,'' Hear. Res. **41**, 137–169.

Perkel, D. H., Gerstein, G. L., and Moore, G. P. (**1967**). ''Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains,'' Biophys. J. **7**, 419–440.

Rodieck, R., Kiang, N. Y. S., and Gerstein, G. L. (**1962**). ''Some quantitative methods for the study of spontaneous activity of single neurons,'' Biophys. J. **2**, 351–368.

Rosen, S. (**1992**). ''Temporal information in speech: acoustic, auditory and

linguistic aspects,'' Philos. Trans. R. Soc. Lond. B Biol. Sci. **336**, 367–373.
Rubinstein, J. T. (**1991**). ''Analytical theory for extracellular electrical stimulation of nerve with focal electrodes. II. Passive myelinated axon,'' Biophys. J. **60**, 538–555.
Rubinstein, J. T., Tyler, R. S., Johnson, A., and Brown, C. J. (**2003**). ''Electrical suppression of tinnitus with high-rate pulse trains,'' Otol. Neurotol. **24**, 478–485.
Rubinstein, J. T., Wilson, B. S., Finley, C. C., and Abbas, P. J. (**1999**). ''Pseudospontaneous activity: stochastic independence of auditory nerve fibers with electrical stimulation,'' Hear. Res. **127**, 108–118.
Schneidman, E., Freedman, B., and Segev, I. (**1998**). ''Ion channel stochasticity may be critical in determining the reliability and precision of spike timing,'' Neural Comput. **10**, 1679–1703.
Shepherd, R. K. (**1986**). ''Cochlear Prosthesis: Safety Investigations,'' Ph.D. thesis, University of Melbourne.
Shepherd, R. K., and Javel, E. (**1997**). ''Electrical stimulation of the auditory nerve. I. Correlation of physiological responses with cochlear status,'' Hear. Res. **108**, 112–144.
Taylor, M. M., and Creelman, C. D. (**1967**). ''PEST: Efficient estimates on probability functions,'' J. Acoust. Soc. Am. **41**, 782–787.
Teich, M. C., and Khanna, S. M. (**1985**). ''Pulse-number distribution for the neural spike train in the cat's auditory nerve,'' J. Acoust. Soc. Am. **77**, 1110–1128.
Tykocinski, M., Shepherd, R. K., and Clark, G. M. (**1995a**). ''Electrophysiologic effects following acute intracochlear direct current stimulation of the guinea pig cochlea,'' Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 68–71.
Tykocinski, M., Shepherd, R. K., and Clark, G. M. (**1995b**). ''Reduction in excitability of the auditory nerve following electrical stimulation at high stimulus rates,'' Hear. Res. **88**, 124–142.
Tykocinski, M., Shepherd, R. K., and Clark, G. M. (**1997**). ''Reduction in excitability of the auditory nerve following electrical stimulation at high stimulus rates. II. Comparison of fixed amplitude with amplitude modulated stimuli,'' Hear. Res. **112**, 147–157.
van den Honert, C., and Stypulkowski, P. H. (**1984**). ''Physiological properties of the electrically stimulated auditory nerve. II. Single fiber recordings,'' Hear. Res. **14**, 225–243.
van den Honert, C., and Stypulkowski, P. H. (**1987**). ''Temporal response patterns of single auditory nerve fibers elicited by periodic electrical stimuli,'' Hear. Res. **29**, 207–222.
Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P. (**1987**). ''Speech waveform envelope cues for consonant recognition,'' J. Acoust. Soc. Am. **82**, 1152–1161.
Van Tasell, D. J., Greenfield, D. G., Logemann, J. J., and Nelson, D. A. (**1992**). ''Temporal cues for consonant recognition: Training, talker generalization, and use in evaluation of cochlear implants,'' J. Acoust. Soc. Am. **92**, 1247–1257.
Walsh, B. T., Miller, J. B., Gacek, R. R., and Kiang, N. Y. S. (**1972**). ''Spontaneous activity in the eighth cranial nerve of the cat,'' Int. J. Neurosci. **3**, 221–236.
Wiesenfeld, K., and Moss, F. (**1995**). ''Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDs,'' Nature (London) **373**, 33–36.
Wilson, B., Finley, C. C., Lawson, D. T., and Zebri, M. (**1997**). ''Temporal representations with cochlear implants,'' Am. J. Otol. **18**, Suppl. 6, S30–S34.
Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). ''Better speech recognition with cochlear implants,'' Nature (London) **352**, 236–238.
Xu, S. A., Shepherd, R. K., Chen, Y., and Clark, G. M. (**1993**). ''Profound hearing loss in the cat following the single co-administration of kanamycin and ethacrynic acid,'' Hear. Res. **70**, 205–215.

Improved temporal coding of sinusoids in electric stimulation of the auditory nerve using desynchronizing pulse trains[a)]

Leonid M. Litvak[b)]
Eaton-Peabody Laboratory and Cochlear Implant Research Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Speech and Hearing Bioscience and Technology Program, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

Bertrand Delgutte
Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

Donald K. Eddington
Cochlear Implant Research Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Research Laboratory of Electronics and Neural Prosthesis Research Center, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

(Received 13 March 2003; accepted for publication 4 August 2003)

Rubinstein *et al.* [Hearing Res. **127**, 108–118 (1999)] suggested that the representation of electric stimulus waveforms in the temporal discharge patterns of auditory-nerve fiber (ANF) might be improved by introducing an ongoing, high-rate, desynchronizing pulse train (DPT). To test this hypothesis, activity of ANFs was studied in acutely deafened, anesthetized cats in response to 10-min-long, 5-kpps electric pulse trains that were sinusoidally modulated for 400 ms every second. Two classes of responses to sinusoidal modulations of the DPT were observed. Fibers that only responded transiently to the unmodulated DPT showed hyper synchronization and narrow dynamic ranges to sinusoidal modulators, much as responses to electric sinusoids presented without a DPT. In contrast, fibers that exhibited sustained responses to the DPT were sensitive to modulation depths as low as 0.25% for a modulation frequency of 417 Hz. Over a 20-dB range of modulation depths, responses of these fibers resembled responses to tones in a healthy ear in both discharge rate and synchronization index. This range is much wider than the dynamic range typically found with electrical stimulation without a DPT, and comparable to the dynamic range for acoustic stimulation. These results suggest that a stimulation strategy that uses small signals superimposed upon a large DPT to encode sounds may evoke temporal discharge patterns in some ANFs that resemble responses to sound in a healthy ear.
[DOI: 10.1121/1.1612493]

PACS numbers: 43.64.Me, 43.64.Nf, 43.64.Pg [WPS]

I. INTRODUCTION

Speech and other sound stimuli contain information in both their slowly varying envelope and their rapidly varying fine-time structure (Rosen, 1992). Yet, many processing strategies used in today's cochlear implants only deliver envelope information and discard the temporal fine structure. For example, continuous interleaved sampling (CIS) strategies (Wilson *et al.*, 1991) use an envelope detector in each frequency channel to derive waveforms used to modulate carrier pulse trains. These envelope detectors discard the information available in the fine-time structure. The SPEAK strategy used with Nucleus implants (Seligman and McDermott, 1995) also discards the fine structure in each stimulated channel.

Smith *et al.* (2002) evaluated the relative importance of envelope and fine-time structure information in auditory perception using acoustic stimuli called ''auditory chimeras.'' The chimeras are sounds which, in each frequency band, have the fine-time structure of one sound, and the envelope of another. Smith *et al.* found that, with four to eight frequency bands, speech comprehension is better using the envelope information than the information in the fine-time structure. However, for the same number of analysis channels, subjects performed melody identification and sound localization tasks better with the fine-time structure than with the envelope information. These results suggest that modifying cochlear implant processing strategies to include fine-time structure information might improve pitch perception. The fine structure may also be essential for taking full advantage of binaural cues delivered by bilateral implants.

Fine structure information might be delivered to cochlear implants in several ways. The simplest method, already used in some processing strategies, is to deliver an

[a)]Portions of this work were presented as a poster at the ARO Midwinter Meeting in St. Petersburg Beach, FL, 2001 and the CIAP Conference in Monterey, CA, 2001.
[b)]Electronic mail: leonid.litvak@advancedbionics.com

analog representation of the stimulus waveform in each frequency channel. Alternatively, the frequency of a periodic stimulus might be encoded by the rate of a pulse train. Although this method has the disadvantage that signals having different waveforms cannot be distinguished if they have the same fundamental frequency, it was used in early Nucleus processors to encode the fundamental frequency of voice (Tong *et al.*, 1980). Finally, fine structure information might be introduced in a CIS strategy by increasing the cutoff frequency of the envelope detector, or even eliminating it altogether. The rate of the carrier pulse trains would also have to be increased so as to sample the resulting high-frequency modulation waveforms without aliasing.

Both single-unit studies in animals and evoked potential studies in implanted human subjects suggest that neither of the above strategies for delivering fine structure information would, *by itself*, produce temporal discharge patterns in the auditory nerve resembling those evoked by acoustic stimuli in a normal ear. Auditory-nerve fibers (ANF) encode the fine-time structure of acoustic stimuli in their temporal discharge patterns for frequencies up to 5 kHz (Rose *et al.*, 1967; Johnson, 1980). For example, in response to a pure tone, neurons fire at random multiples of the stimulus period, in that there may be one, two or more cycles between successive spikes (Rose *et al.*, 1967). The stimulus period is thus represented in the ensemble activity of a population of neurons by a stochastic form of Wever's volley principle, even when the stimulus period is shorter than the neural refractory period. This coding scheme is made possible by the ongoing stochastic release of neurotransmitter at inner-hair-cell synapses, and the modulation of this neurotransmitter release by the receptor potential which tracks each cycle of the stimulus waveform.

Because hair cells and their synapses are missing in deaf ears, temporal discharge patterns produced by electric stimulation are very different from normal acoustic responses. Suprathreshold responses to electric pulse trains or sinusoids with frequencies below 500–800 Hz behave nearly deterministically, typically showing one spike discharge on every stimulus cycle (Moxon, 1967; Hartmann *et al.*, 1984, 1990; van den Honert and Stypulkowski, 1987; Parkins, 1989; Javel *et al.*, 1987; Javel, 1990; Javel and Shepherd, 2000). Such entrainment also occurs for sinusoidally modulated electric pulse trains similar to stimuli produced by CIS processors (Litvak *et al.*, 2001). Moreover, the temporal precision of discharges evoked by such low-frequency electric stimuli is much higher than with acoustic stimulation (Hartmann *et al.*, 1984, 1990; Javel, 1990; Javel and Shepherd, 2000). For example, whereas spikes are distributed over most of one-half cycle in response to a pure tone, they only occupy a small fraction of the stimulus cycle for electric sinusoids, thereby inadequately representing the sinusoidal wave shape (Hartmann *et al.*, 1984; van den Honert and Stypulkowski, 1987). In fact, temporal discharge patterns for electric sinusoidal, triangular and square waves of the same frequency are surprisingly similar considering that these stimuli have very different spectra (van den Honert and Stypulkowski, 1987; Parkins, 1989).

Further issues arise at higher frequencies (>500–800 Hz) where neural refractoriness prevents fibers from discharging on every cycle of the electric stimulus. With both sinusoidal and pulse train stimuli, refractoriness can cause history-dependent changes in response latency, resulting in a neuron discharging at two distinct phases within a stimulus cycle (Parkins, 1989; Javel, 1990; Javel and Shepherd, 2000). Such double-peaked period histograms are not observed for acoustic stimulation above 1000 Hz (Johnson, 1980). Interval histograms are also highly abnormal, demonstrating a tendency for spikes to occur at regular intervals, even for high-frequency (>1000 Hz) stimulation (Parkins, 1989; Dynes and Delgutte, 1992). In some cases, ANFs fire exactly on every other cycle or even at higher multiples of the stimulus period (Javel, 1990; Javel and Shepherd, 2000). This tendency is also apparent in measurements of electric compound action potentials (ECAPs) from cochlear implant users (Wilson *et al.*, 1997). Specifically, in response to a 1-kpps electric pulse train, ECAPs alternate between strong and weak responses to each pulse for up to 100–200 ms after stimulus onset. This alternation suggests that, while most ANFs respond to the first pulse, they are in a refractory state during the second pulse and only respond again to the third pulse. If most fibers fire together on every other cycle, then the auditory nerve population represents half the stimulus frequency rather than the actual frequency. In summary, neural refractoriness, the narrow dynamic ranges, excessively high discharge rates, and exaggerated synchrony combine to make the temporal coding of electric stimuli highly unnatural compared to normal acoustic stimulation, particularly for frequencies above 500 Hz.

Rubinstein *et al.* (1999b) proposed that the temporal coding of stimulus waveforms in cochlear implants might be improved by introducing an ongoing, high-frequency, desynchronizing pulse train (DPT) in addition to the signal produced by the speech processor. The purpose of a DPT is to amplify noise in sodium channels so as to produce more stochastic responses similar to those of spontaneously active fibers. In a companion paper (Litvak *et al.*, 2003b), we recorded ANF responses to a 10-min, 5-kpps DPT. We found that, after 1–2 min of continuous stimulation, the DPT produced activity that, in many fibers, resembled spontaneous activity in a healthy ear. Several types of responses to the DPT were identified. Some fibers (roughly 50%) only responded transiently to the DPT, while the others showed a sustained response throughout 10 min of DPT stimulation. For the sustained responders, the ''pseudo-spontaneous'' activity evoked by the DPT had broadly distributed interspike interval distributions and appeared to be uncorrelated from fiber to fiber. Some interval histograms (25%) had an exponential shape, as does normal spontaneous activity; however, the majority of sustained responses had nonexponential interval histograms, particularly if they discharged at very high rates.

In this paper, we directly test the hypothesis that a DPT improves the representation of the fine-time structure in the temporal discharge patterns of ANFs for sinusoidal stimuli. There are at least three ways in which a DPT could make responses to electric stimuli better resemble acoustic responses in a healthy ear. First, a DPT could desynchronize

stimulus-evoked activity across-fibers. For higher-frequency (>500 Hz) stimuli, desynchronization may allow different neurons to discharge on different stimulus cycles, thereby allowing volley coding of frequency. By simultaneously recording from pairs of fibers, we showed that such desynchronization does in fact occur after a few seconds of DPT stimulation (Litvak *et al.*, 2003b).

Second, a DPT may allow small electric signals to be encoded as modulation of ongoing DPT-evoked pseudospontaneous activity, much as normal acoustic responses of most ANFs are effectively modulations of ongoing spontaneous activity. Such modulations of random activity allow faithful transmission of the stimulus waveform in neural discharges by a process akin to stochastic resonance (Yu and Lewis, 1989; Collins *et al.*, 1995; Wiesenfeld and Moss, 1995). If the stochastic nature of the responses could be restored by the DPT, then a similar mechanism may also improve the coding of the stimulus fine-time structure in electric stimulation.

Finally, computer simulations of auditory-nerve fibers suggest that a DPT may lower the threshold and increase the dynamic range to electric stimulation (Rubinstein *et al.*, 1999a). The increased dynamic range is particularly significant when one considers complex stimuli such as vowels. These waveforms contain peaks of widely different heights in each period. A wide dynamic range might allow fibers to represent all of the waveform peaks in their temporal discharge patterns rather than just the largest peak.

Previous tests of the ideas underlying the DPT (Rubinstein *et al.*, 1999b; Wilson *et al.*, 1998) were based on electric compound auditory potential (ECAP) responses, which provide only indirect evidence of single-unit activity. In addition, these studies used short pulse trains (30 ms), which are a poor model of an ongoing DPT. We found that responses to a DPT only reach a near steady state after 1–2 min of DPT stimulation (Litvak *et al.*, 2003b). In this paper, we directly test the hypotheses underlying the DPT by recording from single fibers from the auditory nerve of deafened cats. We focus on responses that occurred after adaptation to a DPT presented for 10 min. We compare temporal discharge patterns of ANFs for sinusoidal modulations of a DPT with normal acoustic responses to pure tones. We also test the hypothesis that the DPT increases the dynamic range by studying responses to sinusoidally modulated DPTs over a range of modulation depths.

Our scheme for encoding acoustic signals into electric waveforms differs somewhat from that used by Rubinstein *et al.* (1999b) in their original formulation of the DPT idea. In that paper, an analog electric sinusoid was directly superimposed upon a DPT; here we encode the sinusoid as a small modulation of a DPT. This coding scheme assumes that neural responses to a high-rate pulse train with low modulation depth are similar to those elicited by the superposition of a large, unmodulated DPT and a small, highly modulated pulse train as might be produced by a CIS processor (Fig. 1). This assumption is a mathematical identity if the same pulse train is used for both the DPT and the CIS carrier. It may hold more generally if the time constant of the neural membrane is large compared to the intervals between pulses. In this scheme, modulation depth is proportional to the amplitude of the sinusoidal stimulus for small modulations. The differences between the present scheme and the Rubinstein *et al.* (1999b) original strategy will be further considered in Sec. IV.

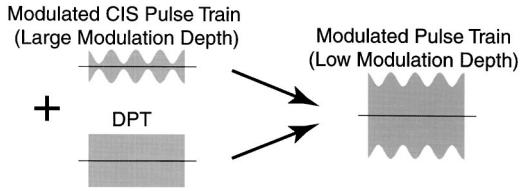


FIG. 1. The stimuli used in our experiments are pulse trains with low modulation depths (right). These stimuli can be thought of as the superposition of a large, unmodulated DPT (bottom left) and a small, fully modulated pulse train (top left) similar to the signals produced by CIS processors.

II. METHODS

The animal preparation, electrical stimulation, and recording methods are described in the companion paper (Litvak *et al.*, 2003b). Briefly, cats were anesthetized with dial in urethane (75 mg/kg), then deafened by co-administration of kanamycin (subcutaneous, 300 mg/kg) and ethacrinic acid (intravenous, 25 mg/kg) (Xu *et al.*, 1993). As described in detail in the companion paper (Litvak *et al.*, 2003b), most of the animals had some residual hearing in the nonimplanted ear. To minimize the effect of residual hair cells in these preparations, we only report responses from neurons with no spontaneous activity in the absence of a DPT. Two intracochlear stimulating electrodes (400 mm Pt/Ir balls) were inserted into the cochlea through the round window. One electrode was inserted approximately 8 mm and was used as the stimulating electrode. The other electrode was inserted just inside the round window and served as the return electrode.

Standard techniques were used to expose the auditory nerve via a dorsal approach (Kiang *et al.*, 1965). We recorded from single units in the auditory nerve using glass micropipettes filled with 3M KCl. For small modulation depths, most of the stimulus artifact could be removed online using a digital signal processor implementing a moving average filter whose length matched the 0.2-ms pulse period. Neural responses were also recorded digitally with a sampling rate of 20 kHz for off-line analysis. Methods used to remove the stimulus artifact from these records are described in Appendix A.

A. Stimuli

We conducted a neural population study by investigating responses to a single DPT level for each animal. To ensure that a large fraction of fibers would respond to the DPT, the DPT level was set at 8–10 dB above ECAP threshold, as described in detail in the companion paper (Litvak *et al.*, 2003b).

We studied responses to small (modulation depth $\leqslant$15%) sinusoidal modulations of the DPT. Figure 3 (top) schematizes the envelope of the electric stimuli. The carrier was a 5-kpps train of biphasic pulses (cathodic/anodic, 25 μs per phase). In order to acquire responses to both the unmodu-

lated DPT and modulations of the DPT for several modulation depths and frequencies, the stimulus was composed of alternating modulated (400 ms) and unmodulated (600 ms) segments.[1] Modulation depth and, in some cases, modulation frequency was changed on each successive segment. Modulation depth was varied from 0.5% to 15%, while modulation frequency was either 104, 417 or 833 Hz. The entire sequence of modulated and unmodulated segments had a 5–12-s period, and was repeated continuously for 10 min or until contact with the fiber was lost.

Modulation was applied such that the mean amplitude of the carrier pulses was the same during modulated and unmodulated segments. Specifically, during modulation, the envelope was defined as $A \cdot (1 + m \cdot \sin(2\pi f_m t))$, where f_m is the modulation frequency, A is the amplitude of the unmodulated DPT, and m is the modulation depth. The modulation period was always an integer multiple of the 0.2-ms carrier period to avoid beating, and the modulation phase was chosen so that the peak of the modulation waveform always coincided with a carrier pulse.

B. Analysis

Responses collected during the unmodulated DPT segments were used to classify each fiber using the same scheme as in the companion paper (Litvak *et al.*, 2003b). Some fibers exhibited only a transient response to the DPT, and adapted to near zero (<5 spikes/s) discharge rate after 100 s of DPT stimulation. We will refer to these fibers as ''transient DPT responders.'' Fibers that responded to the unmodulated segments over the entire stimulus duration will be referred to as ''sustained responders.''

Litvak *et al.* (2003b) further used interval histograms to characterize the temporal discharge patterns of sustained responders. We found that, while some responses had nearly exponential interval histograms, others had strongly nonexponential interval histograms. The degree of ''exponentiality'' of the histogram was quantified using an Interval Histogram Exponential Shape Factor (IH-ExpSF) (Litvak *et al.*, 2001). The IH-ExpSF is computed by first fitting the interval histogram with both a single exponential and, piecewise, by three exponentials. The root mean squared (rms) error of each fit to the data is then determined, and the IH-ExpSF defined as the ratio of the rms error for the piecewise fit to that for the single exponential fit. The IH-ExpSF for samples from a stationary Poisson process is approximately 1.

C. Stochastic threshold model

As a concise way of summarizing the data, we developed a simple functional model of responses of ANFs to modulations of a DPT (Fig. 2). The model takes as input the modulation waveform $m(t)$. For a sinusoidal modulator, $m(t) = m \cdot \sin(2\pi f_m t)$, where f_m is modulation frequency, and m is modulation depth. A spike is produced by the model whenever $m(t)$ crosses a noisy threshold. The threshold is the sum of a deterministic term and a noise term. To account for the refractory properties of ANFs, the deterministic component of threshold $\theta_n(t)$ depends on the time since the preceding spike t_i according to Eq. (1), also used by Bruce *et al.* (1999):

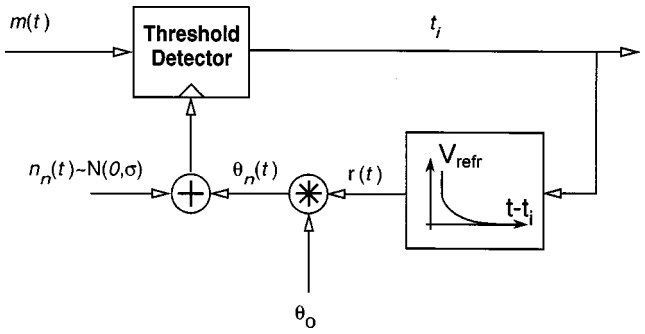


FIG. 2. Stochastic threshold model (STM) of ANF responses to modulations of a DPT. The model takes as input the modulation waveform $m(t)$ and produces a spike whenever the input crosses a noisy threshold. The output of the model is a set of spike times $\{t_i\}$. The threshold is the sum of a Gaussian noise term $n(t)$ and a deterministic term $\theta_n(t)$ which depends on the time since the previous spike. The only free parameters in the model are the resting threshold θ_0 and the noise amplitude σ.

$$\theta_n(t) = \theta_0 \cdot r(t) = \theta_0 \cdot \begin{cases} \infty, & t - t_i < 0.6 \text{ ms}, \\ 1 + 0.97 \exp[-(t - t_i)/1.32 \text{ ms}], & \\ t - t_i \geqslant 0.6 \text{ ms}. & \end{cases} \quad (1)$$

The threshold recovery function $r(t)$ was chosen to fit the absolute and relative refractory periods of electrically stimulated ANFs (Dynes, 1995). For the noise term, we used computer-generated zero-mean, white Gaussian noise with standard deviation σ. Because the model was simulated using 0.2-ms time steps, the noise bandwidth was effectively 2500 Hz. Because the noisy threshold is the critical element of the model, we refer to this model as the stochastic threshold model (STM).

We computed responses of the STM to sinusoidal modulators of different frequencies and modulation depths. The conditional probability that a model neuron fires at time t given that the last spike occurred at t_i is

$$p_f(t) = \Gamma\left(\frac{m}{\sigma} \cdot \sin(2\pi f_m t) - \frac{\theta_0}{\sigma} r(t - t_i)\right), \quad (2)$$

where Γ is the cumulative Gaussian distribution, and $r(t)$ is the recovery function described by Eq. (1). For an unmodulated DPT ($m = 0$), the pseudo-spontaneous discharge rate (the discharge rate during the unmodulated DPT) is entirely determined by the threshold-to-noise ratio θ_0/σ. For nonzero values of m, the discharge rate depends on the ratio m/σ as well. These two ratios entirely determine the model response to sinusoids.

III. RESULTS

A. Basic response characteristics

Our results are based on 68 responses recorded from 62 auditory-nerve fibers in five cats to sinusoidally modulated electric pulse trains. Each record included responses to a

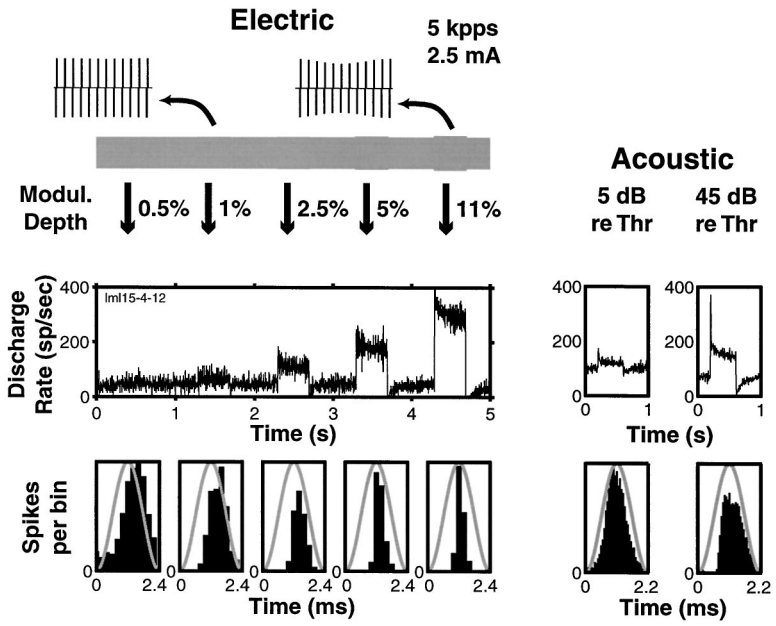


FIG. 3. The top panel shows one cycle of the electric pulse train stimulus (5 kpps, 2.5 mA 0-p) to which a 417-Hz sinusoidal modulation was applied every second for 400 ms. Modulation depth ranged from 0.5% to 11%, and the entire sequence of modulations was repeated every 5 s for 10 min. The left panel of the middle row shows a period histogram (bin width 2.4 ms) locked to the 5-s modulation cycle for one auditory-nerve fiber. For comparison, the right two panels of the middle row show the response pattern of an ANF (CF=650 Hz) from a healthy ear to a 440-Hz pure tone at 5 and 45 dB above threshold. The bottom panels show the period histograms computed from responses during the electric modulations (left) and the acoustic pure tone (right). The gray line shows a sinusoidal waveform for comparison.

series of modulated segments with different modulation depths and, in some cases, different modulation frequencies. For 28 of these records, the fiber responded to the unmodulated DPT segments at a discharge rate above 5 spikes/s throughout the stimulus duration. For the other 40 records, the fiber stopped responding to the unmodulated DPT after 1–2 min of stimulation, although it still responded to some of the modulated segments. The percentage of transient responders in the present data sample is somewhat larger (59% vs 46%) than in the larger sample from ten cats described in the companion paper (Litvak *et al.*, 2003b), and the difference is statistically significant ($p=0.0247$, binomial exact test). We attribute this difference to a combination of biases in data collection (the silent responses of transient responders to unmodulated DPTs were sometimes discarded) and individual differences among animals (some animals had a high proportion of transient responses). In the following, we describe only the responses that occurred after 50 s of stimulation so that the data would reflect the effects of adaptation to the DPT.

Figure 3 shows the response of an auditory-nerve fiber to a DPT that was sinusoidally modulated at 417 Hz for 400 ms every second. Modulation depth was increased from 0.5% to 11% on each successive modulated segment. The entire 5-s cycle of five modulation depths was repeated for 10 min. During the unmodulated segments, the pseudospontaneous discharge rate was 48 spikes/s, and the interspike interval distribution was nearly exponential (IH $-$ ExpSF=0.99).

The left panel in the middle row shows the discharge rate as a function of time from the onset of the 5-s stimulus cycle. Average discharge rate grows monotonically with increasing modulation depth. For modulation depths between 0.5% and 5%, the discharge rate stays below 300 spikes/s, in the range appropriate for responses to tones in a healthy ear (Kiang *et al.*, 1965; Liberman, 1978). At the largest modulation depth (11%), the discharge rate exceeds the rates seen in Liberman's data.

The right two panels in the middle row show the response patterns of a high spontaneous-rate fiber in a healthy ear to 440-Hz tone bursts at 5 and 45 dB above threshold (McKinney and Delgutte, 1999). Consistent with classic descriptions (Westerman and Smith, 1984), the discharge rate decreases rapidly during the first 10 ms of tone-burst stimulation, and this rapid adaptation is followed by slower adaptation with a time constant near 100 ms. In contrast, the responses to modulations of the DPT (left) show a form of slow adaptation, but no sign of rapid adaptation.

The bottom left panel in Fig. 3 shows period histograms locked to the modulator frequency computed from responses during the modulated DPT segments. The period histogram for a modulation depth of 0.5% is already almost fully modulated, indicating exquisite sensitivity to modulation. At this modulation depth, there is little or no increase in discharge rate over that evoked by the unmodulated DPT. Thus, synchrony to the stimulus is already large at a modulation depth that evokes no noticeable increase in average rate over the response to the unmodulated DPT. Similar behavior is found in responses of high-spontaneous ANFs to pure tones (Johnson, 1980).

As the modulation depth increases, responses become more precisely phase locked. For a modulation depth of 1%, the period histogram approaches a half-wave rectified sinusoid, suggesting that the modulator waveform is accurately represented in the temporal discharge patterns. For this modulation depth, the period histogram resembles those in response to a pure tone in a healthy ear (Fig. 3, bottom right). For modulation depths above 2.5%, however, the period histogram consists of a very sharp mode restricted to a small fraction of the stimulus cycle, and does not resemble the response to a pure tone at any level. These hypersynchronized responses resemble responses to sinusoidal electric stimulation without a DPT (Hartmann *et al.*, 1984; van den Honert and Stypulkowski, 1987).

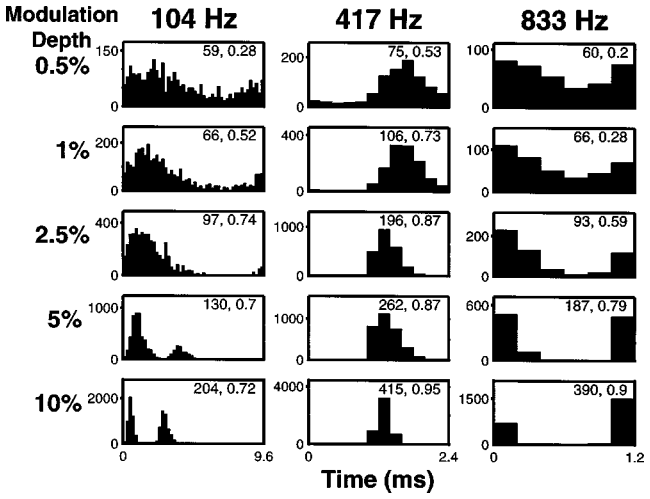


FIG. 4. Period histograms (bin width 0.2 ms) computed from the response of an auditory-nerve fiber to sinusoidally modulated pulse trains at three different frequencies (columns) and five modulation depths (rows). During the unmodulated segments of the DPT, this fiber had a pseudo-spontaneous discharge rate of 54 spikes/s, and its interval histogram had a nearly exponential shape (IH−ExpSF=0.97). Numbers in each panel are the average discharge rate (in spikes/s) and the synchronization index to the modulation. The vertical axis is in spikes/s.

B. Temporal discharge patterns for exponential fibers

Figure 4 shows period histograms computed from responses to 104, 417, and 833 Hz sinusoidal modulations of the DPT for another ANF. Modulation depth was varied from 0.5% to 10%. During the unmodulated segments, this fiber had a pseudo-spontaneous discharge rate of 54 spikes/s and an exponential interspike interval distribution.

For all modulation frequencies, stimuli with modulation depths below 5% evoked average discharge rates below 300 spikes/s, in the range reported for ANFs in normal ears for pure-tone stimulation (Liberman, 1978). In addition, the synchronization index was always below 0.9, consistent with acoustic responses to pure tones. The nearly sinusoidal shape of the period histograms for modulation depths below 2.5% suggests that the modulation waveform is accurately represented for all three modulation frequencies.

At 10% modulation depth, the discharge rates in responses to 417- and 833-Hz modulators exceed those seen in a normal ear. At this high modulation depth, the period histogram for the 104-Hz modulator reveals multiple peaks for each cycle, and this is seen for 5% modulation as well. Unlike the ''peak splitting'' observed in responses of ANFs to low-frequency tones (Johnson, 1980; Kiang and Moxon, 1972) which always occurs on opposite phases, both peaks of the responses to the electric stimulus occur during the same half-cycle of the modulator. Similar double-peaked responses have been reported for sinusoidal electric stimulation without a DPT (Hartmann *et al.*, 1984; van den Honert and Stypulkowski, 1987).

Interspike interval histograms were also examined to determine whether they resemble those evoked by pure tones in a healthy ear. Figure 5 shows the interval histograms that correspond to the period histograms in Fig. 4. Interval histograms for a 440-Hz pure tone in a normal ear are also shown for comparison. Phase locking to the stimulus can be seen in these histograms as the clustering of intervals around integer multiples of the stimulus period (dashed lines).

For modulation depths below 5%, interval histograms for the modulated DPT resemble acoustic responses in that they have exponential envelopes, and show several modes at multiples of the modulation period. One exception is the response to the 417-Hz modulator at 2.5% modulation depth, where the sharp interval mode at twice the stimulus period means that spikes occur on every other stimulus cycle. However, this response pattern is unusual for this modulation frequency and depth. Detailed examination of the interval histograms for modulation depths below 2.5% reveals an additional difference between DPT responses and normal responses to pure tones. Interval histograms for pure tone stimuli always show a mode at the tone's period for frequencies below 1000 Hz (Rose *et al.*, 1967). In contrast, responses to the 417-Hz modulated electric pulse train show no mode at the modulation period; instead, the earliest mode occurs at twice the period. Such lack of a mode at the period has also been observed in response to both electric sinusoids and sinusoidally-modulated pulse trains presented without a DPT (van den Honert and Stypulkowski, 1987; Litvak *et al.*, 2001).

For 10% modulation depth, the interval histograms of

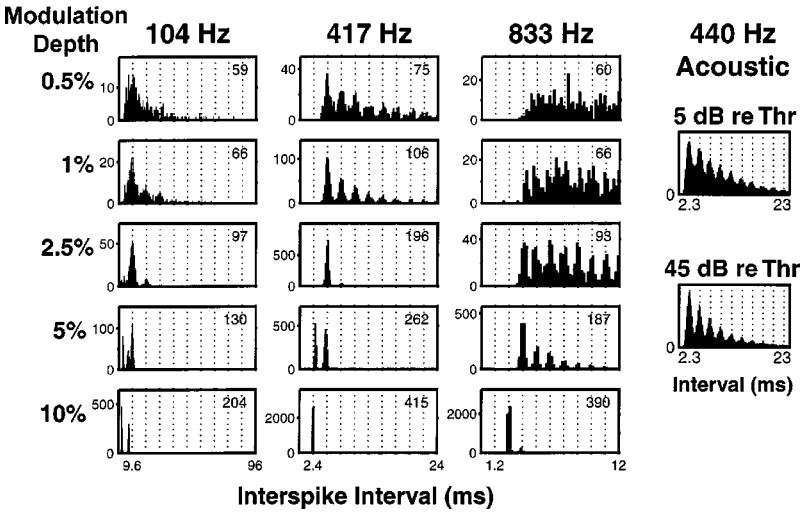


FIG. 5. The first three columns show the interval histograms corresponding to the period histograms in Fig. 4. The numbers in each panel are the average discharge rates in spikes/s. Vertical dashed lines mark multiples of the stimulus period. The vertical axis represents the number of intervals per bin. For comparison, the rightmost column shows interval histograms of an ANF (CF=650 Hz) from a healthy ear for a 440-Hz pure tone at 5 and 45 dB above threshold.

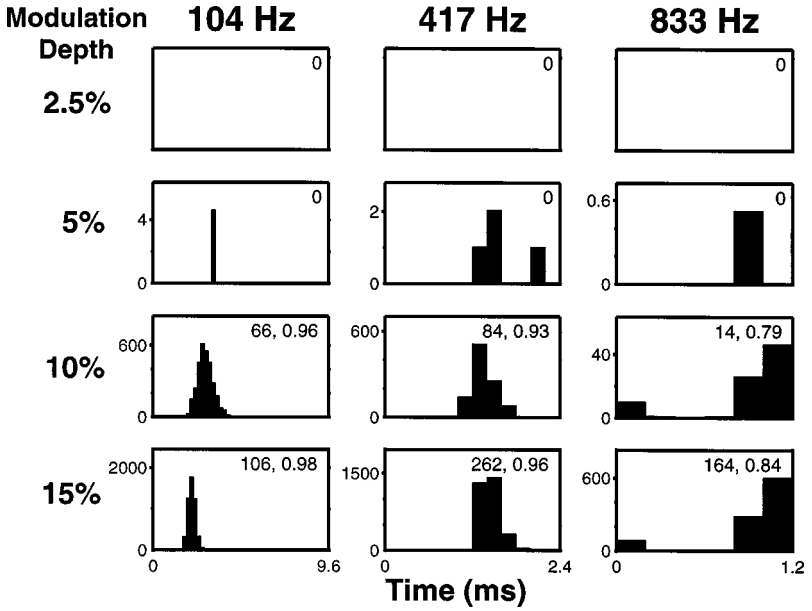


FIG. 6. Period histograms of responses of a fiber that gave a transient response to the unmodulated DPT for three modulation frequencies and four modulation depths. Same format as in Fig. 4.

responses to electric stimulation differ strongly from normal responses to tones. In response to the 104-Hz modulator, the fiber fired twice on each modulation cycle, so that the interval histogram had just two modes, one at 2.3 ms and the other one at 7.3 ms. The sum of these two mode locations equals the 9.6-ms modulation period. For 417-Hz modulation, the interval histogram had a single mode at the modulation period, meaning that a spike occurred exactly once per modulation cycle. Such entrainment is not seen in normal ANF responses to pure tones (Rose *et al.*, 1967), but is common in electric responses without a DPT. For the 833-Hz modulator, the histogram showed a dominant mode at twice the modulation period, indicating that a spike occurred on nearly every other cycle. This distortion is particularly disturbing, because, if it occurs in many fibers, and these fibers fire in synchrony, then the response of the ANF population would represent a signal at half the actual modulation frequency. A tendency for spikes to occur at integer multiples of the stimulus period for frequencies above 500 Hz has also been observed for pulsatile electric stimulation without a DPT (Javel, 1990; Javel and Shepherd, 2000).

Responses of transient responders, a representative example of which is shown in Fig. 6, contrast sharply with those of sustained responders illustrated in Figs. 3–5. This fiber did respond to modulations of the DPT, even though, after adaptation, it showed no spike discharges to the unmodulated DPT (zero pseudo-spontaneous rate). However, the modulation depth necessary to evoke a significant response exceeded 5%, larger than that for any of the sustained responders. In addition, for the 104- and 417-Hz modulators, the synchronization index was invariably higher than that for pure tone responses in a healthy ear. Regardless of modulation depth, the period histogram poorly represented the sinusoidal modulation waveform. Overall, these responses resemble responses to electric sinusoids presented without a DPT (Hartmann *et al.*, 1984; van den Honert and Stypulkowski, 1987).

C. Temporal discharge patterns for nonexponential fibers

For the most part, fibers with nonexponential interval histograms during the unmodulated DPT (Litvak *et al.*, 2001, 2003b) responded to modulations of the DPT in a manner similar to the exponential, sustained responders illustrated in Fig. 3–5, but they also showed some unique features, particularly for very small modulations. For example, the top left panel of Fig. 7 shows the interspike interval computed from responses of a fiber to an unmodulated DPT. The interval histogram is clearly nonexponential and shows a pronounced mode at 3.3 ms. We refer to the location of this mode as the *preferred interval*. We found that preferred intervals systematically influenced the temporal discharge patterns for small modulations of a DPT. The middle panels of Fig. 3 shows the interval (left) and period (right) histogram for responses to 417-Hz modulation of the DPT at 0.5% modulation depth. The period histogram shows that the re-

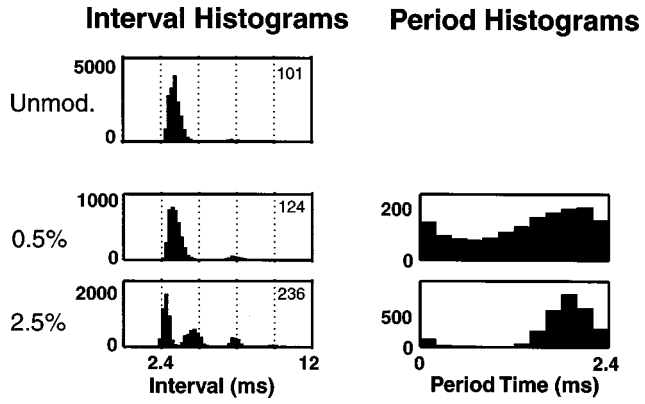


FIG. 7. Interval (left) and period (right) histograms of a fiber's responses to 417-Hz modulations of the DPT at two modulation depths (0.5% and 2.5%). The nonexponential interval histogram (IH−ExpSF=0.51) computed from responses during the unmodulated DPT segments is shown on top. The vertical axis represents number of intervals per bin for the interval histograms, and discharge rate in spikes/s for the period histograms.

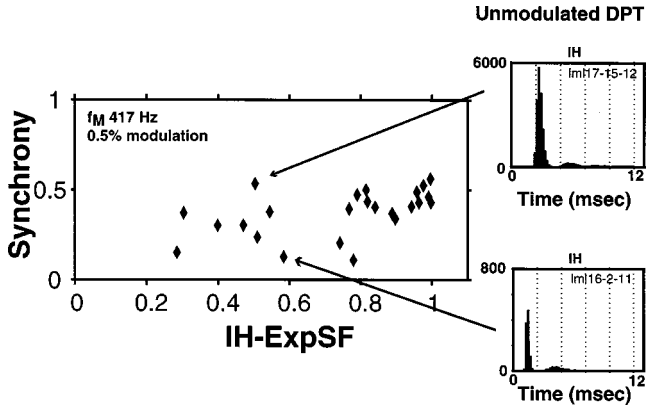


FIG. 8. The left panel shows the synchronization index to a 417-Hz modulator (0.5% modulation depth) as a function of the interval histogram exponential shape factor (IH-ExpSF) for 25 fibers from four cats. The IH-ExpSF was computed based on responses during the unmodulated segments of the DPT. The right panels show interval histograms of responses to the unmodulated DPT for two units with low IH-ExpSF. The vertical dashed lines mark multiples of the modulation period.

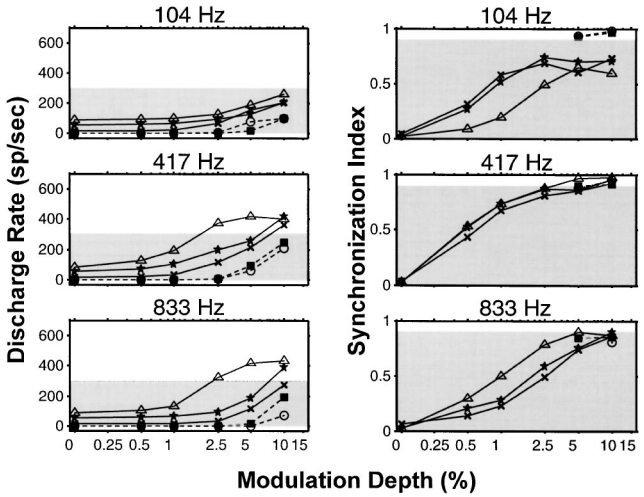


FIG. 9. Average discharge rate (left) and synchronization index (right) of five fibers as a function of modulation depth for three modulation frequencies (rows). Each symbol represents data from one fiber. Transient responders are represented by dashed lines, sustained responders by solid lines. The synchronization index is only plotted if the response includes at least 100 spikes. Shading indicates the range of normal acoustic responses to low-frequency pure tones.

sponse is phase locked to the modulator, yet the first mode of the interval histogram is systematically offset from the modulation period (dotted lines). We refer to the shift of the largest mode in the interval histogram as the *mode offset*. The direction of the mode offset is towards the preferred interval for the unmodulated DPT. A mode offset is also detectable at 2.5% modulation depth (lower left panel), although it is less pronounced than for 0.5% modulation.

The mode offsets in interspike intervals of nonexponential units are accompanied by decreased phase locking to the modulator, when compared to exponential units. At 0.5% modulation depth, responses of the nonexponential fiber in Fig. 7 were weakly phase locked to the modulator (synchronization index of 0.24). In contrast, at the same modulation depth, the response of the exponential unit in Fig. 4 was already nearly fully modulated (SI=0.53). The nonexponential unit also shows reduced phase locking at the 2.5% modulation depth.

Reduced phase locking to a 417-Hz modulator as in Fig. 7 was observed in some, but not all nonexponential fibers. Figure 8 shows the synchronization index to a 417-Hz sinusoidal modulator (0.5% modulation depth) against the interval histogram exponential shape factor (IH-ExpSF) for the unmodulated DPT. For units with IH-ExpSF above 0.8, the synchronization index was uniformly high (between 0.3 and 0.6). In contrast, for units with IH-ExpSF below 0.8, the synchronization index varied widely (from 0.1 to 0.5), and could be as high as those of exponential units in some cases, but much lower in others. The correlation between IH-ExpSF and synchronization index was highly significant ($p=0.007$, permutation test).[2]

The variability in phase locking among nonexponential neurons can be understood from the relationship between the modulation period and the preferred interval for unmodulated DPTs. When the preferred interval is close to the modulation period, synchrony to the modulator is high (Fig. 8, upper inset). In contrast, when the preferred interval does not match the modulation period, synchrony to the modulator is low (lower inset in Fig. 8; see also Fig. 7). This observation suggests that the responses to small modulations of a DPT show an interaction between the stimulus drive and the intrinsic dynamics of the neural membrane giving rise to preferred intervals.

D. Threshold and dynamic range

Because discharge rates in response to sinusoidal modulations increase monotonically with modulation depth, the sound pressure level of acoustic stimuli could in principle be encoded by changes in modulation depth in a DPT-enhanced stimulation strategy (Fig. 1). To evaluate the feasibility of such a scheme, we methodically investigated how average discharge rate and synchronization index to sinusoidal modulations vary with the depth of modulation of the DPT, and determined the range of modulation depths for which these measures lie within the normal acoustic range.

Figure 9 shows the average discharge rate and the synchronization index as a function of modulation depth for five auditory-nerve fibers in response to three modulation frequencies. The left most point in each curve represents the response to the unmodulated DPT. The two fibers having the lowest pseudo-spontaneous rates are transient responders (dashed lines), while the other three are sustained responders (solid lines). Gray shading indicates the approximate normal range of responses to low-frequency pure tones in a healthy ear: average discharge rate below 300 spikes/s (Liberman, 1978), and synchronization index below 0.9 (Johnson, 1980).

For the three sustained DPT responders in Fig. 9, both the average discharge rate and the synchronization index grow over a wide range of modulation depths. With one exception, both response measures are within the normal acoustic range for modulation depths below 5%. At higher modulation depths, most fibers had discharge rates exceeding those of acoustically stimulated fibers. Synchrony for the sustained responders grew rapidly with modulation depth

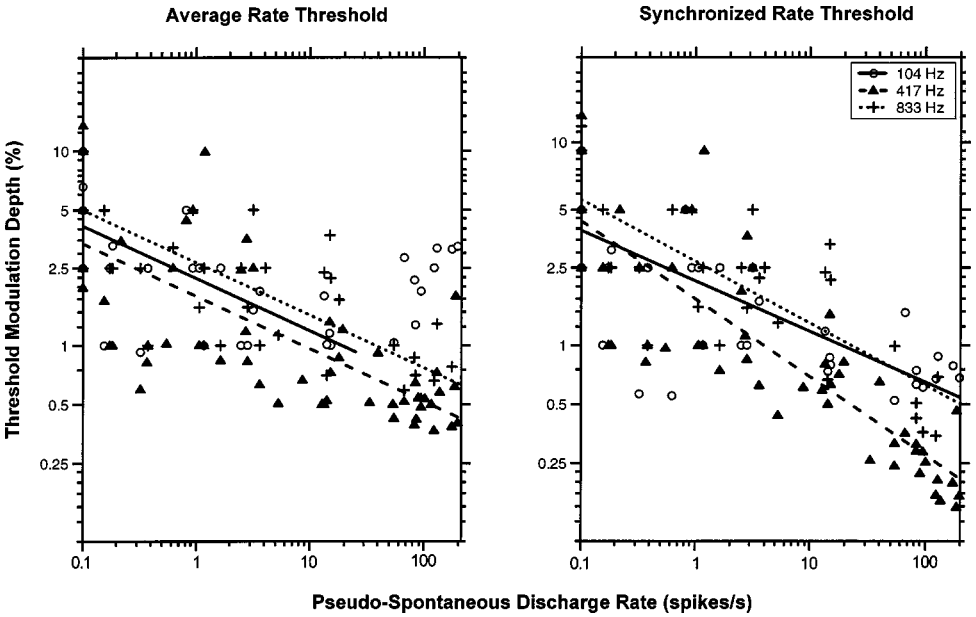


FIG. 10. Modulation detection thresholds based on average discharge rate (left) and synchronized rate (right) as a function of pseudo-spontaneous rate to the DPT for three modulation frequencies. Each point shows data from one fiber. Symbols code modulation frequencies. The straight lines are least-squares fits to the data for each frequency (see Table I). For the average-rate thresholds, the slopes of the regression lines were constrained to be the same for all three frequencies.

from 0.5 to 2.5%, while average rate remained close to the pseudo-spontaneous rate in this range. This behavior resembles that of high-spontaneous fibers in response to pure tones in an intact ear (Johnson, 1980). However, for the 104-Hz modulator, synchrony varied nonmonotonically with modulation depth for two fibers, first increasing then decreasing somewhat. This decrease is due to the appearance of multiple peaks in the period histogram, as shown in Fig. 4 for the largest modulation depth.

For the two transient DPT responders in Fig. 9, the modulation depths necessary to evoke responses were higher than for the sustained responders. In addition, the discharge rates of transient responders grew more rapidly on the logarithmic scale than did those of sustained responders, suggesting that the dynamic ranges are narrower for transient responders. For transient responders, the synchronization index reaches its maximum value as soon as there is a sufficient number of spikes to estimate it reliably, as is the case for low-spontaneous fibers with pure-tone stimulation (Johnson, 1980).

Two criteria were used to define modulation depth thresholds for individual fibers. The average rate threshold is the modulation depth for which the average discharge rate during the 400-ms modulation exceeds the rate during the following 400-ms unmodulated segment for 75% of the stimulus presentations. Similarly, the synchronized rate threshold is the modulation depth for which the synchronized discharge rate (synchronization index times average rate) during the modulation exceeds the synchronized rate during the following unmodulated segment for 75% of the stimulus presentations. This 75% criterion is similar to that attained by adaptive two-alternative choice procedures commonly used in psychophysics for threshold estimation (e.g., Levitt, 1971). Thresholds estimated by our method can therefore be directly compared to psychophysical thresholds. Because the sampling of modulation depths was rather coarse, and most modulation depths tested were above threshold, thresholds were estimated using a special algorithm described in Appendix B.

Figure 10 shows modulation thresholds based on average discharge rate and synchronized rate against the pseudo-spontaneous discharge rate to the unmodulated DPT. Regression lines were fit to the data at each frequency on double logarithmic coordinates (Table I). Confidence intervals for

TABLE I. Dependence of modulation thresholds based on average rate and synchronized rate on pseudo-spontaneous rate (PSR) and modulation frequency (F_m). Thresholds were fit by the equation $\log_{10}(\text{Threshold}) = A + B \log_{10}(\text{PSR})$. The table shows means and 95% confidence intervals for the slope B and the value of the fitted line at 50 spikes/s. Confidence intervals are based on 5000 bootstrap replications of the regression line (Efron and Tibshirani, 1993, Chap. 7). For average-rate thresholds, the slopes of the regression lines were constrained to be the same at all three modulation frequencies because a model with separate slopes did not significantly improve the fit.

	F_m (Hz)	Slope		Threshold at 50 spikes/s	
		Mean	95% C.I.	Mean	95% C.I.
Average rate	104	−0.270	−0.31–0.22	N/A	N/A
	417	same	same	0.62	0.53–0.73
	833	same	same	0.93	0.76–1.14
Synchronized rate	104	−0.260	−0.32–0.20	0.78	0.63–0.91
	417	−0.402	−0.46–0.35	0.36	0.31–0.42
	833	−0.318	−0.39–0.23	0.59	0.59–1.10

the regression parameters were determined using bootstrap replications of the data (Efron and Tibshirani, 1993). For the average-rate thresholds, the slopes of the regression lines were constrained to be the same at all three frequencies because using separate slopes did not significantly improve the mean square error ($p=0.33$).

Both average-rate and synchronized-rate thresholds decrease with increasing pseudo-spontaneous rate for all three frequencies. The slopes of the regression lines range from -0.40 to -0.26 (Table I), meaning that a tenfold increase in pseudo-spontaneous rate results in a 45%–60% drop in threshold. Confidence intervals for the slopes comprise only negative values, indicating that the falling trend is statistically significant. However, for the 104-Hz modulation frequency, average rate thresholds deviate from this trend for pseudo-spontaneous rates above 25 spikes/s; these high-rate data were not included in the regression analysis. These data points may represent a different response regime since multiple spikes per cycle become increasingly common as the discharge rate approaches the stimulus frequency.

A second trend apparent in Fig. 10 is the existence of cross-frequency differences in thresholds. On the average, thresholds are lower for the 417-Hz modulator than for the other two frequencies. This effect is most apparent for sustained DPT responders (pseudo-spontaneous rates above 5 spikes/s). The difference in mean thresholds between 104 and 417 Hz is highly significant for both average and synchronized rate [$p<0.001$, two-sided permutation test for location (Good, 2000, p. 37)], and so is the difference between 833 and 417 Hz ($p<0.002$). Twelve out of 13 sustained responders have lower synchronized rate thresholds for the 417-Hz modulator than for the other two frequencies. Cross-frequency threshold differences are less obvious for transient DPT responders. Regression lines for the synchronized rate thresholds in Fig. 10 tend to come together at low pseudo-spontaneous rates. Permutation tests confirm that transient responders do not show preferred sensitivity to the 417-Hz modulator; in fact, 14 out of 27 fibers are more sensitive to the 104-Hz modulator. Thus, threshold sensitivity appears to be band-pass for sustained DPT responders, and low-pass for transient DPT responders.

A final observation from Fig. 10 is that synchronized-rate thresholds are significantly lower than average-rate thresholds for sustained responders ($p=0.003$, two-sided permutation test), but not for transient responders ($p=0.64$). In this respect, pseudo-spontaneous activity in response to the DPT behaves similarly as true spontaneous activity in that synchronized rate thresholds to pure tones in a healthy ear are lower than average-rate thresholds for high-spontaneous fibers, but not low-spontaneous fibers (Johnson, 1980).

Synchronized rate thresholds of sustained responders in Fig. 10 could be as low as 0.2% modulation depth for the 417-Hz modulator. Because the median DPT level in these experiments was 6 dB *re*: 1 mA (see Table I in Litvak *et al.*, 2003b), the peak amplitude of the modulation waveform was typically only 4 μA at threshold. This is much lower than single-unit thresholds reported for low-frequency sinusoidal electric stimulation, which are typically tens or hundreds of μA (Parkins and Colombo, 1987; Hartmann *et al.*, 1990; Dynes and Delgutte, 1992). Thus the introduction of a DPT is highly effective in improving thresholds for sustained responders.

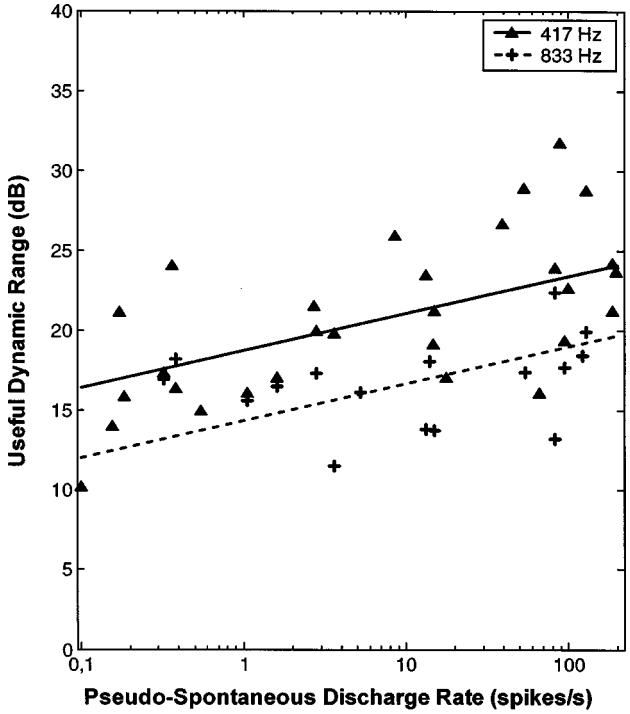


FIG. 11. Useful dynamic range to sinusoidal modulations of a DPT as a function of pseudo-spontaneous rate for modulation frequencies of 417 and 833 Hz. The useful dynamic range is the ratio of the modulation depth that evokes a discharge rate of 300 spikes/s to the synchronized rate threshold. Each point shows data from one fiber. Symbols code modulation frequencies. The straight lines are least-squares fits to the data for each frequency, with the constraint that the slopes be the same for both frequencies (see Table II).

We define the ''useful dynamic range'' of a fiber as the range of modulation depths for which both average discharge rates and synchronization indices resemble those seen for pure-tone stimuli in a healthy ear. The lower limit of the useful range is the modulation threshold based on synchronized discharge rate. The upper limit is the modulation depth that evokes a discharge rate of 300 spikes/s. As shown in Fig. 9, for these discharge rates, most sustained DPT responders have synchronization indices below 0.9, and are therefore within the normal range of synchrony for low-frequency pure tones (Johnson, 1980). The modulation depth that produced a 300 spikes/s discharge rate was estimated by fitting a sigmoid function to the rate versus modulation depth curve. This analysis was not carried out for responses to the 104-Hz modulator because these responses rarely reached 300 spikes/s, and did not resemble normal acoustic responses at high discharge rates in that discharges often occurred at more than one phase of the modulator (e.g., Fig. 4, lower left panel).

Figure 11 shows the useful dynamic range as a function of pseudo-spontaneous discharge rate for the 417- and 833-Hz modulators. Regression lines were fit to the dynamic range at each frequency against the logarithm of the pseudo-spontaneous rate (Table II). The slopes of the regression lines

TABLE II. Dependence of useful dynamic range on pseudo-spontaneous rate (PSR) and modulation frequency (F_m). The dynamic range DR in dB was fit by the equation $DR = A + B \log_{10}(PSR)$. The table shows means and 95% confidence intervals for the slope B and the value of the fitted line at 50 spikes/s. Confidence intervals are based on 5000 bootstrap replications of the regression line (Efron and Tibshirani, 1993, Chap. 7). The slopes of the regression lines were constrained to be the same at both modulation frequencies because a model with separate slopes did not significantly improve the fit.

F_m (Hz)	Slope (dB/decade)		Dynamic range at 50 spikes/s (dB)	
	Mean	95% C.I.	Mean	95% C.I.
417	2.32	1.21–3.43	22.7	20.9–24.6
833	same	same	18.3	16.7–19.9

were constrained to be the same at both frequencies because a regression with separate slopes did not significantly reduce the mean square error ($p = 0.21$). The slope of the regression line is 2.3 dB/decade, meaning that a tenfold increase in pseudo-spontaneous rate increases the dynamic range by 2.3 dB. The statistical significance of this trend is confirmed by the entirely positive 95% confidence intervals for the slope in Table II.

The average useful dynamic range of sustained responders is 23 dB for the 417-Hz modulator, and 17 dB for the 833-Hz modulator. The difference between the two means is highly significant ($p < 0.001$, two-sided permutation test for difference in means). These values of dynamic range are comparable to those observed with pure-tone stimulation in a normal ear (Sachs and Abbas, 1974), and considerably larger than the dynamic ranges observed with sinusoidal electric stimulation without a DPT (van den Honert and Stypulkovsky, 1987; Hartmann *et al.*, 1990; Dynes and Delgutte, 1992). In contrast, transient DPT responders have useful dynamic ranges as low as 10 dB, more in line with electric dynamic ranges reported in the literature. However, Fig. 11 may not give a completely representative picture of the dynamic range of transient responders because it does not include data from the fibers (primarily transient responders) whose discharge rates failed to reach the 300 spikes/s upper limit of the useful dynamic range at the highest modulation depth tested.

While the DPT-induced increase in dynamic range for sustained responders is encouraging, a key question is what fraction of the population of auditory-nerve fibers would be likely to benefit from a DPT at a given level. To address this question, solid lines in Fig. 12 shows the percentage of fibers that gave ''acoustic-like'' responses to the 417- and 833-Hz sinusoidal modulators as a function of modulation depth. By ''acousticlike,'' we mean that the response is above the synchronized rate threshold and the average rate is below 300 spikes/s. Percentages are computed separately for sustained and transient DPT responders. Dotted curves indicate the percentage of fibers that have sustained discharge rates above 300 spikes/s, and are therefore above the normal acoustic range. For modulation depths between 1% and 5%, most sustained responders have responses within the acoustic range at both frequencies. For modulation depths of 10% and above, most sustained DPT responders have discharge rates above the normal acoustic range. On the other hand, modulation depth has to exceed 5% for a majority of transient DPT responders to respond to the modulation. Because approximately 50% of the fibers in our entire sample from ten animals were transient responders (Litvak *et al.*, 2003b), the percentage of acoustic-like responses for this sample (not shown) would be nearly the mean of that for the sustained and transient responders. Overall, a substantial fraction (40%–55%) of the fibers would exhibit acoustic-like responses for modulation depths between 1% and 5%. Of course, these estimates are likely depend on the choice of DPT amplitude.

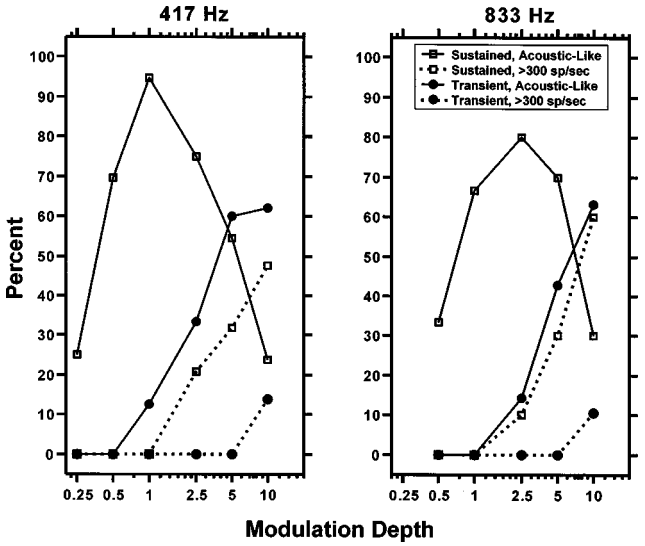


FIG. 12. This figure shows, as a function of modulation depth, the percentage of fibers whose responses to a modulated DPT are within the useful dynamic range (solid lines), as well as the percentage of fibers that exceed the maximal acoustic rate of 300 spikes/s (dashed lines) for modulation frequencies of 417 Hz (left) and 833 Hz (right). Percentages are shown separately for sustained and transient DPT responders. Only responses that were recorded at the standard DPT level in each animal (Table I in Litvak *et al.*, 2003b) are included.

E. Responses of the stochastic threshold model

Several key features of ANF responses to sinusoidally modulated DPTs are predicted by an extremely simple stochastic threshold model (STM). The STM produces a spike whenever the modulation waveform $m(t)$ crosses a noisy threshold (Fig. 2). The response of the model is entirely determined by two dimensionless parameters: m/σ and θ_0/σ. The input-to-noise ratio m/σ determines the effective stimulus level (modulation depth), while the threshold-to-noise ratio θ_0/σ determines the pseudo-spontaneous discharge rate in the absence of modulation. We will show that the relationship between the STM's pseudo-spontaneous rate and its responses to sinusoidal modulators is similar to that seen in the data.

Figure 13 shows the average discharge rate and synchronization index as a function of m/σ for model fibers with different pseudo-spontaneous discharge rates (different values of θ_0/σ). The model predicts several features of the data

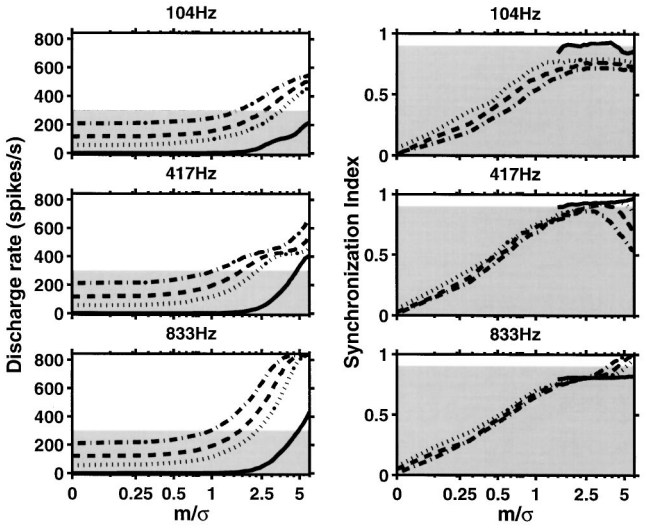


FIG. 13. Average discharge rate (left) and synchronization index (right) of the stochastic threshold model as a function of the normalized modulation depth m/σ for three modulations frequencies (rows). Each curve corresponds to one value of the threshold to noise ratio θ_0/σ: 1.4 (dot-dash), 1.8 (dash), 2.4 (dot), and 4 (solid). The corresponding pseudo-spontaneous discharge rates are 210, 119, 30, and 0.2 spikes/s, respectively.

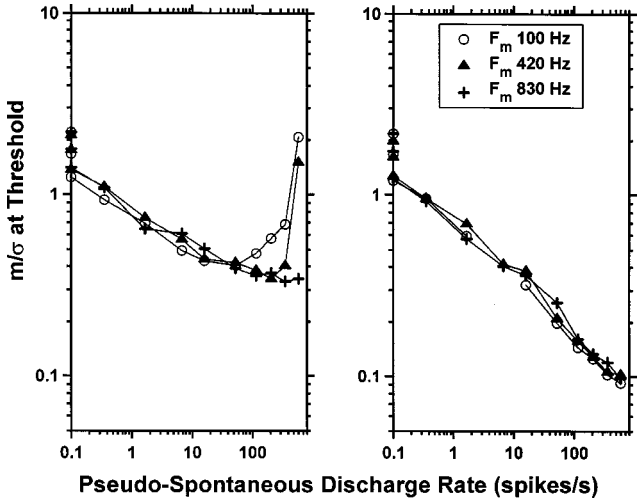


FIG. 14. Normalized modulation depth thresholds m/σ based on average rate (left) and synchronized rate (right) as a function of pseudo-spontaneous rate for the stochastic threshold model. Symbols code modulation frequencies.

shown in Fig. 9: (1) Both average rate and synchrony grow over a wide range of modulation depths. (2) Both model and neural fibers with low (<5 spikes/s) pseudo-spontaneous rates have higher modulation thresholds than fibers with high pseudo-spontaneous rates. (3) For both model and neural fibers with high pseudo-spontaneous rates, synchrony grows rapidly at low modulation depths, while the average rate remains near pseudo-spontaneous. In contrast, fibers with low pseudo-spontaneous rates show high synchrony as soon as the modulation depth reaches threshold. (4) For both model and data, synchrony grows nonmonotonically with modulation depth at 104 Hz due to the occurrence of multiple spikes in each stimulus cycle at high levels. However, the maximum discharge rates exhibited by the model are larger than those seen in the data. In addition, while neural synchrony grows monotonically with modulation depth at 417 Hz, synchrony is somewhat nonmonotonic in the model responses at very high discharge rates.

Figure 14 shows the model average and synchronized rate thresholds as a function of pseudo-spontaneous rate for three modulation frequencies. These predictions can be directly compared to the neural thresholds in Fig. 10. Note, however, that while the neural pseudo-spontaneous discharge rates are always below 200 spikes/s, the model rates were extended to 600 spikes/s. For both model and data, average rate thresholds decrease nearly linearly with the logarithm of pseudo-spontaneous rate at relatively low rates for all three modulation frequencies. The model thresholds behave nonmonotonically, increasing steeply once the pseudo-spontaneous rate exceeds the modulation frequency. Such nonmonotonic dependence of detectability on noise amplitude (which controls the pseudo-spontaneous rate) is a defining characteristic of stochastic resonance (Wiesenfeld and Moss, 1995). Some evidence for a nonmonotonicity can also be discerned in the neural data of Fig. 10 for the 104-Hz modulator in that thresholds for pseudo-spontaneous rates above 40 spikes/s lie above the regression line fit to the data at lower pseudo-spontaneous rates. However, the data show no nonmonotonicity at 417 and 833 Hz, presumably because the pseudo-spontaneous rates always remain well below the modulation frequency.

The model synchronized rate thresholds decrease nearly linearly with the logarithm of pseudo-spontaneous rate for all three modulation frequencies. The synchronized rate thresholds decrease more rapidly than the average rate thresholds. The same trends are apparent in the data for the 417-Hz modulator (Fig. 10). However, only the data show a dependence of thresholds on modulation frequency, with the lowest thresholds at 417 Hz.

Figure 15 shows the model useful dynamic range as a function of pseudo-spontaneous rate for 417- and 833-Hz

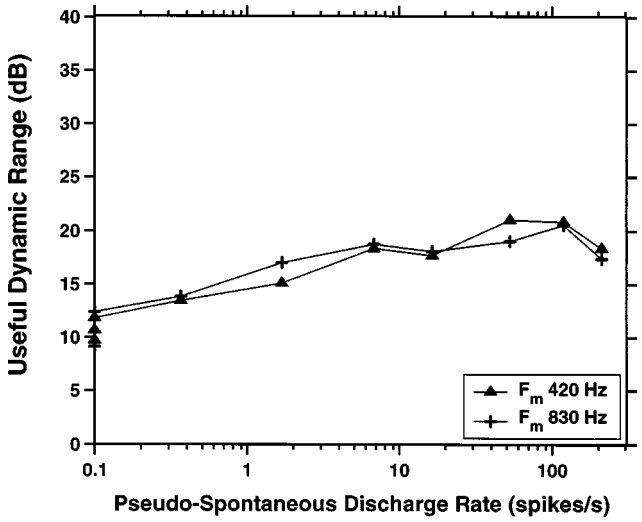


FIG. 15. Useful dynamic range of the stochastic threshold model as a function of pseudo-spontaneous rate for two modulation frequencies. The definition of useful dynamic range is the same as in Fig. 11.

modulators. As for the data, the useful dynamic range is the ratio of the modulation depth that evokes a discharge rate of 300 spikes/s to the synchronized rate threshold. Because the dynamic range is a ratio of modulation depths, there are no free parameters in this plot. The model dynamic range increases linearly with the logarithm of pseudo-spontaneous rate up to 50 spikes/s, then flattens out at about 19 dB for both frequencies. The increase in dynamic range at low pseudo-spontaneous rates is apparent in the data of Fig. 11, but the plateau at high rates is less obvious. The model's plateau dynamic range (19 dB) is intermediate between the mean 23-dB dynamic range in the data for the 417-Hz modulator and the 17 dB observed for the 833-Hz modulator. However, the model does not predict the difference in dynamic ranges between the two modulation frequencies.

Overall, the stochastic threshold model does a good job of predicting the relationship between the pseudo-spontaneous discharge rates of ANFs and their responses to sinusoidal modulators at a given frequency. However, this version of the model cannot account for the frequency dependence of neural sensitivity to modulation.

IV. DISCUSSION

A. Representation of sinusoid in temporal discharge patterns

All auditory-nerve fibers in our study responded to sinusoidal modulations of the DPT. Neurons that had sustained responses to the DPT were the most sensitive to modulations. These neurons showed strong phase locking to the modulator for modulation depths as low as 0.5%. In these neurons, the average discharge rate grew monotonically with modulation depth over a range of 17–23 dB before reaching 300 spikes/s, the upper limit of discharge rates for pure-tone stimulation. The temporal discharge patterns to modulation depths below 2.5%–5% resembled acoustic responses to pure tones. These responses were stochastic, with interspike intervals occurring at random at multiples of the modulator period. The period histograms of these responses were similar to those seen in responses to pure tones, and provided a good representation of the sinusoidal stimulus waveform.

For modulation depths above 5%–10%, sustained DPT responders exhibited higher discharge rates than those seen for acoustic stimulation. For 417-Hz modulation, discharges tended to entrain (occur once per stimulus cycle) while, for 833-Hz modulation, some neurons fired exactly on every other stimulus cycle. These responses were more precisely phase locked than acoustic responses. In most respects, responses to large modulations of the DPT resemble suprathreshold responses to sinusoidal electric stimulation without a DPT (Hartmann *et al.*, 1984; van den Honert and Stypulkowski, 1987).

Fibers that only responded transiently to the unmodulated DPT nevertheless responded with phase-locked spike discharges to sinusoidal modulations. However, larger modulation depths were needed to produce a response in these fibers than in sustained responders. Thus, the introduction of a DPT qualitatively mimics the threshold differences between high-spontaneous and low-spontaneous fibers observed in normal ears (Liberman, 1978). At these large modulation depths, the responses of transient DPT responders were more synchronized than responses to tones in a normal ear, and poorly represented the stimulus waveform. In addition, transient DPT responders had a smaller dynamic range than sustained responders. In this respect, transient responders differ from low spontaneous-rate ANFs in a healthy ear, which have wider dynamic ranges than high spontaneous fibers (Schalk and Sachs, 1980).

The deafening protocol in these experiments was only partly successful in that some animals had residual hearing in the nonimplanted ear (Litvak *et al.*, 2003b). This observation raises the possibility that the acoustic-like responses of sustained responders may reflect hair-cell mediated activity. Although we cannot rule out this possibility in every case, we did observe acoustic-like responses to modulations of the DPT in animals with no residual hearing as well as in animals with substantial hearing. Our measure of residual hearing was very conservative in that it did not take into account any additional hearing loss caused by insertion of the stimulating electrodes into the implanted cochlea.

While the temporal discharge patterns of sustained responders to small sinusoidal modulations of a DPT resemble responses to pure tones in a healthy ear, there are nevertheless differences in the interspike interval distributions between the two modes of stimulation. While the interspike distributions in response to a pure tone always show a mode at the tone period for frequencies below 1 kHz, this mode was often lacking in responses to 417-Hz electric modulations. This lack of a mode at the stimulus period has also been observed for 500-Hz sinusoidal electric stimulation without a DPT (van de Honert and Stypulkowski, 1987). In addition, for neurons having nonexponential interspike interval distributions in responses to unmodulated DPTs, modes in the interval histogram for sinusoidal modulators were often systematically shifted away from multiples of the stimulus period, particularly for very low modulation depths. These shifts are considerably larger than the small, but systematic offsets seen in interspike interval distributions to pure tones in a healthy ear (McKinney and Delgutte, 1999). The mode offsets seen with modulated DPTs seem to reflect an interaction between the stimulus period and the intrinsic dynamics of the neural membrane leading to preferred intervals in response to the unmodulated DPT. Preferred intervals have also been noted for electric stimulation with high-frequency (>1 kHz) sinusoids without a DPT, and found to distort phase locking to the stimulus (Parkins, 1989).

Although interspike interval distributions of sustained DPT responders for small sinusoidal modulations differ somewhat from responses to pure tones in a healthy ear, these distortions may not severely interfere with neural representation of the stimulus frequency. Because there is no evidence for cross-fiber correlation in responses to unmodulated DPTs (Litvak *et al.*, 2003b), responses to small modulations of a DPT may be conditionally uncorrelated from one fiber to the next. By ''conditionally uncorrelated,'' we mean that the occurrence of a spike on a particular stimulus cycle for one fiber does not alter the probability that a spike occurs

on the same cycle in another fiber. By integrating synaptic inputs from several conditionally uncorrelated auditory-nerve fibers, some central auditory neurons might show a mode at the stimulus period in their interval histograms even if such a mode is lacking in their ANF inputs. Similarly, because mode offsets in interspike intervals differ from one fiber to the next and can be either positive or negative (Fig. 8), these offsets might average out in the responses of integrating central neurons so as to produce a mode centered at the stimulus period. Nevertheless, both distortions could degrade somewhat the accuracy of frequency representation. Direct recordings from central neurons (e.g., cochlear nucleus) in response to a modulated DPT stimulus would test whether synaptic integration suffices to overcome the distortions in the coding of the modulation frequency observed in the auditory nerve.

B. Dynamic range

A major finding of the present study is the relatively wide dynamic range (17–23 dB) for sustained DPT responders. In studies of electric stimulation, dynamic range is normally measured by the varying the amplitude of a stimulus without changing its waveform, and determining the range of amplitudes over which the discharge rate changes. Here, we varied the depth of sinusoidal modulations of a DPT, and determined the range of modulation depths over which either average rate or synchronized rate varied (with an upper limit of 300 spikes/s for average rate). As shown in Fig. 1, this method is equivalent to varying the amplitude of a 100%-modulated stimulus superimposed upon a fixed, unmodulated DPT, which is not considered part of the information-bearing stimulus since its function is to evoke pseudo-spontaneous activity. Because modulation depth in percent can be translated into stimulus amplitude in mA by simple multiplication by the DPT amplitude in mA, a modulation-depth dynamic range in dB is numerically equivalent to a conventional dynamic range for the information-bearing part of the stimulus (excluding the fixed DPT).

The most detailed information on dynamic range for electric stimulation comes from the example of rate-level functions for periodic trains of biphasic pulses shown by Javel and his colleagues (Javel *et al.*, 1987; Javel, 1990; Shepherd and Javel, 1997; Javel and Shepherd, 2000). Although no formal statistics are provided, the dynamic ranges for the highest pulse rate tested (800 pps) range from 3 to 12 dB in the examples shown, with a median near 6 dB. The discharge rates of the fibers studied by Javel *et al.* typically approached the 800/s pulse rate at high stimulus amplitudes. Therefore, the dynamic ranges would only be about half as large if a 300 spikes/s upper limit was imposed, as we did in the present study. These values are consistent with those reported for sinusoidal electric stimulation at frequencies of 1 kHz and above (van den Honert and Stypulkowski, 1987; Hartmann *et al.*, 1990; Dynes and Delgutte, 1992). They are also consistent with the 2–4 dB conventional dynamic ranges measured in our earlier study (Litvak *et al.*, 2001) for 4.8-kpps pulse trains sinusoidally modulated at 400 Hz with a 10% depth. The modulation dynamic ranges observed for transient DPT responders (as low as 10 dB) overlap with the values given in the literature for sinusoidal or pulsatile stimuli. However, the 17–23 dB dynamic ranges of sustained DPT responders are clearly higher than any of the values reported in the literature.

The improvement in dynamic range resulting from the introduction of a DPT in sustained responders arises from at least two distinct effects. First, the pseudo-spontaneous activity evoked by the DPT allows synchronized rate thresholds to be about 10 dB lower than the average rate thresholds. In this range of modulation depths, the stimulus sinusoidally modulates the pseudo-spontaneous activity without causing an increase in average rate, as is the case in a spontaneously active fiber with acoustic stimulation.

The second effect giving rise to improved dynamic range with a DPT is that, because rate-level functions are expansive near threshold, the response to the superposition of two stimuli is always greater than the sum of the responses to each of the stimuli. The presence of a DPT therefore allows the superimposed modulated stimulus to evoke an increment in discharge rate over background, even if, by itself, this stimulus would be below threshold. Because rate-level functions become compressive rather than expansive near their saturation, this mechanism is likely to be effective over only a limited range of pseudo-spontaneous rates. Specifically, if the pseudo-spontaneous rate approached the upper limit imposed by neural refractoriness, the threshold for superimposed stimuli would be expected to increase rather than decrease. Such an increase in rate thresholds is observed for the stochastic threshold model at high pseudo-spontaneous rates (Fig. 14). However, this condition did not occur in our data (Fig. 10), presumably because pronounced adaptation over the course of the DPT always kept the pseudo-spontaneous rates well below saturation (Litvak *et al.*, 2003b). Adaptation to the continuous DPT is thus likely to play an indirect role in improving the dynamic range. Whether it also plays a more direct role, for example by decreasing the slopes of rate-level functions, cannot be determined from our data.

The above reasoning suggests that the effectiveness of a DPT in lowering threshold and increasing dynamic range does not depend on either the exact parameters of the DPT, or the particular scheme used for encoding the sinusoidal stimulus, so long as the pseudo-spontaneous rate evoked by the DPT lies in the appropriate range. Thus, a DPT might remain effective if an analog sinusoidal stimulus was simply added to the DPT, as originally proposed by Rubinstein *et al.* (1999b), rather than encoded as a modulation of the DPT. In other words, a DPT may be as effective with an analog stimulation strategy as with the CIS-type strategy used in this paper. Similarly, a lower-rate pulse train (e.g., 1 kpps) might be equally effective as our 5-kpps DPT in improving the dynamic range. However, since ANFs adapt less to low-rate pulse trains than to high-rate trains (Moxon, 1967; van den Honert and Stypulkowski, 1987), more fibers might be near saturation of the input–output function with a low-rate DPT. A low-rate DPT would have the further disadvantage that phase locking to the pulse rate might give rise to pitch sensations that would bear no relation to the sinusoidal stimulus.

In addition, if sinusoidal stimuli are encoded as modulations of the DPT, the low rate of a 1-kpps pulse train would prevent all but very low frequency signals to be encoded without aliasing the modulation waveform. Thus, a 5-kpps DPT seems to be preferable over a lower-rate DPT, particularly with a CIS-type strategy.

C. Comparison with models of electrically stimulated fibers

1. The stochastic threshold model

We presented a stochastic threshold model of ANF responses to modulations of a DPT. Despite its simplicity, the STM accounts for the dependence of both threshold and dynamic range on the pseudo-spontaneous discharge rate evoked by the unmodulated DPT. In addition, the STM quantitatively predicts (within 2–4 dB) the average useful dynamic range of sustained DPT responders. However, the dependence of threshold and dynamic range on modulation frequency is not predicted by the STM. A parsimonious way to account for that dependence would be to introduce a bandpass filter at the input to the model, so as to amplify modulation frequencies near 400 Hz. We show in a companion paper (Litvak *et al.*, 2003a) that such a filter is also necessary to account for ANF responses to complex modulations of a DPT.

It may seem at first sight surprising that the dependence of threshold and dynamic range on pseudo-spontaneous rate is predicted by the very simple stochastic threshold model. We believe that the agreement is not accidental. Many biophysically realistic neural models can be approximated by a driving function, which depends only on the stimulus, and another (possibly noisy) function describing threshold dynamics, which depends only on previous activity (e.g., Hill, 1936). In general, the driving function may depend on the stimulus in a complex, nonlinear way. However, because the stimuli in our study are composed of a large signal (the DPT) perturbed by small modulators, the driving function may be linearized around the operating point imposed by the DPT. Thus, to the extent that the threshold component of the model captures the complex dynamics of neural refractoriness, the stochastic threshold model can be considered an approximation of a more realistic neural model for small modulation depths.

It is worth emphasizing that the DPT is not explicitly represented in the STM because the input to the model is the modulation waveform without a pulse-train carrier. The effect of a DPT is simulated in the model by varying the threshold-to-noise ratio θ_0/σ (Fig. 2), which controls the pseudo-spontaneous rate. The pseudo-spontaneous rate can be increased by either decreasing the resting threshold θ_0 or, equivalently, increasing the neural noise. That the STM can nevertheless predict the effects of a DPT on neural responses to modulations bolsters the argument made in the previous section that the effectiveness of a DPT in improving dynamic range and temporal discharge patterns is not likely to depend on the exact characteristics of the DPT such as its pulse rate.

2. The Rubinstein biophysical model

In many respects, ANF responses to a modulated DPT also resemble the responses of a biophysical model to electric sinusoids in the presence of a 5-kpps conditioner (Rubinstein *et al.*, 1998). At low levels of the electric sinusoid, period histograms for this model resemble the sinusoidal stimulus waveform. We obtained a similar result for responses to small modulations of the DPT. The useful dynamic range of the biophysical model responses was near 20 dB, and changed little with DPT level, so long as the DPT evoked discharge rates between 10 and 250 spikes/s. The measured responses to the modulated DPT showed a similar trend, with no systematic relationship between the pseudo-spontaneous rate and useful dynamic range for discharge rates above 5 spikes/s (Fig. 11). Because Rubinstein *et al.* did not report interval histograms of model responses to the sinusoid, it is unclear whether their model captures the mode shifts, and the missing first mode observed in the data. Although responses of the biophysical model of Rubinstein *et al.* (1999b) to a modulated DPT remain to be studied in detail, the results so far suggest that this model captures the essential features of responses of ANFs to sinusoidal electric stimulation.

D. Function of spontaneous activity

Spontaneous activity in sensory neurons is often considered as noise that imposes fundamental limitations on the performance achievable by an ideal observer in any detection or discrimination task based on the neural discharge patterns (Barlow and Levick, 1969; Siebert, 1965; Werner and Mountcastle, 1965). While this point of view is certainly valid, it does not specify a functional advantage of spontaneous activity that might account for its nearly ubiquitous presence in primary sensory neurons (Retinal ganglion cells: Kuffler, 1953; Rodieck, 1967; Somatosensory afferents: Werner and Mountcastle, 1965; Auditory nerve: Kiang *et al.*, 1965; Vestibular nerve: Goldberg and Fernandez, 1971; Walsh *et al.*, 1972; Olfactory receptors: Chaput and Holley, 1979; Rosparset *et al.*, 1994; Gustatory receptors: Pfaffmann, 1955). Our physiological and modeling results, as well as those of others (Yu and Lewis, 1989; Schneidman *et al.*, 1998; Rubinstein *et al.*, 1999a, b) suggest such an advantage.

Spontaneous activity helps to faithfully encode stimulus waveforms in the temporal discharge patterns of sensory neurons by allowing these waveforms to be represented by small modulations of ongoing activity. Such modulation coding lowers threshold and mitigates the distortions caused by refractoriness in single neurons. Spontaneous activity may also desynchronize stimulus-driven activity across neurons in a population, thereby allowing a volley principle to operate when the stimulus period is shorter than the neural refractory period. In this view, noise resulting from random spontaneous activity is the price paid for the lower thresholds and improved temporal representation of waveforms in the neural population. The performance limitations imposed by noise can, in principle, be reduced by averaging responses

across similarly driven neurons. The net result will be beneficial if the population is large enough and the noise largely uncorrelated across neurons.

E. Comparison with psychophysical modulation detection thresholds

It is of interest to compare our single-fiber modulation detection thresholds with psychophysical thresholds for human cochlear implant listeners (Shannon, 1992). In that study, subjects were stimulated continuously with 500–2000-Hz sinusoids. Beats were produced by presenting another sinusoid differing slightly in frequency. Because our neural threshold criterion closely parallels that used in psychophysics, direct comparison of threshold values is appropriate.

Shannon found that subjects were most sensitive to 100-Hz beats. At these frequencies, a modulation depth of 1% could be detected. These modulation depths are comparable to the most sensitive rate modulation detection thresholds that we observed for 104-Hz modulators, and are close to the mean synchronized rate thresholds. Unlike the neural data, however, sensitivity of human subjects to beats dropped rapidly with increasing beat frequency. The best psychophysical modulation detection thresholds for a 400-Hz modulator were near 3%. In contrast, our rate-based neural detection thresholds were lower for 417- than for 104-Hz modulation, and could be as low as 0.4%.

Shannon (1992) suggested that the drop in the subjects' ability to detect modulations with increasing beat frequency may reflect a central limitation in sensitivity to high-frequency modulations. Because the peripheral neurons encode 400-Hz modulations at least as well as 100-Hz modulations, our data are consistent with this view. However, the differences between our neural data and Shannon's (1992) psychophysical thresholds might also be accounted for by differences in stimuli (sinusoidal modulations of a 5-kpps pulse train versus beats of a 0.5–2 kHz sinusoid). Psychophysical studies with stimuli more similar to those used in this study are needed to determine how well higher-frequency modulations of a DPT can be detected by human listeners.

F. Implications for cochlear implant processors

Our results strongly suggest that a high-rate, ongoing DPT can enhance the representation of the temporal fine structure of sinusoids in auditory nerve responses for frequencies up to at least 800 Hz. The enhancement is maximal for small ($<5\%$) modulations of the DPT and is most apparent in fibers that have sustained responses to the DPT. Specifically, responses of sustained DPT responders to small modulations resemble responses to sound in a healthy ear in at least four respects: (1) the dynamic range is near 20 dB; (2) interspike intervals occur randomly at multiples of the stimulus period; (3) spikes are distributed over most of one-half cycle, thereby giving a faithful representation of the stimulus waveform; and (4) there is a range of stimulus levels over which synchrony grows rapidly with little or no increase in average rate. No other strategy for electric stimulation has been shown to exhibit any of these properties over a wide range of levels and frequencies. These observations suggest that a processing strategy that incorporates a DPT, and that preserves temporal fine structure information in the signals delivered to the stimulating electrodes, may improve performance of cochlear implant users. In particular, improved coding of the sound fine-time structure may lead to improved perception of musical pitch, and to more effective utilization of binaural cues in bilateral implants (Smith *et al.*, 2002).

One stimulation strategy that does preserve the fine-time structure of the signals delivered to the stimulating electrodes is the continuous analog (CA) strategy. Because a DPT may allow for better encoding of the fine-time structure, performance with a CA strategy may be substantially enhanced by a DPT. In current implementations, the CIS strategy discards the fine-time structure at the envelope detector stage. One function of the envelope detector is to produce a strictly positive modulation waveform that can be used to modulate a carrier pulse train. Because negative waveforms can be represented as decreases in DPT amplitude, it may be possible to entirely eliminate the envelope detector stage in DPT-enhanced CIS strategies. The rate of the pulse-train carrier would also have to be increased in order to sample the higher modulation frequencies without aliasing. Thus, a DPT may be beneficial with CIS as well as with CA strategies, so long that the information-bearing signals are small compared to the ongoing DPT.

We found that sustained DPT responders are best at representing the sinusoidal stimuli in their temporal discharge patterns. A comfortable majority of the sustained responders in our data gave acousticlike responses for modulation depths below 5% (Fig. 12). A key question is what fraction of the auditory-nerve population has a sustained response to a given DPT level. In the companion paper (Litvak *et al.*, 2003b), we showed that the DPT needs to be at least 4 dB above ECAP threshold to evoke a sustained response in a sizeable fraction of the neurons, but that DPT levels more than 8 dB above ECAP threshold can evoke long-lasting changes in neural excitability, and may therefore be harmful to the nerve. To stimulate a large number of fibers, while simultaneously meeting the safety constraint, it may be advantageous to present a DPT at a low (4 dB above ECAP threshold) level on several electrodes. Further research is needed to work out the best trade-off between the safety of DPT stimulation and its effectiveness.

The transient responders comprised roughly 50% of the neurons that we recorded from in our sample from ten aninals. Even though a strategy that uses only small modulation depths may stimulate only half of the neurons that are excited electrically, it may still produce substantial benefit by mimicking some aspects of normal ANF responses in the neurons that it does stimulate. On the other hand, even with a large number of fibers showing fairly natural temporal discharge patterns, the central nervous system might not be able to make use of this information if temporal processing depends critically on factors that were not considered here such as phase relationships between response of fibers innervating

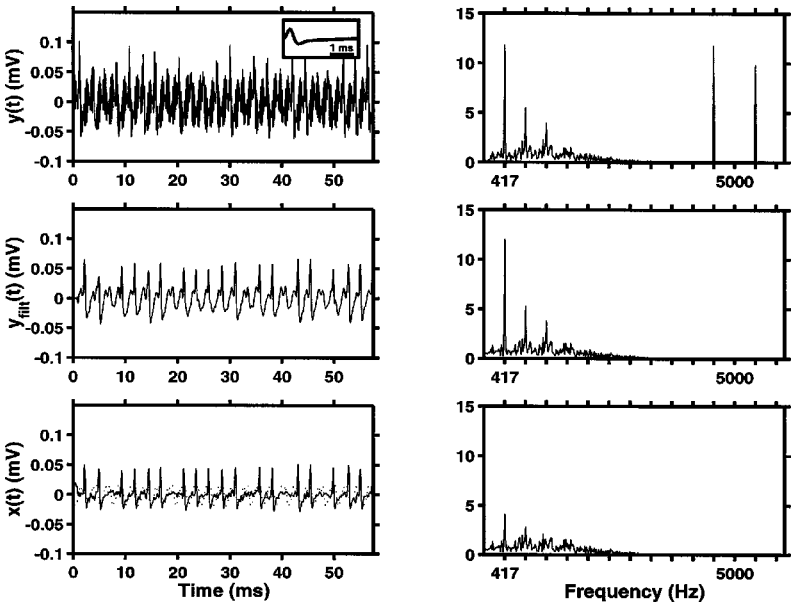


FIG. 16. Algorithm for removing the stimulus artifact from contaminated recordings. The top row shows the waveform (left) and the spectrum (right) of the signal recorded from a microelectrode after it has been passed through a 0.2-ms moving-average filter. The stimulus is a sinusoidally modulated pulse train (modulation frequency: 417 Hz; modulation depth: 5%). The inset shows the spike waveform for an unmodulated pulse train. The middle row shows the waveform and spectrum after low-pass filtering the signal at 3 kHz. The bottom row shows the waveform and spectrum after eight steps of the iterative artifact rejection algorithm described in Appendix A.

different locations (Heinz *et al.*, 2001; Shamma and Klein, 2000).

Our results suggest that a DPT would decrease thresholds to electric stimulation. In the case of CIS, if the CIS carrier and the DPT are in phase, then our results suggest that there would be sufficient information in the auditory nerve to detect modulations as small as 0.5% for 417-Hz modulators, and near 1 to 2% for 104- and 833-Hz modulators. For a DPT level of 2 mA and a modulation depth of 0.5%, the amplitude of the modulation waveform is only 10 μA, which is considerably lower than typical single-unit thresholds reported for sinusoidal electric stimulation. While these results are promising, the improvements in single-fiber thresholds may not necessarily translate into improvements in psychophysical dynamic range because the latter is determined by the distribution of ANF thresholds as well as by the dynamic range of individual fibers. Because a DPT is expected to modify the threshold distribution, its effect on the psychophysical dynamic range is unclear.

V. GENERAL CONCLUSION

We showed that, using a DPT, it may be possible to design strategies for cochlear implants that would realistically code the fine-time structure of sound stimuli in the temporal discharge patterns of electrically stimulated auditory-nerve fibers. If the central nervous system can make use of this information, then these strategies may substantially improve the quality of auditory experience enjoyed by cochlear implant users.

ACKNOWLEDGMENT

The authors thank M. F. McKinney for providing the acoustic data used in this study, Z. M. Smith for help with the physiological experiments, and K. M. Brinsko for assistance in figure preparation. We would also like to acknowledge the tireless efforts of Leslie Liberman without whose surgical skills this work would be impossible. This work was supported through the W.M. Keck Foundation and NIH Grant Nos. DC00361, DC02258, and N01-DC-6-2100.

APPENDIX A: METHOD FOR ARTIFACT REJECTION

Single-unit recordings in response to electric stimulation are contaminated by two distinct signals with can interfere with detection of action potentials: (1) the stimulus artifact proper, representing the voltage drop between the recording electrodes directly due to the stimulating current, and (2) evoked potentials reflecting the summed activity of many neurons recorded in the far field. For convenience, we refer to the superposition of these interfering signals as ''the artifact.''

To remove the artifact from responses off-line, we first filtered the recorded signal using a moving average filter of length equal to the carrier period (0.2 ms). This filter removed the 5-kpps periodic component from the artifact. For modulation depths below 2.5%, the remaining artifact was smaller than the noise in the recording for nearly all of the records. For modulation depths between 2.5% and 15%, however, the artifact was often comparable in amplitude to the spikes. The top row of Fig. 16 shows a typical waveform and spectrum of the recorded signal after processing by the moving-average filter. This signal, which we denote by $y(t)$, is the sum of (1) the artifact waveform $a(t)$, (2) the spike train $x(t)$, and (3) noise $n(t)$. In this example, the modulator frequency is 417 Hz. As expected, the spectrum has its largest peaks at the two components present in the stimulus: $5000-417$ Hz and $5000+417$ Hz. In addition, there are peaks at the 417-Hz modulation frequency and its harmonics 834 and 1251 Hz. These peaks are distortion products of the modulator waveform, and indicate the nonlinear nature of the artifact (including evoked responses). While the higher-frequency components at 5000 ± 417 Hz can easily be removed by low-pass filtering (Fig. 16, second row), the low-frequency distortion products occur in the region where spikes have significant energy, and therefore cannot be simply filtered out.

We developed an iterative algorithm to remove these low-frequency components of the artifact. The algorithm is based on two assumptions: (1) that the artifact $a(t)$ can be represented by a low-order Fourier series of the modulator frequency, and (2) that the spike train $x(t)$ is composed of spikes whose waveform is known within an overall gain. The principle of the algorithm is to refine an estimate of $a(t)$ until the spikes in the estimated neural response $y(t)-a(t)$ maximally resemble spikes recorded during the unmodulated DPT (Fig. 16, upper left panel, inset). While the artifact is always a low-order Fourier series, the spikes, which have rapid rise times and occur with some jitter relative to the modulator, are poorly represented by this Fourier series. Thus, it is unlikely that spikes would be erroneously introduced or removed by this algorithm.

As an initial guess $a^0(t)$ for the artifact waveform, we use the first three Fourier components of the minimum of the recorded signal $y(t)$ over all modulator periods. The rationale is that, because spikes are mostly positive, $y(t)$ is likely to have smaller amplitude during a modulation cycle which contains no spikes than during cycles in which a spike does occur. Nevertheless, the algorithm works even if spikes occur on every cycle, so long as they show sufficient time jitter.

At step n, the algorithm first estimates the noisy neural response $x^n(t)$ using the equation $x^n(t)=y(t)-a^{n-1}(t)$, where $a^{n-1}(t)$ is the artifact estimate at step $n-1$. Next, this noisy response is fit by a model neural response $\hat{x}^n(t)$, which is defined as

$$\hat{x}^n(t)=\sum_j g_j \cdot s(t-\tau_j), \tag{A1}$$

where j is spike number, $s(t)$ is the known spike waveform, while g_j and τ_j are the spike gains and spike times, respectively. The g_j's and τ_j's are determined by cross-correlating $x^n(t)$ with the spike waveform $s(t)$. Spikes are assumed to occur whenever the correlation coefficient is larger than a threshold, which was normally set at 20%, but was increased for particularly noisy recordings. Whenever a spike is detected, its gain g_j is determined from the local maximum of the correlation, while the location of the maximum is taken as spike time τ_j. Finally, the artifact waveform $a^n(t)$ at step n is estimated from the first three Fourier components of $y(t)-\hat{x}^n(t)$.

The algorithm is stopped whenever the difference between $a^n(t)$ and $a^{n-1}(t)$ is less than 0.1% of the spike height. The bottom row of Fig. 16 shows the estimated neural response $x(t)$ after eight iterations. The dotted line is the estimated artifact $a(t)$. In this case, the estimated spikes match those that would be picked by an experienced eye.

The algorithm was tested using synthetic neural recordings in response to 100-, 400-, and 800-Hz modulators. The synthetic recordings were composed of stereotyped spikes similar to those recorded from ANFs, and additive white noise.[3] For the 100- and 400-Hz stimuli, one synthetic spike occurred on every stimulus cycle with uniformly distributed time jitter. For 800 Hz, a spike occurred on every other cycle. Time jitter and noise amplitude were systematically varied from 50 to 400 μs and from 0% to 30% of spike amplitude, respectively. The algorithm was applied to synthetic spike trains comprising responses to 25 response cycles. For each noise and jitter condition, the procedure was repeated 100 times, and the maximum deviation between the estimated and the actual artifact waveforms was computed. For all conditions, this deviation was less than 10% of spike height. The largest deviations were seen for responses to the 800-Hz modulator.

The results of the algorithm were also monitored for real neural recordings, and a record was rejected if the estimated neural signal did not resemble spikes. This occurred in only about 2% of the records and only at the largest modulation depths.

APPENDIX B: METHOD FOR ESTIMATING MODULATION THRESHOLD

Single-fiber modulation threshold was defined in this study as the modulation depth at which either the average discharge rate or the synchronized rate during the sinusoidal modulation exceeded the discharge rate during a following unmodulated DPT segment on 75% of stimulus trials. The most direct method to determine threshold would rely on the percentage of *hit trials*, those in which the rate during the modulation exceeds the rate during the unmodulated DPT. To reliably estimate threshold by this method, the data would need to span the 75% criterion over the range of modulation depths tested. This condition was rarely met, particularly when measuring the very low synchronized rate thresholds for a 417-Hz modulator. A common pattern in the data was a hit percentage between 75% and 95% for the smallest measured modulation depth, and 100% for all larger modulation depths. Examples of this pattern are shown for three of the four fibers in the left panel of Fig. 17. In such cases, it is not possible to reliably fit a curve to the data so as to extrapolate the hit percentage to the 75% threshold criterion.

We therefore developed an alternative method for determining threshold which does not directly depend on hit percentages. The method seeks to utilize data points for which the hit percentage is 100%. To this end, we developed a measure, called *quartile rate difference*, which continues to grow with modulation depth even when the hit percentage reaches 100%. For each modulation depth m, and each trial i, let $r_m(m,i)$ denote the rate during the modulation (rate can be either synchronized or average discharge rate), and let $r_u(m,i)$ denote the rate during the corresponding unmodulated segment. The quantity $r_d(m,i)=r_m(m,i)-r_u(m,i)$ is the rate difference. The quartile rate difference $r_d^q(m)$ is the 25th percentile of $r_d(m,i)$ over all stimulus trials (modulated segments). By definition, at modulation threshold, the rate difference is positive on 75% of the trials (i.e., the rate during the modulation exceeds the rate during the corresponding unmodulated segment), and negative on the other 25% of the trials, so $r_d^q(m)$ must be zero. Thus, to determine threshold, we need to find the modulation depth m_{thr} such that $r_d^q(m_{thr})=0$.[4]

The middle panel of Fig. 17 shows the quartile rate difference as a function of modulation depth for the same data set as in the left panel. The threshold criterion of 0 is shown as a dotted line. Theoretically, as m approaches 0, $r_d^q(m)$ approaches a negative value r_0, which is the 25% percentile

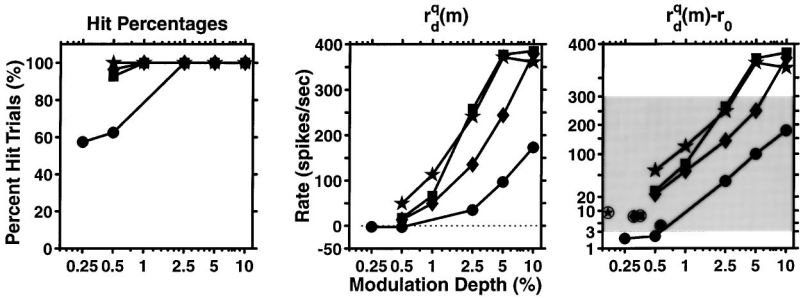


FIG. 17. Method for estimating modulation depth thresholds. The left panel shows the percentage of hit trials against depth of modulation at 417 Hz for four sustained DPT responders. For three of these fibers, the hit percentage was below 100% for only one modulation depth, preventing a direct determination of threshold. The middle panel plots the quartile rate difference $r_d^q(m)$ against modulation depth for the same four fibers. The dotted line indicates the threshold criterion. The right panel shows the same data plotted on an inverted Gaussian vertical scale defined by Eq. (B2). Gray shading indicates the region where the data grow approximately linearly. Circled symbols show the estimated thresholds.

of $r_u(m,i)-r_u(m,j)$ for all pairs of trials (i,j). This value was estimated directly from responses to the unmodulated DPT segments. For 55% of the synchronized rate records, $r_d^q(m)$ crossed zero, and threshold was estimated by linearly fitting $r_d^q(m)$ over a range of modulation depths spanning zero. However, for many records from the more sensitive sustained DPT responders (including three of the four shown in Fig. 17), $r_d^q(m)$ was above 0 even for the smallest modulation depth tested. We were still able to estimate threshold in these sensitive fibers by applying a nonlinear transformation to the data. Specifically, we found that $r_d^q(m)$ could usually be described by the expression

$$r_d^q(m)=r_0+g_\Gamma\cdot\Gamma(\alpha\cdot\ln(m)+T_0), \tag{B1}$$

where $\Gamma(x)$ is a cumulative Gaussian distribution, g_Γ is the maximum discharge rate, and α and T_0 are free parameters fit to the data. The maximum rate g_Γ was set to the modulation frequency for 104- and 417-Hz modulators (when spikes tend to occur on every stimulus cycle at high modulation depths), and to half the modulation frequency for 833 Hz (when spikes tend to occur on every other cycle).

This equation implies that

$$\Gamma^{-1}\left(\frac{r_d^q(m)-r_0}{g_\Gamma}\right) \tag{B2}$$

and $\ln(m)$ are linearly related. The right panel of Fig. 17 shows $r_d^q(m)-r_0$ against $\ln(m)$ for the same data set as in the other panels. The ordinate uses the inverse Gaussian scale defined by Eq. (B2). In the shaded region (between 3 and 300 spikes/s), the data grow nearly linearly, as predicted by the model. Outside the shaded region, the data deviate from linearity, perhaps because of difficulties in accurately estimating a cumulative Gaussian near its saturation. In any case, for most fibers, at least 3–4 data points are located inside the linear, shaded region. This sharply contrasts with the hit percentage (left panel), which, typically, provides only one useful point for estimating threshold.

This method assumes that, over a range of modulation depths, the rate of growth of discharge rate near threshold follows the same cumulative Gaussian function as the rate of growth well above threshold. This assumption was directly verified for all cases when the data did cross the threshold criterion so that extrapolation was unnecessary.

This threshold estimation procedure was applied to 45% of the data records. The free parameters α and T_0 were estimated using nonlinear least squares. For 34% of the total number of records, the fit was sufficiently close so that threshold could be determined by inverting Eq. (B1) for the fitted curve. However, for the remaining 11% of the records, threshold could not be determined by this method because of a poor fit, or lack of data. These records are not included in Fig. 10. Most of these (9% of total) were transient responders whose thresholds were higher than the highest modulation depth tested. For the other 2%, $r_d^q(m)$ was above criterion for all modulation depths; therefore, these fibers' thresholds are likely to be at least as low as those of the more sensitive fibers.

[1]To expedite data collection in the later experiments, we shortened the duration of the unmodulated segments to 300 ms when the preceding modulated segment had a modulation depth below 2.5%.

[2]We showed in the companion paper (Litvak *et al.*, 2003b) that nonexponential units respond to the DPT with higher discharge rates, so the observed correlation might simply reflect these rate differences. However, the synchronization index was not correlated with discharge rate during the unmodulated DPT (correlation coefficient—0.02, $p=0.54$, permutation test).

[3]Mathematical analysis shows that, with our choice of an initial guess, the output of the algorithm is independent of the artifact. Thus, no artifact was added to the model responses.

[4]This problem is ill-defined for transient responders, because $r_d^q(m)$ is zero for small m. To resolve this ambiguity, a small random number (at most corresponding to half a spike) was arbitrarily added to each $r_m(m,i)$ and $r_u(m,i)$ that equaled 0. Because thresholds of transient responders are typically well above the smallest modulation depth tested, this addition has only a minor effect on the estimate of modulation threshold.

Barlow, H. B., and Levick, W. R. (**1969**). "Three factors limiting the reliable detection of light by the retinal ganglion cells of the cat," J. Physiol. (London) **200**, 1–24.

Bruce, I. C., Irlicht, L. S., White, M. W., O'Leary, S. J., Dynes, S., Javel, E., and Clark, G. M. (**1999**). "A stochastic model of the electrically stimulated auditory nerve: pulse-train response," IEEE Trans. Biomed. Eng. **46**, 630–637.

Chaput, M., and Holley, A. (**1979**). "Spontaneous activity of olfactory bulb neurons in awake rabbits, with some observations on the effects of pentobarbital anaesthesia," J. Physiol. (Paris) **75**, 939–948.

Collins, J. J., Chow, C. C., and Imhoff, T. T. (**1995**). "Stochastic resonance without tuning," Nature (London) **376**, 236–238.

Dynes, S. B. C. (**1995**). "Discharge Characteristics of Auditory Nerve Fibers for Pulsatile Electrical Stimuli," Ph.D. thesis, MIT, Cambridge, MA.

Dynes, S. B. C., and Delgutte, B. (**1992**). "Phase locking of auditory-nerve

discharges to sinusoidal electric stimulation of the cochlea,'' Hear. Res. **58**, 79–90.
Efron, B., and Tibshirani, R. (**1993**). *An Introduction to the Bootstrap* (Chapman and Hall, New York).
Goldberg, J. M., and Fernandez, C. (**1971**). ''Physiology of peripheral neurons innervating semicircular canals of the squirrel monkey. I. Resting discharge and response to constant angular accelerations,'' J. Neurophysiol. **34**, 635–660.
Good, P. (**2000**). *Permutation Tests. A Practical Guide to Resampling Methods for Testing Hypotheses* (Springer, New York).
Hartmann, R., Topp, G., and Klinke, R. (**1984**). ''Discharge patterns of cat primary auditory fibers with electrical stimulation of the cochlea,'' Hear. Res. **13**, 47–62.
Hartmann, R., Topp, G., and Klinke, R. (**1990**). ''Response characteristics of nerve fibers to patterned electrical stimulation,'' in *Cochlear Implants. Models of the Electrically Stimulated Ear*, edited by J. M. Miller and F. A. Spelman (Springer-Verlag, New York), pp. 135–160.
Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001**). ''Rate and timing cues associated with the cochlear amplifier: Level discrimination based on monaural cross-frequency coincidence detection,'' J. Acoust. Soc. Am. **110**, 2065–2084.
Hill, A. V. (**1936**). ''Excitation and accommodation in nerve,'' Proc. R. Soc. London, Ser. B **119**, 305–355.
Javel, E. (**1990**). ''Acoustic and Electrical Encoding of Temporal Information,'' in *Cochlear Implants. Models of the Electrically Stimulated Ear*, edited by J. M. Miller and F. A. Spelman (Springer-Verlag, New York), pp. 247–295.
Javel, E., and Shepherd, R. K. (**2000**). ''Electrical stimulation of the auditory nerve. III. Response initiation sites and temporal fine structure,'' Hear. Res. **140**, 45–76.
Javel, E., Tong, Y. C., Shepherd, R. K., and Clark, G. M. (**1987**). ''Responses of cat auditory nerve fibers to biphasic electrical current pulses,'' Ann. Otol. Rhinol. Laryngol. Suppl. **128**, 26–30.
Johnson, D. H. (**1980**). ''The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones,'' J. Acoust. Soc. Am. **68**, 1115–1122.
Kiang, N. Y.-S., and Moxon, E. C. (**1972**). ''Physiological considerations in artificial stimulation of the inner ear,'' Ann. Otol. Rhinol. Laryngol. **81**, 714–730.
Kiang, N. Y. S., Watanabe, T., Thomas E. C., and Clark, L. F. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT, Cambridge, MA).
Kuffler, S. (**1953**). ''Discharge patterns and functional organization of the mammalian retina,'' J. Neurophysiol. **16**, 37–68.
Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Liberman, M. C. (**1978**). ''Auditory-nerve response from cats raised in a low-noise chamber,'' J. Acoust. Soc. Am. **63**, 442–455.
Litvak, L. M., Delgutte, B., and Eddington, D. K. (**2001**). ''Auditory nerve fiber responses to electric stimulation: modulated and unmodulated pulse trains,'' J. Acoust. Soc. Am. **110**, 368–379.
Litvak, L. M., Delgutte, B., and Eddington, D. K. (**2003a**). ''Improved neural representation of vowels in electric stimulation using desynchronizing pulse trains,'' J. Acoust. Soc. Am. **114**, 2099–2111.
Litvak, L. M., Smith, Z. M., Delgutte, B., and Eddington, D. K. (**2003b**). ''Desynchronization of electrically evoked auditory-nerve activity by high-frequency pulse trains of long duration,'' J. Acoust. Soc. Am. **114**, 2066–2078.
McKinney, M. F., and Delgutte, B. (**1999**). ''A possible neurophysiological basis of the octave enlargement effect,'' J. Acoust. Soc. Am. **106**, 2679–2692.
Moxon, E. C. (**1967**). ''Electric stimulation of the cat's cochlea: a study of discharge rates in single auditory nerve fibers,'' M.S. thesis, MIT, Cambridge, MA.
Parkins, C. W. (**1989**). ''Temporal response patterns of auditory nerve fibers to electrical stimulation in deafened squirrel monkeys,'' Hear. Res. **41**, 137–169.
Parkins, C. W., and Colombo, J. (**1987**). ''Auditory nerve single neuron thresholds to electrical stimulation from scala tympani electrodes,'' Hear. Res. **31**, 267–285.
Pfaffmann, C. (**1955**). ''Gustatory nerve impulses in rat, cat and rabbit,'' J. Neurophysiol. **18**, 429–440.
Rodieck, R. W. (**1967**). ''Maintained activity of cat retinal ganglion cells,'' J. Neurophysiol. **30**, 1043–1071.
Rose, J. E., Brugge, J. R., Anderson, D. J., and Hind, J. E. (**1967**). ''Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey,'' J. Neurophysiol. **30**, 769–793.
Rosen, S. (**1992**). ''Temporal information in speech: acoustic, auditory and linguistic aspects,'' Philos. Trans. R. Soc. Lond. B Biol. Sci. **336**, 367–373.
Rospars, J. P., Lansky, P., Vaillant, J., Duchamp-Viret, P., and Duchamp, A. (**1994**). ''Spontaneous activity of first- and second-order neurons in the olfactory system,'' Brain Res. **662**, 31–44.
Rubinstein, J. T., Abbas, P. J., and Miller, C. A. (**1998**). ''The neurophysiological effects of simulated auditory prosthesis simulation,'' NIH Q. Prog. Rep. No 8, University of Iowa.
Rubinstein, J. T., Miller, C. A., Abbas, P. J., and Wilson, B. S. (**1999a**). ''Emulating physiologic firing patterns of auditory neurons with electrical stimulation,'' Abstr. Assoc. Res. Otolaryngol. **22**, 8.
Rubinstein, J. T., Wilson, B. S., Finley, C. C., and Abbas, P. J. (**1999b**). ''Pseudospontaneous activity: stochastic independence of auditory nerve fibers with electrical stimulation,'' Hear. Res. **127**, 108–118.
Sachs, M. B., and Abbas, P. J. (**1974**). ''Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli,'' J. Acoust. Soc. Am. **56**, 1835–1847.
Schalk, T. B., and Sachs, M. B. (**1980**). ''Nonlinearities in auditory-nerve fiber responses to bandlimited noise,'' J. Acoust. Soc. Am. **67**, 903–913.
Schneidman, E., Freedman, B., and Segey, I. (**1998**). ''Ton channel stochasticity may be critical in determining the reliability and precision of spike timing,'' Neural Comput. **10**, 1679–1703.
Seligman, P., and McDermott, H. (**1995**). ''Architecture of the Spectra 22 Speech Processor,'' Ann. Otol. Rhinol. Laryngol. **104** (Suppl. 166), 139–141.
Shamma, S., and Klein, D. (**2000**). ''The case of the missing pitch templates: how harmonic templates emerge in the early auditory system,'' J. Acoust. Soc. Am. **107**, 2631–2644.
Shannon, R. V. (**1992**). ''Temporal modulation transfer functions in patients with cochlear implants,'' J. Acoust. Soc. Am. **91**, 2156–2164.
Shepherd, R. K., and Javel, E. (**1997**). ''Electrical stimulation of the auditory nerve. I. Correlation of physiological responses with cochlear status.'' Hearing Res. **108**, 112–144.
Siebert, W. (**1965**). ''Some implications of the stochastic behavior of auditory neurons,'' Kybernetik **2**, 206–215.
Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). ''Chimaeric sounds reveal dichotomies in auditory perception,'' Nature (London) **416**, 87–90.
Tong, Y. C., Clark, G. M., Seligman, P. M., and Patrick, J. F. (**1980**). ''Speech processing for a multiple-electrode cochlear implant hearing prosthesis,'' J. Acoust. Soc. Am. **68**, 1897–1899.
van den Honert, C., and Stypulkowski, P. H. (**1987**). ''Temporal response patterns of single auditory nerve fibers elicited by periodic electrical stimuli,'' Hear. Res. **29**, 207–222.
Walsh, B. T., Miller, J. B., Gacek, R. R., and Kiang, N. Y. S. (**1972**). ''Spontaneous activity in the eighth cranial nerve of the cat,'' Int. J. Neurosci. **3**, 221–236.
Werner, G., and Mountcastle, V. B. (**1965**). ''Neural activity in mechanoreceptive cutaneous aferents: Stimulus-response relations, Weber functions, and information transmission,'' J. Neurophysiol. **28**, 359–397.
Westerman, L. A., and Smith, R. L. (**1984**). ''Rapid and short-term adaptation in auditory nerve responses,'' Hear. Res. **15**, 249–260.
Wiesenfeld, K., and Moss, F. (**1995**). ''Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDs,'' Nature (London) **373**, 33–36.
Wilson, B., Finley, C. C., Lawson, D. T., and Zebri, M. (**1997**). ''Temporal representations with cochlear implants,'' Am. J. Otol. **18**, Suppl. 6, S30–S34.
Wilson, B., Lawson, D., Zerbi, M., Finley, C., and van den Honert, C. (**1998**). *Speech Processors for Auditory Prostheses*, Final NIH Report, Research Triangle Institute.
Xu, S. A., Shepherd, R. K., Chen, Y., and Clark, G. M. (**1993**). ''Profound hearing loss in the cat following the single co-administration of kanamycin and ethacrynic acid,'' Hear. Res. **70**, 205–215.
Yu, X., and Lewis, E. R. (**1989**). ''Studies with spike initiators: Linearization by noise allows continuous signal modulation in neural networks,'' IEEE Trans. Biomed. Eng. **36**, 36–43.

Improved neural representation of vowels in electric stimulation using desynchronizing pulse trains[a)]

Leonid Litvak[b)]
Eaton-Peabody Laboratory and Cochlear Implant Research Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Speech and Hearing Bioscience and Technology Program, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

Bertrand Delgutte
Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

Donald Eddington
Cochlear Implant Research Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Research Laboratory of Electronics and Neural Prosthesis Research Center, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

(Received 13 March 2003; accepted for publication 4 August 2003)

Current cochlear implant processors poorly represent sound waveforms in the temporal discharge patterns of auditory-nerve fibers (ANFs). A previous study [Litvak *et al.*, J. Acoust. Soc. Am. **114**, 2079–2098 (2003)] showed that the temporal representation of sinusoidal stimuli can be improved in a majority of ANFs by encoding the stimuli as small modulations of a sustained, high-rate (5 kpps), desynchronizing pulse train (DPT). Here, these findings are extended to more complex stimuli by recording ANF responses to pulse trains modulated by bandpass filtered vowels. Responses to vowel modulators depended strongly on the discharge pattern evoked by the unmodulated DPT. ANFs that gave sustained responses to the DPT had period histograms that resembled the modulator waveform for low ($<5\%$) modulation depths. Spectra of period histograms contained peaks near the formant frequencies. In contrast, ANFs that gave a transient (<1 min) response to the DPT poorly represented the formant frequencies. A model incorporating a linear modulation filter, a noisy threshold, and neural refractoriness predicts the shapes of period histograms for both types of fibers. These results suggest that a DPT-enhanced strategy may achieve good representation of the stimulus fine structure in the temporal discharge patterns of ANFs for frequencies up to 1000 Hz. It remains to be seen whether these temporal discharge patterns can be utilized by cochlear implant subjects.
[DOI: 10.1121/1.1612494]

PACS numbers: 43.64.Nf, 43.64.Pg [WPS]

I. INTRODUCTION

In a healthy ear, the fine-time structure of acoustic stimuli is represented in the temporal discharge patterns of auditory-nerve fibers (ANFs) for frequencies up to 5 kHz (Johnson, 1980). For a complex sound, the auditory nerve is able to represent several frequency components simultaneously in the temporal discharge patterns. For example, with a sustained vowel stimulus, both the information in the envelope and the information in the fine-time structure (e.g., formant frequencies) are represented in the temporal patterns (Delgutte, 1984; Miller and Sachs, 1984; Young and Sachs, 1979). Future cochlear implant devices may attempt to reproduce these temporal discharge patterns in the deaf cochlea. Doing so may provide more information to the implantees about the incoming sounds than is available through existing implants, which often provide only envelope information.

Although it would be conceptually straightforward to include fine-time structure information in the electric stimuli generated by a cochlear implant processor, and, in fact, some processors such as those using compressed analog strategies already do so, physiological studies suggest that the fine structure will be poorly represented in the temporal discharge patterns of ANFs for any of these schemes. Responses to vowels delivered electrically occur primarily at the largest peak of the vowel waveform over a wide range of vowel amplitudes (van den Honert and Stypulkowski, 1987; Knauth *et al.*, 1994). Thus, only the pitch-related envelope, and not the formant-related fine-time structure, is coded reliably in ANF temporal discharge patterns. This is true even if vowels are amplitude compressed to the psychophysical dynamic range (Knauth *et al.*, 1994). This poor temporal representation reflects basic characteristics of electric stimulation, including a small dynamic range, very high discharge rates amplifying refractory effects, and excessive synchrony of responses across fibers.

[a)]Portions of this work were presented as a poster at the ARO Midwinter Meeting in St. Petersburg Beach, FL, 2001 and the CIAP Conference in Monterey, CA, 2001.
[b)]Electronic mail: leonid.litvak@advancedbionics.com

Rubinstein *et al.* (1999) proposed that the temporal coding of the waveforms of electric stimuli might be improved by introducing an ongoing, high-frequency, desynchronizing pulse train (DPT) over the stimulus delivered by the processor. The purpose of a DPT is to desynchronize the spike activity across auditory nerve fibers, much as spontaneous activity does in a normal ear, thereby expanding the dynamic range and reducing temporal distortions due to neural refractoriness. In a companion paper (Litvak *et al.*, 2003a), we recorded ANF responses to sinusoidal modulations of long-duration (10-min), 5-kpps DPTs. We found that, for small modulation depths (<5%), temporal discharge patterns of some fibers to sinusoidal modulators resembled responses to pure tones in a healthy ear. Specifically, the spike discharges of fibers that gave sustained responses to the unmodulated DPT were stochastic, occurring every one, two or more cycles, and provided a more accurate representation of the sinusoidal stimulus waveform than responses to either electric sinusoids (Hartmann *et al.*, 1984; van den Honert and Stypulkowski, 1987) or modulated pulse trains (Litvak *et al.*, 2001) presented without a DPT. In addition, both the average discharge rates and the synchronization indices grew monotonically with modulation depth, together providing a total dynamic range of about 20 dB, comparable to that found for acoustic stimulation with pure tones. The relatively large dynamic range to sinusoidal modulations suggests that complex waveforms having peaks of differing heights might be faithfully represented in the temporal discharge patterns of sustained responders with a DPT. To test this hypothesis, we recorded responses of ANFs to DPTs modulated by filtered vowels. We focused on vowels because of the perceptual importance of these stimuli, and because their representation in a normally functioning ear has been extensively characterized (Delgutte, 1984; Miller and Sachs, 1984; Young and Sachs, 1979).

In the companion paper (Litvak *et al.*, 2003a), we proposed a simple stochastic threshold model (STM) that describes ANF responses in the presence of a DPT. For sinusoidal modulations of the DPT, the model successfully predicted the dependence of threshold and dynamic range on the discharge rate evoked by the unmodulated DPT, which we refer to as ''pseudo-spontaneous rate.'' We suggested an elaboration of the stochastic threshold model that might also predict responses to complex modulators by incorporating a linear modulation filter. The modulation filter is introduced to account for differences in sensitivity across modulation frequency. To test this idea, we estimated the modulation filter from the responses to vowels, and compared the ANF responses to vowel modulators with predictions of the modified STM.

As in the companion paper on sinusoidal stimuli (Litvak *et al.*, 2003a), the vowels were encoded as small modulations of a high-rate pulse train [Fig. 1(a)]. We assume that neural responses to a high-rate pulse train with a low modulation depth are similar to those elicited by the superposition of an ongoing, unmodulated DPT and a highly modulated pulse train (as might be produced by a CIS processor). This assumption is necessarily true if the same pulse carrier is used for both the DPT and the CIS stimulus. It may hold more generally if the time constant of the neural membrane is large compared to the intervals between DPT pulses. In either case, this scheme maps the amplitude of the CIS-type stimulus onto the modulation depth of the DPT. This scheme differs from a standard, high-rate CIS strategy in that (1) the unmodulated DPT is large and continuously present; (2) the modulation depths are always small (<15%); and (3) there is no envelope detector and no half-wave rectification of the input signal, so both positive and negative deflections of the acoustic waveforms are represented in the modulated pulse train.

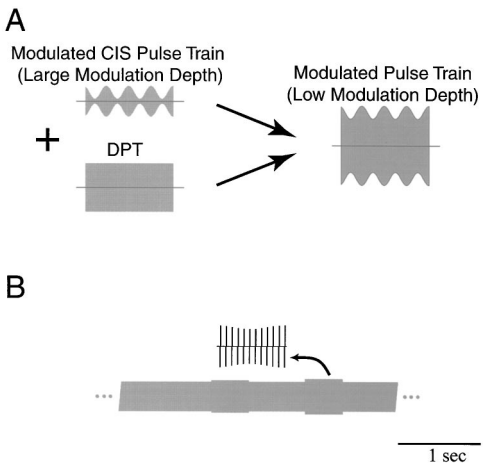


FIG. 1. (a) Implementation of the DPT-enhanced stimulation strategy. The left panel shows two pulse train stimuli that might be generated by a cochlear implant processor: one, strongly modulated, is the CIS signal, and the other, unmodulated, is the DPT. We assume that the DPT-enhanced stimulus can be modeled by a carrier of the same frequency as the DPT that is more weakly modulated than the original CIS stimulus (right). (b) Schematic representation of the electric stimuli used in this study. The stimulus consists of a sustained DPT, which is modulated every second for 400 ms. The modulation waveform and/or modulation depth are changed on each successive modulated segment. The entire cycle of modulation waveforms and modulation depths repeats for up to 10 min.

II. METHODS

The animal preparation, electrical stimulation, and recording methods have been described in a companion paper (Litvak *et al.*, 2003b). Briefly, cats were anesthetized with dial in urethane (75 mg/kg), then deafened by coadministration of kanamycin (subcutaneous, 300 mg/kg) and ethacrinic acid (intravenous, 25 mg/kg) (Xu *et al.*, 1993). Two intracochlear stimulating electrodes (400 mm Pt/Ir balls) were inserted into the cochlea through the round window. One electrode inserted approximately 8 mm was used as the stimulating electrode. The return electrode was inserted just inside the round window. To minimize the possible effects of residual hearing in our preparation, we only report responses from neurons having no spontaneous activity in absence of a DPT [see Litvak *et al.* (2003b) for details].

Standard techniques were used to expose the auditory nerve via a dorsal approach and record from single fibers using glass micropipettes filled with 3M KCl (Kiang *et al.*, 1965). For small modulation depths, most of the stimulus artifact could be removed online using a digital signal pro-

cessor implementing a moving average filter whose length matched the 0.2-ms pulse period. Neural responses were also recorded digitally at a sampling rate of 20 kHz for off-line analysis. To remove the artifact from these records, we used an extension of the iterative method described in the companion paper for sinusoidal modulators (Litvak *et al.*, 2003a). This modified method is described in Appendix A.

A. Stimuli

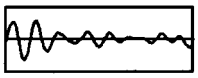
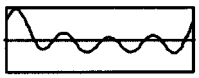
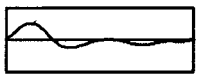
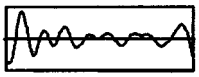
We conducted a neural population study by investigating responses to a single DPT level in each animal. The DPT level was set at 8–10 dB above the electric compound action potential threshold, so that a large fraction of the ANF population would respond to the DPT (Litvak *et al.*, 2003b).

The DPT was a 5-kpps pulse train composed of biphasic (cathodic/anodic) pulses (25 μs per phase). In order to acquire responses to both the unmodulated DPT and to modulations of the DPT for several modulation waveforms and depths, the stimulus consisted of alternating modulated (400 ms) and unmodulated (600 ms) segments [Fig. 1(b)].[1] Modulation depth and/or modulation waveform was changed on each successive modulated segment. Modulation waveforms were derived from synthetic vowels as described below. Modulation depth ranged from 0.5% to 15%. Modulation was applied such that the mean amplitude of the carrier pulses was the same during modulated and unmodulated segments. The entire sequence of modulated and unmodulated segments had a duration of 13–16 s, and was repeated continuously either for 10 min or until contact with a fiber was lost.

The modulation waveforms were derived from synthetic vowels having a 100-Hz fundamental frequency created by Assmann and Summerfield (1990) using a series formant synthesizer (Klatt, 1980). A single 10-ms period of each vowel was extracted, and digitally resampled at 40 kHz. From each vowel, two modulation waveforms were produced by filtering the vowel period into two frequency bands (0 to 500 Hz, and 500 to 1000 Hz). This filtering roughly mimics that performed by multi-channel cochlear implant processors. Although some of the vowels used by Assman and Summerfield (1990) had second formants as high as 2.2 kHz, we did not analyze formants above 1000 Hz because higher frequencies cannot be represented with a 5-kpps carrier pulse rate without significant aliasing. Filtering was accomplished by computing the discrete Fourier transform of the vowel period, zeroing the harmonics outside of the pass band, and then computing the inverse Fourier transform. The filtered vowel periods were concatenated to produce 400-ms modulators. Finally, only band-vowel combinations that contained at least one formant were investigated. Table I gives the characteristics of the four vowel-band combinations used in this study.

Two definitions of modulation depth were used to specify the amplitude of the vowel stimuli. In early experiments, modulation depth was defined based on the peak amplitude of the modulation waveform. Specifically, if $e(t)$ represents a filtered vowel waveform, the DPT envelope was defined as $A\cdot(1+m\cdot e(t)/\max(|e(t)|))$, where m is the modulation depth and A is the amplitude of the unmodulated DPT.

TABLE I. Vowel stimuli used in this study. The fundamental frequency was 100 Hz. Band 1 refers to the frequency range 0–500 Hz, band 2 to 500–1000 Hz.

Sound	Band	Formants	Waveform
/ɔ/	2	F1 = 750 Hz	
/ɜ/	1	F1 = 450 Hz	
/u/	1	F1 = 250 Hz	
/u/	2	F2 = 850 Hz	

With this definition, the rms amplitude of the envelope can differ by 1.5 dB between vowels at a given modulation depth. In later experiments, we redefined modulation depth so as to equalize the rms amplitudes of the modulation envelopes across vowels. Specifically, for the first vowel in the sequence (the low-pass /ɔ/), modulation depth was set as above. The other vowels were then normalized so that, for a given modulation depth, the rms amplitude of each modulation waveform matched that of the first vowel. Specifically, the modulation envelope was defined as $A\cdot(1+m\cdot e(t)\cdot K/\mathrm{rms}(e(t)))$, where K is the ratio of rms to peak amplitude for the lowpass /ɔ/.

The differences between the two definitions of modulation depth are small compared to the variability in ANF responses for a given modulation depth. For example, for a 10% modulation depth based on rms amplitude, the actual modulation depths for the other vowels (under the original definition) would be 10.7%, 11.9%, and 11.1%. In fact, analysis of responses to vowel modulators revealed no consistent dependence of discharge rate on the definition of modulation depth. For the purpose of this paper, therefore, we pool the data from the two definitions together when describing population responses.

B. Data analysis

Responses collected during the unmodulated DPT segments were used to classify ANFs into two groups as in the companion paper (Litvak *et al.*, 2003b). Some fibers exhibited only a transient response to the DPT, and adapted to a pseudo-spontaneous discharge rate near zero (<5 spikes/s) after a minute of DPT stimulation. We refer to these fibers as ''transient DPT responders.'' Fibers that responded throughout the unmodulated segments with a pseudo-spontaneous rate above 5 spikes/s are referred to as ''sustained responders.''

In a normally functioning ear, each ANF represents the frequency components of the vowel stimulus close to its characteristic frequency in its temporal discharge patterns (Delgutte and Kiang, 1984; Young and Sachs, 1979). To determine whether the fine-time structure of vowels is also coded in responses to a modulated DPT, period histograms were computed locked to the 10-ms period of the vowel. To

eliminate the effect of possible phase locking to the 5-kpps carrier, the histograms were computed using 0.2-ms bins. The magnitude and phase of the response to each vowel harmonic were determined from the discrete Fourier transform of the period histogram. If the period histogram is scaled to represent the instantaneous discharge rate in spikes/s, then each Fourier component also has units of spikes/s, and represents the synchronized rate, i.e., the average discharge rate multiplied by the synchronization index to that component (Young and Sachs, 1979).

To formally characterize the transformation between the modulation waveform and the neural response in the frequency domain, we estimated neural modulation transfer functions (MTFs). A MTF expresses, as a function of frequency, the complex ratio (magnitude and phase) of the modulation in the neural response to the modulation in the stimulus. Traditionally, MTFs are estimated from responses to sinusoidal modulators (e.g., Møller, 1974). At frequency f_m, the MTF magnitude is twice the synchronization index of the response to a sinusoidal modulator at f_m divided by the modulation depth of the stimulus, and the MTF phase is the mean phase of the response minus the phase of the stimulus. Alternatively, if the system behaves linearly, the MTF can be estimated directly from the Fourier transform of the period histogram for a complex stimulus (Møller, 1973; Møller and Rees, 1986). MTFs were estimated from responses to vowel modulators by computing, for each vowel harmonic, twice the ratio of the synchronization index at the harmonic frequency to the same component of the vowel's Fourier transform (e.g., Young and Sachs, 1979). In the case of sinusoidal modulators, this definition is identical to the traditional one. We used bootstrap techniques to derive confidence intervals for the MTF of each fiber (Efron and Tibshirani, 1993). These techniques allow confidence intervals to be determined for arbitrary functions of the data without requiring specific assumptions about the probability distributions of the data. Specifically, we resampled the spike times with replacement to generate 500 bootstrap replications of the spike train. The MTF was computed for each bootstrap spike train, and confidence intervals were estimated by determining, for each modulation frequency, a range that includes 99% of the bootstrap MTF values. To test the linearity assumption, we compared MTFs computed from responses of the same fiber to different vowel modulators. For a linear system, the MTF should be independent of the stimulus. We also compared the average MTF estimated from vowel responses to the estimates at three modulation frequencies (104, 417, and 833 Hz) based on responses to sinusoidal modulators described in the companion paper (Litvak *et al.*, 2003a).

C. Stochastic threshold model

In a companion paper (Litvak *et al.*, 2003a), we introduced a simple functional stochastic threshold model (STM) of ANF responses in the presence of a DPT (Fig. 2). The model takes as input the modulator waveform $m(t)$, which is compared to a noisy threshold. A spike is produced by the model whenever $m(t)$ crosses the threshold. The threshold is the sum of a deterministic term, which includes both an absolute and a relative refractory period, and a Gaussian noise term. In the absence of any input, the STM has a pseudo-spontaneous discharge rate which is determined by the ratio of the noise standard deviation σ to the resting threshold θ_0. We showed that this model successfully predicts the dependence of threshold and dynamic range of responses to sinusoidal modulations on the pseudo-spontaneous discharge rate at a given modulation frequency. However, the model failed to account for the dependence of ANF sensitivity on modulation frequency. To account for this dependence, we introduce a linear modulation filter as a first stage of the model (Fig. 2). We will show that this filter can be estimated from ANF responses to complex modulators. Specifically, we will use neural MTFs derived from vowel responses to estimate the model's modulation filter.

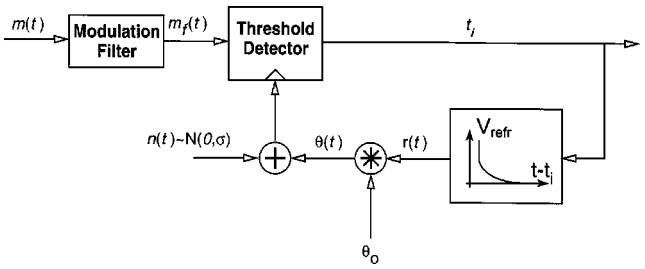


FIG. 2. Block diagram of the stochastic threshold model for predicting ANF responses to arbitrary modulation waveforms applied to the DPT. The model takes as input the modulation waveform $m(t)$, which is passed through a modulation filter. It produces a spike whenever the filtered input crosses a noisy threshold. The threshold is the sum of a Gaussian noise term $n(t)$ and a deterministic term $\theta(t)$ which depends on the time since the previous spike. The model outputs the set of spike times $\{t_i\}$. Free parameters of the model are the modulation filter, the resting threshold θ_0, and the noise standard deviation σ.

Because the output of the STM is in the form of a spike train, the STM has an MTF which can be estimated using the same methods as for actual neural data. A necessary condition for using the neural MTF as an estimate of the STM's modulation filter is that the model's own MTF matches the underlying filter. We tested this assumption by computing MTFs from spike trains produced by the model in response to vowel stimuli. When the model's pseudo-spontaneous rate was set to 50 spikes/s, the model's empirical MTF differed in shape from the underlying modulation filter. However, if the model's pseudo-spontaneous rate was increased to 300 spikes/s, there was good agreement between the model MTF and the underlying modulation filter (not shown). For this reason, we used only data from ANFs with high pseudo-spontaneous rates ($>$130 spikes/s) in estimating the STM's modulation filter. Although such fibers form only a small fraction of our sample, we will show that the STM filter derived from responses of these fibers can also predict responses of fibers with lower pseudo-spontaneous rates.

After setting the STM's modulation filter, we computed model responses to vowel modulators with different modulation depths, and compared them to ANF responses to a modulated DPT. The modulation filter was applied directly in the frequency domain for each harmonic of the stimulus; therefore the model in its present form can only predict

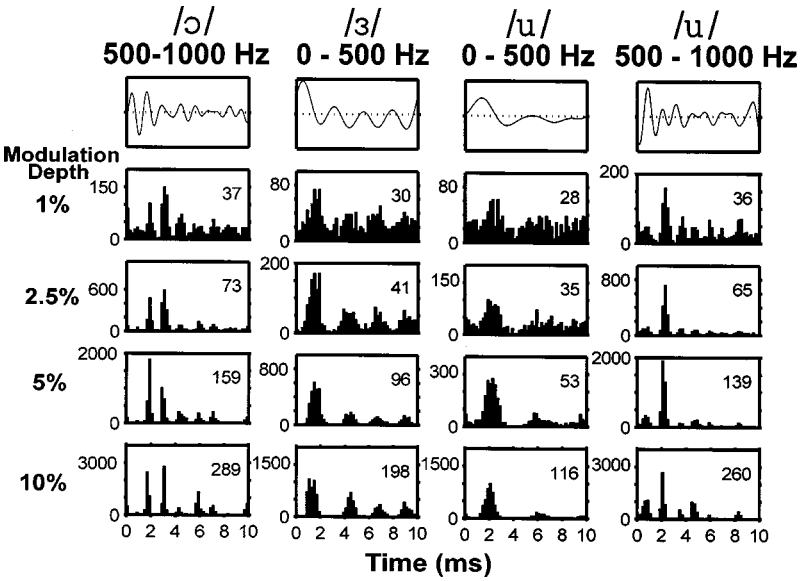


FIG. 3. Period histograms of responses to vowel modulators for a sustained DPT responder (pseudo-spontaneous rate: 28 spikes/s). Top row: waveforms of the four filtered vowel modulators. Lower four rows: Period histograms of responses to each vowel for four modulation depths. The ordinate in each panel is in spikes/s. The numbers in each panel indicate the average discharge rate during the vowel modulator in spikes/s.

responses to periodic stimuli with a 100-Hz fundamental. The model was simulated using 0.2-ms time steps.

III. RESULTS

Our results are based on 36 responses recorded from 31 ANFs in four cats to DPTs modulated by vowels. Each record was composed of responses to four filtered vowels at four or five modulation depths ranging from 1% to 15%. In 17 of these records, the fiber responded to the DPT at a rate exceeding 5 spikes/s throughout the stimulus duration; in the other 19 records, the fibers stopped responding to the unmodulated DPT after 1–2 min of stimulation, although it still responded to modulations of the DPT. The percentage of transient responses in this data set (53%) did not significantly differ ($p=0.26$, binomial exact test) from the 46% found in a larger data set from ten animals (Litvak *et al.*, 2003b).

A. Temporal discharge patterns

Figure 3 shows period histograms computed from the responses of one fiber to each of the four filtered-vowel modulators at four different modulation depths. During the unmodulated DPT segments, this sustained responder had a pseudo-spontaneous discharge rate of 28 spikes/s. For all modulation depths, the period histograms broadly resemble the modulation waveforms. In particular, modes in the period histograms match peaks of the modulation waveform. The intervals between peaks in the modulation waveforms are related to the vowel formant frequencies. This result suggests that, with a DPT, the responses to modulated vowels code the vowel formant frequencies.

Although the peaks of the modulation waveform are represented for all modulation depths, the best resemblance between the modulation waveform and the period histograms occurs for modulation depths below 5%. For example, the time between the two largest peaks in the waveform of the filtered /ɔ/ is 1.2 ms. This interval matches the distance between the two largest modes of the corresponding period histogram for modulation depths between 1% and 5%, but not for 10% where the interval between modes increases to 1.4 ms. Similar distortions can be seen in responses to the

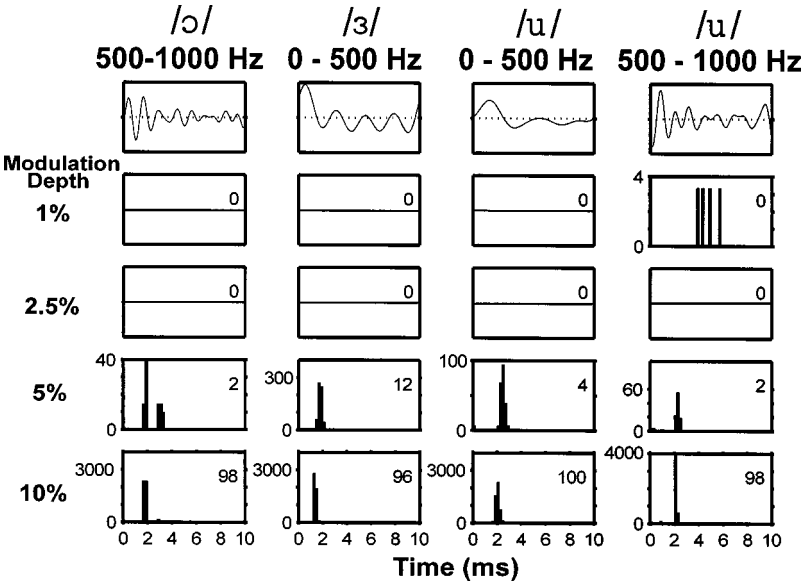


FIG. 4. Period histograms of responses to the vowel modulators for a transient DPT responder (pseudo-spontaneous rate: 0 spike/s). Same format as in Fig. 3.

vowel /u/. In addition, some of the smaller peaks in the modulation waveform are only represented at the lower modulation depths.

In contrast to sustained DPT responders, transient responders poorly represented the complex modulation waveforms. Figure 4 shows the response of a fiber that responded only transiently to the unmodulated DPT. The fiber did not respond significantly to the vowel modulator for modulation depths below 5%. This observation is consistent with the lower sensitivity of transient DPT responders to sinusoidal modulations (Litvak *et al.*, 2003a). At modulation depths of 5% and 10%, the fiber did respond with high synchrony to the modulator, but spikes occurred almost exclusively in response to the largest peak in the modulation waveform. One exception is the response to the second largest peak for the filtered vowel /ɔ/ at 5% modulation. These responses resemble responses to vowels presented electrically without a DPT (van den Honert and Stypulkowski, 1987; Knauth *et al.*, 1994).

To quantify the resemblance between the modulation waveform and the period histogram of ANF responses, we computed the correlation coefficient between the two. To account for possible neural delay, the modulation waveform was shifted successively in steps of 0.2 ms (up to 2 ms), until the largest stimulus-response correlation was obtained. Figure 5 shows the maximum stimulus–response correlation against pseudo-spontaneous discharge rate for modulation depths of 2.5% (top) and 10% (bottom). At 2.5% modulation, all but two responses from sustained DPT responders (pseudo-spontaneous rates above 5 spikes/s) had correlations above 0.6, with a median correlation of 0.79, implying excellent representation of the vowel waveform. In contrast, transient DPT responders tended to have lower correlations (median 0.65). The difference in correlation between transient and sustained responders is significant [$p<0.001$, permutation test (Good, 2000, p. 37)]. At 10% modulation depth, the correlation for both sustained and transient responders was lower (median 0.55) than at 2.5% depth, and the difference between the two depths is highly significant ($p<0.001$, permutation test). At this higher modulation depth, correlations for transient and sustained responders were more similar. Thus, the representation of the vowel waveform in the ANF temporal discharge patterns is better at the 2.5% modulation depth than at the 10% modulation depth, and, at the lower modulation depths, better in sustained responders than in transient responders.

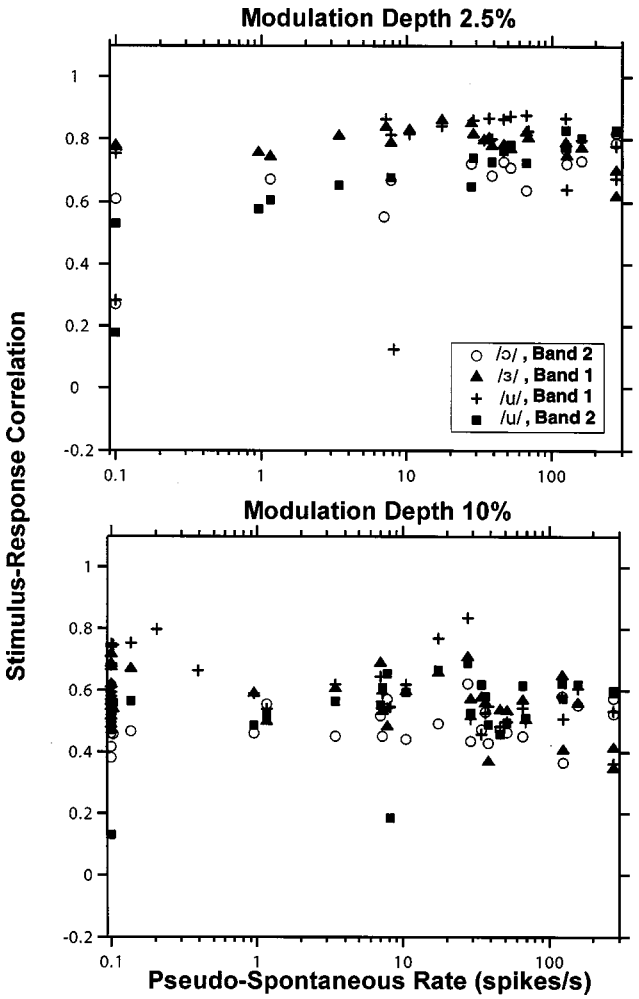


FIG. 5. Maximum correlation coefficient between the period histogram of ANF responses and the vowel waveform as a function of pseudo-spontaneous rate for 2.5% (top) and 10% (bottom) modulation depths. Symbols code the correlation coefficients for each vowel. Data are only included if the response contains at least ten spikes and the bootstrap 90% confidence intervals for the correlation coefficient are below 0.1.

B. Representation of formant frequencies

The high correlations of responses to vowel modulators with the stimulus waveform at low modulation depths strongly suggest, but do not directly show, that the spectral peaks of the modulation waveform are well represented in the temporal discharge patterns. In the case of vowels, these spectral peaks correspond to formant frequencies, which are important cues for vowel identification (Peterson and Barney, 1952). To determine whether the spectral peaks of the modulation waveform are accurately represented in the ANF temporal patterns, we computed the discrete Fourier transform of the period histogram (Young and Sachs, 1979). Transforms were computed from histograms containing two stimulus cycles. Because neural responses to the periodic vowels contain both a stimulus-locked, periodic component and an aperiodic, noise component, the amplitudes of the Fourier components located between vowel harmonics (i.e., at odd multiples of 50 Hz) give a measure of the noise floor (Young and Sachs, 1979).

Magnitude spectra of individual fiber responses depended strongly on the pseudo-spontaneous discharge rate. While there were notable differences in the shape of the spectrum between transient and sustained DPT responders, within each group the spectra were more similar. Figure 6 shows the average magnitude spectra for both bands of the vowel /u/ [0–500 Hz (left) and 500–1000 Hz (right)] at two modulation depths (2.5% and 10%). Separate averages were computed for transient (top) and sustained (bottom) responders. Average spectra for the other two vowels showed the same general trends, and are not shown.

In general, peaks in the average response spectra for sustained DPT responders qualitatively match those in the stimulus spectrum. In particular, a first-formant peak at 200

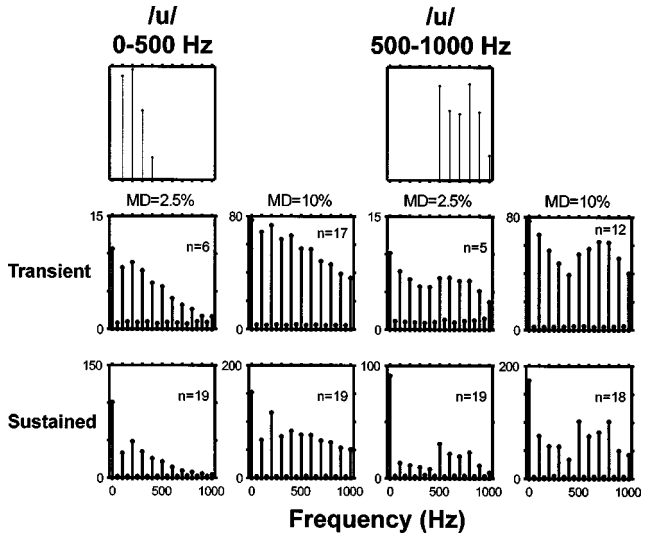


FIG. 6. Average of period histograms for two vowel modulators. The top row shows the spectra of the two vowels. Peaks in these spectra occur at the harmonics of the 100-Hz fundamental that are closest to the a formant frequency: 200 Hz for low-pass /u/, and 800 Hz for high-pass /u/. The bottom two rows show the average spectra of period histograms in response to both vowels for modulation depths of 2.5% (left) and 10% (right). The spectra were averaged separately for transient DPT responders (pseudo-spontaneous rates below 5 spikes/s; middle row), and sustained responders (bottom row). Numbers in each panel indicate the number of fibers whose responses are included in the average. Responses had to contain at least ten spikes to be included. The ordinates of each panel are in spikes/s.

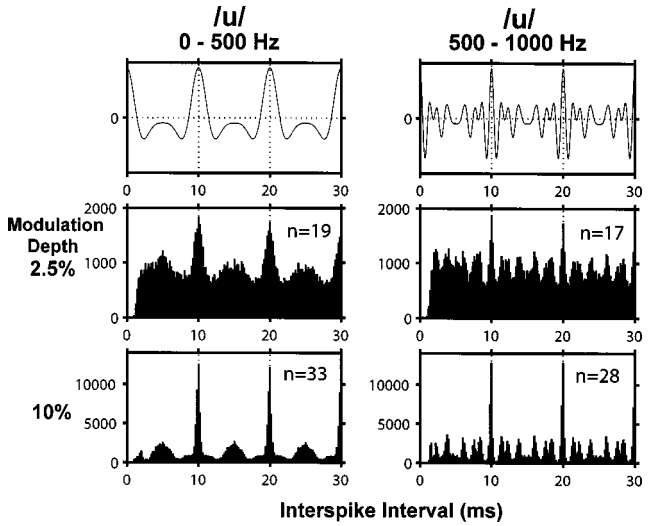


FIG. 7. Pooled autocorrelation histograms for two vowel modulators. The top row shows the autocorrelation functions of both vowels. Dashed lines show multiples of the stimulus period. The bottom two rows show pooled autocorrelation histograms of the electric responses for modulation depths of 2.5% (middle row) and 10% (bottom row). Numbers in each panel indicate the number of fibers whose responses were included in the pooled autocorrelation histogram. Responses had to contain at least ten spikes to be included. Ordinates in each panel are numbers of intervals per bin.

Hz is apparent in responses to the low-pass /u/ at both modulation depths, and a local maximum at 800 Hz gives a representation of the second formant for the high-pass /u/. However, at both modulation depths, the response spectra show components well above the noise floor at frequencies for which the stimulus contains no energy (above 500 Hz for low-pass /u/, and below 500 Hz for high-pass /u/), indicating a nonlinear transformation between the stimulus and the response. These distortion components are particularly apparent at the higher modulation depth. Part of the nonlinearity may just be half-wave rectification, which is clearly apparent in the period histograms of Fig. 3.

The average response spectra for transient DPT responders gave a poorer representation of the stimulus spectra than those of sustained responders. Although there is a small local maximum at 200 Hz for the low-pass /u/ at both modulation depths, it barely exceeds the noise floor. For the high-pass /u/ there may be a small second-formant-related peak near 800 Hz at 10% modulation depth, but not at 2.5% depth. However, for both stimuli and both modulation depths, the presence of the peaks well above the noise floor at the harmonics of the fundamental indicates that the response is locked to the fundamental frequency. For the high-pass /u/, the largest non-dc component of the average spectrum is the fundamental frequency. Thus, transient DPT responders represent primarily the fundamental in their temporal discharge patterns.

C. Representation of fundamental frequency

Because analyses based on period histograms require that the stimulus period be known, they cannot be used to directly test how well the vowel fundamental is represented in ANF temporal discharge patterns. Arbitrary response periodicities can be revealed using all-order interval (a.k.a. autocorrelation) histograms (Cariani and Delgutte, 1996a, b; Møller, 1970). We computed the autocorrelation histograms of individual fibers, and then summed them over the entire fiber population (including both transient and sustained responders) to create ''pooled'' autocorrelation histograms. Figure 7 shows the pooled autocorrelation histograms for both filtered /u/ stimuli at modulation depths of 2.5% (middle row) and 10% (bottom row). Autocorrelation histograms for the other two vowels show the same general trends and are not shown.

All pooled histograms have a pronounced mode at the vowel period of 10 ms, suggesting that the fundamental is robustly coded in the temporal patterns of ANF responses. In addition, the fine structures of the histograms are very different for the two stimuli. The smaller modes in the histogram match peaks in the autocorrelation function of the filtered vowel waveform (top row). Because minor peaks in the stimulus autocorrelation are related to formants, this result shows that the pooled autocorrelation histogram codes the formant frequencies in its pattern of secondary peaks. A similar result has been reported for pooled autocorrelation histograms computed from ANF responses to vowels presented acoustically (Cariani and Delgutte, 1996a, b).

D. Linearity of responses

The modulation transfer function (MTF) is defined as the ratio of modulation in the neural response to the modulation of the stimulus. Traditionally, MTFs are measured using sinusoidal modulators (e.g., Møller, 1974; Delgutte *et al.*, 1998). However, for a system that is nearly linear, the MTF can in principle also be estimated using complex, periodic modulators that contain several frequency components. We

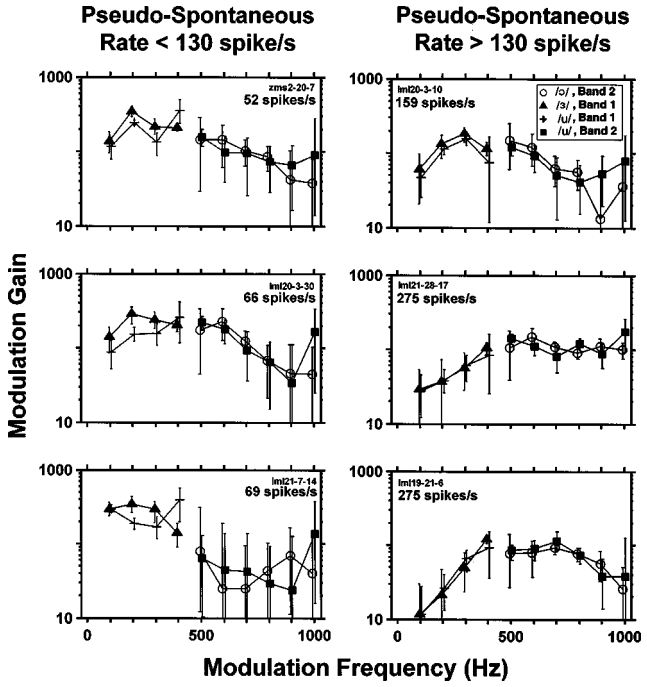


FIG. 8. Neural MTFs estimated from responses to vowel modulators at 1% modulation depth for three fibers with pseudo-spontaneous rates below 130 spikes/s (left), and three with pseudo-spontaneous rates above 130 spikes/s (right). Each panel shows the neural MTF magnitudes estimated from responses of one fiber whose pseudo-spontaneous rate is indicated inside the panel. Symbols code estimates based on different vowels. The MTF could only be estimated at frequencies for which the vowel stimulus has energy. Error bars represent the 99% confidence intervals computed by bootstrapping ANF spike trains (see Sec. II). For the three fibers on the right, the intrinsic periodicities (defined as the inverse of the largest mode in the interval histogram of responses to the unmodulated DPT) were, from top to bottom, 234, 537, and 383 Hz.

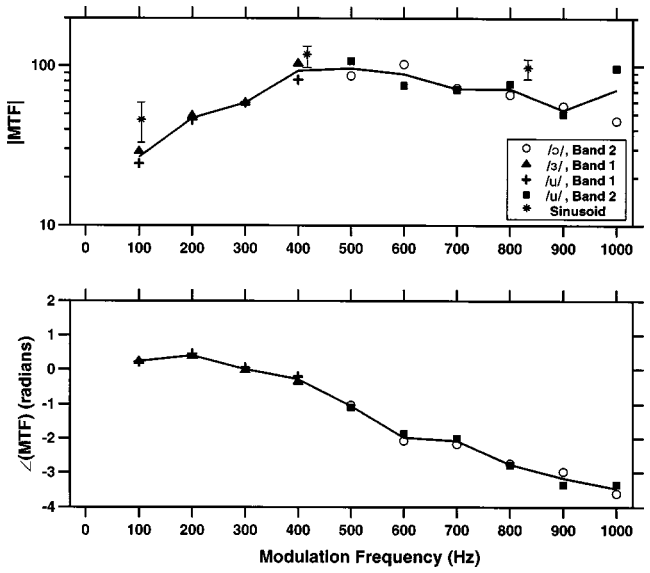


FIG. 9. Magnitude (top) and phase (bottom) of the average neural MTF estimated from responses to vowel modulators with 1% modulation depth. The black solid line shows the grand average MTF, which was computed by averaging individual MTFs across the four vowel modulators for ANFs with pseudo-spontaneous rates above 130 spikes/s. The symbols show the average MTF based on responses to individual vowels. The stars indicate the average, and the error bars 99% confidence intervals of the MTF magnitude estimated from responses to sinusoidal modulators with 1% modulation depth (Litvak *et al.*, 2003a).

tested the linearity assumption by computing MTFs from responses to different vowels. In a linear system, the MTF estimate should be independent of the modulator waveform used to estimate it.

Figure 8 shows the MTFs estimated from responses to vowels at 1% modulation depth for six ANFs, three with pseudo-spontaneous rates below 130 spikes/s (left), and three with rates above 130 spikes/s (right). Estimates based on responses to different vowels are denoted by different symbols. For each vowel, the MTF can only be estimated for frequencies at which the modulator has spectral energy.

For this 1% modulation depth, the modulation gain of all fibers is strikingly large, exceeding 100 at frequencies near 400–500 Hz. This means that a stimulus with a 1% modulation depth produces a fully modulated response, consistent with our earlier observations for sinusoidal modulators (Litvak *et al.*, 2003a).

MTFs computed from responses of fibers with pseudo-spontaneous discharge rates above 130 spikes/s (right panels in Fig. 8) are nearly independent of the vowel used to compute the MTF. This can be seen from the overlapping confidence intervals for the different vowels. The independence of the MTF from the vowel stimulus is consistent with the linearity assumption. However, there are statistically significant differences in MTFs across fibers.

For fibers with pseudo-spontaneous discharge rates below 130 spikes/s (Fig. 8, left column), MTF estimates clearly depend on the vowel stimulus. These responses are not consistent with linearity. The most noticeable nonlinearity is half-wave rectification, which is apparent in period histograms (e.g., Fig. 3). The dependence of MTF shape on stimulus spectrum does not necessarily imply that a model that includes a linear modulation filter cannot predict the responses of these fibers if the model also includes nonlinear elements (as the STM does). The data do show, however, that neural MTFs cannot be used to estimate the model's modulation filter when the pseudo-spontaneous rate is below 130 spike/s.

Figure 9 shows the magnitude and the phase of the average MTF computed from the responses of all fibers with pseudo-spontaneous rates above 130 spikes/s. The average MTF obtained by combining responses to all the vowels (solid lines) resembles the MTFs obtained by averaging data from each individual vowel (symbols) over each half of the spectrum, except at 1000 Hz. The average MTF magnitude has a bandpass shape, with a center frequency near 400–600 Hz, and a steeper cutoff on the low-frequency side than on the high-frequency side. The stars show the MTF magnitudes estimated at three frequencies from responses to sinusoidal modulators described in the companion paper (Litvak *et al.*, 2003a). These sinusoid-based estimates follow the general shape of the vowel-based MTF, although with a slightly higher gain. The lower modulation gain with vowel stimuli may reflect the synchrony suppression known to result from half-wave rectification of complex waveforms (Greenwood, 1986). The MTF phase shown in the bottom of Fig. 9 is approximately linear for frequencies between 200 and 1000 Hz, indicating a constant group delay of 700 μs.

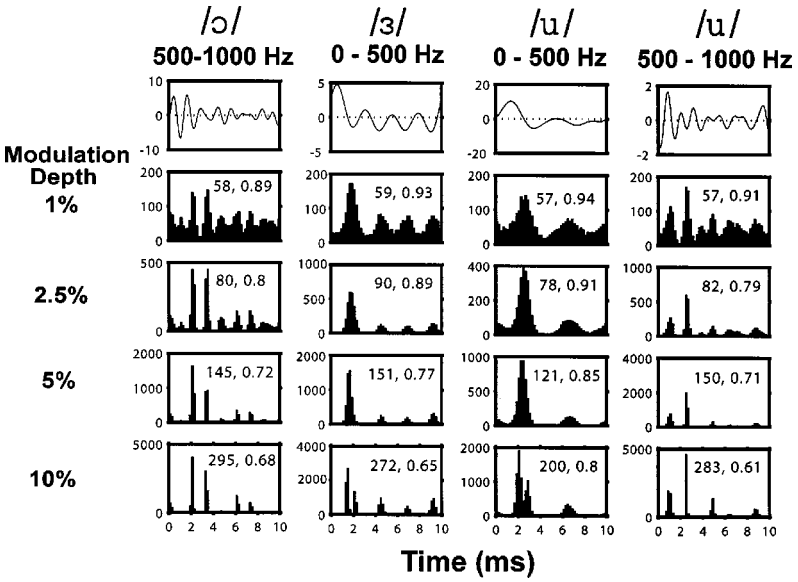


FIG. 10. Period histograms of responses of a model fiber with a high pseudo-spontaneous rate (50 spikes/s, $\theta_0/\sigma=2.2$) to the vowel stimuli. Same format as in Fig. 3. The second number in each panel represents is the stimulus-response correlation as defined as in Fig. 5, while the first number is the average discharge rate during the vowel stimulus.

E. Predictions of the stochastic threshold model

In a companion paper (Litvak *et al.*, 2003a), we introduced a stochastic threshold model (STM) for predicting responses of ANFs in the presence of a DPT. We showed that the STM successfully predicts many aspects of the responses to sinusoidal modulations of the DPT. Here, we examine whether the STM can also predict responses to complex modulators by comparing measured and model responses to vowel modulators.

The parameters of the STM are the resting threshold θ_0, the noise standard deviation σ, and the modulation filter. The threshold-to-noise ratio θ_0/σ specifies the pseudo-spontaneous discharge rate. To use the STM with arbitrary stimulus waveforms, both the modulation filter and σ must be specified. We chose the average neural MTF shown by the solid lines in Fig. 9 as the modulation filter. The noise amplitude σ was set to 1.7 in order to match the MTF gain of a model fiber with a pseudo-spontaneous rate of 300 spikes/s with the gain of the neural MTF in Fig. 9 at 400 Hz.

Figure 10 shows the period histograms of responses of a model fiber with a pseudo-spontaneous rate of 50 spikes/s to all four vowel modulators for modulation depths ranging from 1% to 10%. The predicted period histograms have many features in common with measured histograms for sustained DPT responders (Fig. 3). Both encode the peaks of the modulator waveform as modes of the period histogram. The locations and the numbers of major peaks are similar in measured and model histograms. Moreover, the correlations between the vowel waveform and the model response (second number inside each panel) are in the range observed for ANF sustained responders (Fig. 5). For both model and neural responses, the greatest correlations and the smallest discharge rates are observed for the low-pass /u/. However, the stimulus-response correlations are somewhat lower for the data than for the model at the higher modulation depth.

Figure 11 shows the responses of a model fiber with a pseudo-spontaneous rate near 0 spikes/s to the four vowel modulators. The responses of this model fiber to vowel

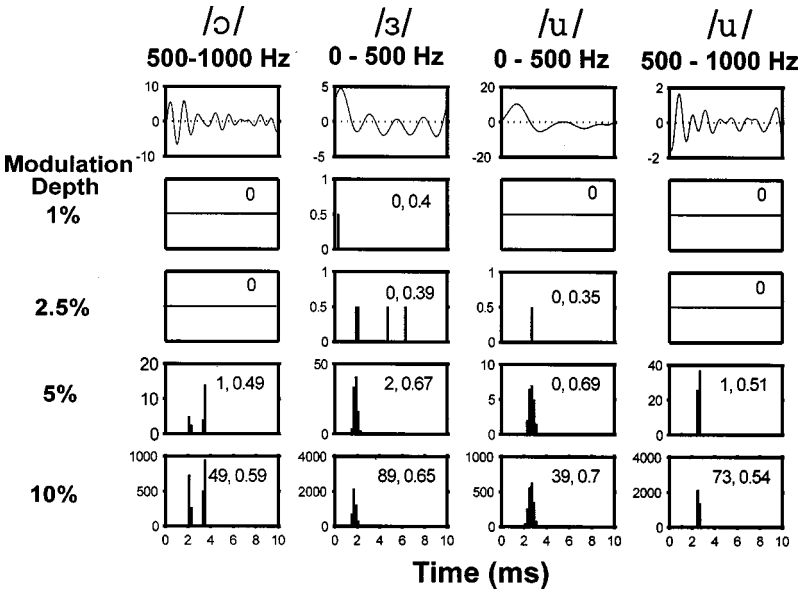


FIG. 11. Period histograms of responses of a model fiber with a low pseudo-spontaneous rate (0.0014 spikes/s, $\theta_0/\sigma=5$) to the vowel stimuli. Same format as in Fig. 10.

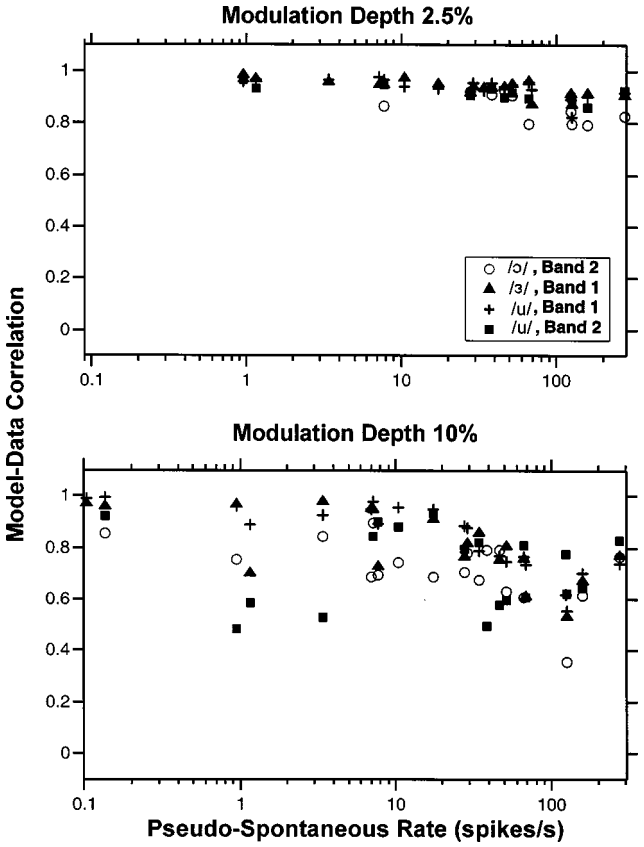


FIG. 12. Correlation coefficient between the period histograms of STM responses to vowel modulators and those of ANF responses as a function of pseudo-spontaneous discharge rate for modulation depths of 2.5% (top) and 10% (bottom). Each data point shows the correlation coefficient between the period histogram for one ANF and the histogram for a model fiber having the same pseudo-spontaneous rate. To account for differences in neural delay across fibers, we allowed a ±0.4-ms shift between the two responses. Symbols code responses to different vowels. Data are only included if the response contains at least ten spikes and the bootstrap 90% confidence intervals for the correlation coefficient are below 0.1.

modulations resembles those of transient DPT responders (e.g., Fig. 4). Both the model fiber and transient DPT responders require large modulation depths to produce a response. For both the model and neural data, responses tend to occur only at the largest peak of the vowel waveform. These responses, therefore, are most accurate in representing the fundamental and poorly represent the formant frequencies. However, responses of both the model fiber and the transient DPT responder do represent the two largest peaks of the high-passed /ɔ/.

To quantify how well the STM predicts the temporal discharge patterns of ANFs in response to vowel modulators, we computed the correlation coefficient between the period histogram of the neural response and the period histogram predicted by the STM. The STM was fit to each fiber by adjusting the threshold θ_0 so as to match the model fiber's pseudo-spontaneous rate to the data. Figure 12 shows the correlation between measured and model responses for modulation depths of 2.5% (top) and 10% (bottom) against the pseudo-spontaneous rate. Larger modulation depths were used for this correlation analysis than the 1% depth used for estimating the modulation filter because the higher discharge rates at higher modulation depths allow more accurate estimates of the period histograms and, therefore, the correlations. For both modulation depths, the model–data correlations are high, significantly better than the stimulus–response correlations in Fig. 5 ($p<0.001$, permutations test). For the lower modulation depth, the median model–data correlation is 0.93. Although the modulation filter was estimated from responses of fibers with pseudo-spontaneous rates above 130 spikes/s, the correlations at the lower modulation depth are nearly independent of pseudo-spontaneous rate. This result, as well as the ability of the STM to improve correlation, suggests that the model captures some of the nonlinearities in the neural responses. The correlations are somewhat lower for the 10% modulation depth (median 0.78). Thus, the STM is best at predicting responses to low modulation depths, which are likely to be more linear.

IV. DISCUSSION

A. Representation of vowel modulators in ANF responses

In response to filtered vowels, auditory-nerve fibers that gave sustained responses to the unmodulated DPT had complex period histograms whose modes matched peaks in the modulation waveform. Fourier and autocorrelation analyses revealed that sustained responders represent both the fundamental and the formant frequencies in their temporal discharge patterns. Because peaks in the vowel waveform can differ substantially in amplitude, only neurons with a wide dynamic range would be able to simultaneously represent several of these peaks in their temporal discharge patterns. The ability of sustained DPT responders to represent the entire vowel waveform is therefore consistent with the relatively wide (17–23 dB) dynamic range demonstrated previously using sinusoidal modulators (Litvak *et al.*, 2003a).

The closest resemblance between the period histograms of sustained responders and vowel waveforms occurred for modulation depths below 5%. At larger modulation depths, the period histogram showed distortions in the representation of the fine-time structure. In addition, for modulation depths above 5%, the average discharge rates could exceed those seen acoustically.

As with sinusoidal stimuli, the representation of vowels is more accurate in sustained DPT responders than in transient responders. A typical response pattern for a transient DPT responder is a single spike per stimulus period occurring at the modulator peak. For the modulation depths studied, transient DPT responders are poor at representing the formant frequency. They do, however, accurately represent the fundamental. Responses of transient DPT responders resemble responses to vowels presented electrically without a DPT (Knauth *et al.*, 1994; van den Honert and Stypulkowski, 1987).

Reproducing the exact spatio-temporal pattern of ANF activity in a healthy ear in response to acoustic stimuli would require an ability to stimulate narrower portions of the cochlea than is possible with current cochlear implants. However, because the pooled autocorrelation histogram of responses to complex modulators resembles the autocorrelation of the stimulus, reproducing the pooled interspike

interval distribution of acoustic responses may be achievable through manipulation of the modulation waveform in a DPT-enhanced strategy. It remains to be seen whether implanted subjects can make use of such interspike interval information.

B. Modulation transfer functions

We have hypothesized that responses of sustained responders to low modulation depths can be characterized by a linear relationship between the modulation waveform and the period histogram. We tested the linearity assumption by estimating neural MTFs for individual fibers using responses to different vowels. Under linearity, the estimated MTF should be independent of the vowel used to derive the estimate. We found that, for a modulation depth of 1%, linearity held for the small fraction of fibers with pseudo-spontaneous rates above 130 spikes/s. As expected, linearity did not hold as well for fibers with lower pseudo-spontaneous rates. Because these responses showed significant half-wave rectification even at 1% modulation depth (e.g., Fig. 3), it is possible that linearity would hold for a larger number of fibers at smaller modulation depths. However, accurately measuring period histograms to smaller modulations would require very long data collection times.

While consistent with linearity, MTFs estimated from fibers with pseudo-spontaneous rates above 130 spikes/s differed somewhat from one fiber to the next. We have previously reported that fibers with high pseudo-spontaneous rates tend to have preferred interspike intervals in response to the DPT, and that these intrinsic periodicities interact with periodicities in the modulation waveform in determining responses to modulations of a DPT (Litvak *et al.*, 2003a,b). One possibility is that these intrinsic periodicities, which differ from fiber to fiber, may be partly responsible for differences in MTFs among fibers with high pseudo-spontaneous rates. Since synchrony to a sinusoidal modulator is highest when the modulation period coincides with the intrinsic period (Litvak *et al.*, 2003a), this hypothesis predicts that the MTF gain should be above average when the modulation period coincides with the intrinsic period. Qualitative analysis of three fibers with pseudo-spontaneous rates above 130 spikes/s suggests that this may indeed be the case (see caption of Fig. 8). However, we do not have sufficient data to conclusively test this hypothesis.

The average neural MTF estimated from the responses of fibers with pseudo-spontaneous rates above 130 spikes/s had approximately the same frequency dependence as the average modulation gain at three frequencies for the sinusoidal modulators investigated in the companion paper (Litvak *et al.*, 2003a). The estimated MTF showed a bandpass characteristic (passband 400–600 Hz). In the passband, the modulation gain was extremely high (near 100), meaning that a stimulus with 1% modulation depth produces a fully modulated response. Above the passband, the response dropped off very gently with frequency, suggesting that, with a DPT, ANFs could represent formant frequencies in their temporal discharge patterns up to at least 1000 Hz.

The neural mechanisms responsible for the bandpass characteristic of the MTFs are unclear. The high-frequency drop off is consistent with low-pass filtering in the neural membrane. Adaptation of electric responses (Dynes and Delgutte, 1992; Killian, 1994; Litvak *et al.*, 2003a) may cause the drop in the magnitude of the MTF for very low frequencies (Delgutte *et al.*, 1998). Another mechanism for the low-frequency drop off may be a partial excitation block for high-frequency stimulation. For example, Litvak *et al.* (2001a) noted that modulation of a 5-kpps carrier often produced an increase in discharge rate, even though the peak amplitude of the modulated pulse train was the same as that of the unmodulated carrier, and therefore the rms amplitude of the modulated stimulus was lower. The mechanism for excitation block was hypothesized to be the interaction between the dynamics of sodium channels and high-frequency stimulation. One possibility is that the sodium inactivation gate is never able to recover with high-frequency stimulation.

C. Stochastic threshold model

In the companion paper (Litvak *et al.*, 2003a), we showed that the stochastic threshold model (STM) quantitatively predicts many features of single ANF responses to sinusoidal modulations of a DPT, including growth of synchrony and discharge rate with modulation depth, and the dependence of modulation threshold and dynamic range on pseudo-spontaneous rate. In order to extend the STM to arbitrary modulation waveforms, a modulation filter needs to be specified for the model. Because responses to sinusoidal modulators were only measured at three modulation frequencies, these data did not allow us to estimate the modulation filter with sufficient resolution. However, we determined that the modulation filter can be estimated directly from responses to vowel modulators. This approach assumes that (1) the system behaves linearly and that (2) the measured MTF accurately reflects the shape of the underlying modulation filter. We tested the linearity assumption and found that it held for 1% modulation depth in fibers with pseudo-spontaneous rates above 130 spikes/s. To test the second assumption, we measured the MTF of the STM and compared it to its modulation filter. We found that, for fibers with high pseudo-spontaneous rates, the measured MTFs from model fibers resembled the modulation filter. This result confirms that, if the STM is applicable to the responses of ANFs, then the modulation filter can indeed be estimated from measured neural MTFs.

With the modulation filter of Fig. 9, the STM approximately predicted the shapes of period histograms in response to vowel modulators. At a modulation depth of 2.5%, the median correlation coefficient between neural and model period histograms was 0.93. Interestingly, the model was as successful at predicting responses of fibers with low pseudo-spontaneous rates as those of fibers with high pseudo-spontaneous rates, even though only data from the latter were used in deriving the model's modulation filter. This finding suggests that the model captures some of the nonlinearities in the neural response, including half-wave rectification. However, model predictions were poorer for 10% modulation depth than for 2.5% depth. Responses of ANFs to stimuli with large modulation depths are strongly influenced by the fiber's refractory properties because these

stimuli drive the fiber at high discharge rates. By adjusting the refractory characteristics of the model to better fit those of ANFs, it might be possible to make the STM better approximate the responses at large modulation depths.

It is worth emphasizing that the STM does not explicitly represent the DPT. The input to the model is the modulation waveform, not a modulated DPT, and the effect of the DPT is only mimicked by adjusting the resting threshold so as to set the model's pseudo-spontaneous rate. The success of this very simple model in predicting responses to complex stimuli further supports the argument made in the companion paper (Litvak *et al.*, 2003a) that the exact details of a DPT may not be essential in improving stimulus coding in temporal discharge patterns of ANFs, so long as it evokes a sustained pseudo-spontaneous discharge rate and the stimulus is a small signal compared to the DPT.

Our results suggest that, with the estimated modulation filter, and at least for small modulation depths, the STM can accurately predict the discharge probability of ANFs for arbitrary modulator waveforms. The predictive capability of the STM makes it a useful tool for investigating the psychophysics of electric hearing. In particular, the STM can be useful in interpreting psychophysical responses of human cochlear implant subjects to small stimuli superimposed upon a large DPT. The STM could also be useful for modeling responses of neurons in the central nervous system to electric stimulation, where a model of activity in the auditory nerve inputs is necessary.

V. CONCLUSION

Together, the results of this paper and its companions (Litvak *et al.*, 2003a, b) show that a desynchronizing pulse train can improve the representation of both sinusoidal and complex stimuli in the temporal discharge patterns of auditory-nerve fibers that have a sustained response to the DPT. While these results are encouraging, it is not possible to predict the behavioral consequences of this improved representation from single-unit experiments alone. Key unanswered questions include whether a DPT could be safe yet stimulate a sufficient number of ANFs to make a difference, and whether the central nervous system would be able to make use of the more natural temporal information provided by a DPT. The present work provides physiological guidelines for designing DPT-enhanced stimulus strategies whose efficacy can be tested by psychophysical experiments in implanted subjects. Whatever their clinical outcome, such experiments will certainly provide clues as to how the temporal information available in the auditory nerve is processed by the central nervous system.

ACKNOWLEDGMENTS

The authors would like to acknowledge tireless efforts of Leslie Liberman without whose surgical skills this work would be impossible. We also thank Z. M. Smith for assistance with data collection, and K. M. Brinsko for help in figure preparation. This work was supported through W. M. Keck Foundation and NIH Grant Nos. DC00361, DC02258, and N01-DC-6-210.

APPENDIX A: ARTIFACT REJECTION METHOD

To remove the artifact in response to vowel modulators, we used a generalization of the method used for sinusoidal modulators in the companion paper (Litvak *et al.*, 2003a). Briefly, the recorded microelectrode signal was first processed through a low-pass digital filter (3 kHz cutoff) to remove the linear component of the stimulus artifact. The remaining artifact was significantly below the spikes for most responses except at the largest modulation depth (10%–15%). To further cancel the residual artifact, we used an iterative algorithm to find the closest fit between the filtered, recorded signal and a model signal consisting of the sum of synthetic spikes and a periodic artifact template composed of N harmonics of the stimulus period (see Litvak *et al.*, 2003a for details). This periodic template was subtracted from the recorded waveform to derive the estimated spike train.

Because the stimuli used in this study are more complex than the sinusoids used in the companion paper, the number of harmonics N in the artifact template had to be increased from 3 to 16. Because some of these harmonics lie in the 1–2-kHz frequency region where spikes have most of their energy, the derived artifact template could in principle include the spike waveform as well as the artifact proper if the response is highly synchronized. To detect whether this condition occurred, we simultaneously monitored the estimated spike waveforms, the artifact template, and the difference between the artifact template and the scaled artifact template estimated from the responses to the same stimulus at half the modulation depth. A record was accepted if either the estimated artifact was at most one-third of the peak spike amplitude, or the difference between the artifact templates at both modulation depths did not contain any spikes.

[1]To expedite data collection, in later experiments we shortened the duration of the unmodulated segments to 300 ms when the preceding modulated segment had a modulation depth below 2.5%, and therefore produced low discharge rates.

Assmann, P. F., and Summerfield, Q. (**1990**). "Modeling the perception of concurrent vowels: vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Cariani, P. A., and Delgutte, B. (**1996a**). "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience," J. Neurophysiol. **76**, 1698–1716.

Cariani, P. A., and Delgutte, B. (**1996b**). "Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch," J. Neurophysiol. **76**, 1717–1734.

Delgutte, B. (**1984**). "Speech coding in the auditory nerve: II. Processing schemes for vowel-like sounds," J. Acoust. Soc. Am. **75**, 879–886.

Delgutte, B., and Kiang, N. Y. S. (**1984**). "Speech coding in the auditory nerve: I. Vowel-like sounds," J. Acoust. Soc. Am. **75**, 866–878.

Delgutte, B., Hammond, B. M., and Cariani, P. A. (**1998**). "Neural coding of the temporal envelope of speech: Relation to modulation transfer functions," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Reese, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 595–603.

Dynes, S. B. C., and Delgutte, B. (**1992**). "Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea," Hear. Res. **58**, 79–90.

Efron, B., and Tibshirani, R. (**1993**). *An Introduction to the Bootstrap* (Chapman and Hall, New York).

Good, P. (**2000**). *Permutation Tests. A Practical Guide to Resampling Methods for Testing Hypotheses* (Springer, New York).

Greenwood, D. D. (**1986**). ''What is 'Synchrony suppression'?'' J. Acoust. Soc. Am. **79**, 1857–1872.
Hartmann, R., Topp, G., and Klinke, R. (**1984**). ''Discharge patterns of cat primary auditory fibers with electrical stimulation of the cochlea,'' Hear. Res. **13**, 47–62.
Johnson, D. H. (**1980**). ''The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones,'' J. Acoust. Soc. Am. **68**, 1115–1122.
Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve*. (Cambridge, MA: The MIT Press).
Killian, M. J. P. (**1994**). ''Excitability of the Electrically Stimulated Auditory Nerve,'' Ph.D. thesis, University of Utrecht, Netherlands
Klatt, D. H. (**1980**). ''Software for cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 971–995.
Knauth, M., Hartmann, R., and Klinke, R. (**1994**). ''Discharge pattern in the auditory nerve evoked by vowel stimuli: a comparison between acoustical and electrical stimulation,'' Hear. Res. **74**, 247–258.
Litvak, L. M., Delgutte, B., and Eddington, D. K. (**2001**). ''Auditory nerve fiber responses to electric stimulation: modulated and unmodulated pulse trains,'' J. Acoust. Soc. Am. **110**, 368–379.
Litvak, L. M., Delgutte, B., and Eddington, D. K. (**2003a**). ''Improved temporal coding of sinusoids in electric stimulation of the auditory nerve using desynchronizing pulse trains,'' J. Acoust. Soc. Am **14**, 2079–2098.
Litvak, L. M., Smith, Z. M., Delgutte, B., and Eddington, D. K. (**2003b**). ''Desynchronization of electrically evoked auditory-nerve activity by high-frequency pulse trains of long duration,'' J. Acoust. Soc. Am. **14**, 2066-2078.
Miller, M. I., and Sachs, M. B. (**1984**). ''Representation of voice pitch in discharge patterns of auditory-nerve fibers,'' Hear. Res. **14**, 257–279.
Møller, A. R. (**1970**). ''The use of correlation analysis in processing neuroelectric data,'' in *Progress in Brain Research, Computers and Brains*, edited by J. P. Schade and J. Smith (Elsevier, Amsterdam), pp. 87–99.
Møller, A. R. (**1973**). ''Statistical evaluation of the dynamic properties of cochlear nucleus units using stimuli modulated with pseudorandom noise,'' Brain Res. **57**, 443–456.
Møller, A. R. (**1974**). ''Responses of units in the cochlear nucleus to sinusoidally amplitude-modulated tones,'' Exp. Neurol. **45**, 104–117.
Møller, A. R., and Rees, A. (**1986**). ''Dynamic properties of the responses of single neurons in the inferior colliculus of the rat,'' Hear. Res. **24**, 203–215.
Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of vowels,'' J. Acoust. Soc. Am. **24**, 175–184.
Rubinstein, J. T., Wilson, B. S., Finley, C. C., and Abbas, P. J. (**1999**). ''Pseudospontaneous activity: stochastic independence of auditory nerve fibers with electrical stimulation,'' Hear. Res. **127**, 108–118.
van den Honert, C., and Stypulkowski, P. H. (**1987**). ''Temporal response patterns of single auditory nerve fibers elicited by periodic electrical stimuli,'' Hear. Res. **29**, 207–222.
Xu, S. A., Shepherd, R. K., Chen, Y., and Clark, G. M. (**1993**). ''Profound hearing loss in the cat following the single co-administration of kanamycin and ethacrynic acid,'' Hear. Res. **70**, 205–215.
Young, E. D., and Sachs, M. B. (**1979**). ''Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory nerve fibers,'' J. Acoust. Soc. Am. **66**, 1381–1403.

An approximate transfer function for the dual-resonance nonlinear filter model of auditory frequency selectivity

Enrique A. Lopez-Poveda[a)]
Instituto de Neurociencias de Castilla y León, Universidad de Salamanca, Avenida Alfonso X El Sabio s/n, 37007 Salamanca, Spain and Centro Regional de Investigaciones Biomédicas, Facultad de Medicina, Universidad de Castilla-La Mancha, 02071 Albacete, Spain

(Received 13 February 2003; revised 25 June 2003; accepted 14 July 2003)

The dual-resonance nonlinear filter [Meddis *et al.*, J. Acoust. Soc. Am. **109**, 2852–2861 (2001)] was presented as a digital time-domain algorithm to model nonlinear auditory frequency selectivity. This report extends previous work by presenting an approximate analytic transfer function that allows calculating and analyzing its level-dependent frequency-domain response. The transfer function is derived on the assumption that the filter behaves linearly for any given input amplitude. It matches accurately the response (gain and phase) of the digital filter for tones. Practical uses for the transfer function are suggested. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1605389]

PACS numbers: 43.66.Ba [WPS]

I. INTRODUCTION

The dual-resonance nonlinear (DRNL) filter is an algorithm capable of reproducing an important number of physiological (Meddis *et al.*, 2001; Sumner *et al.*, 2002, 2003a, 2003b) and psychophysical (Lopez-Poveda and Meddis, 2001) phenomena pertaining to nonlinear auditory frequency selectivity. It was originally designed digitally in the *time* domain to extend its validity for complex, broadband stimuli and to facilitate its application to real-time systems. However, filters are better understood in terms of their *frequency*-domain response. This is particularly true for the DRNL filter as its best frequency, tuning, gain, and phase characteristics change with the amplitude of its input signal. This report extends previous work by presenting an *analytic* transfer function for evaluating the amplitude-dependent frequency response (gain and phase) of the DRNL filter based on its response to *tones*.

Nonlinear filters alter the spectral content of the input waveform, and thus do not have proper transfer functions (Smith, 2002). Indeed, deriving an approximate transfer function of the DRNL filter has been possible because all but one of its components (Fig. 1) are linear and time invariant. The nonlinear stage [$U(f)$ in Fig. 1] applies a memoryless, time-*varying* gain that depends on the instantaneous amplitude of its input waveform. Although its exact transfer function does not exist, an approximation is made here to obtain it. It consists of treating the nonlinear stage as a time-*invariant* gain that depends on the *peak* amplitude of its input. This approximation is equivalent to assuming that the DRNL filter is *linear* for any given input level. Below it is shown that this approximation preserves, with good accuracy, the nonlinear gain and phase properties of the digital DRNL filter in response to tones, although it sacrifices the effect of the original nonlinear gain stage on the spectral content of its output.

The proposed transfer function may have a number of applications. The assumption on which it has been derived (linear behavior of the filter for any given input level) is commonly made in psychophysics for measuring auditory filter shapes (Moore, 1998, Chap. 3), and in physiology for measuring basilar-membrane (BM) isointensity and input/output curves (Robles and Ruggero, 2001). Therefore, it may be used to model these types of data. The transfer function demonstrates the contribution of every parameter of every component stage of the DRNL filter to its frequency response. This knowledge facilitates tuning the numerous parameters of the filter to reproduce specific data sets.

The transfer function may also be used to investigate the extent that suppression and distortion phenomena affect auditory filter shapes derived from notch-noise data (Moore, 1998). An estimate could be obtained by comparing the transfer function against the response of the digital DRNL filter to the notch-noise stimuli, both computed with identical parameters tuned for the transfer function to match the filter shapes. The difference in response may be attributed to suppression and distortion present during data collection, as these effects will be modeled by the digital DRNL filter (Meddis *et al.*, 2001) but not by the transfer function.

The transfer function allows computing the response of the DRNL filter for tones much more rapidly than its digital implementation. Therefore, it may be particularly useful for applications that require evaluating the frequency response of filter *banks*; for example, during the development of speech processing strategies for auditory prostheses based on the DRNL filter (e.g., Wilson *et al.*, 2002). It may also be used to compute level-dependent excitation patterns from auditory filter shapes modeled with the DRNL filter using the method of Glasberg and Moore (1990).

II. THE TRANSFER FUNCTION

The transfer function of a filter is the ratio of the Fourier transform of its output signal, $y(t)$, to the Fourier transform of the input signal, $x(t)$ (Hartmann, 1998, p. 195). For the DRNL filter (Fig. 1), this can be written as

a)Electronic mail: ealopezpoveda@usal.es

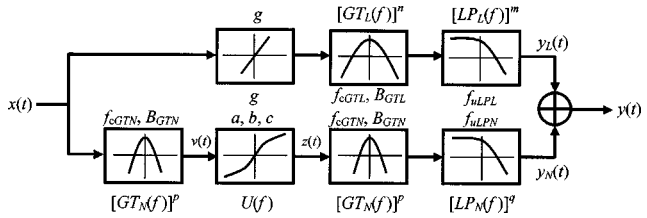


FIG. 1. Architecture of the dual-resonance nonlinear filter. The symbols between the top and the bottom processing paths inform of the parameters of each filter stage. The transfer function of each (GT or LP) filter or gain stage is given above (top path) or below (bottom path) its corresponding block. See the main text for details.

$$H_{\mathrm{DRNL}}(f)=\frac{\mathcal{F}[y(t)](f)}{\mathcal{F}[x(t)](f)}, \tag{1}$$

where $\mathcal{F}$ means Fourier transform. The output from the DRNL filter is the sum of the outputs from its linear, $y_L(t)$, and nonlinear, $y_N(t)$, paths. The input to both paths is the same and equal to $x(t)$. Therefore, by virtue of the linear properties of the Fourier transform, Eq. (1) can be rewritten as

$$H_{\mathrm{DRNL}}(f)=\frac{\mathcal{F}[y_L(t)+y_N(t)](f)}{\mathcal{F}[x(t)](f)}=H_L(f)+H_N(f). \tag{2}$$

That is, the transfer function of the DRNL filter can be expressed as the sum of the individual transfer functions of its two paths, $H_L(f)$ and $H_N(f)$.

A. The transfer function of the linear path

The linear path (Fig. 1) consists of a linear gain, *g*, followed by a cascade of *n first*-order gammatone (GT) filters followed by a cascade of *m second*-order Butterworth low-pass (LP) filters, all with unit gain in their passbands. All these elements are linear and time invariant. Therefore, the transfer function of the linear path, $H_L(f)$, can be expressed as the product of the individual transfer functions of its three stages

$$H_L(f)=g\cdot[GT_L(f)]^n\cdot[LP_L(f)]^m, \tag{3}$$

where $GT_L(f)$, and $LP_L(f)$ denote the transfer functions of the individual GT and LP filters in the linear path, respectively. These are described in the following two sections.

1. The transfer function of the first-order gammatone filter

The transfer function of a *first*-order GT filter with unit gain at its center frequency, f_c (Hz), is as follows (Stone, 1995; Hartmann, 1998):

$$GT(f)=k\cdot\left[\frac{B}{B+j(f-f_c)}+\frac{B}{B+j(f+f_c)}\right], \tag{4}$$

where $j=\sqrt{(-1)}$, B is *half* the 3-dB bandwidth (Hz) of the *first*-order GT filter (Hartmann, 1998), and k is a constant that makes $GT(f)$ have unit gain at f_c

$$k=\frac{1}{|1+(B/B+2jf_c)|}. \tag{5}$$

For convenience, $GT(f)$ can be expressed using complex polar notation as

$$\mathrm{GT}(f)=R_{\mathrm{GT}}(f)\cdot\exp[j\phi_{\mathrm{GT}}(f)], \tag{6}$$

where $R_{\mathrm{GT}}(f)=|\mathrm{GT}(f)|$ is the gain of the GT filter, and $\phi_{\mathrm{GT}}(f)=\arg[\mathrm{GT}(f)]$ its phase.

2. The transfer function of the second-order Butterworth low-pass filter

The transfer function of a second-order Butterworth low-pass filter, $\mathrm{LP}(f)$, can be expressed using complex polar notation as

$$\mathrm{LP}(f)=R_{\mathrm{LP}}(f)\cdot\exp[j\phi_{\mathrm{LP}}(f)] \tag{7a}$$

(Oppenheim *et al.*, 1999), where the gain, $R_{\mathrm{LP}}(f)$, and the phase, $\phi_{\mathrm{LP}}(f)$, of the filter are

$$R_{\mathrm{LP}}(f)=\left[\frac{1}{(1-f_R^2)^2+2f_R^2}\right]^{0.5}, \tag{7b}$$

$$\phi_{LP}(f)=\arctan\left(\frac{-\sqrt{2}f_R}{1-f_R^2}\right), \tag{7c}$$

with $f_R=f/f_u$, and f_u being the 3-dB-down cutoff frequency (Hz) of the filter.

Therefore, the transfer function of the linear path can be obtained by substituting Eqs. (6) and (7a) into Eq. (3)

$$H_L(f)=g\cdot[R_{\mathrm{GTL}}(f)]^n\cdot[R_{\mathrm{LPL}}(f)]^m\cdot\exp\{j[n\phi_{\mathrm{GTL}}(f)+m\phi_{\mathrm{LPL}}(f)]\}. \tag{8}$$

The subscripts GTL and LPL refer to the GT and LP filters in the linear path, respectively.

B. The transfer function of the nonlinear path

The nonlinear path (Fig. 1) is a cascade of *p* first-order GT filters, followed by a *compressive* nonlinear gain, followed by another cascade of *p* GT filters (identical to the first one), followed by a cascade of *q* second-order Butterworth LP filters. Assuming that the transfer function for the nonlinear gain, $U(f)$, exists (see below), the transfer function of the nonlinear path can be expressed as the product of the transfer functions of its stages

$$H_N(f)=U(f)\cdot[\mathrm{GT}_N(f)]^{2p}\cdot[\mathrm{LP}_N(f)]^q, \tag{9}$$

where $\mathrm{GT}_N(f)$ and $LP_N(f)$ denote the transfer functions of the individual GT and LP filters in the nonlinear path, respectively.

Only the transfer function of the nonlinear gain, $U(f)$, remains to be described.

1. The transfer function of the nonlinear gain

Meddis *et al.* (2001) defined the time-domain form of the nonlinear gain as

$$z(t)=\mathrm{sign}[v(t)]\cdot\min\{a\cdot|v(t)|,b\cdot|v(t)|^c\}, \tag{10}$$

where $v(t)$ and $z(t)$ are the input and output waveforms to/from the nonlinear gain (Fig. 1), *a* and *b* are gain parameters ($\geqslant 0$), and *c* is the compression exponent ($0<c\leqslant 1$). It is, therefore, a memoryless (instantaneous), time-*varying*

nonlinear gain. As such, it is *not* possible to derive its analytic transfer function (Smith, 2002).

In practice (e.g., Meddis *et al.*, 2001; Lopez-Poveda and Meddis, 2001), the frequency response of the *digital* DRNL filter is evaluated by examining the *peak* amplitude and the phase of its output waveform in response to sinusoids. For sinusoidal inputs, an approximation can be made that allows derivation of an analytic transfer function for the nonlinear-gain stage. Let $u(t)$ denote the approximated time-domain function, which is as follows:

$$u(t)=\min(a,bV^{c-1})\cdot v(t). \tag{11}$$

$u(t)$ applies an instantaneous time-*invariant* gain that depends only on the *peak* amplitude, V, of the input sinusoid, $v(t)$. The error of the approximation is zero when V is less than or equal to the compression-threshold amplitude $V_c=(b/a)^{1/(1-c)}$, as $u(t)=z(t)$. However, when $V>V_c$ the original nonlinear gain, $z(t)$, applies instantaneous compression that alters the shape of the input waveform [Fig. 2(a)] and hence its spectral content [Fig. 2(b)]. This property is *not* preserved by the approximated gain, $u(t)$, which remains a pure sinusoid. Nevertheless, the frequency and phase of $u(t)$ are identical to those of the fundamental frequency of $z(t)$. Furthermore, its amplitude is equal to the *peak* amplitude of $z(t)$ [Fig. 2(a)]. In summary, $u(t)$ maintains the properties of the original nonlinear gain, and hence of the DRNL filter, regarding the gain and the phase of its response to sinusoids (see Sec. III below).

An important advantage of Eq. (11) is that it has an exact transfer function

$$U(f)=\min(a,bV^{c-1}). \tag{12}$$

In the DRNL filter (Fig. 1), V is the product between the peak amplitude of the input sinusoid to the DRNL filter, X, and the gain of the first GT-filter cascade in the nonlinear path, $[R_{\mathrm{GTN}}(f)]^p$. Hence, V depends on the frequency *and* on the peak amplitude of the input tone to the DRNL filter

$$V(f,X)=X\cdot[R_{\mathrm{GTN}}(f)]^p. \tag{13}$$

The transfer function of the approximated nonlinearity is obtained by replacing $V(f,X)$ into Eq. (12)

$$U(f,X)=\min\{a,b[R_{\mathrm{GTN}}(f)]^{p(c-1)}X^{c-1}\}. \tag{14}$$

The notation $U(f,X)$ makes explicit that the transfer function depends on the frequency of the input tone to the DRNL filter *and* on its peak amplitude (X).

The frequency transfer function of the nonlinear path can now be derived by substituting Eqs. (6), (7a), and (14) into Eq. (9)

$$\begin{aligned}H_N(f,X)=&\min\{a,b[R_{\mathrm{GTN}}(f)]^{p(c-1)}X^{c-1}\}\cdot[R_{\mathrm{GTN}}(f)]^{2p}\\&\cdot[R_{\mathrm{LPN}}(f)]^q\cdot\exp\{j[2p\phi_{\mathrm{GTN}}(f)\\&+q\phi_{\mathrm{LPN}}(f)]\}.\end{aligned} \tag{15}$$

C. The transfer function of the DRNL filter

It follows from Eq. (2) that the transfer function of the DRNL filter, $H_{\mathrm{DRNL}}(f)$, can be expressed in polar form as

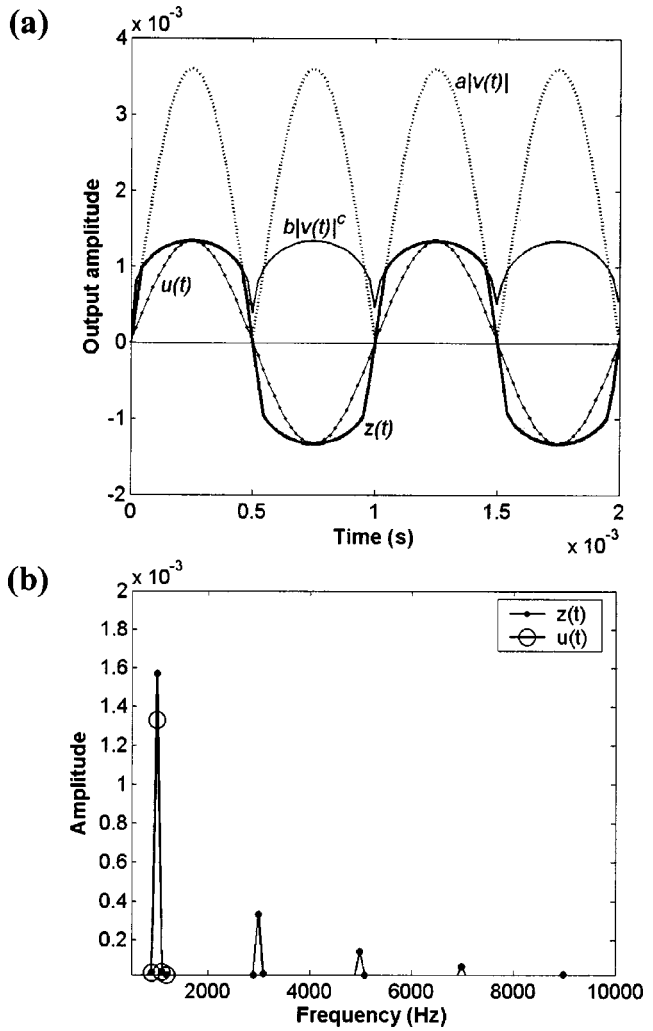


FIG. 2. (a) Comparison of the output waveforms from the original [Eq. (10)] and the approximated [Eq. (11)] nonlinear gains for an input waveform, $v(t)$, undergoing compression ($V>V_c$). The output waveform from the original nonlinear gain, $z(t)$ (thick continuous line), is equal to $av(t)$ at times when $a|v(t)|<b|v(t)|^c$, but equal to $b|v(t)|^c$ otherwise. It is periodical but not a pure sinusoid. Hence, it contains odd distortion harmonics [illustrated in panel (b)]. The output from the approximated nonlinear gain, $u(t)$, has the same peak amplitude *and* phase as $z(t)$, but is purely sinusoidal and hence shows *no* distortion harmonics [illustrated in panel (b)]. (b) Amplitude spectra of signals $z(t)$ and $u(t)$ shown in panel (a).

$$H_{\mathrm{DRNL}}(f)=R_{\mathrm{DRNL}}(f)\cdot\exp[j\phi_{\mathrm{DRNL}}(f)], \tag{16}$$

where

$$R_{\mathrm{DRNL}}(f)=[R_L^2+R_N^2+2R_LR_N\cos(\phi_N-\phi_L)]^{0.5}, \tag{17a}$$

$$\phi_{\mathrm{DRNL}}(f)=\arctan\left(\frac{R_L\sin\phi_L+R_N\sin\phi_N}{R_L\cos\phi_L+R_N\cos\phi_N}\right). \tag{17b}$$

$R_L(f)$ and $R_N(f)$ are the magnitudes of $H_L(f)$ and $H_N(f)$, respectively, and $\phi_L(f)$ and $\phi_N(f)$ their phases, all of which can be easily worked out from Eqs. (8) and (15).

It is noteworthy that Eqs. (17a) and (17b) show that when $R_L(f)\ll R_N(f)$, the gain and the phase of the DRNL filter are those of its nonlinear path. However, the linear path dominates the response when $R_L(f)\gg R_N(f)$. They also show that when $R_L(f)=R_N(f)$ *and* $(\phi_N-\phi_L)=N\pi(N=1,3,5,...)$, the gain of the DRNL filter is equal to zero and a notch occurs in the filter's frequency response, as noted by Meddis *et al.* (2001) and Lopez-Poveda and Meddis (2001).

III. EVALUATION

The validity of the transfer function was tested by comparing its output with the response of the digital, time-domain implementation[1] of the DRNL filter for identical sinusoidal inputs and for a large number of DRNL-filter parameter sets. The results shown in Fig. 3 are only an example. They are based on parameters reported in Meddis *et al.* (2001, Table I) for modeling basilar-membrane responses for the case L113 of Ruggero *et al.* (1997). These parameters are reproduced in Table I. The frequency of the sinusoids ranged from $0.25\times f_{c\mathrm{GTN}}$ to $2\times f_{c\mathrm{GTN}}$ in steps of $0.05\times f_{c\mathrm{GTN}}$. Their amplitude corresponded to sound-pressure levels ranging from 0 to 100 dB in steps of 20 dB, but it was scaled down to typical values of stapes velocity (in units of m/s), to match the expected order of magnitude for the DRNL-filter input. The scalar was fixed at 1.5×10^{-5} (m/s/Pa) across frequencies.

The amplitude and the phase responses of the digital DRNL filter were measured by applying a standard sine wave fit algorithm (Händel, 2000) to its output. The input sinusoids had a duration of 10 ms and were sampled at a rate of 10^6 Hz. The sampling rate was made so large to minimize any possible error due to sampling.

Figures 3(a) and (b) show a close match in the gain and the phase responses between the digital and the analytical implementations. Figure 3(c) shows that the error is small, particularly for low-and high-input amplitudes, where the DRNL filter behaves linearly.

Although not illustrated here, the discrepancy between the transfer function and the digital evaluations increases when lower sampling rates are used. The discrepancy is qualitatively more important in the phase response of the filter. It is attributed to sampling and is most prominent for the GT filters than for the low-pass filters, especially for high f_c's and for frequencies remote from f_c (see the footnote).

Increasing the amount of compression (by decreasing exponent c of the compressive nonlinearity) hardly reduces the accuracy of the transfer function. However, its match with the digital evaluation improves when c increases, as the filter behaves more linearly and the negative effects of the approximation diminish.

Remarkably, the transfer function took 0.16 s of CPU time to compute, whereas its digital counterpart took 30.4 s (both computed in MATLAB™ 6.5). Obviously, the long time required for evaluating the digital DRNL filter is the result of using an excessively high sampling rate. However, the digital implementation still took 5.5 s when the sampling rate was ten times smaller (10^5 Hz).

IV. AN EXAMPLE APPLICATION OF THE APPROXIMATE ANALYTIC TRANSFER FUNCTION: MODELING BASILAR-MEMBRANE RESPONSES TO PURE TONES

Previous dedicated reports have shown that the digital DRNL filter reproduces to a good approximation BM responses to pure tones (Meddis *et al.*, 2001; Sumner *et al.*, 2002, 2003b; Lopez-Najera *et al.*, 2003). For this purpose, they compared the *peak* amplitude of the output waveform

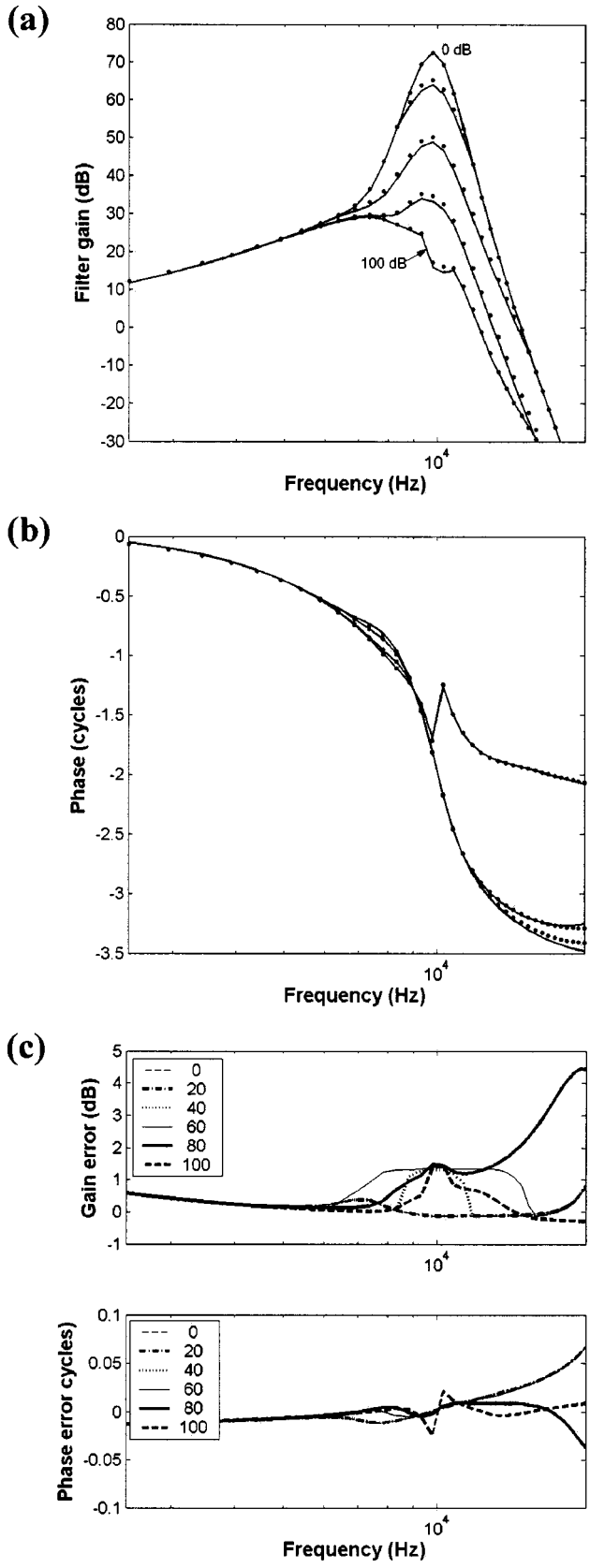


FIG. 3. A comparison between the response of the transfer function (continuous line) and the digital DRNL filter (dots). Different lines illustrate the response to different input amplitudes corresponding to levels ranging from 0 to 100 dB SPL in 20-dB steps. The amplitudes were further scaled down by 1.5×10^{-5} across frequencies (see the main text for details). (a) Gain (dB). Note the overlap between the lines for 0- and 20-dB input levels. (b) Phase (cycles). (c) Difference (digital minus transfer function) between the results obtained with the two methods.

TABLE I. Parameters for the DRNL filter used to produce the results shown in this report (reproduced from Table I of Meddis *et al.*, 2001, set L113).

Parameter	Value
Linear path	
n	2
m	4
g	100
f_{cGTL} (Hz)	8 000
B_{GTL} (Hz)	2 000
f_{uLPL} (Hz)	8 000
Nonlinear path	
p	3
q	3
a	12 000
b	0.057
c	0.25
f_{cGTN} (Hz)	9 800
B_{GTN} (Hz)	1 400
f_{uLPN} (Hz)	9 800

from the filter in response to pure tones of various frequencies and levels against corresponding experimental data (e.g., Meddis *et al.*, 2001). Above, it has been shown that the approximate analytic transfer function resembles the level-dependent frequency response of the digital filter as measured from its peak response to pure tones. Therefore, it may be used to model BM iso-intensity or input/output curves.

An example is shown in Fig. 4, where the analytic transfer function (thick continuous line) is compared against BM iso-intensity curves (thin dotted lines) reported by Ruggero *et al.* (1997, Fig. 9). The response of the digital DRNL filter is also shown (thick dashed lines) for comparison. Both the digital and analytic versions of the filter were computed as described in Sec. III. This time, however, the experimental frequency response of the stapes was used as the input (crosses in Fig. 4). This was taken from Fig. 9 of Ruggero *et al.* (1997) and was assumed to grow linearly with level. Both filter evaluations were computed only for those frequencies for which experimental data were available. Although no attempt was made to adjust the original parameters [provided by Meddis *et al.* (2001) to model this specific data set], the fit of the analytic transfer function is reasonable and comparable to that of the digital filter. The total Euclidean distance to the data was comparable for both evaluations: 57.1 dB for the digital, and 58.4 dB for the analytical version.

The procedure for optimizing the parameters of the approximate transfer function to model other data sets would be identical to that for the digital DRNL filter. The latter is described in detail elsewhere (Meddis *et al.*, 2001; Lopez-Poveda and Meddis, 2001). This may seem surprising at first, given that the digital filter is intrinsically nonlinear for each level, whereas the proposed transfer function is linear (but different for each level). However, it is noteworthy that the procedures of Meddis *et al.* (2001) and Lopez-Poveda and Meddis (2001) were based on adjusting the digital filter to reproduce experimental frequency responses in the form of input/output (or their equivalent iso-intensity) curves, such as those shown in Fig. 4. They then showed that other nonlinear properties (such as level-dependent frequency selectivity and phase, distortion, or suppression) *emerge* naturally from the filter as a result of its characteristic dual-resonance architecture. It follows from their procedure that the analytic transfer function may be used as a fast tool to optimize the parameters of the digital DRNL filter.

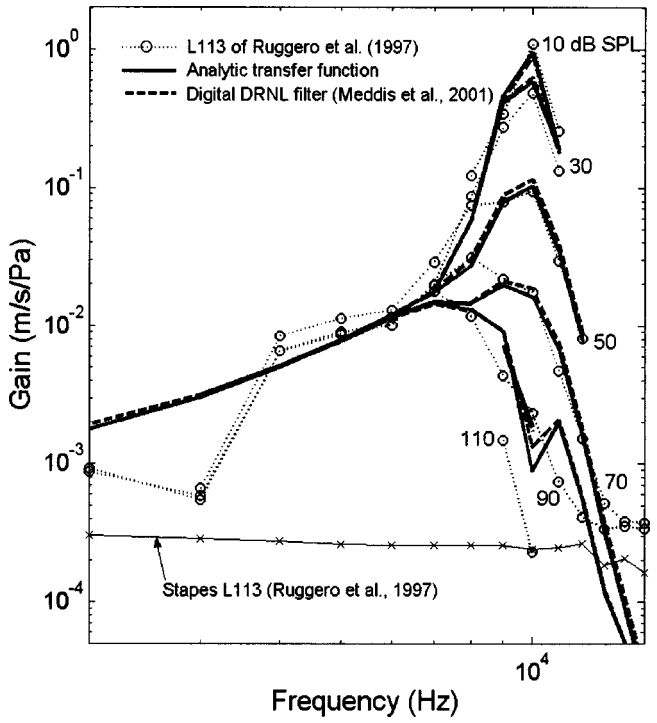


FIG. 4. A comparison of the approximate analytic transfer function (thick continuous line) with the iso-intensity gain BM data (dotted line, open circles) of Fig. 9 of Ruggero *et al.* (1997). The response of the digital DRNL filter is also shown (thick dashed line). Both versions of the DRNL filter were evaluated with identical parameters (given in Table I). They were computed for the same frequencies for which experimental data were available and for input levels ranging from 10 to 110 dB SPL in steps of 20 dB. These are indicated by the numbers next to each curve. Note that the fit of the analytic transfer function is comparable to that for the digital DRNL filter (see the main text for details).

It must be acknowledged, however, that Lopez-Najera *et al.* (2003) suggest that a more realistic nonlinear behavior is achieved when the parameters are optimized to fit *simultaneously* the *amplitude* and *phase* aspects of the BM response, and not only the amplitude aspect as Meddis *et al.* (2001) or Lopez-Poveda and Meddis (2001) did. The analytic transfer function is still valid for this purpose, as it also allows calculating the phase response of the digital DRNL filter.

V. DISCUSSION

The transfer function has been derived on an approximation (see above) that maintains the gain and the phase properties of the digital DRNL filter *almost* intact, but disregards the effect of the nonlinearity on the spectral content of its output [Fig. 2(b)]. That is, the transfer function does not preserve the distortion harmonics observed in output signal from the digital nonlinearity for a sinusoidal input. This defect is minimized by the fact that the second cascade of GT filters in the nonlinear path attenuates any high-order har-

monics generated by the original nonlinearity. The degree of attenuation depends on the order and the bandwidth of these filters.

It is noteworthy that our approach for developing the approximate transfer function of the nonlinear gain, $U(f)$, does not require the particular form of the nonlinearity used in the DRNL filter. Any other nonlinear gain that does not deviate too much from a linear function over the range of input amplitudes of interest could also be ''linearized.''

The approximate transfer function does not preserve suppression and distortion phenomena characteristic of the digital DRNL filter when it is operating in its compression region. Therefore, it must be used with care for evaluating the output spectrum of the digital DRNL filter in response to *multitonal* or *broadband* stimuli whose peak amplitude (after allowing for filtering through the first GT cascade in the nonlinear path), V, exceeds the compression threshold amplitude, V_c. However, it provides accurate spectrum estimates for *any* stimuli such that $V<V_c$. For stimuli with very large amplitudes, generally such that $V\gg(g/b)^{1/(c-1)}$, the output from the linear path, R_L, is much larger than that from the nonlinear path, R_N. In this case, the DRNL filter behaves almost linearly and its approximate analytic transfer function also provides a good estimate of the output spectrum from the digital filter in response to *any* stimuli.

The DRNL filter, and more clearly its approximate analytic transfer function, suggest that auditory filters may be described by the added output from two parallel resonances, both intrinsically independent of level. It is the relative contribution of each resonance to the total filter output, controlled by its memoryless nonlinear gain, that confers the DRNL filter level-dependent frequency selectivity and gain. This scheme resembles de Boer's two-component EQ-NL theory (de Boer, 1997). According to this theory, the BM impedance is composed of two components, both intrinsically independent of level; one corresponds to the impedance of the ''passive'' BM, and one ''extra'' impedance that de Boer relates to the outer hair cells. The relative contribution of each component to the total BM impedance is controlled by a level-dependent factor whose role resembles that of the memoryless nonlinear gain of the DRNL filter. Interestingly, de Boer's theorem also asserts the existence of linear approximations to nonlinear auditory filters in certain circumstances (e.g., for wideband random noise of very low amplitude).

ACKNOWLEDGMENTS

I thank Ray Meddis and Reinhold Schatzer for their invaluable suggestions, and Alberto Lopez-Najera for his comprehensive reviewing of the formulation. I am also grateful to two anonymous reviewers for their comments on earlier versions of the manuscript. The author carried out this work on a research contract of the ''Ramón y Cajal'' Program of the Spanish Ministry of Science and Technology. Work supported by Instituto de Salud Carlos III (FIS PI020343 and G03/203).

[1]The digital DRNL filter was implemented in MATLAB™ as described in the Appendix of Lopez-Poveda and Meddis (2001). It uses digital IIR GT filters derived from the continuous-time GT impulse response by applying the impulse-invariance technique (Stone, 1995). Slightly different results from those shown in Fig. 3 would have been obtained if digital GT filters derived with the bilinear transformation had been used (R. Schatzer, personal communication, 2003). Slaney (1993) discusses different digital implementations of the GT filter. The bilinear GT filter provides a closer match to the *phase* of its transfer function for all frequencies and for the two sampling rates tested (10^5 and 10^6 Hz). Its *gain* is also closer to the transfer function for all conditions except for the lower sampling rate (10^5 Hz) *and* frequencies much higher than f_c, where the impulse-invariance GT filter performs better. MATLAB™ implementations of the digital DRNL filter and of its approximate transfer function are available from the author.

de Boer, E. (**1997**). ''Connecting frequency selectivity and nonlinearity for models of the cochlea,'' Aud. Neurosci. **3**, 377–388.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

Händel, P. (**2000**). ''Properties of the IEEE-STD-1057 four-parameter sine wave fit algorithm,'' IEEE Trans. Instrum. Meas. **49**, 1189–1193.

Hartmann, W. M. (**1998**). *Signals, Sound, and Sensation* (AIP, Springer, New York).

Lopez-Poveda, E. A., and Meddis, R. (**2001**). ''A human nonlinear cochlear filterbank,'' J. Acoust. Soc. Am. **110**, 3107–3118.

Lopez-Najera, A., Meddis, R., and Lopez-Poveda, E. A. (**2003**). ''A computational algorithm for computing nonlinear auditory frequency selectivity: Further studies,'' in Proceedings of the 13th International Symposium on Hearing, Dourdan, France.

Meddis, R., O'Mard, L. P., and Lopez-Poveda, E. A. (**2001**). ''A computational algorithm for computing nonlinear auditory frequency selectivity,'' J. Acoust. Soc. Am. **109**, 2852–2861.

Moore, B. C. J. (**1998**). *Cochlear Hearing Loss* (Whurr, London).

Oppenheim, A. V., Shafer, R. W., and Buck, J. R. (**1999**). *Discrete-time Signal Processing*, 2nd ed. (Prentice-Hall, Englewood Cliffs, NJ).

Robles, L., and Ruggero, M. A. (**2001**). ''Mechanics of the mammalian cochlea,'' Physiol. Rev. **81**, 1305–1352.

Ruggero, M. A., Rich, N. C., Recio, A., Shyamla Narayan, S., and Robles, L. (**1997**). ''Basilar membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.

Slaney, M. (**1993**). ''An efficient implementation of the Patterson–Holdsworth auditory filter bank,'' Apple Computer Technical Report #35. Apple Computer Inc.

Smith, J. O. (**2002**). ''Introduction to Digital Filters,'' Center for Computer Research in Music and Acoustics (CCRMA), Stanford University. Web-published at http://www-ccrma.stanford.edu/~jos/filters/.

Stone, M. A. (**1995**). ''Spectral enhancement for the hearing impaired,'' Ph.D. dissertation. University of Cambridge, UK.

Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (**2002**). ''A revised model of the inner hair cell and auditory-nerve complex,'' J. Acoust. Soc. Am. **111**, 2178–2188.

Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (**2003a**). ''Adaptation in a revised inner-hair cell model,'' J. Acoust. Soc. Am. **113**, 893–901.

Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (**2003b**). ''A nonlinear filter-bank model of the guinea-pig cochlear nerve,'' J. Acoust. Soc. Am. **113**, 3264–3274.

Wilson, B. S., Brill, S. M, Cartee, L. A., Cox, J. H., Lawson, D. T., Schatzer, R., and Wolford, R. D. (**2002**). ''Speech processors for auditory prostheses,'' Final report, NIH project N01-DC-8-2105.

Spectral pattern, harmonic relations, and the perceptual grouping of low-numbered components

Brian Roberts[a)]
School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, England

Jeffrey M. Brunstrom
Department of Human Sciences, Loughborough University, Loughborough, Leicestershire LE11 3TU, England

(Received 13 September 2002; revised 8 May 2003; accepted 14 July 2003)

Mistuning a harmonic increases its salience and produces an exaggerated change in its pitch. The effects on component grouping of spectral pattern, global pitch, and local harmonicity were explored using these phenomena. Stimuli were either harmonic (F0=200 Hz) or frequency shifted by 25% of F0. Component 1 or 2 was replaced by one of a set of sinusoidal probes in the same spectral region. Listeners either matched the probe pitch by adjusting the frequency of a pure tone (experiments 1 and 3) or matched the probe loudness by adjusting the level of a tone of identical frequency (experiment 2). Probe positions corresponding to greatest perceptual fusion were estimated from the variations in pitch shift and loudness across frequency. Both measures gave similar estimates. For harmonic stimuli, fusion was greatest at harmonic values. For shifted stimuli, fusion was greatest close to the suboctave (225 Hz) and the frequency (450 Hz) of component 2. The latter value moved downwards to near 433 Hz (2:3 ratio with component 3) when component 1 was removed. Together, these results indicate that the lowest component of a shifted complex is grouped by local harmonicity, whereas the higher components are grouped by common spectral spacing. Global pitch did not influence component grouping.
[DOI: 10.1121/1.1605411]

PACS numbers: 43.66.Ba, 43.66.Cb, 43.66.Hg [MRL]

I. INTRODUCTION

A. The relationship between component grouping and global pitch

Periodic complex tones are usually heard as single entities with a characteristic pitch corresponding to the fundamental frequency (F0). Hence, it has long been assumed that spectral fusion (the integration of the individual partials of a complex tone into a unified percept) and the perception of global pitch (the pitch corresponding to F0) arise from a common mechanism that responds to harmonic relations between partials. Two observations in particular lend credence to this assumption. First, a mistuned harmonic in an otherwise periodic sound typically makes a reduced contribution to global pitch (Moore *et al.*, 1985; Darwin *et al.*, 1994) and stands out as a salient, separate, acoustic event (Moore *et al.*, 1986; Hartmann *et al.*, 1990; Brunstrom and Roberts, 1998). Second, the global pitch of a frequency-shifted complex tone (derived from a harmonic stimulus by adding a constant frequency increment to each component) can be predicted well in terms of the best fit of a harmonic template (e.g., Duifhuis *et al.*, 1982) that weights more heavily information derived from components falling in the so-called dominant region for pitch (typically, harmonics 3–5; see Ritsma, 1967; Moore *et al.*, 1985).

Over the last decade or so, this orthodoxy has been challenged by exploration of the perceptual organization of complex tones composed of a variety of spectral structures. Three findings are of particular note. First, the mistuning of a component in frequency-shifted stimuli, or stimuli with a moderate degree of spectral stretch (i.e., a progressive increase in spacing between consecutive partials with increasing component number), results in changes in the salience of that component like those seen for harmonic stimuli (Roberts and Brunstrom, 1998, 2001). Second, the fusion of individual components in a frequency-shifted stimulus is best predicted by the common pattern of spectral spacing rather than the global pitch (Brunstrom and Roberts, 2000). Third, a single even harmonic added to an otherwise odd-harmonic complex is typically more salient than its odd neighbors, because it is not consistent with the regular pattern of spectral spacing defined by them (Roberts and Bregman, 1991; Roberts and Bailey, 1993a, b, 1996a, b; Roberts, 1998). In contrast, the added even harmonic does not make a reduced contribution to the computation of global pitch (Ciocca, 1999).

Roberts and Brunstrom (1998) first proposed the notion of separate templates for global pitch and for spectral fusion. They envisaged the spectral-grouping template[1] to be sensitive (within a limited range) to deviations from equal spacing between consecutive components, rather than directly sensitive to harmonic relations. Roberts and Brunstrom (2001) developed this proposal in terms of an autocorrelogram representation of auditory responses. In particular, it was speculated that global-pitch perception relies on a cross-channel *aggregation* of periodicity information akin to the summary autocorrelation function (SACF; Meddis and Hewitt, 1991; Meddis and O'Mard, 1997), but that the perceptual segregation of a mistuned component relies on a cross-channel *com-*

a)Electronic mail: b.roberts@bham.ac.uk

 0001-4966/2003/114(4)/2118/17/$19.00

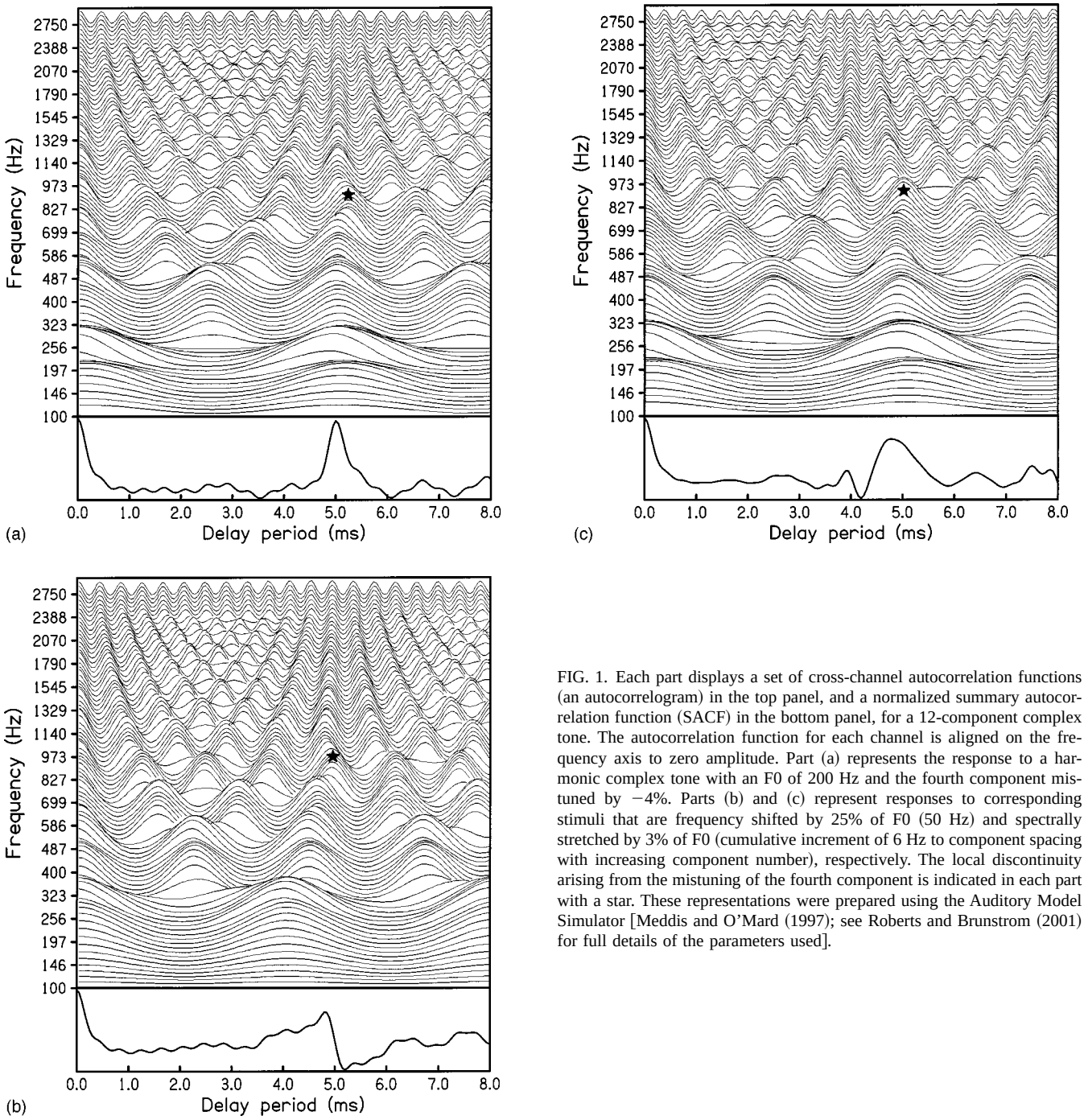


FIG. 1. Each part displays a set of cross-channel autocorrelation functions (an autocorrelogram) in the top panel, and a normalized summary autocorrelation function (SACF) in the bottom panel, for a 12-component complex tone. The autocorrelation function for each channel is aligned on the frequency axis to zero amplitude. Part (a) represents the response to a harmonic complex tone with an F0 of 200 Hz and the fourth component mistuned by −4%. Parts (b) and (c) represent responses to corresponding stimuli that are frequency shifted by 25% of F0 (50 Hz) and spectrally stretched by 3% of F0 (cumulative increment of 6 Hz to component spacing with increasing component number), respectively. The local discontinuity arising from the mistuning of the fourth component is indicated in each part with a star. These representations were prepared using the Auditory Model Simulator [Meddis and O'Mard (1997); see Roberts and Brunstrom (2001) for full details of the parameters used].

parison of periodicity information. Parts (a)–(c) of Fig. 1 each present an autocorrelogram (top panel) and a normalized SACF (bottom panel) for exemplar harmonic, shifted, and stretched stimuli, respectively. Each of these stimuli contains a fourth component that has been mistuned from its original value by −4%. For harmonic stimuli, the alignment of common periods across frequency channels produces a characteristic vertical ''spine'' (at the pitch period of the stimulus) in the autocorrelogram. The modification of a harmonic stimulus by applying increasing degrees of frequency shift or spectral stretch causes this spine to become increasingly curved, because the originally common periods become misaligned across channels, but the spine maintains a smooth trajectory. It can be seen that mistuning a single component in all three cases introduces a discontinuity that appears as a ''kink'' in the trajectory of the spine (marked with a star). In principle, this local discontinuity could be detected by cross-channel comparison.

B. Why explore the grouping of low-numbered components?

Roberts and Brunstrom's (2001) qualitative model of how a mistuned component might segregate perceptually from an otherwise regularly structured complex tone owes much to the empirical study of the perceptual organization of frequency-shifted complex tones (Roberts and Bailey, 1996b; Roberts and Brunstrom, 1998; Brunstrom and Roberts,

2000). However, it should be noted that the most detailed of these studies (Brunstrom and Roberts, 2000) focused on a restricted range of components (6–8). In contrast, little attention has been paid to the grouping of the lowest-numbered components. Unlike components 6–8, these partials are highly resolved (Shackleton and Carlyon, 1994). Furthermore, they deviate most from harmonic relations with their neighbors in frequency-shifted stimuli.[2] What little is known about the grouping of the lowest-numbered components indicates that their role in determining the perceptual cohesion of complex tones merits a thorough examination. This evidence is now reviewed:

(1) Roberts and Bailey (1996a) observed that clarity ratings (a measure of salience) for all the higher components in an odd-harmonic complex increased markedly when the fundamental component was removed (harmonics up to the ninth were tested). This result suggests a reduction in the overall perceptual cohesion of the complex. It also highlights the distinction between the factors affecting the grouping of a given partial and the contribution of that partial to the ''perceptual glue'' that holds together the whole complex as a coherent entity. No study of this kind has been carried out using consecutive-harmonic stimuli.
(2) Brunstrom and Roberts (2000) explored the perceptual cohesion of harmonic and shifted stimuli using a hit-rate measure. Hit rate was defined as the proportion of occasions on which listeners adjusted the frequency of a pure tone near to that of a mistuned target rather than near to that of any other partial (Hartmann *et al.*, 1990). Brunstrom and Roberts (2000) noted that shifted stimuli, though perceptually coherent, are less so than otherwise similar harmonic stimuli (difference in hit rate of at least 10%). The available data for component 2, the lowest tested by Roberts and Brunstrom (1998, 2001), indicate a tendency for hit rates to be much lower when it is mistuned in a shifted rather than in a harmonic complex. This effect of spectral context was marginal for the higher resolved components tested, suggesting that it may reflect the increasing deviation of the lowest components in shifted stimuli from harmonic relations.[2] The pattern of errors made when attempting to match the pitch of a mistuned target is consistent with this interpretation (Roberts and Brunstrom, 2001). Specifically, it was found that the main effect of increasing the degree of frequency shift for a complex tone (from 0% to 50% of F0) was to increase the likelihood of matching a low nontarget component, especially component 1.
(3) Roberts and Bailey (1996b) observed that frequency shifting an odd-harmonic complex progressively up to 30% increased the clarity rating for the first component by more than that of any other. Note that an upward frequency shift to the odd harmonics of F0 is equivalent to a downward frequency shift to the consecutive harmonics of 2F0.

Given these observations, our study has focused in depth on the grouping of the first two spectral components. This involved measuring the effect of mistuning these components from their original frequencies on the strength of their spectral fusion with the other components of a complex tone.

C. Partial-pitch shifts and the grouping of low-numbered components

Hartmann *et al.* (1990) discovered that, in addition to an increase in salience, mistuning a harmonic produces small but systematic mismatches between its frequency and the preferred matching frequency of an adjustable pure tone. These mismatches, known as *partial-pitch shifts*, are typically exaggerations of the mistuning imposed on the partial (i.e., upward pitch shifts for positive mistunings, and downward shifts for negative mistunings). Hartmann and Doty (1996) took measures across a range of mistunings to build up a *pitch-shift profile* for each component tested. Beyond a certain degree of mistuning, which varied with component number, these profiles typically showed saturation or supersaturation effects (i.e., the magnitude of the pitch shift peaked and then stabilized or declined).

de Cheveigné (1997) proposed that these pitch shifts arose through an interaction between the noisy internal representation of a mistuned partial and the operation of a harmonic template (Duifhuis *et al.*, 1982). The template is envisaged as possessing a series of ''slots'' at harmonic multiples of its F0 that act to inhibit the individual pitches of in-tune partials (Brunstrom and Roberts, 1998). For a small amount of mistuning, the degree of segregation of the component is assumed to be greater when its frequency is represented as further away from the center of a template slot. This leads to a skewed distribution in estimates of its pitch that is increasingly under-represented near the slot center, and hence displaces its mean pitch away from the slot center in both directions. The pitch-shift effect will saturate, and then decline, with increasing mistuning as the component falls outside the slot's range of influence.

Lin and Hartmann (1998) condensed Hartmann and Doty's procedure by measuring pitch shifts for components mistuned from harmonic values by +8% and −8% and then computing the difference between them. This value was called the pitch-shift gradient and was typically positive even for mistuned harmonics isolated from the rest of the complex by gaps of one or more harmonic positions. Given that harmonics other than the immediate neighbors could induce pitch shifts on the mistuned partial, Lin and Hartmann (1998) also concluded that these shifts arise centrally from an interaction between the mistuned partial and a slot in a harmonic template (cf. de Cheveigné, 1997).

Roberts and Brunstrom (1998) and Brunstrom and Roberts (2001) observed that similar partial-pitch shifts are associated with mistuned partials in frequency-shifted stimuli, suggesting that they arise from the process of spectral fusion rather than from the computation of global pitch. Furthermore, Hartmann and Doty's (1996) data suggest that it should be possible to estimate the frequency of a maximum in fusion from a pitch-shift profile. This is unnecessary in the harmonic case, for which the in-tune frequencies are predefined, but offers a potentially useful tool for exploring the perceptual organization of other kinds of complex tone. Lin and Hartmann's (1998) finding that mistuned low harmonics,

especially the fundamental component, are associated with the largest pitch shifts suggests that this measure is particularly suitable for an investigation of the grouping of the first two spectral components in harmonic and frequency-shifted stimuli.

II. HYPOTHESES AND PREDICTIONS

We now consider three alternative hypotheses that predict different values for the points of maximum spectral fusion in the region encompassing components 1 and 2 for frequency-shifted stimuli. The values quoted are based on stimuli with an F0 of 200 Hz and a frequency shift of 25% of F0 (i.e., a frequency increment of 50 Hz for each partial):

(1) If a common pattern of spectral spacing determines the grouping of components 1 and 2 (cf. Brunstrom and Roberts, 2000, for components 6–8), then the predicted points of maximum fusion are 250 and 450 Hz.
(2) Following Brunstrom and Roberts (2000), the global pitch of our frequency-shifted stimuli can be estimated from the data of Patterson (1973). In his study, listeners adjusted the F0 of a harmonic stimulus (pulse train) to match as closely as possible the perceived global pitch(es) of a set of frequency-shifted stimuli. Stimuli corresponding closely to ours were associated with a primary pitch of about 211 Hz and a secondary pitch of about 94 Hz. The lower value arises from an octave error, such that the shifted stimulus is processed as approximating a set of odd harmonics. Our stimuli differ from those used by Patterson in that the frequency of either component 1 or 2 is mistuned from its original value. However, Patterson found that components 3–5 are weighted most strongly in the perception of pitch, and it is clear from his findings that our manipulation of the lowest two components will have little effect on these estimates of global pitch. If the grouping of components 1 and 2 is influenced by the primary pitch, then the predicted points of maximum fusion are 211 and 422 Hz (i.e., one octave higher). Any effects of the secondary, suboctave, pitch should be manifest at harmonic multiples of 94 Hz falling in the general region of the components tested (i.e., at 188, 282, 376, and 470 Hz). Finding fusion maxima at harmonic multiples of the global pitches would support the arguments of Bregman (1990) and Hartmann (1996) regarding spectral grouping.
(3) Effects of harmonicity need not necessarily be mediated by global-pitch processing, but may instead be more local in character. Demany and Semal (1988) reported that listeners are sensitive to an octave relationship between two simultaneous and dichotically presented pure tones. Specifically, they demonstrated that listeners were better able to distinguish in-phase from out-of-phase frequency modulation 2-Hz rate of the two components when those components had a starting ratio of exactly 1:2 rather than 1:2±100 cents. Therefore, it is possible that an octave relationship between components 1 and 2 may influence their grouping in our frequency-shifted stimuli. Specifically, when the first component is fixed at 250 Hz, the greatest fusion of the probe in the upper range may occur when the probe frequency is 500 Hz. A similar argument can be made for a probe in the lower range at 225 Hz when the second component is fixed at 450 Hz.

III. EXPERIMENT 1

The aim was to use partial-pitch shifts to determine whether the auditory grouping of the lowest two partials in a frequency-shifted complex tone is governed by spectral pattern or by (one or more) other factors. An answer to this question should help to elucidate further the mechanism responsible for spectral fusion and segregation in complex-tone perception.

A. Method

1. Subjects

Six individuals volunteered for this study. All were young adults who reported normal hearing and who were paid for their assistance.

2. Overview and conditions

Harmonic complex tones consisted of the first 12 components of a nominal F0 of 200 Hz (actual value chosen randomly from a rectangular distribution with a width of ±10% around 200 Hz, quantized in steps of 1%). Frequency-shifted complex tones were derived from their harmonic counterparts by incrementing the frequency of each harmonic by 25% of F0 (i.e., 50 Hz when F0=200 Hz). For both harmonic and shifted stimuli, either component 1 or 2 was removed and replaced by one of a set of sinusoidal probes in the same spectral region (the other components of the complex constitute the *spectral frame*). The task of the listeners was to match the pitch of the probe as accurately as possible by adjusting the frequency of a pure tone.

In the harmonic conditions, the two sets of probes corresponded to frequencies of ±2.5%, ±5%, ±10%, and ±15% with respect to component positions 1 and 2. The range was great enough to encompass any saturation and supersaturation effects in the associated pitch-shift profiles (Hartmann and Doty, 1996). For a nominal F0 of 200 Hz, the lower and higher sets of probes spanned 170 to 230 Hz and 340 to 460 Hz, respectively. The sets of probes used in the shifted conditions were greater in range and more finely spaced in frequency to help distinguish between the alternative predictions. For a nominal F0 of 200 Hz, the lower set of probes spanned 170–300 Hz (component 1 = 250 Hz) in 10-Hz steps, and the upper set spanned 360–540 Hz (component 2 = 450 Hz) in 20-Hz steps, with a finely spaced region (420–500 Hz) in 10-Hz steps.

Spectral positions associated with maximum perceptual fusion were estimated for each probe set from the profile describing the pitch shifts across that set. How these estimates were computed is explained in Sec. III A 5. The expected values are known *a priori* for the harmonic case, and so the estimates computed using the data for the harmonic conditions can be used to assess the veracity of this measure. Predicted values for the shifted case depend on the factor assumed to determine spectral fusion, as discussed earlier

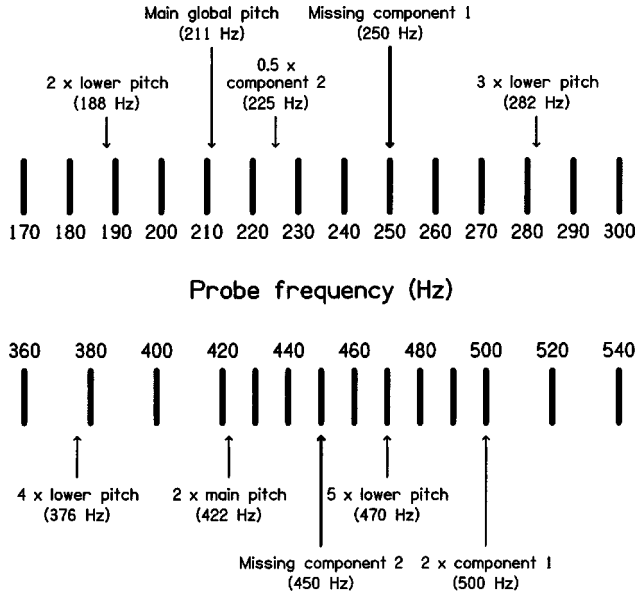


FIG. 2. Schematic summarizing predicted points of maximum fusion for the shifted stimuli used in experiment 1, according to the three hypotheses under investigation. The upper part of the figure illustrates the set of probes plus predicted values for condition S1, and the lower part illustrates the set of probes plus predicted values for condition S2. A nominal F0 of 200 Hz is assumed. The different hypotheses predict fusion maxima based on spectral pattern (frequencies of the missing components), global pitch (harmonics of the main or suboctave pitch), and local harmonic relations (one octave above or below the frequency of an adjacent component).

(see Sec. II). These values are summarized in Fig. 2 for the two frequency-shifted conditions (assuming an F0 of 200 Hz and a frequency shift of 50 Hz).

3. Stimulus synthesis and parameters

Each trial consisted of a repeating cycle of a complex tone (spectral frame plus probe) followed by an adjustable pure tone. Both tones were 540 ms in duration, including linear onset and offset ramps of 20 ms each. The within- and between-cycle silent intervals were 400 and 800 ms, respectively. These values were sufficient to exclude the possibility of pitch shifts on the adjustable tones resulting from prior stimulation in the same frequency region (Rakowski and Hirsh, 1980). All of the components were in sine phase and of equal amplitude, each set to 60 dB SPL. The adjustable pure tone was set to 57 dB SPL. The matching range of the adjustable pure tones was limited to within $\pm 3\%$ of the probe frequency and was quantized in steps of 0.1 Hz. This range was chosen because it was broad enough to accommodate the magnitude of any expected partial-pitch shift (see Hartmann *et al.*, 1990; Hartmann and Doty, 1996; Roberts and Brunstrom, 1998), yet narrow enough to prevent listeners from matching the pure tone to an adjacent component in the spectral frame. The initial frequency of the pure tone was set randomly within the range of adjustment.

All stimuli were generated using MITSYN software (Henke, 1997). Stimuli were synthesized at a sampling rate of 16 kHz and played back via a 16-bit digital-to-analog converter (Data Translation DT2823). They were low-pass filtered (Fern Developments EF16X module; corner frequency=5.2 kHz, roll-off=100 dB/oct) and presented binaurally over Sennheiser HD 480-13II earphones. The levels of the stimuli were set using a programmable attenuator and calibrated using a sound-level meter (Brüel and Kjaer, type 2209) connected to the earphones by an artificial ear (type 4153). Stimuli were played to the listeners in a double-walled sound-attenuating chamber (Industrial Acoustics 1201A).

4. Procedure

Subjects were instructed to match the adjustable tone to the pitch of a salient pure tone (the probe) embedded in the complex. The adjustment was made using a trackball mouse. If the upper or lower limit of the range was reached during an adjustment, the message ''top of range'' or ''bottom of range'' was displayed, as appropriate. The buttons on the mouse were used to begin and end each trial. The number of cycles within each trial was not restricted, and subjects were encouraged to be as accurate as possible in making their adjustments.

The two factors (type of spectral frame, probe range) were combined to give four conditions: harmonic and lower range (H1), harmonic and upper range (H2), shifted and lower range (S1), shifted and upper range (S2). Each condition was run as four blocks of trials (H1, H2 = 4 repetitions $\times$ 8 probe values = 32 trials; S1, S2 = 4 repetitions $\times$ 14 probe values = 56 trials). A new randomized order was used in each block. Experimental sessions focused on either component 1 (H1 followed by S1) or component 2 (H2 followed by S2), giving a total of 32+56=88 trials per session. Harmonic stimuli were always run first in a session to optimize the accuracy of matches for the shifted stimuli. Each session typically lasted about an hour, and three sessions of the same type were run (on different days) to obtain 12 judgments per stimulus for each subject (grand total=88$\times$3$\times$2=528 trials). Three subjects began with the sessions focusing on component 1 (H1–S1) and the other three began with the sessions focusing on component 2 (H2–S2).

Subjects were trained using a reduced set of harmonic stimuli in which component 2 was replaced by one of six probes set to $\pm 2.5\%$, $\pm 5\%$, and $\pm 10\%$ from the in-tune frequency. Listeners were first familiarized with the task by judging two blocks of stimuli (12 trials). They then judged a further six blocks of stimuli (72 trials) for practice. The entire session took about an hour. The acceptance criterion for the main study was a pitch shift at least 0.5% greater for the $+5\%$ probe than for the -5% probe (based on the findings of Hartmann and Doty, 1996). All subjects met this criterion.

5. Data analysis

For each subject, the frequency of the pitch match for a given stimulus was defined as the mean of the 12 judgments made. A partial-pitch shift is defined as the percentage deviation of the frequency of the pitch match to a probe from the true frequency of that probe. This definition follows the practice of earlier studies (Hartmann and Doty, 1996; Lin and Hartmann, 1998; Roberts and Brunstrom, 1998; Brunstrom and Roberts, 2001).

We developed an approach to automating the estimation of a fusion maximum from a pitch-shift profile that involved fitting a cubic polynomial [the lowest-order function that can fit effectively a profile of the form seen by Hartmann and Doty (1996)] and then computing the frequency at the inflection. The inflection frequency corresponds to the midpoint between the frequencies of the minimum and maximum of the cubic function fitted to the profile. The zero crossing (corresponding to a veridical frequency match to the target component) is not a suitable measure. This is because factors like asymmetric masking (Terhardt, 1971) may also influence the pitch of the probe, leading to an offset in the profile with respect to the zero crossing. This issue is considered further in Sec. III C. The effectiveness of the inflection as a measure of points of maximum fusion in shifted stimuli can be assessed using data from the harmonic conditions.

B. Results

Figure 3 displays the mean matches and intersubject standard errors when expressed in terms of pitch shifts. The top and bottom panels show the profiles for conditions H1 and H2, respectively. Both profiles are well described in terms of a cubic polynomial fit to the mean data. The inflection frequencies for these functions are 201.7 and 403.6 Hz, respectively, which are both within 1% of the corresponding in-tune frequencies (200 and 400 Hz).

Figure 4 displays mean pitch shifts, with standard errors, for conditions S1 (top panel) and S2 (bottom panel). It was necessary to exclude data from the extremes of the probe range to obtain a good cubic fit ($r^2>0.7$) for condition S1. This is perhaps unsurprising given the greater range of probes used in this condition. The fit was restricted to the range between the two probe positions below that giving the most negative pitch shift and above that giving the most positive shift. The inflection frequencies are 226.2 Hz for condition S1 and 449.3 Hz for condition S2. For comparison, owing to the restricted range used for the cubic fit to the S1 data, a quintic fit to the entire probe range was computed. The inflection frequency is identical to the midpoint between the frequency of the maximum and minimum for a cubic fit, but not for a quintic fit, and so the midpoint frequency is quoted instead. This value (226.0 Hz) is very close to that for the cubic fit.

The S2 inflection is very close to the frequency of the excised second component (450 Hz), which is consistent with the spectral-pattern hypothesis. In contrast, the S1 inflection is very close to one octave below the frequency of component 2 (i.e., 225 Hz) which is consistent with the local-harmonicity hypothesis. Furthermore, the position predicted on the basis of common spectral pattern (250 Hz) is almost at the maximum of the cubic fit, and the position predicted on the basis of the primary global pitch (211 Hz) is towards the minimum. There is no clear evidence that any other predicted value for conditions S1 and S2 is represented in their pitch-shift profiles.

To assess the significance of differences between the computed inflection frequencies and the values predicted for the shifted conditions by our three hypotheses, r^2 values were computed for a set of cubic fits to the data that were

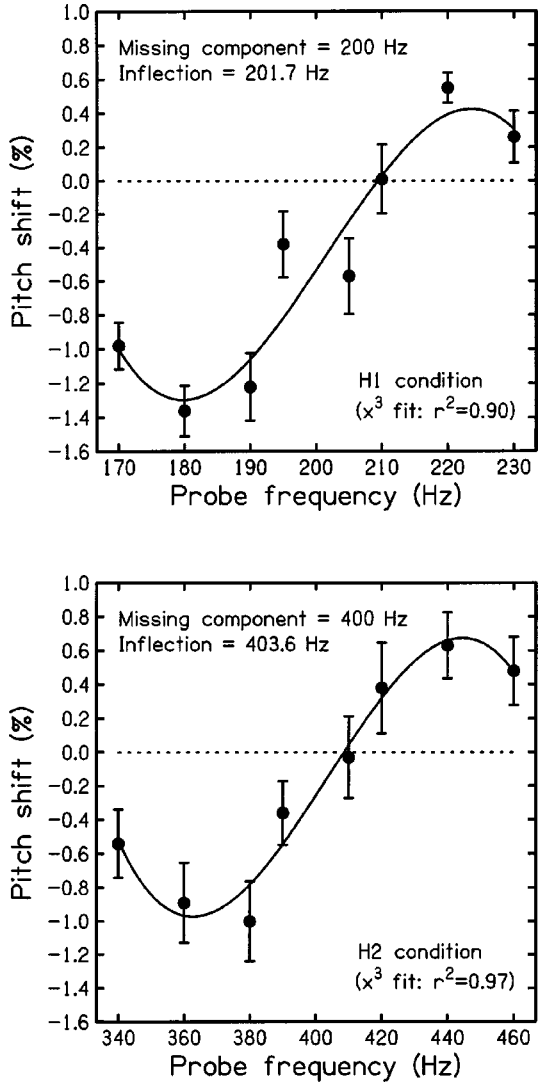
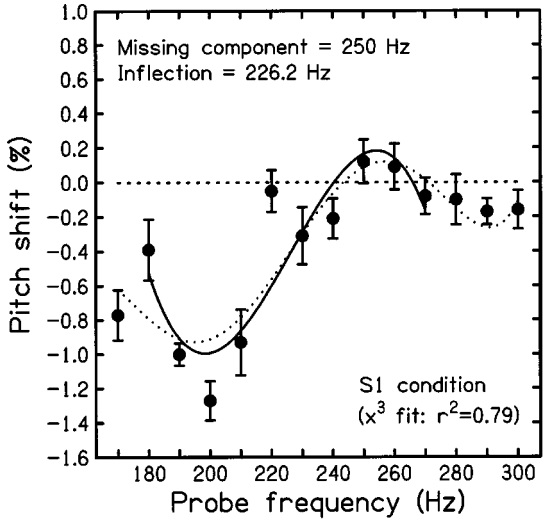


FIG. 3. Results for the harmonic conditions in experiment 1. The top and bottom panels indicate mean pitch shifts across six subjects (with intersubject standard errors) for H1 and H2, respectively. A nominal F0 of 200 Hz is assumed. The horizontal dashed lines indicate veridical matches to the frequency of the probe. Each panel shows the best cubic polynomial fit to the pitch-shift profile across the full probe range, and indicates the inflection frequency and r^2 value for this function.

constrained to have inflection frequencies (in 1-Hz steps) encompassing the predictions. Figure 5 shows the dependency of r^2 value on inflection frequency for the S1 condition (top panel) and the S2 condition (bottom panel). The predicted frequencies according to our three hypotheses are marked on these functions. It can be seen that the r^2 value is strongly dependent on inflection frequency in both conditions. By analogy with stepwise regression (see, e.g., Howell, 1997), the difference between an r^2 value based on a full model (unconstrained fit) and a reduced model (constrained fit to a predicted inflection frequency) can be tested for significance. For the S1 condition, only the value predicted by the local-harmonicity hypothesis (225 Hz) did not differ significantly from the full model [188 Hz: $F(1,7)=8.77$, $p<0.025$; 211 Hz: $F(1,7)=7.69$, $p<0.05$; 225 Hz: $F(1,7)=0.20$, ns; 250 Hz: $F(1,7)=7.00$, $p<0.05$]. For the S2 condition, only the value predicted by the spectral-pattern hypothesis (450 Hz) did not differ significantly from the full model [376 Hz:



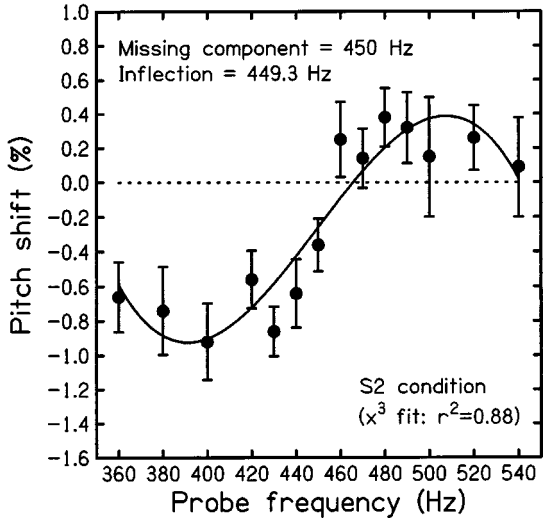
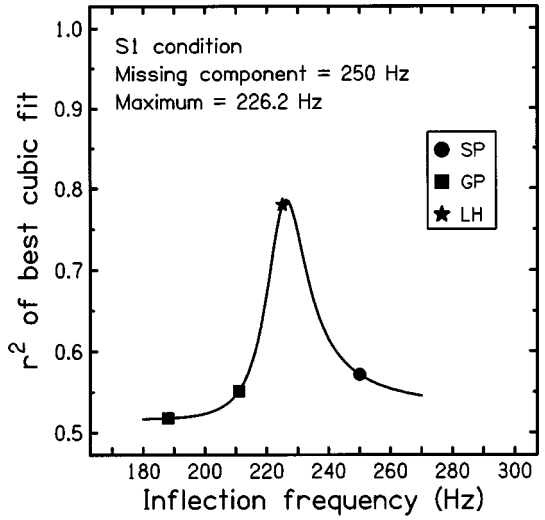


FIG. 4. Results for the frequency-shifted conditions in experiment 1. The top and bottom panels indicate mean pitch shifts across six subjects (with intersubject standard errors) for S1 and S2, respectively. A nominal F0 of 200 Hz is assumed. The horizontal dashed lines indicate veridical matches to the frequency of the probe. Each panel shows the best cubic polynomial fit to the pitch-shift profile (restricted probe range for S1, full probe range for S2), and indicates the inflection frequency and r^2 value for this function. For comparison, the top panel also shows the best quintic polynomial fit (dotted line) to the full probe range. The quintic fit ($r^2=0.74$) has a midpoint frequency of 226.0 Hz.



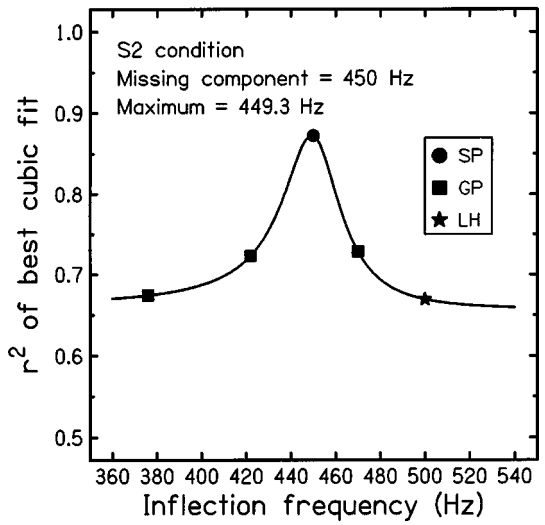


FIG. 5. Analysis of cubic fits to the data for the frequency-shifted conditions in experiment 1. The top and bottom panels display a function indicating how the r^2 value for a constrained cubic fit to the data varies with inflection frequency for S1 and S2, respectively. A nominal F0 of 200 Hz is assumed. r^2 values for inflection frequencies predicted by the spectral-pattern hypothesis (SP), the global-pitch hypothesis (GP), and the local-harmonicity hypothesis (LH) are marked on the r^2 function by circles, squares, and stars, respectively. Note that the SP hypothesis provides the best account for the S2 data, but that the LH hypothesis provides the best fit for the S1 data.

$F(1,11)=17.24$, $p<0.01$; 422 Hz: $F(1,11)=12.99$, $p<0.01$; 450 Hz: $F(1,11)=0.00$, ns; 470 Hz: $F(1,11)=12.47$, $p<0.01$; 500 Hz: $F(1,11)=17.67$, $p<0.01$].

All the pitch-shift profiles show evidence of a negative offset from the zero crossing. This effect is especially marked for the H1 and S1 profiles.

C. Discussion

The close correspondence between the inflection frequencies obtained for the harmonic conditions and the in-tune frequencies of components 1 and 2 supports the idea that our approach is an effective way of measuring points of maximum spectral fusion in a complex tone.

The results for experiment 1 suggest that the grouping of component 2 in a frequency-shifted complex is determined by regularity of spectral pattern, as has been demonstrated previously for higher-numbered components (6–8; Brunstrom and Roberts, 2000). This finding extends the generality of their findings. First, grouping-by-pattern has been demonstrated for a component that is highly resolved from its neighbors (Plomp, 1964). Second, this finding implies that we can interpolate our model to encompass the principal components in the dominant region for pitch (i.e., components 3–5; Ritsma, 1967).

More surprising is the finding that component 1 behaves differently from component 2. The inflection frequency for the S1 profile indicates a point of maximum fusion corresponding to a simple octave relationship between the first

two components. In other words, the grouping of component 1 appears to be governed by the harmonicity of its relationship with its only immediate neighbor. The asymmetry between the results for the two components is emphasized by the paucity of evidence in the S2 profile for a fusion maximum one octave above component 1 (i.e., at 500 Hz). The grouping of neither component appears to be related to the global pitches evoked by the shifted complex (cf. Brunstrom and Roberts, 2000). Earlier work using stimuli composed of odd harmonics plus a single even harmonic is consistent with the finding that an octave relationship is important for the grouping of components 1 and 2. Specifically, the characteristic higher salience of an added even harmonic is marginal or absent for harmonic 2 (e.g., Roberts and Bregman, 1991; Roberts, 1998), but reappears if the octave relationship is destroyed by frequency shifting or spectrally stretching the stimulus (Roberts and Bailey, 1996b).

One issue arising is the need to reconcile the finding for component 1 in our frequency-shifted stimulus with the data presented by Brunstrom and Roberts (2001) for monaural presentation of a similar stimulus. A mistuning of $\pm 4\%$ was imposed on the frequency of the first component (250 Hz) and the direction of the resulting pitch shift (higher or lower) was recorded. Mistuning by $+4\%$ (260 Hz) produced an upward pitch shift and mistuning by -4% (240 Hz) produced a downward shift. This was taken as indicating that the grouping of component 1 in shifted stimuli was governed by spectral pattern (i.e., the point of maximum fusion was 250 Hz). However, comparison with the current data set (Fig. 4, top panel) suggests that this interpretation arose from an artifact of sampling. It can be seen that a probe at 240 Hz produces a pitch shift below the zero crossing and a probe at 260 Hz produces one above. The pitch-shift difference between these probes is small, about 0.3%, but probably big enough to account for the earlier result. This highlights the risk of using a limited sampling density (cf. Lin and Hartmann, 1998) when points of maximum fusion are not known *a priori*. The profiling approach pioneered by Hartmann and Doty (1996) can protect again this kind of outcome.

The marked negative offset from the zero crossing for the H1 and S1 pitch-shift profiles is consistent with the asymmetric-masking hypothesis proposed by Terhardt (1971). Given that there is no component present below the probe, only the upper skirt of the probe's excitation pattern is exposed to partial masking effects. However, this hypothesis predicts a positive offset for the H2 and S2 profiles, rather than the small negative offset observed, because the upward spread of masking would have made the component below the probe a more effective masker (see Moore, 2003). This indicates that some other factor (either additional to, or instead of, asymmetric masking) must have influenced the pitch matches made by listeners. Whatever the cause, the presence of these offsets illustrates why the zero crossing should not be used to estimate points of maximum fusion.

The underlying assumption of the method we have employed is that maxima in spectral fusion can be estimated from the shape of the pitch-shift profile associated with mistuning a component to varying degrees. Support for our interpretation would be obtained if an alternative measure of salience can be shown to vary in a way consistent with these pitch-shift data. Experiment 2 attempted to verify our findings using this approach.

IV. EXPERIMENT 2

Measuring the salience of a component requires careful attention when designing the task. The ability of listeners to hear out a component from a complex tone is likely to be influenced by both frequency resolution and perceptual segregation. However, these factors are neither synonymous nor entirely independent of one another (Hartmann *et al.*, 1990; Roberts, 1998). Studies of frequency resolution usually try to optimize the conditions for analytical listening by cueing repeatedly the frequency region to be attended and by providing trial-by-trial feedback (e.g., Plomp, 1964). In contrast, the perceptual segregation of a component implies that it will capture a listener's attention automatically—a phenomenon known as ''pop out.'' Recent studies using brain event-related potentials (Alain *et al.*, 2001, 2002) have shown that the perception of a mistuned harmonic as a separate entity is associated with a biphasic potential. The early negative wave (known as object-related negativity) is present during both passive and active listening, but the later positive wave is present only during active listening. Alain *et al.* argue that the negative and positive waves reflect an automatic ''pop out'' process (primitive grouping) and an attentional process (schema-driven selection), respectively (see Bregman, 1990).

The crux of the problem of measuring component salience is in essence a ''quantum psychophysics'' effect—namely, the act of making the observation will affect the outcome. Some studies have attempted to minimize this effect by delaying the cue to a target component until after the target has been heard (Martens, 1981, 1984; Roberts, 1998), so that ''pop out'' governs the judgment rather than focused attention. This approach can have a great effect on the outcome. For example, listeners in the studies of Martens (1981) heard a harmonic complex with an incremented component followed by a pure tone of the same frequency. For an F0 of 200 Hz, it was found that harmonic 1 or 2 typically needed a boost of 15–25 dB to be heard out as a pure tone, whereas a boost of just a few dB was required to hear out a harmonic above the fourth. This result is the opposite of what one might expect based on frequency resolution. Presumably, the difference in outcome is influenced by the effects of perceptual fusion—i.e., the lowest harmonics fuse well and so are less likely to segregate spontaneously.[3] In contrast, Martens (1984) found that a mistuned harmonic could often be heard even when presented at a level below that of the surrounding components.

The use of hit rate as a measure of the segregation of a mistuned harmonic was pioneered by Hartmann *et al.* (1990). This approach is based on the idea that a greater proportion of matches will be made to a target component as its salience increases, relative to that of the others. The measure is most effective when target components from across the spectrum are interspersed randomly from trial to trial (e.g., Hartmann *et al.*, 1990; Roberts and Brunstrom, 1998), because listeners cannot learn simply to ignore salient nontarget components on the basis of their pitch height. Hit rates

are less likely to be an effective measure when testing is restricted to a narrow region of the spectrum, particularly when the components are well resolved and hence very likely to be heard out when cued. This limitation clearly applies to an exploration of a restricted spectral region around component 1 or 2.

Here we present an alternative approach to measuring the salience of a probe, which is based on its loudness relative to that of an isolated pure tone of the same frequency. Specifically, listeners are required to adjust the level of the pure tone to match the loudness of the probe. It is necessary for listeners to make repeated comparisons to obtain consistent measures, and so it is inevitable that their attention will be cued to the probe by the pure tone. However, our main concern is to ensure that our estimates are not influenced by sequential capturing of the probe by the adjustable tone, which is likely to occur when tones are heard in rapid succession (see Bregman and Campbell, 1971). Guarding against this effect is considered especially important in view of the findings of van Noorden (1975). He observed, using rapid alternating sequences (tone onset-to-onset times = less than 100 ms), that the sequential capture of a partial by a corresponding pure tone (heard as the ''double tempo'' of that tone) is maximal when they are most similar in perceived rather than physical level. Therefore, this experiment used relatively long silent intervals between consecutive sounds.

A. Method

Six individuals participated, three of whom had previously taken part in experiment 1. Only aspects of the stimuli and task that differed from experiment 1 are described here. Each trial consisted of a repeating cycle of an adjustable pure tone followed by a complex tone (spectral frame plus probe). Each trial began with the pure tone to direct listeners' attention to the probe. The frequency of the pure tone was identical to that of the probe. The task of the listeners was to adjust the level of the pure tone, using a trackball mouse, to match the loudness of the probe as accurately as possible. The range of adjustment was +5 dB to −20 dB, relative to the probe, quantized in 0.25-dB steps. The starting level was set to the middle of the range (−7.5 dB). The silent interval between successive tones was set to 1 s. This was to ensure that the adjustable tone did not compete with the spectral frame to group with the probe (cf. Bregman and Pinker, 1978), which would otherwise reduce the effect of the frame on the loudness of the probe. The nominal F0 of the stimuli was not roved, but was instead fixed at 200 Hz.

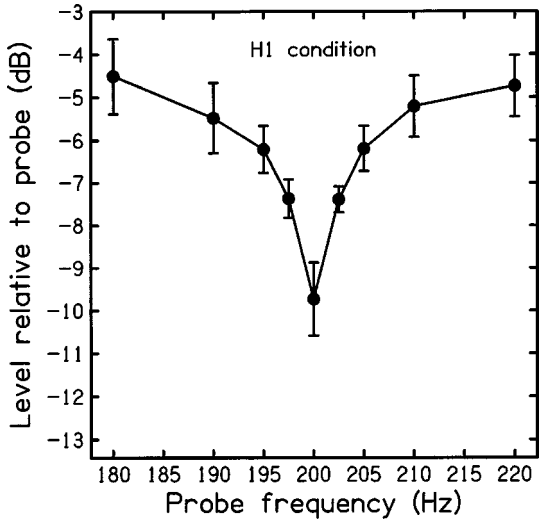
The sets of probes used were narrower in range than those used in experiment 1. The ranges used in experiment 1 reflected both the expected scope of the pitch-shift profiles and, in the shifted conditions, the need to encompass several different predictions about the location of the inflection. Given the clear findings of experiment 1, the probes were also more finely spaced—especially in the regions around the inflection frequencies. Also, it was not considered necessary to include more than one harmonic multiple of the suboctave pitch within the probe ranges tested here. In the harmonic case, the two sets of probes corresponded to frequencies of 0%, $\pm 1\frac{1}{4}$%, $\pm 2\frac{1}{2}$%, ±5%, and ±10%, with respect to components 1 and 2. The lower and higher sets of probes spanned 180–220 Hz and 360–440 Hz, respectively. In the shifted case, the lower set of probes spanned 195–265 Hz in 5-Hz steps, with a central region (220–230 Hz) in $2\frac{1}{2}$-Hz steps, and the upper set spanned 390–530 Hz in 10-Hz steps, with a central region (440–460 Hz) in 5-Hz steps.

The harmonic stimuli were tested in single ten-block sessions (H1, H2 = 10 repetitions × 9 probe values = 90 trials each), and the shifted stimuli were tested in two five-block sessions (S1, S2 = 5 repetitions × 17 probe values = 85 trials each). Therefore, ten judgments were obtained per stimulus for each subject. Each session typically lasted about an hour. The order of conditions was H1–H2–S1–S2, H2–H1–S2–S1, and S1–H1–H2–S2 for subjects 1–3, respectively. These sequences were each run in reverse order for the remaining three subjects. Each session was immediately preceded by one block of stimuli for practice (9 trials for H1 and H2, 17 trials for S1 and S2). Previously, all listeners had acquired experience of level adjustment during the pilot phase of the experiment. There was no direct screening of subjects for this task, but all had successfully completed the pitch-matching training described in experiment 1.

B. Results

Figure 6 displays mean levels (with intersubject standard errors) of the pure tone, relative to the probe level, when adjusted to equal the probe loudness. The top and bottom panels show the profiles for conditions H1 and H2, respectively. A repeated-measures analysis of variance for each condition showed that the main effect of probe frequency on matching level was highly significant [H1: $F(8,40)=11.76$, $p<0.001$; H2: $F(8,40)=11.34$, $p<0.001$]. Both profiles are nearly symmetrical and show a characteristic minimum in matching level (and hence in probe salience) that corresponds exactly to the harmonic frequencies of 200 and 400 Hz, respectively. Hence, it is reasonable to assume that these minima reflect specific effects of spectral grouping on component loudness. All probes were matched in loudness by pure tones that were lower in level. Each profile is quite flat towards the extremes of the range, and so the attenuation on the loudness-matched adjustable tone for a mistuning of ±10% can reasonably be taken as an estimate of ceiling loudness for comparison with the minimum. In both harmonic conditions, the attenuation at ceiling is about 4–5 dB in magnitude. The minima in salience for conditions H1 and H2 are about 5 and 7 dB, respectively, below ceiling.

Figure 7 displays mean relative levels, with standard errors, for conditions S1 (top panel) and S2 (bottom panel). A repeated-measures analysis of variance for each condition showed that the main effect of probe frequency on matching level was highly significant [S1: $F(16,80)=3.82$, $p<0.001$; S2: $F(16,80)=5.41$, $p<0.001$]. Pairwise comparisons were restricted to those between the data for the probe frequency corresponding to the minimum in matching level and the data for the probe frequencies closest to the values predicted by our three hypotheses. Given the small number



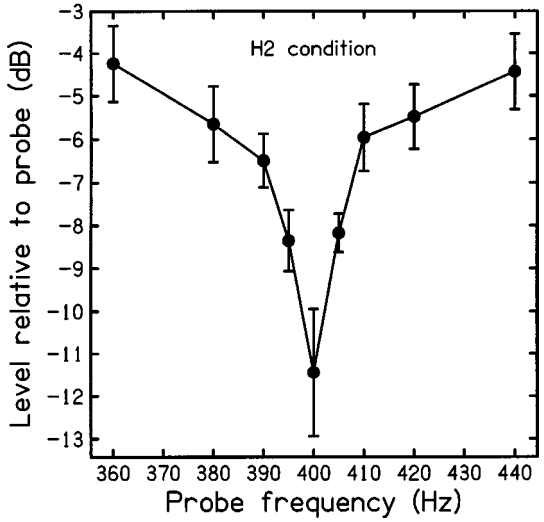
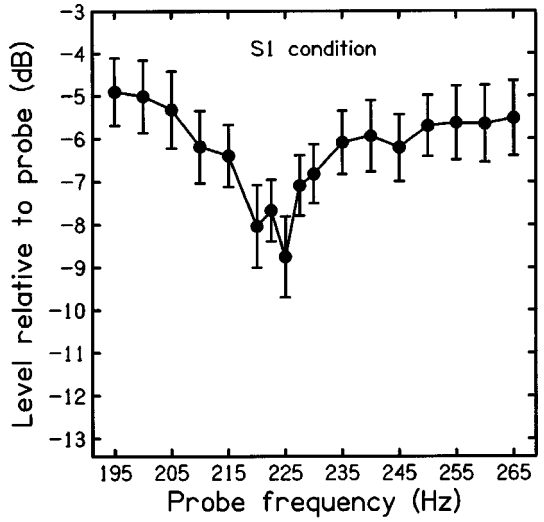


FIG. 6. Results for the harmonic conditions in experiment 2. The level of the adjustable pure tone is indicated, relative to that of the probe, when set to match the loudness of the probe. The top and bottom panels indicate mean levels across six subjects (with intersubject standard errors) for H1 and H2, respectively.



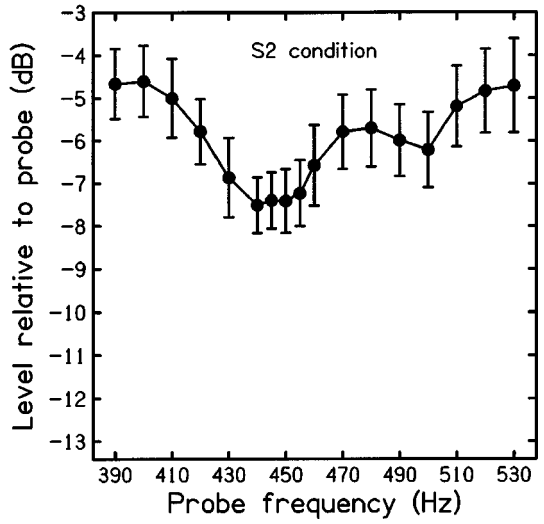


FIG. 7. Results for the frequency-shifted conditions in experiment 2. The level of the adjustable pure tone is indicated, relative to that of the probe, when set to match the loudness of the probe. The top and bottom panels indicate mean levels across six subjects (with intersubject standard errors) for S1 and S2, respectively.

of comparisons required from the large set of possible comparisons, they were computed using the restricted least-significant-difference test (Snedecor and Cochran, 1967).

The profile for condition S1 is somewhat shallower than that seen for condition H1—the minimum is only about 3–$3\frac{1}{2}$ dB below the adjustments made at the extremes of the probe range. The minimum in salience occurs at 225 Hz, exactly one octave below the frequency of component 2. This is fully consistent with our interpretation of the pitch-shift profile for this condition in experiment 1. Changing the probe frequency from 225 to 250 Hz (spectral-pattern prediction) significantly increased the matching level ($p = 0.015$), although changing it from 225 to 210 Hz (close to primary global-pitch prediction) did not ($p = 0.097$). There is no evidence in the profile for a secondary minimum in salience at either of these values.

The profile for condition S2 is less symmetrical than for S1 and also shallower (only $2\frac{1}{2}$–3 dB below ceiling). The primary minimum is at 440 Hz, but the profile is almost flat (variation <0.3 dB) for the four probes spanning 440 to 455 Hz. This broad trough encompasses the frequency predicted by the spectral-pattern hypothesis (450 Hz), which is consistent with our interpretation of the pitch-shift profiles in experiment 1. Changing the probe frequency from 440 to 420 Hz (close to primary global-pitch prediction) or to 470 Hz (secondary global-pitch prediction) led to a significant increase in the matching level ($p = 0.017$ and $p = 0.020$, respectively). Changing the probe frequency from 440 to 500 Hz (local-harmonicity prediction) also led t a significant increase in the matching level ($p = 0.028$), but changing it from 440 to 450 Hz (spectral-pattern prediction) did not ($p = 0.778$). There is no evidence for a secondary minimum in salience at either of the values related to global pitch. However, there is a secondary minimum in the S2 profile at 500 Hz. Although it is too small to be considered a robust finding, it is noteworthy that this frequency is exactly one octave above that of component 1.

C. Discussion

The minima in probe loudness observed in experiment 2 generally correspond closely with the inflection frequencies of the pitch-shift profiles observed in experiment 1. This finding provides further support for our interpretation of these pitch-shift data. Once again, it is evident that component 2 exhibits grouping by spectral pattern, whereas component 1 exhibits maximum fusion when in a simple octave relationship with component 2. Furthermore, there is little or no evidence to suggest that the grouping of either component is related to the global pitches evoked by the shifted complex (cf. Brunstrom and Roberts, 2000). The occurrence of a secondary minimum in the S2 profile at 500 Hz suggests that an octave interval between components 1 and 2 may influence the grouping of component 2 as well as that of component 1. However, this effect is at best a small one, and is certainly secondary to the effect of grouping-by-pattern for component 2.

The exact values of attenuation on the adjustable tone for equal-loudness judgments undoubtedly depend on the specific design of the task. However, comparisons between profiles can be revealing. Three points are of particular note:

First, the minima in salience are roughly twice as deep (in dB) for probes heard in harmonic than in frequency-shifted contexts. This finding is consistent with the commonly made observation that frequency-shifted complex tones sound much more unified than those with irregular spectra, but not quite as fused as those with harmonic spectra. In this regard, it is interesting to note that the *pitch-shift range* (difference between the maximum and minimum shift values for the cubic fit) in experiment 1 was greater for H1 (1.7%) and H2 (1.6%) than for S1 (1.2%) and S2 (1.3%). This adds weight to the notion that both pitch and loudness measures reflect a common underlying process.

Second, the relative bandwidth of the trough in the salience profile is considerably broader for probes heard in shifted than in harmonic contexts. Our pilot work using clarity rating had previously indicated broader minima for the shifted stimuli, particularly around component 2, and so this effect is not peculiar to the level-adjustment paradigm. Again, this finding is consistent with the idea that shifted stimuli are less cohesive perceptually than harmonic stimuli.

Third, fusion in the harmonic context appears to have a greater effect on the salience of component 2 than on that of component 1 (compare the depth of the minima for conditions H1 and H2 in Fig. 6). Given that probes at harmonic frequencies will not exhibit any ''pop out'' effect, this difference may reflect the poorer resolution of component 2. This interpretation is consistent with van Noorden's (1975) findings that (*i*) maximum capture of a harmonic by a corresponding pure tone in a rapid alternating sequence occurs when the two have the same loudness rather than level, and (*ii*) maximum sequential capture requires an increasing attenuation of the pure tone with increasing harmonic number. Alternatively, a change in the depth of the minimum may reflect a change in the extent to which the identity of an in-tune component is actively inhibited (cf. Bregman, 1990; Brunstrom and Roberts, 1998). These two accounts are difficult to distinguish based on the data presented here. Nonetheless, it is notable that Fig. 7 shows the reverse pattern for fusion in the shifted context (i.e., the minimum is deeper for condition S1 than for S2). Given that the relative resolution of components 1 and 2 would have differed little between harmonic and shifted contexts, this finding suggests that a change to the shifted context may weaken the perceptual suppression of component 2 more than that of component 1.

One further issue that merits discussion is a potential confound in the design of the task. Listeners were asked to match the loudness of the probe by adjusting the level of a pure tone of the same frequency. Given that a probe is subject to a pitch shift that varies with position in relation to the spectral frame, there will have been a variable discrepancy in pitch between the probe and the adjustable tone across the probe range. We are confident that any effects of this factor on our results will have been negligible. First, the magnitude of the largest pitch shifts observed was small (rarely beyond $\pm 2\%$). Second, the observed minima in loudness for the harmonic stimuli aligned exactly with the in-tune frequencies, despite the fact that the zero crossing for the pitch-shift profile did not align with these values (cf. Terhardt, 1971).

V. EXPERIMENT 3

Both pitch- and loudness-matching measures of the perceptual organization of frequency-shifted stimuli have indicated that the grouping of component 1 differs from that of the higher components. Although it is true for shifted stimuli that component 1 departs from harmonic relations with its neighbors by more than any other component,[2] this factor seems unlikely to account for the difference. This is because component 2 shows unequivocal grouping by the pattern of spectral spacing, despite its still considerable departure from harmonic relations with the other components. An account based on differences in frequency resolution also seems unlikely. Although the separation of component 1 (250 Hz) from component 2 (450 Hz) in our shifted stimuli was very large (3.24 ERBs; Glasberg and Moore, 1990), the separation of component 2 from component 3 (650 Hz) was also substantial (2.40 ERBs).

Two other factors might plausibly underlie the atypical behavior of component 1. One hypothesis is that component 1 in a shifted stimulus has a special status akin to that of the fundamental component in a harmonic stimulus. For example, we noted earlier that Roberts and Bailey (1996a) presented findings consistent with a major role for the fundamental component in the perceptual cohesion of harmonic stimuli. Alternatively, the lowest component may be special because it defines the lower spectral edge of a complex tone, whereas each interior component is bracketed by neighboring partials. This is true whether or not the lowest component present in the complex also happens to be component 1.

The two hypotheses can be distinguished by measuring the point of maximum fusion for component 2 in the presence and absence of component 1. If component 1 has a special status by acting in a manner akin to a fundamental component, then this manipulation should not affect the frequency of the fusion maximum in our shifted stimuli (even if it affects the extent of the fusion). However, making component 2 the lowest-numbered partial in the complex[4] should

cause it to default to grouping by harmonicity if bracketing is important. The second and third harmonics of a periodic complex tone form a simple interval with one another known as the fifth. For a nominal F0 of 200 Hz and a 50-Hz frequency shift, component 3 has a frequency of 650 Hz, and so changing component 2 from 450 to 433 Hz will restore this 2:3 ratio. Demany and Semal (1992) extended their earlier work with dichotically presented pure-tone dyads by showing that listeners are sensitive not only to the octave, but also to other simple intervals like the fifth. Therefore, we predicted that removal of component 1 from the spectral frame would leave the point of maximum fusion unchanged if the first hypothesis is correct, but would lead to a downward movement of that point (by about 15–20 Hz) if the second hypothesis is correct.

In principle, the effect of removing component 1 on the spectral fusion of component 2 might be investigated using either the pitch-matching or the loudness-matching paradigm. However, a fairly high degree of resolution is needed to distinguish the two hypotheses. Given the rather broad and flat-bottomed loudness profile observed in experiment 2 for component 2 of our shifted stimuli, loudness matching seemed unlikely to offer the resolution required. In contrast, the pitch-shift profiles observed in experiment 1 were fairly steep in the vicinity of the inflection frequency, indicating a relatively high degree of resolution. Therefore, we returned to using the pitch-matching paradigm in our final experiment.

A. Method

The same six individuals who took part in experiment 2 also participated in this experiment. As for experiment 1, listeners were required to adjust the frequency of a pure tone to match accurately the pitch of a probe embedded in a spectral frame. Only the spectral region around component 2 in frequency-shifted stimuli was explored in this experiment. The stimuli for condition 1 were identical in all respects to those for condition S2 in experiment 1. The stimuli for condition 2 differed only in that component 1 was removed from the spectral frame. Hence, the probe was the lowest component of the complex tone in condition 2. The three subjects who had taken part in experiment 1 were only tested on condition 2. Their data for condition S2 in experiment 1 served as their data for condition 1 here. The other subjects were tested first on condition 1 and then on condition 2. Each condition comprised 168 trials (2 sessions $\times$ 6 repetitions $\times$ 14 probe values), giving 12 judgments per stimulus for each subject. Each session was immediately preceded by one block of stimuli (14 trials) for practice. All subjects had some prior experience of pitch matching.

B. Results

Figure 8 displays mean pitch shifts, with intersubject standard errors, for condition 1 (component 1 = present) in the top panel and for condition 2 (component 1 = absent) in the bottom panel. Both pitch-shift profiles show clear evidence of a negative offset from the zero crossing. A good cubic fit was obtained across the full range of probes for

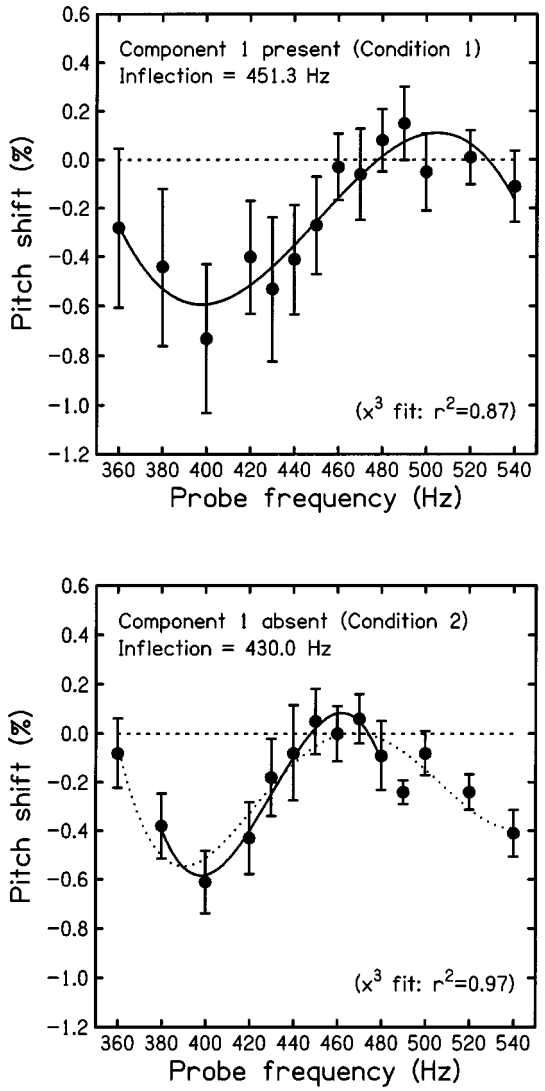


FIG. 8. Results for experiment 3. The top and bottom panels indicate mean pitch shifts across six subjects (with intersubject standard errors) when component 1 is present in or absent from the spectral frame, respectively. A nominal F0 of 200 Hz is assumed. The horizontal dashed lines indicate veridical matches to the frequency of the probe. Each panel shows the best cubic polynomial fit to the pitch-shift profile (full probe range for top panel, restricted probe range for bottom panel), and indicates the inflection frequency and r^2 value for this function. For comparison, the bottom panel also shows the best quintic polynomial fit (dotted line) to the full probe range. The quintic fit ($r^2=0.82$) has a midpoint frequency of 429.0 Hz.

condition 1. However, it was necessary to exclude data from the extremes of the probe range to obtain a good cubic fit ($r^2>0.7$) for condition 2. The inflection frequencies for conditions 1 and 2 are 451.3 and 430.0 Hz, respectively. For comparison, owing to the restricted range used for the cubic fit to the data from condition 2, a quintic fit to the entire probe range was computed. The midpoint frequency for this function (429.0 Hz) was very close to that for the cubic function.

As for condition S2 in experiment 1, the inflection for condition 1 is very close to the frequency of the excised second component (450 Hz), suggesting that the point of maximum fusion in this region is governed by the pattern of spectral spacing. The overall form of the profile is also similar to that observed earlier in condition S2, except that the pitch shifts are smaller and the intersubject variability is

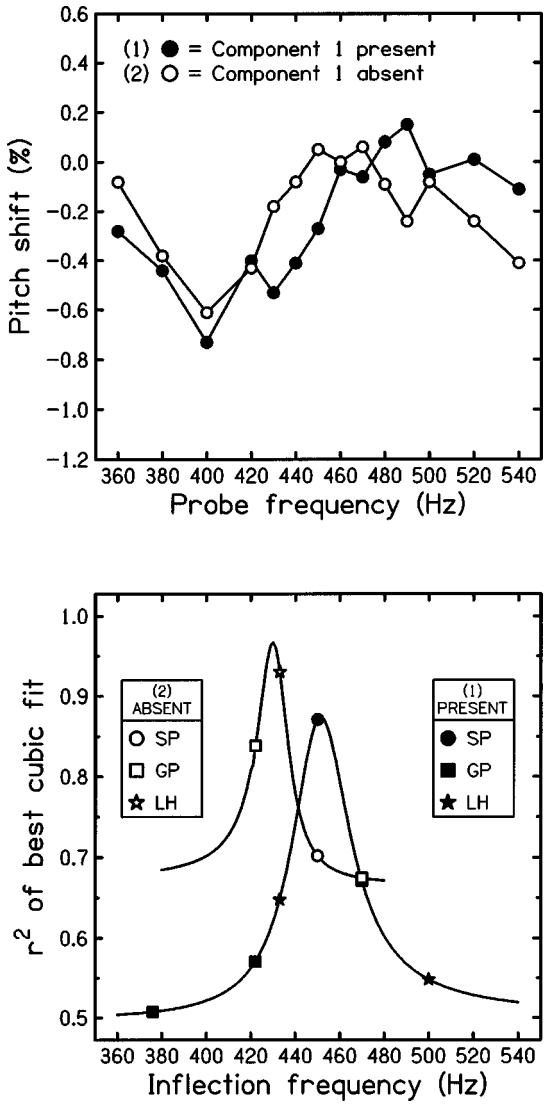


FIG. 9. Results for experiment 3. The top panel shows mean pitch shifts when component 1 is present (filled symbols) or absent (open symbols) from the spectral frame. A nominal F0 of 200 Hz is assumed. The horizontal dashed line indicates veridical matches to the frequency of the probe. The bottom panel shows functions indicating how the r^2 value for a constrained cubic fit to the data varies with inflection frequency for the present condition (filled symbols) or absent condition (open symbols), respectively. r^2 values for inflection frequencies predicted by the spectral-pattern hypothesis (SP), the global-pitch hypothesis (GP), and the local-harmonicity hypothesis (LH) are marked on the r^2 functions by circles, squares, and stars, respectively. Note that the SP hypothesis provides the best account for the present-condition data, but that the LH hypothesis provides the best fit for the absent-condition data.

greater in the current data set. Compared to condition 1, the data for condition 2 show less intersubject variability, but the mean pitch shifts are of similar magnitude. The most notable difference between the conditions is that the removal of component 1 from the spectral frame caused the inflection frequency to move downwards. Its value is very close to a 2:3 ratio with component 3 (i.e., 433 Hz). There is no evidence that harmonic multiples of the global pitches are represented in the pitch-shift profile for either condition.

To emphasize that the change in the inflection frequency is not an artifact of the curve-fitting procedure used, Fig. 9 (top panel) reproduces the mean pitch shifts for both conditions in a single panel without fitted functions. Note the small but definite leftward displacement of the central region of the pitch-shift profile when component 1 is removed from the spectral frame. Following the approach used in experiment 1, Fig. 9 (bottom panel) shows the dependency of r^2 value on inflection frequency for a set of cubic fits to the data for condition 1 (filled symbols) and for condition 2 (open symbols). The predicted frequencies according to our hypotheses (spectral pattern, global pitch, local harmonicity) are marked on these functions. r^2 value is strongly dependent on inflection frequency in both conditions.

For condition 1, only the spectral-pattern prediction (450 Hz) did not differ significantly from the full model [376 Hz: $F(1,11)=31.95$, $p<0.01$; 422 Hz: $F(1,11)=26.45$, $p<0.01$; 433 Hz: $F(1,11)=19.73$, $p<0.01$); 450 Hz: $F(1,11)=0.26$, ns; 470 Hz: $F(1,11)=17.72$, $p<0.01$; 500 Hz: $F(1,11)=28.46$, $p<0.01$]. For condition 2, all of the predicted values differed significantly from the full model [422 Hz: $F(1,6)=23.27$, $p<0.01$; 433 Hz: $F(1,6)=6.73$, $p<0.05$; 450 Hz: $F(1,6)=48.18$, $p<0.01$; 470 Hz: $F(1,6)=53.09$, $p<0.01$]. However, the difference is only marginally significant for the local-harmonicity prediction (433 Hz), and fails to reach significance if the inflection is reduced by only one step to 432 Hz. In this regard, it should be noted that the F value is highly sensitive to small changes in r^2 value when the full model is an excellent fit (here, $r^2=0.97$).

C. Discussion

The central section of the pitch-shift profile is displaced downwards in frequency when probes in the region of component 2 are tested in the absence of component 1. The size of this displacement, as measured by the change in the inflection frequency for a cubic fit to the profile, is very close to that predicted by the second hypothesis. Namely, the point of maximum fusion for component 2 changes from a value consistent with grouping-by-pattern when component 1 is present to one consistent with grouping-by-harmonicity (2:3 interval) when component 1 is removed. This finding suggests that the pattern of spectral spacing governs perceptual fusion only for components that are bracketed by at least one component above and below in frequency.

One might speculate that this principle extends to higher-numbered components as the ones below are removed progressively. However, testing this idea empirically becomes more difficult with increasing component number (n) for two reasons. First, the interval of the form $(n-1):n$ between the probe and the component immediately above it becomes progressively less simple with increasing number (i.e., 2:3, 3:4, 4:5, etc.), and this is likely to reduce our sensitivity to the tuning of the interval. Second, the frequency difference between the harmonic position and the pattern position declines as a percentage, which soon makes it impossible to distinguish whether a point of maximum fusion in the appropriate frequency region corresponds to one or other position. Focusing instead on measuring our sensitivity to a common low-order interval (e.g., one octave) between the probe and its neighbor for all values of n would not provide a simple alternative approach. This is because it is unknown

whether or not the lowest partial of a harmonic complex can act as an ''anchor'' for a probe one-octave below it when that probe is not a harmonic of the same F0 (as would be the case when the lowest partial is an odd harmonic). Without an answer to this question for harmonic stimuli, it would be difficult to interpret any evidence of a sensitivity to this interval in frequency-shifted stimuli.

More tractable for future study is the question of how the spectral fusion of component 2 of a shifted stimulus is influenced by the higher components. In particular, how many consecutive components above it need to be removed before the point of maximum fusion changes from the frequency of the missing component 2 to one octave above the frequency of component 1 (i.e., for our stimuli, from 450 to 500 Hz)? An answer to this question should help to illuminate the relative contributions of neighboring regions and spectrally remote regions to the perceptual organization of components in frequency-shifted stimuli.

The negative offset from the zero-crossing evident in the pitch-shift profiles is of similar magnitude for both conditions. This is inconsistent with an explanation in terms of asymmetric masking (Terhardt, 1971), which would predict a large negative offset only in condition 2 (component 1 = absent). Furthermore, the loudness profiles from experiment 2 are quite symmetrical for component 1 in both harmonic and shifted contexts. If partial masking of the probe by component 2 were a factor, then one might expect probes mistuned upwards to be less salient than those mistuned downwards by the same amount. In this regard, it is notable that both Peters *et al.* (1983) and Hartmann and Doty (1996) failed to find evidence for Terhardtian pitch shifts on partials in harmonic complex tones.

The cause of the negative offset observed both here and in experiment 1 is unclear, but it may be related to the small variation in the pitch of a pure tone that occurs with changes in level (Verschuure and van Meeteren, 1975). In particular, the pitch of a pure tone typically decreases slightly with increasing level below about 2 kHz. Given that our adjustable tone was set to 3 dB below the level of the probe, one might expect a pitch match to an in-tune probe to be made with a slightly lower frequency. Presumably, such an effect on probe pitch would be independent of, and additive with, the effect of mistuning the probe.

VI. GENERAL DISCUSSION

A. Summary of findings

Changes in the pitch and loudness of a sinusoidal probe, as its position is varied in relation to a spectral frame, have been used to explore the grouping of components 1 and 2 in harmonic and frequency-shifted complex tones. Specifically, the position of maximum fusion within each probe range has been inferred from the inflection frequency of the pitch-shift profile and from the frequency of the minimum in the loudness profile. These two measures have produced highly similar estimates and have been validated by their close correspondence to values known *a priori* for the harmonic stimuli. Our results for the frequency-shifted stimuli indicate that the grouping of component 2 is governed by the common pattern of spectral spacing shared by the other components, but that the grouping of component 1 is governed by the harmonicity of its local relationship to component 2. The former result confirms and extends the findings of Brunstrom and Roberts (2000) obtained by manipulating components 6–8, which are resolved only to a limited extent (Plomp, 1964). The latter result establishes that different grouping principles can operate in different spectral contexts. The finding for shifted stimuli that the fusion maximum for component 2 changes to a simple harmonic interval with its neighbor (2:3) when component 1 is removed indicates that it is not component 1 *per se* that is grouped differently from the others, but rather the lowest partial that is present in the complex tone. This suggests that grouping by spectral pattern operates only for components that are bracketed by others conforming to that pattern. Instead, edge components appear to default to local grouping by harmonicity.

B. The role of combination tones

Before the implications of these findings are discussed, the extent to which distortion products may have influenced our observations merits consideration. Brunstrom and Roberts (2000) first observed that a sinusoidal probe in a spectral gap in a frequency-shifted stimulus is more difficult to hear out when its frequency is equal to one of the missing components (6–8). These minima in hit rate were believed to reflect the degree of spectral fusion between the probe and the rest of the complex. However, it was noted that cubic difference tones generated by pairs of partials would have fallen at the frequencies of the missing components and may have been prominent (e.g., Goldstein, 1967; Smoorenburg, 1972). To explore whether or not the hit-rate minima resulted from phase cancellation between a probe and a cubic difference tone at the frequency of the missing eighth component, Brunstrom and Roberts (2000) tested a small set of probes in this frequency region with starting phases of 0°, 90°, 180°, and 270°, relative to the sine-phase complex. The hit-rate minimum in this region remained aligned with the missing component for all conditions, and so was not attributable to phase cancellation. However, this approach did not rule out the possibility that the hit-rate minimum arose instead from an increase in the prominence of probes at neighboring positions arising from beating with a combination tone at the frequency of the missing component.

Pressnitzer *et al.* (2001) have since reported psychophysical measurements of the simple difference tones generated by a set of unresolved harmonics (15–25) of 100 Hz, added in cosine phase. The distortion spectrum was composed of reintroduced lower harmonics. The most intense difference tone was the first, corresponding to the missing F0, which was about 10–15 dB below the level of the primaries. The second and subsequent difference tones declined progressively in level. It seems likely that the higher components of a shifted stimulus would similarly generate difference tones at frequencies corresponding to the lower components. Given that our current experiments focused on the grouping of components 1 and 2, such difference tones may have influenced our findings. However, the results for our shifted stimuli are not compatible with an explanation of this

kind. For a nominal F0 of 200 Hz and a frequency shift of 50 Hz, the most intense difference tone would have been generated at 250 Hz, yet the point of maximum fusion corresponded to 225 Hz. In turn, this implies that the point of maximum fusion at 450 Hz cannot be attributed to the less intense difference tone generated at that frequency. Furthermore, the change in frequency of the fusion maximum for component 2 when component 1 is removed from the complex clearly cannot be accounted for in terms of distortion products. Instead, we assert that our findings reflect the outcome of an auditory-grouping process.

C. Analysis and conjecture

None of our findings have indicated any direct relationship between the global pitch(es) evoked by a shifted complex (Patterson, 1973) and the perceptual organization of its constituent partials. This result provides further support for the proposal by Ciocca (1999) and by Brunstrom and Roberts (2000) that different grouping constraints govern the perceptual segregation of individual components and the operation of pitch-perception processes. Some earlier, introspective, observations by Demany and Semal (1992) also suggest such a distinction. These authors noted that the sensitivity of a given listener to the tuning of a simple harmonic interval between two simultaneous and dichotically presented pure tones did not appear to depend on to whether or not these tones evoked a clear sensation of the missing fundamental (cf. Houtsma and Goldstein, 1972). They further noted that Demany *et al.* (1991) had found, using similar stimuli, that thresholds for detecting mistuning from a one-octave interval and for discriminating differences in pitch (presumably global pitch) varied in different ways as a function of frequency register (defined by the frequency of the lower component). These observations are interesting in that they support the notion that different grouping constraints govern spectral fusion and pitch perception for harmonic, as well as shifted, stimuli.

It is convenient to describe our frequency-shifted stimuli in terms of spectral regularity, but it should be noted that uniformity of beats between neighboring partials may also play a role in determining the perceptual cohesion of these stimuli. Support for this idea comes from the recent finding by Treurniet and Boucher (2001a, b) that the effectiveness of a complex tone as a simultaneous masker for a noise probe depends on the spectral structure of the masker. Their first study found that masked thresholds were as much as 10 dB lower for harmonic maskers than for similar inharmonic maskers that were derived from them by perturbing the component frequencies to nearby values. This effect was reduced or abolished by increasing the frequency separation of the partials by selective omission from the original set, which suggests that intermodulations play a key role in the effect. Their second study showed that frequency-shifted maskers gave rise to thresholds similar to those for otherwise equivalent harmonic maskers. Treurniet and Boucher (2001b) concluded that the reduction in masked threshold must arise from uniform modulation rates across channels (envelope cues) rather than from a common periodicity in the fine structure of the waveform across channels.

Listeners are able to hear out mistuned components up to at least the tenth in harmonic (Hartmann *et al.*, 1990) and in shifted stimuli (Roberts and Brunstrom, 1998). Given the marginal resolution of these components (Shackleton and Carlyon, 1994), it seems highly likely that a cross-channel comparison of intermodulation rates provides a salient cue for segregating them. A cue of this kind would also have influenced the findings of Brunstrom and Roberts (2000), who explored the grouping of components 6–8. Indeed, the influence of level, duration, and phase on the ability of listeners to detect a mistuned fourth harmonic (Hartmann, 1988) suggests that intermodulations play a role even for components traditionally regarded as resolved (Plomp, 1964). However, it seems less likely that beating provided a salient cue for the perceptual segregation of the highly resolved components investigated in the current study.

Excitation patterns computed for our shifted stimuli using Glasberg and Moore's (1990) program show that the amount of excitation that spreads from component 1 (250 Hz) to component 2 (450 Hz) is about -30 dB, relative to the excitation generated by component 2 itself. Hence, the modulation in the output of a channel centered on component 2 would have been negligible. The outputs of channels centered between components 1 and 2 would have shown greater modulation, but the depth of the valley in this region of the excitation pattern is about 15 dB. While it cannot be ruled out entirely that beating in the outputs of these channels had an important influence on our results, three findings suggest that this was not the case. First, Roberts and Bailey (1993a) reported that the higher salience of an added even harmonic in an otherwise odd-harmonic complex was robust in the presence of noise that was undoubtedly intense enough to mask any intermodulations between the low-numbered partials. Second, Lin and Hartmann (1998) found a substantial positive pitch-shift gradient for mistuning the fundamental component by $\pm 8\%$, even when harmonics 2 and 3 were removed from the complex. Any intermodulations generated by such an isolated partial would have been negligible. Third, Brunstrom and Roberts (2001) observed small but reliable partial-pitch shifts, for both harmonic and shifted stimuli, when the mistuned component was presented in the opposite ear to the spectral frame. Therefore, it seems that a comprehensive account of the grouping of concurrent spectral components requires some form of pattern recognition.

The different grouping behavior of the first component in shifted stimuli is germane to the observation that such stimuli, though perceptually coherent, are less so than harmonic stimuli (e.g., Brunstrom and Roberts, 2000). Specifically, it seems likely that the perceptual segregation of the first component from our shifted stimuli reduces the cohesion of the rest of the complex, in the light of Roberts and Bailey's (1996a) finding that the salience of the higher partials in odd-harmonic stimuli increases considerably when the fundamental component is removed. The physical (as opposed to perceptual) removal of the first component would cause the second one to become an edge component that would in turn cause it to segregate perceptually from the shifted complex. This ''fraying'' from the lower edge of the complex would presumably continue (with the removal of

further components) until the mistuning between the edge frequency and the nearest harmonic interval with the adjacent component fell below about 1% (around the fourth component for our 50-Hz shifted stimuli). This kind of grouping behavior further highlights the limitations of the template metaphor, which has already been challenged by evidence that the effect of a set of harmonics on the fusion of an individual component elsewhere in the spectrum attenuates with distance (Lin and Hartmann, 1998; Brunstrom and Roberts, 1998), and by the implausibility of the idea that a set of templates exists specifying various degrees of frequency shift or spectral stretch (Roberts and Brunstrom, 2001).

In terms of Roberts and Brunstrom's (2001) qualitative model of spectral fusion, the new findings can be incorporated by assuming that the curved trajectory of the autocorrelogram spine in the auditory response to a shifted stimulus [see Fig. 1(b)] can be interpolated, but not extrapolated, across channels. Hence, mistuning the lowest component of a shifted stimulus would not produce a detectable local discontinuity in relation to the trajectory established by the patterning of the higher components. Instead, it is assumed that fusion of the lowest component with the rest of the complex depends on common delay periods in the autocorrelation functions associated with the lowest component and its immediate neighbor. Common delay periods arise when the frequencies of these components form a simple harmonic interval (e.g., 1:2 or 2:3).

D. Concluding remarks

We have shown that the pattern of partial-pitch shifts across a set of sinusoidal probes can be used to identify positions of maximum spectral fusion in a complex tone with a high degree of precision. This approach can be usefully supplemented by a more direct measure of salience based on loudness judgments. The experiments presented here clearly show that the principle of grouping by common spectral pattern extends to the low-numbered components of a complex tone. These components are too highly resolved for a mechanism based solely on a cross-channel comparison of intermodulation rates to provide a general account of the phenomenon. It has also been established that grouping-by-pattern breaks down at the lower spectral edge of a shifted stimulus in favor of grouping-by-harmonicity. There remains no evidence that the global pitch evoked by a complex tone influences the fusion of its partials. While it is true that the stimuli we employed are highly unnatural, they can reveal features of the grouping constraints governing the perceptual cohesion of multicomponent sounds that cannot be revealed by stimuli composed of consecutive harmonics. Developing a quantitative and physiologically plausible model of this process presents a major challenge for grouping theorists.

ACKNOWLEDGMENTS

Special thanks go to Tim Yates for his advice on polynomial fits and to Andrew Schofield for his advice on inferential statistics. We are grateful to Roger Cartwright and Paul Rock for their assistance in recruiting and testing our listeners. We also thank Bill Hartmann, Steve Holmes, Marjorie Leek, and the anonymous reviewers for their suggestions, advice, and criticism.

[1]For convenience, we use the term ''template'' to describe auditory responsiveness to spectral pattern, without intent to imply a spectral (as opposed to spectrotemporal) mechanism or a set of predefined internal templates specifying various degrees of shift and stretch. Rather, the auditory response is likely to be central but stimulus driven, as suggested by Roberts and Brunstrom (2001).

[2]Adding a constant frequency increment to each component of a harmonic complex results in a progressively smaller percentage change as component number increases. Hence, the higher components in a shifted complex more closely resemble true harmonics.

[3]Note that this finding does not necessarily reflect the relative saliences of the different harmonics when they have not been boosted in level. Specifically, it is possible that across-frequency differences in sensitivity to changes in spectral envelope may also have influenced the outcome.

[4]Nonlinearities in the auditory system may result in distortion products below the frequency of the lowest primary component (Goldstein, 1967; Smoorenburg, 1972; Pressnitzer *et al.*, 2001). This possibility, and the implications for our findings, are considered in Sec. VI B.

Alain, C., Arnott, S. R., and Picton, T. W. (**2001**). ''Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials,'' J. Exp. Psychol. Hum. Percept. Perform. **27**, 1072–1089.

Alain, C., Schuler, B. M., and McDonald, K. L. (**2002**). ''Neural activity associated with distinguishing concurrent auditory objects,'' J. Acoust. Soc. Am. **111**, 990–995.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).

Bregman, A. S., and Campbell, J. (**1971**). ''Primary auditory stream segregation and perception of order in rapid sequences of tones,'' J. Exp. Psychol. **89**, 244–249.

Bregman, A. S., and Pinker, S. (**1978**). ''Auditory streaming and the building of timbre,'' Can. J. Psychol. **32**, 19–31.

Brunstrom, J. M., and Roberts, B. (**1998**). ''Profiling the perceptual suppression of partials in periodic complex tones: Further evidence for a harmonic template,'' J. Acoust. Soc. Am. **104**, 3511–3519.

Brunstrom, J. M., and Roberts, B. (**2000**). ''Separate mechanisms govern the selection of spectral components for perceptual fusion and for the computation of global pitch,'' J. Acoust. Soc. Am. **107**, 1566–1577.

Brunstrom, J. M., and Roberts, B. (**2001**). ''Effects of asynchrony and ear of presentation on the pitch of mistuned partials in harmonic and frequency-shifted complex tones,'' J. Acoust. Soc. Am. **110**, 391–401.

Ciocca, V. (**1999**). ''Evidence against an effect of grouping by spectral regularity on the perception of virtual pitch,'' J. Acoust. Soc. Am. **106**, 2746–2751.

Darwin, C. J., Ciocca, V., and Sandell, G. J. (**1994**). ''Effects of frequency and amplitude modulation on the pitch of a complex tone with a mistuned harmonic,'' J. Acoust. Soc. Am. **95**, 2631–2636.

de Cheveigné, A. (**1997**). ''Harmonic fusion and pitch shifts of mistuned partials,'' J. Acoust. Soc. Am. **102**, 1083–1087.

Demany, L., and Semal, C. (**1988**). ''Dichotic fusion of two tones one octave apart: Evidence for internal octave templates,'' J. Acoust. Soc. Am. **83**, 687–695.

Demany, L., and Semal, C. (**1992**). ''Detection of inharmonicity in dichotic pure-tone dyads,'' Hear. Res. **61**, 161–166.

Demany, L., Semal, C., and Carlyon, R. P. (**1991**). ''On the perceptual limits of octave harmony and their origin,'' J. Acoust. Soc. Am. **90**, 3019–3027.

Duifhuis, H., Willems, L. F., and Sluyter, R. J. (**1982**). ''Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception,'' J. Acoust. Soc. Am. **71**, 1568–1580.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

Goldstein, J. L. (**1967**). ''Auditory nonlinearity,'' J. Acoust. Soc. Am. **41**, 676–689.

Hartmann, W. M. (**1988**). ''Pitch perception and the segregation and integration of auditory entities,'' in *Auditory Function*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 623–645.

Hartmann, W. M. (**1996**). ''Pitch, periodicity, and auditory organization,'' J. Acoust. Soc. Am. **100**, 3491–3502.

Hartmann, W. M., and Doty, S. L. (**1996**). ''On the pitches of the components of a complex tone,'' J. Acoust. Soc. Am. **99**, 567–578.
Hartmann, W. M., McAdams, S., and Smith, B. K. (**1990**). ''Hearing a mistuned harmonic in an otherwise periodic complex tone,'' J. Acoust. Soc. Am. **88**, 1712–1724.
Henke, W. L. (**1997**). *MITSYN: A coherent family of high-level languages for time signal processing*, software package (Belmont, MA).
Houtsma, A. J. M., and Goldstein, J. L. (**1972**). ''The central origin of the pitch of complex tones: Evidence from musical interval recognition,'' J. Acoust. Soc. Am. **51**, 520–529.
Howell, D. C. (**1997**). *Statistical Methods for Psychology*, 4th ed. (Wadsworth, Belmont, CA).
Lin, J.-Y., and Hartmann, W. M. (**1998**). ''The pitch of a mistuned harmonic: Evidence for a template model,'' J. Acoust. Soc. Am. **103**, 2608–2617.
Martens, J.-P. (**1981**). ''Audibility of harmonics in a periodic complex,'' J. Acoust. Soc. Am. **70**, 234–237.
Martens, J.-P. (**1984**). ''Comment on 'Algorithm for extraction of pitch and pitch salience from complex tonal signals,','' J. Acoust. Soc. Am. **75**, 626–628.
Meddis, R., and Hewitt, M. J. (**1991**). ''Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.
Meddis, R., and O'Mard, L. (**1997**). ''A unitary model of pitch perception,'' J. Acoust. Soc. Am. **102**, 1811–1820.
Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, London).
Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1985**). ''Relative dominance of individual partials in determining the pitch of complex tones,'' J. Acoust. Soc. Am. **77**, 1853–1860.
Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1986**). ''Thresholds for hearing mistuned partials as separate tones in harmonic complexes,'' J. Acoust. Soc. Am. **80**, 479–483.
Patterson, R. D. (**1973**). ''The effects of relative phase and the number of components on residue pitch,'' J. Acoust. Soc. Am. **53**, 1565–1572.
Peters, R. W., Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Pitch of components of complex tones,'' J. Acoust. Soc. Am. **73**, 924–929.
Plomp, R. (**1964**). ''The ear as a frequency analyzer,'' J. Acoust. Soc. Am. **36**, 1628–1636.
Pressnitzer, D., and Patterson, R. D. (**2001**). ''Distortion products and the pitch of harmonic complex tones,'' in *Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht, The Netherlands), pp. 84–91.
Rakowski, A., and Hirsh, I. J. (**1980**). ''Poststimulatory pitch shifts for pure tones,'' J. Acoust. Soc. Am. **68**, 467–474.
Ritsma, R. J. (**1967**). ''Frequencies dominant in the perception of the pitch of complex sounds,'' J. Acoust. Soc. Am. **42**, 191–198.
Roberts, B. (**1998**). ''Effects of spectral pattern on the perceptual salience of partials in harmonic and frequency-shifted complex tones: A performance measure,'' J. Acoust. Soc. Am. **103**, 3588–3596.
Roberts, B., and Bailey, P. J. (**1993a**). ''Spectral pattern and the perceptual fusion of harmonics. I. The role of temporal factors,'' J. Acoust. Soc. Am. **94**, 3153–3164.
Roberts, B., and Bailey, P. J. (**1993b**). ''Spectral pattern and the perceptual fusion of harmonics. II. A special status for added components?'' J. Acoust. Soc. Am. **94**, 3165–3177.
Roberts, B., and Bailey, P. J. (**1996a**). ''Regularity of spectral pattern and its effects on the perceptual fusion of harmonics,'' Percept. Psychophys. **58**, 289–299.
Roberts, B., and Bailey, P. J. (**1996b**). ''Spectral regularity as a factor distinct from harmonic relations in auditory grouping,'' J. Exp. Psychol. Hum. Percept. Perform. **22**, 604–614.
Roberts, B., and Bregman, A. S. (**1991**). ''Effects of the pattern of spectral spacing on the perceptual fusion of harmonics,'' J. Acoust. Soc. Am. **90**, 3050–3060.
Roberts, B., and Brunstrom, J. M. (**1998**). ''Perceptual segregation and pitch shifts of mistuned components in harmonic complexes and in regular inharmonic complexes,'' J. Acoust. Soc. Am. **104**, 2326–2338.
Roberts, B., and Brunstrom, J. M. (**2001**). ''Perceptual fusion and fragmentation of complex tones made inharmonic by applying different degrees of frequency shift and spectral stretch,'' J. Acoust. Soc. Am. **110**, 2479–2490.
Shackleton, T. M., and Carlyon, R. P. (**1994**). ''The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination,'' J. Acoust. Soc. Am. **95**, 3529–3540.
Smoorenburg, G. F. (**1972**). ''Audibility region of combination tones,'' J. Acoust. Soc. Am. **52**, 603–614.
Snedecor, G. W., and Cochran, W. G. (**1967**). *Statistical Methods*, 6th ed. (Iowa U.P., Ames, Iowa).
Terhardt, E. (**1971**). ''Pitch shifts of harmonics, an explanation of the octave enlargement phenomenon,'' in *Proceedings of the 7th International Congress on Acoustics, Budapest, Vol. 3*, pp. 621–624.
Treurniet, W., and Boucher, D. R. (**2001a**). ''A masking level difference due to harmonicity,'' J. Acoust. Soc. Am. **109**, 306–320.
Treurniet, W., and Boucher, D. R. (**2001b**). ''Masked threshold difference due to masker harmonicity: Uniform modulation rates of auditory filter outputs versus periodicity in waveform fine structure,'' J. Acoust. Soc. Am. **110**, 1267–1270.
van Noorden, L. P. A. S. (**1975**). ''Temporal coherence in the perception of tone sequences,'' doctoral thesis, Eindhoven University of Technology, The Netherlands.
Verschuure, J., and van Meeteren, A. A. (**1975**). ''The effect of intensity on pitch,'' Acustica **32**, 33–44.

Modulation masking produced by complex tone modulators[a)]

Jesko L. Verhey,[b)] Stephan D. Ewert,[c)] and Torsten Dau[c)]
Carl von Ossietzky Universität Oldenburg, Medizinische Physik, D-26111 Oldenburg, Germany

(Received 14 March 2002; revised 24 July 2003; accepted 4 August 2003)

Thresholds were measured for detecting sinusoidal amplitude modulation in the presence of a complex-tone masker modulation. Both modulations were applied to the same sinusoidal carrier. Two different masker modulations were used: (*i*) a pair of components beating at the difference frequency and (*ii*) a three-tone complex producing a sinusoidal amplitude modulation of the modulation depth at the difference frequency between adjacent components. Both maskers show a periodicity in the waveform that is not contained in the envelope spectrum itself but can be observed when the envelope of the envelope, referred to as the ''venelope'' [Ewert *et al.*, J. Acoust. Soc. Am. **112**, 2921–2931 (2002)], is calculated. For a signal frequency equal to the masker-venelope periodicity, modulation depth at threshold was measured as a function of the signal phase relative to the phase of the masker-venelope component. Signal frequencies of 5, 30, and 90 Hz were used. It was found that masking was phase dependent for all three signal frequencies. Thresholds were lower for the in-phase condition, where maxima in the signal waveform coincided with maxima in the masker-venelope waveform, than for the antiphase condition. The maximum threshold difference was 15 dB. The results are in contrast to recent data [Moore *et al.*, J. Acoust. Soc. Am. **106**, 908–918 (1999)], where lowest thresholds were reported for the antiphase condition in a similar experiment. The present data are in line with the idea that a nonlinearity prior to a modulation filterbank extracts the venelope of the masker modulator. However, a compressive nonlinearity such as that associated with the processing on the basilar membrane cannot account for the empirical findings, since it predicts the opposite phase effect. [DOI: 10.1121/1.1612489]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Mk [NFV]

I. INTRODUCTION

The concept of critical bands, first introduced by Fletcher (1940), has been successfully used to explain detection performance in the audio-frequency domain. It assumes that the auditory system analyzes the signal by a bank of overlapping bandpass filters and that the output of the filter giving the highest signal-to-noise ratio determines threshold. This filter is usually centered at or close to the signal frequency. More recent studies have shown that the auditory system may use a similar mechanism to analyze envelope fluctuations (amplitude modulations). Psychoacoustical modulation detection experiments using narrow-band noise carriers (Fleischer, 1982, 1983; Dau *et al.*, 1997a; Lorenzi *et al.*, 2001a), modulation masking data with narrow-band noise masker modulators (Houtgast, 1989; Bacon and Grantham, 1989; Ewert and Dau, 2000; Ewert *et al.*, 2002), and physiological data (Langner and Schreiner, 1988; Palmer, 1995; Schulze and Langner, 1997, 1999) suggest that the auditory system may also perform a frequency-selective analysis in the envelope-frequency domain.

Although several researchers have proposed such a mechanism (e.g., Martens, 1982; Houtgast, 1989; Bacon and Grantham, 1989; Dau *et al.*, 1997a), there are only a few studies where quantitative predictions with a model including a modulation filterbank were presented. Dau *et al.* (1997a) expanded their original signal-processing model (Dau *et al.*, 1996) by introducing a modulation filterbank at the output of each peripheral filter. It was shown that this model can account for a large variety of detection and masking data in both the audio-frequency and modulation-frequency domains (Dau *et al.*, 1997a, b; Verhey, 1999; Verhey *et al.*, 1999; Derleth and Dau, 2000; Ewert and Dau, 2000; Verhey, 2002).

Such a model has also been shown to account for modulation detection data using complex masker modulators (Dau, 1996; Dau *et al.*, 1997a). The detection threshold for sinusoidal signal modulation was measured in the presence of a complex-tone modulator consisting of the third to seventh harmonics of 30 Hz, as a function of the signal-modulation frequency. The phase of the masker components was chosen randomly. It was found that the modulation depth at threshold increased monotonically with decreasing spectral difference between the signal modulation and the lowest component of the masker-modulation tone complex. Thus, the threshold for 30-Hz signal modulation (the ''missing fundamental'') was not higher than those for slightly lower or slightly higher modulation frequencies, indicating that masking was not determined by the repetition period of the masker-modulator waveform. The data support the idea that the masked threshold is directly related to the energy falling within the passband of the modulation filter centered at the signal-modulation frequency (Dau *et al.*, 1997a).

More recently, Moore *et al.* (1999) found that the detec-

[a)]Portions of this article were presented at the 141st Meeting of the Acoustical Society of America, 4–8 June 2001, Chicago, IL [J. Acoust. Soc. Am **109**, 2466(A) (2001)].

[b)]Electronic mail: jesko.verhey@uni-oldenburg.de
[c)]Current address: Centre for Applied Hearing Research, Acoustic Technology, Ørsted·DTU, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark; Electronic mail: [se,td]@oersted.dtu.dk

tion threshold for 5-Hz signal modulation was affected by the presence of a spectrally remote pair of masker modulators. Masker modulators of 40 and 45 Hz (and higher) were used, beating at the rate of the signal modulation. They found that the signal-modulation depth at threshold depends on the phase of the signal modulation relative to the beat cycle. At first sight, this seems to be in contrast to the concept of modulation filters, which is based on the assumption that modulation frequencies can only interact if they excite a common modulation filter. However, Moore *et al.* (1999) argued that the compressive nonlinearity associated with the processing on the basilar membrane (e.g., Ruggero, 1992) introduces a physical distortion component in the envelope at the beat frequency which interacts with the signal-modulation frequency. They showed that their data could be predicted with a model assuming a compressive nonlinearity prior to a modulation filterbank.

To characterize slow variations of the modulation depth in more general terms, Ewert *et al.* (2002) suggested the ''venelope,'' which is defined as the Hilbert envelope of the ac-coupled Hilbert envelope of a stimulus. They measured an effect in the envelope-frequency domain which is comparable to the ''asymmetry of masking'' effect in the audio-frequency domain (Hellman, 1972; Hall, 1997; Derleth and Dau, 2000; Verhey, 2002). Thresholds were markedly lower when a narrow-band noise was masked by a sinusoidal masker than in the reversed condition. By analogy with the interpretation of the asymmetry of masking effect in the audio-frequency domain, where amplitude modulations provide important detection cues (Derleth and Dau, 2000; Verhey, 2002), Ewert *et al.* (2002) concluded that the venelope fluctuations provide an additional detection cue in corresponding experiments in the envelope domain.

Recently, Lorenzi *et al.* (2001a, b) measured the detectability of a sinusoidal variation applied to the modulation depth of a ''carrier'' amplitude modulation. It was shown that, in some conditions, the detection thresholds for the resulting venelope fluctuations were close to those for envelope fluctuations.

Taken together, the above studies demonstrate the perceptual salience of the venelope of the stimuli. The data are in line with the idea that some nonlinearity extracts the venelope and combines it with the internal representation of the envelope. Moore *et al.* (1999) suggested compression as a possible nonlinearity, although they noted that other nonlinearities might be involved. Shofner *et al.* (1996) proposed three different physiological mechanisms to account for their findings that units in the ventral cochlear nucleus showed a response to the beating of a two-tone complex modulation: compression, saturation, and threshold. Ewert *et al.* (2002) proposed a black-box approach, assuming that the auditory system may extract the venelope of the stimulus in some way without specifying how this might be realized neurally. However, none of the previous studies investigated explicitly which of the proposed mechanisms accounts best for the psychophysical data.

The main goal of the present study was to investigate the underlying process for the extraction of the venelope in the auditory system. In order to further specify the nonlinear mechanism that may be responsible for the extraction of the venelope, phase-sensitive experiments like those described in the Moore *et al.* (1999) study were necessary. Moore *et al.* (1999) showed an effect of the masker-venelope phase on the threshold of a signal modulation with the same frequency as the venelope frequency only for one frequency (5 Hz), one modulation depth for each masker component (0.3), and only masker components that were harmonics of the signal-modulation frequency. The first experiment in the present study was performed to examine the role of signal frequency, spectral distance between signal and masker and the harmonic relation between the signal frequency and the masker components. The experimental paradigm was comparable to the one presented in Moore *et al.* (1999). The signal-modulation threshold was measured as a function of the relative phase between masker venelope and signal modulation for three different signal-modulation frequencies (5, 30, and 90 Hz) and various spectral distances between masker components and signal modulation. The signal modulation always matched the venelope frequency. In addition, the effect of the relative phase between masker venelope and signal modulation was measured using a three-tone complex masker. Such a masker has a well-defined venelope spectrum, consisting of only a single component at the imposed venelope frequency, compared to the more complex harmonic venelope spectrum of the two-tone masker used in the first experiment. A sinusoidal venelope applied to high-frequency modulation was also used in Lorenzi *et al.* (2001a, b). The effect of the phase was measured for various masker venelope depths in order to determine the lower limit of the interaction between envelope and venelope.

A further goal of the present study was to investigate whether the differences in the findings of the studies of Dau *et al.* (1997a) and of Moore *et al.* (1999) can be explained by the differences in the experimental parameters used in the two studies. Dau *et al.* (1997a) used a higher beating frequency, a higher number of masker components, and a lower modulation depth for the masker components than Moore *et al.* (1999). Another difference was that the phase relation between the masker and the signal was random in Dau *et al.* (1997a) while it was fixed in Moore *et al.* (1999). In the present study, detection thresholds for signal modulations of 20, 30, and 40 Hz were measured in the presence of a two-tone complex masker modulation beating at 30 Hz. The phase relation between the signal modulation and the masker venelope was randomized.

In the second part of the study, predictions of models including envelope extraction, some nonlinearity, and a modulation filterbank were compared to the present experimental data. The models differ in the nonlinearity assumed for the extraction of the venelope component prior to the modulation filterbank. Models of temporal processing that do not contain a modulation filterbank but, instead, a modulation low-pass filter as the smoothing device (e.g., Viemeister *et al.*, 1979), are not considered explicitly in the present study. These models cannot account for modulation masking data, such as those obtained in, e.g., Bacon and Grantham (1989), Houtgast (1989), and Ewert and Dau (2000).

II. METHODS

A. Subjects

Five subjects (two males, three females) ranging in age between 25 and 31 years participated in the experiments. One of them was the author JV. All subjects had normal audiograms (i.e., absolute threshold in quiet $\leqslant$15 dB HL) and no previous history of any hearing problems. Except for a signal frequency of 5 Hz, all experiments were performed with the same three subjects (CH, JV, KT). Two of those three subjects (JV, KT) and one (MK) or two (MK, GG) additional subjects participated in the experiments with a signal frequency of 5 Hz. All subjects had experience in psychoacoustic experiments. Four subjects were members of the ''Medizinische Physik'' at the University of Oldenburg. Subject KT was a paid volunteer.

B. Apparatus and stimuli

The stimuli were generated digitally at a sampling rate of 32 kHz. Stimulus generation and presentation were controlled with the software package SI developed at the Universität Göttingen, using a Silicon Graphics workstation (INDY), which also sampled the listener's responses and controlled the procedure. The stimuli were D/A converted (16 bits), and then preamplified and low-pass filtered at 16 kHz with a computer-controlled audiometric amplifier. Stimulus presentation was diotic via Sennheiser HD25 headphones. The subjects were seated in a double-walled sound-attenuating booth.

A sinusoidal carrier at 5 kHz was used in all experiments. The carrier was 600 ms in duration including 20-ms cosine-squared rise/fall times. The carrier level was fixed at 70 dB SPL. Modulation was applied during the whole period of the carrier. The expression describing the masker-alone waveform $W_m(t)$ is

$$W_m(t)=C(t)(1+M_m(t)) \tag{1}$$

and the expression describing the masker+signal waveform $W_{m+s}(t)$ is

$$W_{m+s}(t)=C(t)(1+M_s(t)+M_m(t)), \tag{2}$$

where $C(t)$ is the sinusoidal carrier, $M_m(t)$ is the masker modulator, and $M_s(t)$ is the signal modulator. The sinusoidal signal modulation is described by the following expression:

$$M_s(t)=m_s \sin(2\pi f_s t+\phi_s), \tag{3}$$

where m_s is the signal-modulation depth, f_s is the signal-modulation frequency, and ϕ_s is the starting phase. Three different masker modulators were used:

In the first experiment, a pair of sinusoidal modulators was used. The expression describing the two-tone masker is

$$M_m(t)=m_m \sin(2\pi f_{m_1}(t+t_0))+m_m \sin(2\pi f_{m_2}(t+t_0)), \tag{4}$$

where m_m is the modulation depth of each masker component, f_{m_1} and f_{m_2} are the frequencies of the two components, and t_0 is a constant time shift. The spectral distance between the two components was set to be equal to the frequency f_s of the signal modulation, i.e., $f_s=f_{m_2}-f_{m_1}$, with $f_{m_2}>f_{m_1}$. The modulation depth m_m of each masker component was set to 0.3. This corresponds to a modulation depth of about -10 dB in terms of $20\log(m)$. The time shift t_0 was set to $0.75/f_s$. Except for the time shift, the equation for this masker is the same as the one used in Moore *et al.* (1999). As an example, the waveform of a two-tone complex masker modulator with $f_{m_1}=180$ Hz and $f_{m_2}=210$ Hz is shown in Fig. 1 (top panel). The pair of modulation components produces beats at the difference frequency of 30 Hz. The venelope of the two-tone complex masker (dashed curve) is a rectified sinusoid with a frequency equal to half the beat frequency. Thus, the spectrum of the venelope has a fundamental-frequency component at 30 Hz but also contains higher-order harmonics which are integer multiples of the fundamental frequency. Note that 30 Hz is not contained in the envelope spectrum itself, which only contains f_{m_1} and f_{m_2}. The time shift t_0 was introduced in order to obtain a masker modulation with a venelope which has a maximum at the same time as the signal modulation with a starting phase of 0° (lower middle panel of Fig. 1). Accordingly, the maximum of the venelope coincides with the minimum of the signal having a starting phase of 180° (bottom panel of Fig. 1).

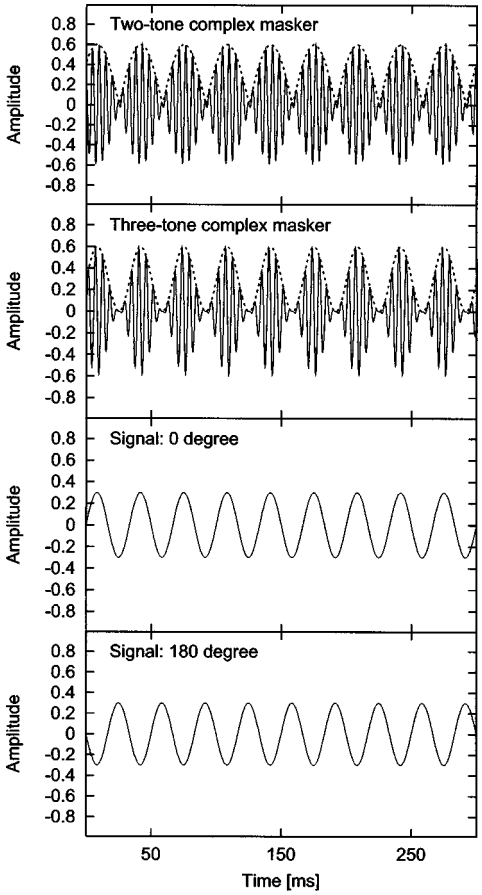


FIG. 1. The upper two panels show the temporal waveform of the two-tone complex masker (top panel) and the three-tone complex masker (upper middle panel). In addition, the dashed line shows the venelope for the two maskers. The bottom two panels show two realizations of the signal modulation: 0° starting phase (lower middle panel) and 180° starting phase (bottom panel).

In the second experiment, the masker modulation consisted of a modulation component at f_m that was sinusoidally modulated in its modulation depth. The equation describing the masker modulator is

$$M_m(t) = m_m \sin(2\pi f_m t)(1+\sin(2\pi f_s t)), \quad (5)$$

where m_m is the mean modulation depth and f_s is the signal-modulation frequency. Thus, the masker modulation was a tone with a frequency f_m which was 100% sinusoidally amplitude modulated at a rate that was equal to the signal-modulation frequency. In contrast to the two-tone complex, the venelope of the masker contains only a component at the signal frequency. As an example, the masker modulator waveform for $f_m = 180$ Hz and $f_s = 30$ Hz is shown in Fig. 1 (upper-middle panel). The venelope is a sinusoid with a frequency of 30 Hz (dashed line in the upper-middle panel of Fig. 1).

In the third experiment, the masker modulation was again a two-tone complex. However, in contrast to the masker used in the first experiment [Eq. (4)], a random starting phase of the two masker components was used for every stimulus presentation.

C. Procedure

For all experimental conditions, a three-alternative forced-choice (3AFC) procedure was used to measure the detection threshold for sinusoidal amplitude modulation in the presence of a complex-tone masker modulation. Both modulations were imposed on the same sinusoidal carrier. A trial consisted of three intervals separated by 500-ms silent intervals. In one randomly chosen interval, the carrier was modulated with the sum of the signal and masker modulation. In the two other intervals the carrier was modulated with the masker modulation alone. The subjects were instructed to choose the interval that was different from the two other intervals. Visual feedback indicating a correct or false response was provided after each trial. The signal-modulation depth, m_s, was varied according to a one-up two-down procedure. This procedure estimates the 70.7% point of the psychometric function (Levitt, 1971). The step size was 4 dB [in terms of $20\log(m_s)$] for the first two reversals, was reduced to 2 dB for the following two reversals, and set to 1 dB for the remaining eight reversals that made up one run. The threshold was determined by calculating the median of the signal-modulation depth (in dB) at these eight reversals. The procedure was repeated at least three times for each stimulus configuration and each subject. The final threshold estimate was taken as the average across the last three runs.

III. RESULTS

A. Effect of signal-modulation phase in two-tone complex masking

Figure 2 shows individual data and intraindividual standard deviations (upper panels and lower left panel) together with the mean data and interindividual standard deviations (lower right panel) for a signal modulation of 30 Hz. The upper two panels and the lower left panel show the individual data for different two-tone complex masker modulations: 90 and 120 Hz (upper left, ○), 97 and 127 Hz (upper right, □), and 180 and 210 Hz (lower left, ◇). The dotted, dashed, and dashed-dotted lines indicate data for the different subjects CH, JV, and KT, respectively. The horizontal lines indicate the individual unmasked thresholds for the 30-Hz signal modulation, i.e., the modulation depth at threshold in the absence of the masker modulation. The lower right panel shows the mean data. The same symbols as for the individual data are used to indicate the thresholds for the different modulation maskers. The horizontal line indicates the unmasked threshold for the 30-Hz signal modulation averaged across all three subjects. For the two-tone complex masker modulation with the components 90 and 120 Hz, threshold was measured between 0° and 270° in 45° steps. For the other two-tone complex masker modulations, thresholds were measured in the same range in 90° steps.

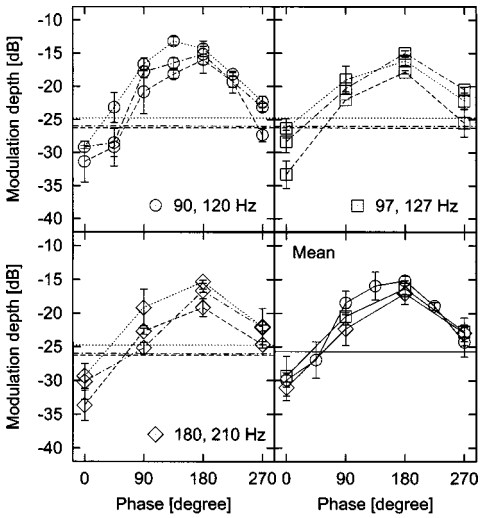


FIG. 2. Modulation depth at threshold as a function of the signal starting phase for 30-Hz signal modulation in the presence of a two-tone complex masker modulation. The horizontal line indicates the modulation depth at threshold in the absence of the modulation masker (unmasked threshold). The two top panels and the left bottom panel show individual data and intraindividual standard deviations for three two-tone complex masker modulators: 90 and 120 Hz (top left, ○), 97 and 127 Hz (top right, □), 180 and 210 Hz (bottom left, ◇). The data for the different subjects are indicated by different line types: CH (dotted), JV (dashed), KT (dashed dotted). The right bottom panel shows mean data and interindividual standard deviations for the three different two-tone complex masker modulators.

Similar unmasked thresholds for 30-Hz signal modulation were obtained for the three subjects. In general, thresholds in the presence of the masker modulation were also very similar for the different subjects. For all subjects, threshold was lowest for a signal-modulation phase of 0°, i.e., when the signal modulation was in phase with the venelope. Thresholds for this ''in-phase condition'' were below the unmasked threshold. They increased as the phase increased. For the masker modulators with components at 97 and 127 Hz (upper right), and 180 and 210 Hz (lower left), all three subjects obtained thresholds above their individual unmasked threshold for all other signal-modulation phases. For those two maskers, all subjects showed a maximum threshold at 180°, referred to as the ''antiphase condition.'' For the data with the masker-modulator components at 90 and 120

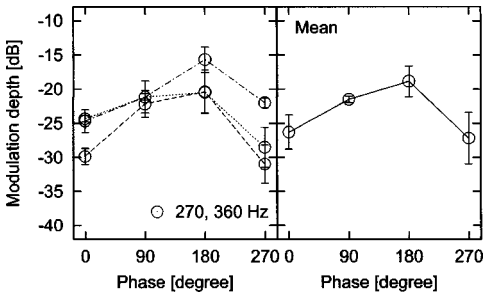


FIG. 3. Modulation depth at threshold as a function of the signal starting phase for 90-Hz signal modulation in the presence of a two-tone complex masker modulation (270 and 360 Hz). The left panel shows individual data and intraindividual standard deviations for three subjects: CH (dotted), JV (dashed), KT (dashed dotted). The right panel shows mean data and interindividual standard deviations.

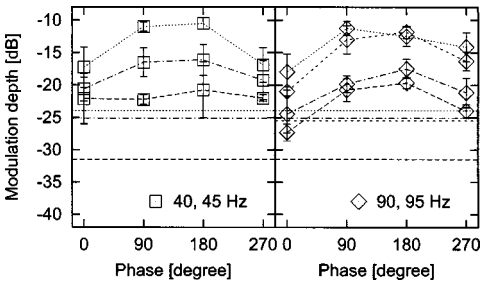


FIG. 4. Modulation depth at threshold as a function of the signal starting phase for 5-Hz signal modulation in the presence of a two-tone complex masker modulation. In addition, the horizontal lines indicate the modulation depth at threshold in the absence of a masker modulation. The two panels show individual data and intraindividual standard deviations for two two-tone complex masker modulations: 40 and 45 Hz (left, □), 90 and 95 Hz (right, ◇). The data for the different subjects are indicated by different line types: GG (short dashed), JV (long dashed), MK (dotted), and KT (dashed dotted).

Hz, only two of the three subjects (JV and KT) showed a maximum at 180°, whereas the third subject (CH) obtained a maximum threshold at 135°. It is possible that this subject might also have shown a maximum at 135° for the other two modulators if threshold had been measured for this signal-modulation phase.

On average, almost the same thresholds were measured for the four signal-modulation phases (0°, 90°, 180°, and 270°) used for all three different masker modulators (lower right panel). The maximum difference between the threshold curves is only 4 dB (at 90°). The maximum phase effect, i.e., the threshold difference between the in-phase and the antiphase condition, is 15 dB.

Figure 3 shows individual data (left panel) and mean data (right panel) for a signal modulation of 90 Hz. The masker-modulator components were in this case 270 and 360 Hz. The individual data for the different subjects are indicated by the same line types as used in Fig. 2. The unmasked threshold was not obtained for this modulation frequency. However, data in the literature indicate that thresholds are similar for 30- and 90-Hz modulators using a 5-kHz sinusoidal carrier (Dau, 1996; Kohlrausch *et al.*, 2000). As for 30 Hz, all subjects showed the highest threshold for the antiphase condition. Assuming that the subjects had the same unmasked thresholds at 30 and 90 Hz, mean thresholds for 0° and 270° were slightly below the unmasked threshold, whereas thresholds for the other two signal-modulation phases were above the unmasked threshold.

Figure 4 shows individual data for a signal modulation of 5 Hz. The left panel shows the data for a two-tone masker modulation with frequency components at 40 and 45 Hz. Such a complex was used by Moore *et al.* (1999). The right panel shows thresholds for masker components at 90 and 95 Hz, frequencies not used by Moore *et al.* (1999). The horizontal line indicates the individual unmasked threshold for GG (short dashed), JV (long dashed), MK (dotted), and KT (dashed dotted).[1] The same line types are used for the corresponding masked thresholds.

In contrast to the threshold curves obtained with 30- and 90-Hz signal modulation, subjects showed masked thresholds for the 5-Hz signal modulation that were consistently higher than their unmasked threshold for all signal phases. For the masker modulation with higher components (right panel of Fig. 4), thresholds were lowest in the in-phase condition. For three of the four subjects (GG, JV, and KT) the maximum threshold occured for the antiphase condition. For one subject (MK), the threshold at 90° was slightly higher than at 180°. For the masker modulation with components at 40 and 45 Hz (left panel of Fig. 4), two of the three subjects showed a similar threshold curve as for the masker with components at 90 and 95 Hz. However, the phase effect for three (GG, JV, and KT) of the four subjects was considerably smaller (up to 7 dB) for the masker modulation with components at 40 and 45 Hz. About the same thresholds were obtained for 90° and 180°. Only one subject (MK) obtained the same maximum phase effect for both sets of masker frequencies.

In general, fairly large individual differences of up to 11 dB were found. Thus, mean data were not calculated for the 5-Hz thresholds.

B. Effect of signal phase in three-tone complex masking

Figure 5 shows individual data (upper panels) and mean data (lower panels) for three subjects. Modulation depth at threshold for 30-Hz signal modulation is shown as a function of the signal-modulation phase for a masker-modulation depth of 0.3 (left panels), and as a function of the masker-modulation depth m_m, in dB, for the antiphase and the in-phase conditions (right panels). The masker consisted of a 180-Hz component that was modulated at 30 Hz. The horizontal lines indicate unmasked thresholds for 30-Hz signal modulation. The same line types are used as in Fig. 2 to indicate the thresholds for the different subjects.

The threshold curve as a function of the signal-modulation phase is very similar to that obtained with the two-tone complex maskers (Fig. 2). On average, thresholds are highest at 180° signal-modulation phase (antiphase condition) and lowest at 0° signal-modulation phase (in-phase condition). For 0° and 45°, thresholds are below unmasked threshold. For all other signal-modulation phases thresholds are above unmasked threshold. In contrast to the data obtained with the two-tone complex at 90 and 120 Hz, this is

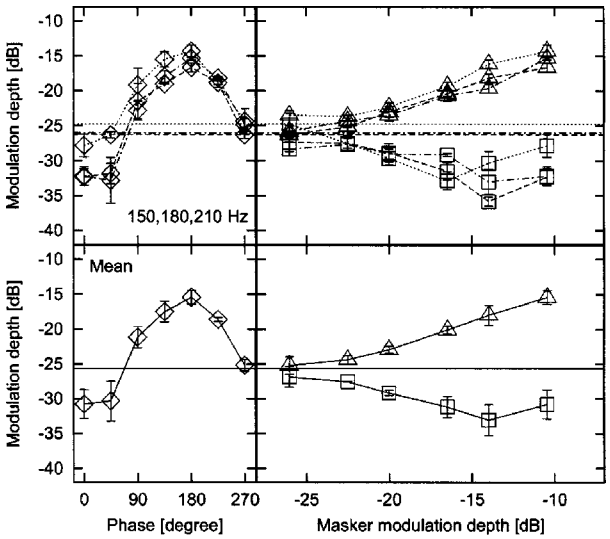


FIG. 5. Modulation depth at threshold for a 30-Hz signal modulation in the presence of a three-tone complex modulation masker (150, 180, and 210 Hz). In addition the horizontal lines indicate the threshold in the absence of a masker modulation. The left panels show the threshold as a function of the starting phase. The right panels show the threshold as a function of the modulation depth of the highest component of the three-tone complex (see text for details). The different symbols indicate different signal starting phases: 0° (□) and 180° (△). The top panels show individual data and intraindividual standard deviations of three subjects. The same line types are used as in Fig. 2 to indicate the data for the different subjects. The bottom panels show mean data and interindividual standard deviations.

true for the data of all subjects. On average, the maximum phase effect is 15 dB.

The right panel of Fig. 5 shows the threshold as a function of the masker-modulation depth for the antiphase (△) and the in-phase (□) conditions. The masker-modulation depth of the rightmost data points (about −10 dB) corresponds to that used in the left panel. For all subjects, thresholds in the antiphase condition decrease with decreasing masker-modulation depth. For the in-phase condition, thresholds decrease for masker modulation depths between −26 and −17 dB. On average, the same threshold was obtained for masker-modulation depths of −17 and −10 dB. For all subjects, the threshold difference between the antiphase and the in-phase condition decreases as the masker-modulation depth decreases. Note that, on average, a phase effect of 6 dB can still be observed at −20 dB, where the masker-modulation depth of the 180-Hz component is only 6 dB above its threshold.

C. Effect of signal-modulation frequency in random-phase tone-complex masking

Figure 6 shows individual data and intraindividual standard deviations (left panel) and mean data and interindividual standard deviations (right panel) as a function of signal-modulation frequency. The masker consisted of two components: 90 and 120 Hz. The phase of each component of the masker was chosen randomly. Thus, this condition is similar to the third experiment in Dau *et al.* (1997a), as described in the Introduction.

Approximately the same thresholds were obtained for all three subjects (differences between subjects were smaller

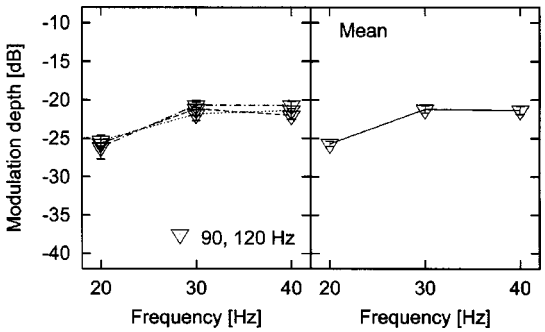


FIG. 6. Modulation depth at threshold as a function of signal-modulation frequency. The masker modulation had components at 90 and 120 Hz. The left panel shows individual data and intraindividual standard deviations. The same line types are used as in Fig. 2 to indicate the data for the different subjects. The right panel shows mean data and interindividual standard deviations.

than 2 dB). The lowest threshold was obtained for a signal-modulation frequency of 20 Hz. On average, the same thresholds were obtained at 30 and 40 Hz. Assuming the same unmasked threshold for the three signal-modulation frequencies, thresholds are, on average, 5 dB higher than the unmasked threshold at 30 and 40 Hz, and equal to the unmasked threshold at 20 Hz. In agreement with results presented in Dau (1996) and Dau *et al.* (1997a), no clear peak was found at 30 Hz.

IV. MODELING THE RESULTS

The data are compared to predictions of a model of amplitude-modulation processing in the auditory system. Three versions of the model are evaluated here differing in the nonlinear processing of the signal's envelope. The aim of the present modeling is to assess the extent to which a specific kind of nonlinearity can account for the results. The model does not specify the exact neural realization of the process. The mechanisms in the auditory system may be described by a combination of these simple nonlinearities. This is discussed in Sec. V.

A. General model structure

The first stage of all models evaluated here is a linear gammatone filter that simulates the bandpass characteristic of the basilar membrane (Patterson *et al.*, 1987). The center frequency was set equal to the carrier frequency (5 kHz). Thus, as in Moore *et al.* (1999), off-frequency listening was not considered. Ewert *et al.* (2002) showed that the envelope in an off-frequency filter tuned to the lowest sideband originating from the modulation contains a weak venelope component even without nonlinear processing. However, the size of this component in the filter tuned to either the lowest or highest sideband would increase as the frequency of the masker components increases. In contrast, the amount of the phase effect as observed in the data was independent of the spectral distance between masker modulation and signal modulation (Sec. III A). Thus, this component was considered to be negligible for the current model predictions.[2] In the second stage, the Hilbert envelope of the signal at the output of the filter is extracted. A model-specific nonlinearity produces a component at the venelope frequency. These non-

linear stages are described in detail in the next section. The output of the nonlinear stage is analyzed by a modulation filter centered at the signal-modulation frequency. This stage is used since several psychophysical and physiological data sets indicate the existence of such modulation-frequency selectivity in the auditory system (see Sec. I). Gaussian noise is added to the internal envelope representation in order to introduce a lower limit of resolution. The internal noise is assumed to be independent for every sample point and its variance is adjusted to account for the unmasked (reference) threshold for each of the three model versions. It was then kept constant for all model predictions. In order to reduce the effect of onset and offset, cosine-squared rise-fall envelopes of the same length as used for signal generation are applied to the internal representation of the stimuli. The decision device calculates the root-mean square (rms) value of the output of the modulation filter for the three different stimuli in a trial, i.e., the two reference intervals and the stimulus containing the additional signal modulation. The signal is considered to be detected if the signal interval gives the largest rms value. Only the rms value was considered since Moore and Sek (2000) showed that other measures such as the min-max value or the crest factor are not consistent with their data on the perception of three-component amplitude modulation. Thresholds are predicted using the same adaptive procedure as for the experimental data. The final threshold estimate is averaged across the threshold predictions for ten runs of the adaptive procedure.

B. Model-specific parameters

Three different model versions were used. The models differed in the nonlinearity that introduces the venelope component. Since the empirical data were similar for the broad range of modulation frequencies tested (5 to 90 Hz), only instantaneous (fast-acting) nonlinearities were considered here. It is possible, however, that the nonlinear mechanisms involved in auditory processing of amplitude modulation are to some degree time-varying, as discussed by Moore and Sek (2000). Figure 7 shows the temporal envelope (left column) and the output of the different nonlinearities (other columns) for the masker alone (top row), the in-phase signal added to the masker (middle row), and the antiphase signal added to the masker (bottom row). The masker was a three-tone complex modulation with components at 150, 180, and 210 Hz as used in the second experiment. The mean masker-modulation depth was −10 dB. The modulation depth of the 30-Hz signal modulation was set to −10 dB. Each panel shows a 100-ms interval of the steady-state part of the stimuli. The lower graph in each panel of the second to fourth column shows the output of the modulation filter tuned to the signal-modulation frequency.

(*i*) Compression model: The envelope $\mathrm{env}(t)$ was compressed by raising it to the power of 0.4. This value was also used by Moore *et al.* (1999). The output $\mathrm{out}(t)$ of this stage is given by the following equation:

$$\mathrm{out}(t) = (\mathrm{env}(t))^{0.4}. \tag{6}$$

This nonlinearity introduces a weak component in the filter at the signal frequency as shown in the top panel of the second column (lower graph) of Fig. 7. The component is 180° out of phase relative to the masker venelope [see also Ewert *et al.* (2002), their Fig. 6]. Consequently, the modulation depth at the output of the modulation filter at the signal frequency is larger for an antiphase signal (bottom panel) than for an in-phase signal (middle panel).

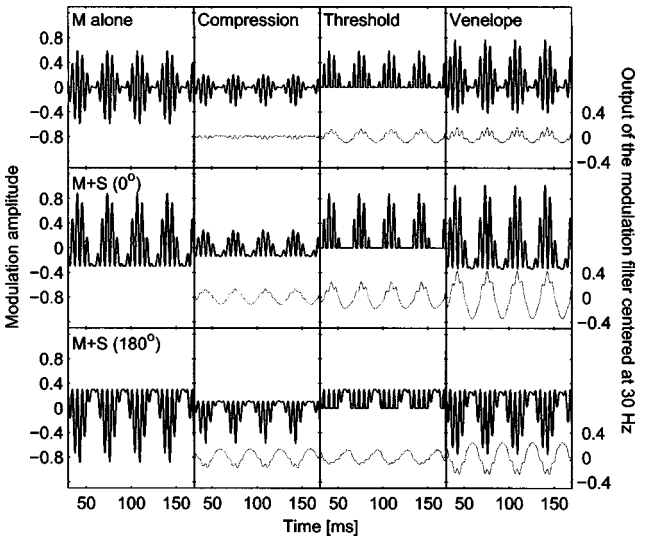


FIG. 7. Each panel of the left-hand column shows the temporal envelope of the three-tone complex masker modulation (150, 180, and 210 Hz) alone (top row) and the 30-Hz signal modulation added to the masker at a modulation depth of −10 dB. The signal starting phase was either 0° (middle row) or 180° (bottom row). The panels of the other columns show the output of the nonlinearity for three different models: compression model (second column), threshold model (third column), and the venelope model (right-hand column). In addition, the lower graph in each of these panels shows the output of the modulation filter centered at the signal frequency following the nonlinearity. The output was shifted downwards by 0.8 (see right-hand axis).

(*ii*) Threshold model: A constant value $\mathrm{mean}_t(\mathrm{env}(t))$ equal to the mean of the steady-state portion of the internal envelope representation was used as threshold. The output $\mathrm{out}(t)$ is equal to the maximum of the envelope $\mathrm{env}(t)$ and this threshold value:

$$\mathrm{out}(t) = \max(\mathrm{env}(t), \mathrm{mean}_t(\mathrm{env}(t))). \tag{7}$$

Thus, for the steady-state portion of the stimuli, the output of this stage is equal to the half-wave rectified version of the ac-coupled envelope as shown in the third column of Fig. 7 (upper graphs). Shofner *et al.* (1996) proposed a threshold as a possible physiologically motivated nonlinearity in order to account for the presence of a spectral peak at the envelope beat frequency in their physiological data (see Sec. I). The nonlinearity introduces a component at the signal frequency that is in phase with the masker venelope (lower graph in the top panel of the third column) and thus facilitates an in-phase signal (middle panel).

(*iii*) Venelope model: The venelope $\mathrm{ven}(t)$ is generated by calculating the envelope of the ac-coupled envelope of the stimuli. The output $\mathrm{out}(t)$ of this stage is the sum of the venelope (weighted by a factor of 0.3) and the original envelope:

$$\mathrm{out}(t) = \mathrm{env}(t) + 0.3\mathrm{ven}(t). \tag{8}$$

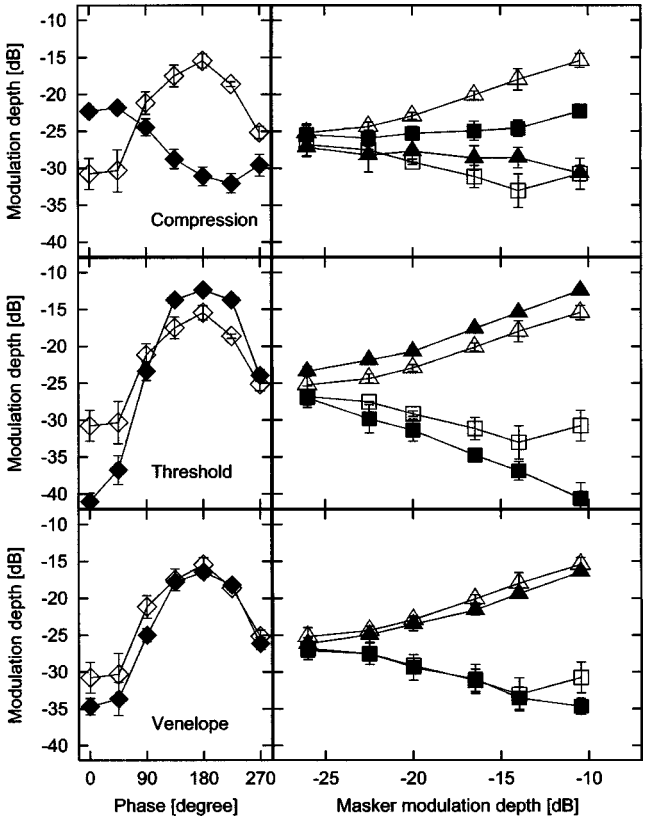


FIG. 8. Modulation depth at threshold for a 30-Hz signal modulation in the presence of a three-tone complex masker modulation (150, 180, and 210 Hz). The left panels show the threshold as a function of the starting phase. The right panels show the threshold as a function of the modulation depth of the highest component of the three-tone complex. Filled symbols indicate simulated data. The open symbols indicate the mean measured data as shown in the lower panel of Fig. 5. The top panels show simulated data for the model including a compressive nonlinearity. The middle panels show predictions of the model with a threshold at the overall level of the signal. The bottom panels show threshold obtained with the model which extracts the venelope explicitly.

This procedure was proposed by Ewert *et al.* (2002). The factor of 0.3 was chosen in order to quantitatively account for the masking effects as observed in the present data. The right column of Fig. 7 shows the output of this nonlinear stage. The nonlinear stage introduces a component at the signal frequency which is in phase with the masker venelope (upper panel). Thus, the in-phase signal is facilitated (middle panel).

C. Model predictions

In order to investigate if the models are able to quantitatively predict the effect of the phase relation between signal modulation and masker venelope at threshold, model predictions for the three-tone complex masker were compared to the experimental data. Using the data for the three-tone complex masker, it is also possible to investigate if the model can predict how masker-modulation depth affects the maximum phase effect.

Figure 8 shows the model predictions (closed symbols) together with the mean measured data (open symbols, see Fig. 5). The left panels of Fig. 8 show that all models predict thresholds that depend on signal phase. However, the compression model predicts the opposite phase effect, i.e., the highest threshold is at 0° in contrast to the data (upper panels of Fig. 8). In addition, the predicted difference between maximum and minimum threshold is too small for this model version. The threshold model and the venelope model show a phase effect similar to that observed in the measured data. However, the threshold model considerably overestimates the maximum phase effect (middle panels of Fig. 8). The model predicts a 29-dB effect while the data only show a 15-dB effect. The data in the right panel show that this model overestimates the phase effect for all masker-modulation depths. The venelope model (lower panels) provides the best fit to the data. For all masker-modulation depths except for −10 dB, the difference between measured and predicted data is smaller than 2 dB. For −10 dB, the model overestimates the effect by about 5 dB.

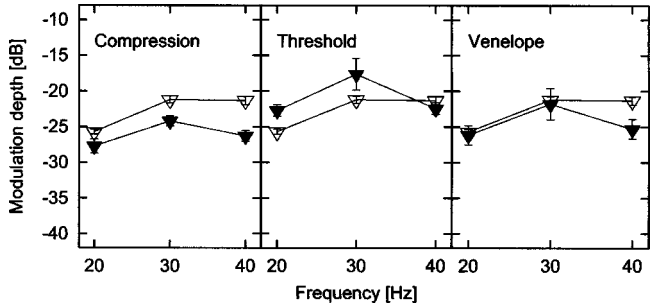


FIG. 9. Modulation depth at threshold as a function of signal-modulation frequency. The masker modulation consisted of 90 and 120 Hz. The filled symbols indicate the model predictions for the model including a compressive nonlinearity (left), the model with a threshold at the overall level (middle), and the model that extracts the venelope explicitly (right). Open symbols indicate the mean measured data as shown in the right panel of Fig. 6.

Figure 9 shows the predicted thresholds (filled symbols) for the random-phase conditions, together with the mean measured data, replotted from Fig. 6. Simulated thresholds obtained with the compression model, the threshold model, and the venelope model are shown in the left, middle, and right panels, respectively. All models predict an increase in threshold of 4 dB between 20 and 30 Hz. For the compression model, thresholds are always 2–3 dB lower than the measured thresholds. For the threshold model, predicted thresholds are 3 dB above the measured thresholds for 20 and 30 Hz and 1 dB below the measured threshold for 40 Hz. The venelope model accounts for the thresholds for 20 and 30 Hz, while for 40 Hz, the predicted threshold is about 4 dB lower than the mean measured threshold. All models predict a maximum at 30 Hz that is not observed in the measured data.

V. DISCUSSION

The experimental results of the present study are consistent with the general finding of Moore *et al.* (1999) that the detection threshold for a signal modulation can be affected by the presence of a tone-complex masker beating at the rate of the signal modulation. The results support the hypothesis that a nonlinear mechanism in the auditory system introduces a distortion component at the beat frequency in the internal

representation of the envelope. The results could be explained by assuming that this envelope is then processed by a bank of overlapping modulation bandpass filters as proposed by Dau *et al.* (1997a, 1999) and Ewert and Dau (2000), where the signal and the distortion component interact within the filter tuned to the signal frequency. For a signal modulation of 30 Hz, the same magnitude of the phase effect was found irrespective of the spectral distance between signal modulation and masker components (Fig. 2). The findings are in agreement with recent studies by Lorenzi *et al.* (2001a, b) and Ewert *et al.* (2002) where it was shown that the detectability of envelope fluctuations is affected by the presence of envelope fluctuations of the same frequency range. The data of these studies and the present study consistently show that the envelope and the envelope interact at some stage of the auditory pathway. The findings are consistent with the idea that both are processed within the same modulation filter.

A. Relation to the data from Moore *et al.* (1999)

Moore *et al.* (1999) reported that the thresholds for a 5-Hz signal modulation in the presence of masker-modulator components at 40 and 45 Hz are higher than the unmasked threshold, irrespective of the signal phase. The 5-Hz data of the present study for a comparable condition (left panel of Fig. 4) are also above threshold, i.e., they confirm the data of Moore *et al.* (1999) in this respect. However, the direction of the phase effect in the 5-Hz data of this study is completely different. Moore *et al.* (1999) obtained the highest detection threshold when the maxima in the signal modulation were in phase with the maxima in the masker modulation. In contrast, in all conditions where the threshold curve showed a distinct maximum in the present study, this was obtained for the antiphase condition, and not for the in-phase condition as reported by Moore *et al.* (1999). The same result was found for more distant masker components (right panel of Fig. 4).

The results for the two other signal-modulation frequencies tested here (30 and 90 Hz) showed the highest thresholds for the antiphase condition, in contrast to the results of Moore *et al.* (1999). For these signal frequencies, thresholds fell below the unmasked threshold (negative masking) for some signal phase conditions, in contrast to the findings for 5 Hz.

In summary, all data from the present study consistently show the opposite phase effect to that reported by Moore *et al.* (1999). The reason for the different results in Moore *et al.* remains unclear to the authors of the present study.[3]

B. Effect of the random masker phase

For the random-phase two-tone complex masker with masker components at 90 and 120 Hz, thresholds were the same for signal frequencies of 30 and 40 Hz (see Fig. 6). The threshold for a signal frequency of 20 Hz was about 4 dB lower. The lack of a prominent peak at the period of the masker venelope (30 Hz) is in agreement with the data of Dau *et al.* (1997a) using a random-phase five-tone complex masker. Dau *et al.* (1997a) concluded from their results that modulation masking is only determined by the spectral content of the masker modulator rather than by the temporal periodicity of the masker. However, the thresholds for 20, 30, and 40 Hz reported here do not increase continuously as would be expected if only the masker components themselves and not the venelope component(s) determined threshold. Note that the range of signal frequencies is much smaller than used in Dau *et al.* (1997a). Thus, the effect of signal frequency will be much smaller in this restricted range of signal frequencies than reported in Dau *et al.* for the range from 20 to 120 Hz. However, their average data in the same range (20 to 40 Hz) are very similar. The threshold increases considerably between 20 and 30 Hz, whereas hardly any increase is observed between 30 and 40 Hz. Thus, the data in the present study and in Dau *et al.* (1997a) may also be interpreted as showing a slightly elevated threshold for a signal-modulation frequency equal to the masker venelope frequency. The lack of a clear masking effect at the masker venelope frequency, where the phase relation between the masker venelope and the signal envelope is random, can be explained from the amount of masking observed for the different fixed-phase conditions in the first experiment. For the anti-phase condition, thresholds were elevated in the presence of the two-tone complex masker compared to the unmasked threshold; for the in-phase condition, a threshold decrease was observed (negative masking). Thus, for random phase, the presence of the masker will hamper the detection of the signal modulation in some trials and will be beneficial for signal detection in other trials. The masking effect expected for the random-phase masker will be close to the mean threshold for the different phase conditions. The mean threshold for the different phase conditions (0, 90, 180, and 270°), taken from the first experiment (with masker components at 90 and 120 Hz), is -21.9 dB, while the threshold for the random-phase masker is -21.2 dB at a signal frequency of 30 Hz. The venelope model can account for the reduced effect of the masker venelope in random-phase conditions by simulating the experiment explicitly, i.e., by using the same random-phase masker and the same psychoacoustical procedure as were used in gathering the psychoacoustical data (see Fig. 9). However, it seemed to be equally important that the model can quantitatively predict the phase effect for the deterministic masker. The compression model underestimates the phase effect, which is also reflected in the smaller masking effect for the random masker. The threshold model overestimates the effect of phase on the threshold and also overestimates the masking effect for the random masker. All models predict a decrease in threshold between 30 and 40 Hz which is not observed in the measured data. The reason for the elevated measured threshold at 40 Hz is unclear and needs further investigation.

C. Effect of the harmonic relation between signal and masker components

Lorenzi *et al.* (2001b) found noticeable notches in the second-order temporal modulation transfer function (TMTF) when the carrier of the second-order modulation was around twice the venelope frequency. They proposed that this may be connected to a similar effect first observed by Strickland and Viemeister (1996) in the envelope-frequency domain. Strickland and Viemeister (1996) showed that, using one

sinusoidal masker modulator, modulation depth at threshold can depend on the relative phase between the signal and the masker if the masker modulation and the signal modulation are harmonically related. The present data for the 30-Hz signal modulation in the presence of a two-tone complex masker modulation indicate that the same maximum phase effect was found for harmonically related masker components (e.g., 90 and 120 Hz) and masker components with no harmonic relation to the signal modulation (97 and 127 Hz, see Fig. 2). Thus, an exact harmonic relation between venelope carrier frequency and signal-modulation frequency does not seem to be a crucial prerequisite for the interaction between signal modulation and masker venelope.

D. Role of signal frequency

The data in the present study showed that (*i*) the maximum amount of the phase effect was almost the same for the signal frequencies 5 and 90 Hz (see Figs. 4 and 3), and that (*ii*) the amount of masking was strongly dependent upon the relative phase between signal modulation and masker venelope. While in the present study, signal and masker modulation were applied to the same carrier, Sheft and Yost (1997) did similar experiments where signal and masker modulation were applied to different carriers. They found a different pattern of results: the modulation detection interference (MDI) between the signal modulation and the beating of the masker components was reduced when the signal frequency (and beat rate) was increased from 4 to 10 Hz. Furthermore, MDI studies have shown a general lack of consistent phase effect across subjects (Yost and Sheft, 1989; Bacon and Konrad, 1993). At present it is not clear what mechanisms are responsible for the effects found in the MDI studies. It is possible that a later processing stage, where information is combined across frequency, is responsible for the (modulation) frequency-dependent masking effects. This needs to be clarified in future studies.

E. Role of compression

Moore *et al.* (1999) showed (for a 5-Hz signal modulation masked by a two-tone complex masker) that a compressive nonlinearity can predict the effects observed in their data. They showed that compression causes a threshold increase if the maxima in the masker venelope coincide with the maxima in the signal-modulation waveform whereas it predicts a threshold decrease compared to the unmasked threshold for an antiphase signal modulation. Figure 8 (upper panels) showed similar predictions for a 30-Hz signal modulation masked by a three-tone complex masker modulation using a similar model in the present study. The compression model was shown to be able to predict an interaction between the masker venelope and the signal modulation, because the compression introduces a distortion component at the difference frequency. This distortion component has its minima at the positions of the maxima in the masker venelope (see Fig. 7). The distortion component cancels the signal modulation and hence increases threshold if the signal is in phase with the masker venelope. For the antiphase signal, the distortion component is in phase with the signal modulation and enhances detectability.

However, the experimental data in the present study show a threshold minimum in the in-phase condition. Thus, the simulated threshold curve obtained for the compression model is 180° out of phase relative to the experimental data. As a consequence, it does not seem to be plausible that the phase effect is connected to a compressive nonlinearity as, for example, observed at the level of the basilar membrane. This is also consistent with recent findings by Tandetnik *et al.* (2001). They concluded on the basis of their results with hearing-impaired listeners that cochlear damage and the associated loss or reduction of the compressive nonlinearity in the input–output function of the basilar membrane [see Moore (1995) for a review] has little effect on the perception of the venelope. Ewert (2002) repeated the experiments of the present study using hearing-impaired listeners. The results were similar to the results for normal-hearing listeners reported here, again indicating little or no role of basilar-membrane compression in venelope processing.

F. Role of other nonlinearities

The other two model approaches used in the present study follow the ideas from previous studies (Shofner *et al.*, 1996; Ewert *et al.*, 2002). A model that includes a threshold can account for the direction of the phase effect (Fig. 8, middle panels). However, the maximum threshold difference produced by phase changes is considerably overestimated for a large masker-modulation depth. Additional simulations with an altered threshold (not shown) indicated that it is possible to predict a maximum phase effect for the largest masker-modulation depth that is similar to the measured phase effect. However, such a shift of the threshold does not lead to a better overall agreement with the data. Particularly, with a smaller threshold, the model can no longer predict a phase effect for small masker-modulation depths. A higher threshold decreases the sensitivity of the model to the signal modulation and thus the model can no longer predict the unmasked thresholds.

The venelope model (Ewert *et al.*, 2002) shows the best overall agreement with the data (Fig. 8, bottom panel). In particular, the maximum amount of the phase effect can be accounted for very well for almost the whole range of venelope-carrier levels. In this model, the venelope was extracted separately, attenuated, and then added to the envelope prior to the modulation filterbank. The attenuation within the model is responsible for the magnitude of the phase effect. It was set to 0.3 to match the current results. The predictions would be similar to those of the threshold model if the venelope were not attenuated before summation. This suggests that a separate venelope extraction stage is needed to account for the data. However, venelope extraction by means of a second-order envelope as proposed in Ewert *et al.* (2002) should be considered as a mathematical simplification of the underlying physiological mechanism. More realistically, one could imagine the extraction of the venelope by a threshold mechanism as described in the threshold-model version. The difference to that model remains the separate pathway fol-

lowed by a weighted sum of the output of this pathway and the envelope pathway. The output out(t) of this nonlinearity would be described by the following equation:

$$\text{out}(t) = \text{env}(t) + c\ \max(\text{env}(t), \text{mean}_t(\text{env}(t))), \qquad (9)$$

where c is the weighting factor. The simplest possible physiological realization of this model approach might be a three-unit circuit: one unit receives excitatory inputs from two cells with a similar best frequency, one unit with a high threshold extracts the envelope and one unit with a lower threshold encodes the complete envelope.

Note that, theoretically, it is also possible to model the phase effect by introducing an expansive nonlinearity. Such a nonlinearity was proposed by, e.g., Zhang and Zeng (1997) to account for their loudness data with cochlear-implant persons.

G. Role of internal noise

Within the model, detectability was limited by an internal noise with a constant level that was added prior to the calculation of the rms value (see Sec. IV). Thus, it is assumed that the auditory system is sensitive to a certain fixed increment irrespective of the excitation level. This was also assumed for deterministic carriers in the modulation-filterbank model (Dau *et al.*, 1997a). In addition, the model predictions of the phase effect as a function of the masker level show that this is a reasonable approach for simulating the data for low masker levels (and therefore low envelope amplitudes), where thresholds for the in- and antiphase condition are symmetric around the unmasked threshold. However, for high masker levels this symmetry seems to disappear in the data. In the in-phase condition and for masker levels of -17 and -10 dB similar thresholds were obtained, while the current model predicts a further decrease in threshold (see bottom left panel of Fig. 8). Similar thresholds for large masker-modulation depths can be modeled if it is assumed that the internal noise changes from having a constant rms value at low levels to a rms value which is proportional to the excitation level at higher levels. Experiments on modulation-depth discrimination are in agreement with a different sensitivity for low levels (up to about 10 dB above unmasked thresholds) than for higher levels (e.g., Wakefield and Viemeister, 1990; Ewert and Dau, 2003). A smooth transition could be achieved by assuming that the limiting internal noise is a weighted sum of a constant internal noise and a level-dependent (modulation-depth dependent) internal noise. Such a model approach is currently under investigation, using data from normal-hearing and hearing-impaired listeners.

VI. SUMMARY AND CONCLUSIONS

(i) Thresholds for detecting signal modulation with frequencies of 5, 30, and 90 Hz were affected by the presence of tone-complex masker modulations that had a variation in the modulation depth (venelope) at the signal-modulation rate. The thresholds were consistently lower in the in-phase condition, where the maxima in the signal modulation coincide with the maxima in the masker venelope, than in the antiphase condition. This phase effect was the same for all signal-modulation frequencies tested.

(ii) The phase effect was independent of the spectral distance between signal and masker in the envelope spectrum. Similar results were obtained for a two-tone complex masker beating at the rate of the signal modulation and a three-tone complex masker with a sinusoidal venelope.

(iii) The observed phase effect persisted for masker-modulation components which were only 6 dB above their own threshold.

(iv) For the 30-Hz signal modulation in the presence of the three-tone complex masker, the threshold increase for the in-phase condition and the decrease for the antiphase condition were approximately symmetric around the unmasked threshold. This explains the finding that hardly any masking effect was observed in the random-phase masker condition. This finding is also in line with a previous study by Dau *et al.* (1997a), where a random-phase tone-complex masker was used.

(v) A compressive nonlinearity, as proposed by Moore *et al.* (1999), cannot account for the data. Compared to the experimental data, the threshold curve predicted using a compressive nonlinearity shows the opposite phase effect and underestimates the amount of interference.

(vi) The phase effect and the amount of interference between the signal modulation and the venelope fluctuation can be accounted for by the venelope model, as proposed in Ewert *et al.* (2002). An important feature of this model is the assumption of a separate pathway for the venelope followed by the weighted sum of the output of this pathway and the envelope pathway prior to a modulation filterbank.

ACKNOWLEDGMENTS

This research was supported by the Deutsche Forschungsgemeinshaft (DFG). We thank Neal Viemeister and Brian Moore for their many helpful suggestions during the review process.

[1]For subject KT, quiet threshold was not measured for 5-Hz signal modulation. The quiet threshold shown here is averaged across the data for frequencies of 1.5, 2, 4, and 8 Hz, taken from another unpublished study. In this unpublished study, the modulation was imposed on a sinusoidal carrier with the same carrier frequency as used in the present study. The carrier level was 65 dB SPL.

[2]These assumptions hold for symmetric and linear filters as used in the present model predictions. However, for filters with asymmetric shape and frequency-dependent compression within their passband off-frequency information might have to be considered.

[3]Recently, Moore mentioned in a personal communication that he and his colleagues could not replicate their original results presented in Moore *et al.* (1999) but instead found similar phase effects as described in the present study.

Bacon, S. P., and Grantham, D. W. (**1989**). ''Modulation masking: Effects of modulation frequency, depth, and phase,'' J. Acoust. Soc. Am. **85**, 2575–2580.

Bacon, S. P., and Konrad, D. L. (**1993**). ''Modulation detection interference under conditions favoring within- or across-channel processing,'' J. Acoust. Soc. Am. **93**, 1012–1022.
Dau, T. (**1996**). ''Modeling auditory processing of amplitude modulation,'' Ph.D. thesis, Universität Oldenburg, BIS-Verlag Oldenburg. An online version is available at the Internet site ⟨http://www.bis.uni-oldenburg.de/bisverlag/daumod96/daumod96.html⟩.
Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). ''Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers,'' J. Acoust. Soc. Am. **102**, 2892–2905.
Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). ''Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration,'' J. Acoust. Soc. Am. **102**, 2906–2919.
Dau, T., Püschel, D., and Kohlrausch, A. (**1996**). ''A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure,'' J. Acoust. Soc. Am. **99**, 3615–3622.
Dau, T., Verhey, J., and Kohlrausch, A. (**1999**). ''Intrinsic envelope fluctuations and modulation-detection thresholds for narrowband noise carriers,'' J. Acoust. Soc. Am. **106**, 2752–2760.
Derleth, R. P., and Dau, T. (**2000**). ''On the role of envelope fluctuation processing in spectral masking,'' J. Acoust. Soc. Am. **108**, 285–296.
Ewert, S. D. (**2002**). ''Auditory spectro-temporal processing in the envelope-frequency domain: experiments and models,'' Ph.D. thesis, Universität Oldenburg, BIS-Verlag Oldenburg. An online version is available at the Internet site ⟨http://docserver.bis.uni-oldenburg.de/publikationen/dissertation/2003/eweaud02/pdf/eweaud02.pdf⟩.
Ewert, S. D., and Dau, T. (**2000**). ''Characterizing frequency selectivity for envelope fluctuations,'' J. Acoust. Soc. Am. **108**, 1181–1196.
Ewert, S. D., and Dau, T. (**2003**). ''External and internal limitations in amplitude-modulation processing,'' J. Acoust. Soc. Am. (submitted).
Ewert, S. D., Verhey, J. L., and Dau, T. (**2002**). ''Spectro-temporal processing in the envelope-frequency domain,'' J. Acoust. Soc. Am. **112**, 2921–2931.
Fleischer, H. (**1982**). ''Modulationschwellen von Schmalbandrauschen,'' Acustica **51**, 154–161.
Fleischer, H. (**1983**). ''Modulation thresholds of narrow noise bands,'' in *Proceedings of the 11th ICA*, Paris, pp. 99–102.
Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–65.
Hall, J. L. (**1997**). ''Asymmetry of masking revisited: Generalization of masker and probe bandwidth,'' J. Acoust. Soc. Am. **101**, 1023–1033.
Hellman, R. P. (**1972**). ''Asymmetry of masking between noise and tone,'' Percept. Psychophys. **11**, 241–246.
Houtgast, T. (**1989**). ''Frequency selectivity in amplitude-modulation detection,'' J. Acoust. Soc. Am. **85**, 1676–1680.
Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). ''The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers,'' J. Acoust. Soc. Am. **108**, 723–734.
Langner, G., and Schreiner, C. (**1988**). ''Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanism,'' J. Neurophysiol. **60**, 1799–1822.
Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Lorenzi, C., Simpson, M. I. G., Millman, R. E., Griffiths, T. D., Woods, W. P., Rees, A., and Green, G. G. R. (**2001a**). ''Second-order modulation detection thresholds for pure-tone and narrow-band noise carriers,'' J. Acoust. Soc. Am. **110**, 2470–2478.
Lorenzi, C., Soares, C., and Vonner, T. (**2001b**). ''Second-order temporal modulation transfer functions,'' J. Acoust. Soc. Am. **110**, 1030–1038.
Martens, J. (**1982**). ''A new theory of multitone masking,'' J. Acoust. Soc. Am. **72**, 397–405.
Moore, B. C. J., ed. (**1995**). *Hearing, Handbook of Perception and Cognition*, 2nd ed. (Academic, San Diego).
Moore, B. C. J., and Sek, A. (**2000**). ''Effects of relative phase and frequency spacing on the detection of three-component amplitude modulation,'' J. Acoust. Soc. Am. **108**, 2337–2344.
Moore, B. C. J., Sek, A., and Glasberg, B. R. (**1999**). ''Modulation masking produced by beating modulators,'' J. Acoust. Soc. Am. **106**, 908–918.
Palmer, A. R. (**1995**). ''Neural signal processing,'' in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).
Patterson, R. D., Nimmo-Smith, J., Holdsworth, J., and Rice, P. (**1987**). ''An efficient auditory filterbank based on the gammatone function,'' paper presented at a meeting of the IOC Speech Group on Auditory Modelling at RSRE.
Ruggero, M. A. (**1992**). ''Physiology and Coding of Sound in the Auditory Nerve,'' in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay, Springer Handbook of Auditory Research (Springer, New York), Chap. 2, pp. 34–93.
Schulze, H., and Langner, G. (**1997**). ''Periodicity coding in the primary auditory cortex of the Mongolian gerbil (*Meriones unguicilatus*): two different coding strategies for pitch and rhythm,'' J. Comp. Physiol., A **181**, 651–663.
Schulze, H., and Langner, G. (**1999**). ''Auditory cortical responses to amplitude modulation with spectra above frequency reseptive fields: evidence for wide spectral integration,'' J. Comp. Physiol., A **185**, 493–508.
Sheft, S., and Yost, W. A. (**1997**). ''Modulation detection interference with two-component masker modulation,'' J. Acoust. Soc. Am. **102**, 1106–1112.
Shofner, W. P., Sheft, S., and Guzman, S. J. (**1996**). ''Response of ventral cochlear nucleus units in the chinchilla to amplitude modulation by low-frequency, two-tone complexes,'' J. Acoust. Soc. Am. **99**, 3592–3605.
Strickland, E. A., and Viemeister, N. (**1996**). ''Cues for discrimination of envelopes,'' J. Acoust. Soc. Am. **99**, 3638–3646.
Tandetnik, S., Garnier, S., and Lorenzi, C. (**2001**). ''Measurement of first- and second-order modulation detection thresholds in listeners with cochlear hearing loss.,'' Br. J. Audiol. **35**, 355–364.
Verhey, J. L. (**1999**). ''Psychoacoustics of spectro-temporal effects in masking and loudness perception,'' Ph.D. thesis, Universität Oldenburg, BIS-Verlag Oldenburg, ISBN 3-8142-0662-2: An online version is available at the internet site <http://docserver.bis.uni-oldenburg.de/publikationen/bisverlag/verpsy99/inhalt.html>
Verhey, J. L. (**2002**). ''Modeling the influence of inherent amplitude fluctuation simultaneous masking experiments,'' J. Acoust. Soc. Am. **111**, 1018–1025.
Verhey, J., Dau, T., and Kollmeier, B. (**1999**). ''Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model,'' J. Acoust. Soc. Am. **106**, 2733–2745.
Viemeister, N. F., (**1979**) ''Temporal modulation transfer function based on modulation thresholds,'' J. Acoust Soc. Am. **66**, 1364–1380.
Wakefield, G. H., and Viemeister, N. F. (**1990**). ''Discrimination of modulation depth of sinusoidal amplitude modulation (SAM) noise,'' J. Acoust. Soc. Am. **88**, 1367–1384.
Yost, W. A., and Sheft, S. (**1989**). ''Across-critical-band processing of amplitude modulated tones,'' J. Acoust. Soc. Am. **85**, 848–857.
Zhang, C., and Zeng, F.-G. (**1997**). ''Loudness of dynamic stimuli in acoustic and electric hearing,'' J. Acoust. Soc. Am. **102**, 2925–2934.

A measure of internal noise based on sample discrimination

Walt Jesteadt,[a] Lance Nizami, and Kim S. Schairer
Center for Hearing Research, Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 8 November 2002; revised 18 July 2003; accepted 28 July 2003)

Internal noise is often inferred from the difference between observed performance and optimum performance in detection and discrimination tasks. It can be measured directly in some cases by observing the extent to which a change in external variability impacts performance. In the studies reported here, external variability was added to an intensity discrimination task by adding a Gaussian random variable with zero mean to the overall level presented in each interval of a two-interval forced-choice task. The standard deviation of the random variable was set to half the mean difference between the levels in the two intervals, resulting in $d'_{\text{ideal}}=2$. As the mean difference and the corresponding standard deviation of the random variable decreased in size, performance was increasingly limited by internal noise, permitting a reliable estimate of internal noise to be obtained. This can be viewed as a sample discrimination task, with one component per sample. In the first study, performance was measured using 2-kHz tones presented at an average level of 70 dB SPL, with mean differences between distributions ranging from 0.1 to 2.2 dB in steps of 0.3 dB. The distributions were either Gaussian in level or in power. Conditions with no external variability were used to obtain a psychometric function. In the second study, performance was measured using 2-kHz tones presented at average levels of 50 and 90 dB SPL, with mean differences ranging from 0.4 to 2.2 dB in steps of 0.6 dB. In both studies, the measure of internal noise was highly reliable and in good agreement with the intensity difference limen (DL) estimated from the psychometric function. Analyses suggest that this measure could be used to estimate the mean difference between the decision distributions as well as the amount of internal noise in cases where the mean difference between the distributions is unknown. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1610456]

PACS numbers: 43.66.Fe, 43.66.Dc, 43.66.Ba [MRL]

I. INTRODUCTION

The optimum level of performance in detection and discrimination tasks can be predicted from knowledge of the statistical properties of the stimuli and application of standard decision-theory models (e.g., Green, 1960b). Performance of human observers is generally 5 or 6 dB poorer than ideal performance and this difference, which is relatively independent of the properties of the *external noise*, is attributed to inefficiency in the decision process (Tanner and Birdsall, 1958) or to some other source of *internal noise* (deBoer, 1966; Green, 1960a; Swets *et al.*, 1959). Internal noise represents the sum of all limitations of the observer, including transduction noise, memory variance, and criterion variance. It might be better to use the terms external and internal variability rather than noise, because the external variability is not restricted to noise waveforms and the internal variability is not associated with the perception of a noise. The terms internal and external noise have a 40-year history in the psychoacoustics literature, however, and we will continue to use them here. Although our focus is on psychophysical measurement of internal noise regardless of its source, an extensive literature on the stochastic properties of auditory-nerve responses suggests that a significant portion of the internal noise is peripheral in origin (e.g., Carney, 1994; Colburn, 1973, 1977; Javel and Viemeister, 2000; Siebert, 1968; Zhang *et al.*, 2001).

Internal noise is assumed in virtually all models of detection and discrimination and is often inferred from the difference between observed performance and optimum performance. Spiegel and Green (1981) reviewed and compared two direct measures of internal noise. The first, which we refer to as a correlation measure, is based on the relation between subjects' responses and specific properties of the stimulus waveforms (e.g., Green, 1964; Spiegel and Green, 1981; Watson, 1962). To the extent that the subject always responds the same way given the same features in the stimulus, performance is controlled by the external input, and internal noise is assumed to be low. The second measure explored by Spiegel and Green (1981), which we refer to as an additivity measure, is based on the effect of reducing the variability of the external noise (Siegel, 1979; Swets *et al.*, 1959). Swets *et al.* provide a detailed description of the logic of the latter method, which we will use here. Given two independent sources of noise, external noise, σ_e, introduced by the experimenter and internal noise, σ_i, associated with the subject, detection of a stimulus difference, Δ, is given by

$$d' = \Delta / \sqrt{(\sigma_e^2 + \sigma_i^2)}. \tag{1}$$

Spiegel and Green (1981) determined the effect of external variability by comparing performance when the exter-

[a]Electronic mail: jesteadt@boystown.org

nal noise samples were independent from one observation interval to the next to performance when the same noise sample was repeated across observation intervals. This provides a more direct measure of internal noise than can be obtained by comparing observed and ideal performance, in that the amount of external noise is manipulated within each condition. This yields an estimate of internal noise that is independent of the normal threshold measurement.

Differences in internal noise are not the source of all differences in performance across conditions. Detection of tones in noise is typically poorer, for example, when the noise is gated with the tone than when the noise is continuous (Wier *et al.*, 1977). Although this difference might be attributed to greater internal noise in the gated conditions, Spiegel and Green (1981) found that internal noise was actually lower in the gated conditions than in the continuous conditions. This is an example of the value of independent measurement of internal noise.

In fitting functional models to data, internal noise is often assumed to be uniform across all conditions (e.g., Patterson, 1976; Plack and Oxenham, 1998). In other cases, multiple factors are assumed to contribute to internal noise and only some of these factors are assumed to vary as a function of the properties of the stimuli or as a function of the procedure. Oxenham and Buus (2000), for example, accounted for different effects of duration on level discrimination of sinusoids when the stimuli were the same frequency or differed in frequency in terms of differences in internal noise. They assumed a duration-dependent internal noise that was uniform across conditions, but that had less impact on performance in the presence of a larger ''central noise'' encountered in the across-frequency conditions. They assumed that the internal noises combined as in Eq. (1) but they inferred the properties of the internal noise sources from the threshold data, rather than estimating the total internal noise using one of the measures described by Spiegel and Green (1981). This approach is common in studies of intensity discrimination, in which performance is assumed to be determined solely by different sources of internal noise (e.g., Durlach and Braida, 1969).

Measures of internal noise that rely on reduction of the variability in the external noise cannot be applied to intensity discrimination or to conditions with sinusoidal maskers because there is no external variability to be reduced. The obvious solution to this problem is to introduce external noise and observe the extent to which the *increase* in external variability impacts performance. This can be done by introducing random variation in the level of the sinusoidal masker or the pedestal in an intensity discrimination task. Limited data for such conditions can be found in the literature on sample discrimination (Berg and Robinson, 1987; Lutfi, 1989).

In a sample discrimination task, stimuli are drawn at random from alternative distributions that vary in amplitude, frequency, or duration and the subject is asked to identify the distribution from which the sample was drawn. In conditions that are most relevant to the current study, Lutfi (1989) presented a 250-Hz or 4000-Hz sinusoid whose level was drawn from a normal distribution with a mean of either 65 or 70 dB SPL and a standard deviation of 5 dB. One level was drawn from each distribution for each trial and the two levels were presented in random order in a two-interval, forced-choice (2IFC) procedure. The subject's task was to specify which interval contained the tone drawn from the distribution having the higher mean intensity. The performance of the ideal observer in this task is determined by the ratio of the mean difference between the distributions, Δ, to the common standard deviation of the distributions, σ. Because the distributions in this case are distributions of levels, we will use ΔL to refer to the mean difference. Lutfi found performance that was close to $d'_{\mathrm{idcal}}=1$ for all three of his subjects at both frequencies. In subsequent conditions using only 250-Hz tones, Lutfi found that reducing ΔL to 2.5 dB resulted in performance less than $d'_{\mathrm{ideal}}=0.5$, presumably because external noise was no longer as large with respect to internal noise. Setting both ΔL and σ to 10 dB restored near-optimal performance. Drawing tones from distributions normally distributed in power or intensity, rather than in level, also resulted in performance that was close to $d'_{\mathrm{ideal}}=1$.

Lutfi (1989) chose values of ΔL that were large with respect to internal noise, because he was concerned with how subjects combine and use information provided by multiple stimuli. To obtain estimates of internal noise itself, it is necessary to use smaller values of ΔL and σ_e that cover the range of potential values of σ_i. Different combinations of ΔL and σ_e should generate a pattern of d' values that leads to a consistent estimate of σ_i. To the extent that different values of the intensity DL within a group of subjects are determined by different amounts of internal noise, the size of the DL determined in conditions without stimulus variability should correlate with σ_i, the internal-noise estimate. Differences in performance across conditions for an individual subject may or may not be accompanied by differences in internal noise.

Another issue to be considered in specifying the stimuli is the form of the external-noise distribution. Equation (1) is invariant for all scalar transforms from external to internal units, and the limited range of stimuli within any given intensity-discrimination condition makes the equation relatively insensitive to nonlinear transforms as well. If the internal units are assumed to be a linear transform of power, and internal noise is assumed to be Gaussian in those units, then ideally the external noise should be Gaussian in power. If the internal units are assumed to be a linear transform of dB, however, then ideally the external noise should be Gaussian in dB. Lutfi obtained data for external noise distributions Gaussian in either power or dB and did not see a difference between them in terms of how well the model fitted the data. Equation (1) is not very sensitive to differences in the distributions of σ_e and σ_i. The psychometric function for intensity discrimination is probably the most sensitive measure of the form of the internal noise distribution and there is no consensus regarding its form (e.g., Buus and Florentine, 1991; Moore *et al.*, 1999). In this study, we made a direct comparison of the two forms of external variability using variability in intensity from interval to interval that was Gaussian either in level or in power.

In the first experiment, we obtained estimates of d' in an intensity discrimination task for eight values of ΔL for ped-

estal levels centered on 70 dB SPL. External variability, σ_e, was set to 0 to obtain a psychometric function and was then set to $\frac{1}{2}\Delta L$ to obtain estimates of σ_i. In these latter conditions $d'_{\text{ideal}}=2$. Values of σ_e reflected variability from interval to interval in the level of the pedestal. The first experiment included one series of conditions in which that variability was Gaussian in dB and a second series of conditions in which it was Gaussian in power. In the second experiment, we obtained data for four values of ΔL for conditions where $\sigma_e=0$ and for the same values of ΔL combined with values of $\sigma_e=\frac{1}{2}\Delta L$, for pedestal levels centered on 50 and 90 dB SPL.

II. EXPERIMENT 1: VALIDATION OF A SAMPLE-DISCRIMINATION MEASURE OF INTERNAL NOISE

The goal of the first experiment was to obtain an estimate of σ_i for intensity discrimination using a two-interval sample discrimination task where values of σ_e and ΔL were comparable in size to σ_i. Values of σ_e reflected variability from interval to interval in the level of the pedestal, whereas values of ΔL reflected the mean difference in level between the intervals. In the majority of conditions, values of σ_e and ΔL covaried, such that $\sigma_e=\frac{1}{2}\Delta L$ and $d'_{\text{ideal}}=2$. In the first set of conditions, the variability represented by σ_e was Gaussian in dB. In the second set, the same values of ΔL and σ_e were used, but the variability was Gaussian in power. In the third set, values of ΔL remained the same, but σ_e was set to 0 to obtain a psychometric function. Fits to the psychometric-function data provided a traditional measure of the limits of performance in intensity discrimination that could be compared to the estimates of σ_i obtained by fitting Eq. (1) to the sample-discrimination data.

A. Subjects

The participants were seven paid student volunteers recruited from Creighton University. Five were female (average age=19.8 years) and two were male (average age=20 years). All had immediate previous experience in intensity discrimination, having spent several months doing level discriminations of very brief 500-Hz tones. Subjects' detection thresholds for 200-ms tones with 5-ms $\cos^2$ ramps, determined using a 2IFC adaptive procedure, were lower than 15 dB SPL bilaterally at 0.5, 1, 2, and 4 kHz.

B. Stimuli and apparatus

The stimuli in all conditions were 2-kHz tones of 300-ms duration, with 10-ms $\cos^2$ ramps. All stimuli were generated digitally at a sampling rate of 50 kHz using an array processor (TDT AP2) and a 16-bit digital-to-analog converter (TDT DD1). The equipment was configured to permit up to four subjects to be tested simultaneously using adaptive procedures. The output of the DD1 was low-pass filtered at 10 kHz (TDT FT6), then routed to four parallel attenuators (TDT PA4), one per subject. Each attenuator output was routed to a mixer that had only one active input and the output of each mixer was routed to a headphone buffer (TDT HB6), then to a remote passive attenuator in each of the sound-treated rooms (see below), and finally to a Sennheiser HD 250 Linear II headphone. All stimuli were presented monaurally to the left ear.

C. Procedure

Subjects were tested as a group of three (KM, JB, TC) and a group of four (AH, AS, ES, SH) in separate sections of a double-walled Industrial Acoustics sound-treated room that had single walls between sections. The subject groups were run in 2-h or 3-h sessions, with a 10-min break between the first and second hours, and a 15-min break between the second and third hours, if a third hour was used.

Each stimulus block consisted of 100 two-interval, forced-choice (2IFC) trials, with each trial consisting of a 300-ms warning interval, two 350-ms observation intervals separated by 400 ms. Following the response of the final subject, there was an 800-ms interval before the beginning of the next trial. Subjects were given visual markers for the warning and observation intervals on the message window of a small keypad that they used to indicate their responses. For stimuli drawn from overlapping Gaussian distributions, the correct answer was defined in terms of the distributions from which the levels were drawn, not the actual levels presented on that trial. Thus, subjects could potentially choose the more intense stimulus, but be incorrect as to the distribution from which it was drawn. Because feedback might be confusing or annoying in these conditions, no feedback was given in any of the conditions.

Two tones differing in level by ΔL *on average* were presented on each trial and the subjects were asked to indicate which interval contained the higher-level tone. The value of ΔL was fixed within a block of trials. Across blocks, ΔL ranged from 0.1 to 2.2 dB in steps of 0.3 dB. The two distributions of levels for each condition were symmetric around 70 dB SPL. For $\Delta L=0.1$ dB, the two levels were 69.95 and 70.05 dB; for $\Delta L=2.2$ dB, the two levels were 68.9 and 71.1 dB. Thus, the unbiased criterion point for the subject was 70 dB SPL in all conditions.

The levels of the tones were drawn from trial-to-trial for any given block of trials according to one of three stimulus presentation conditions. In the ''constant-level'' condition, there was no external variability and each trial in a given block contained the same two levels, $70-(\Delta L/2)$ and $70+(\Delta L/2)$, presented in random order. In the ''Gaussian-level'' condition, the tone levels were drawn separately for each interval from two Gaussian distributions of level in dB SPL. The means of the lower and upper distributions were $70-(\Delta L/2)$ and $70+(\Delta L/2)$, respectively. The common standard deviation of the two distributions, σ_e, was always $\frac{1}{2}\Delta$. For $\Delta L=2.2$ dB, for example, the level in one interval was selected at random from a distribution with a mean of 68.9 dB and a standard deviation of 1.1 dB, while the level for the other interval was selected from a distribution with a mean of 71.1 dB and standard deviation of 1.1 dB. The Gaussian distributions were limited to a range of ± 3 standard deviations.

In the ''Gaussian-power'' condition, the tone levels were drawn separately interval-by-interval from two distributions that were Gaussian in power, but that were otherwise com-

TABLE I. Summary of Gaussian power conditions used in experiment 1. The intent was to convert the mean of the lower distribution to power, divide it by the ratio of I/σ_e to obtain a standard deviation in power, and use that standard deviation for both distributions. This would have resulted in $d'_{\text{deal}} = 2$ for all values of ΔL. Because the standard deviations of the upper distributions were computed independently using the same ratio of I/σ_e, but with the higher mean power, those standard deviations were higher than intended. This resulted in values of d' ideal that were lower than intended and unequal across conditions.

Delta L	Mean level of lower distribution	Mean level of upper distribution	I/σ_e	d'_{deal}
0.1	69.95	70.05	85.86	1.98
0.4	69.8	70.2	20.73	1.91
0.7	69.65	70.35	11.44	1.83
1.0	69.5	70.5	7.72	1.76
1.3	69.35	70.65	5.73	1.68
1.6	69.2	70.8	4.49	1.61
1.9	69.05	70.95	3.64	1.53
2.2	68.9	71.1	3.03	1.46

parable to the distributions used in the Gaussian-level condition. The means of the lower and upper power distributions corresponded to SPLs of $70-(\Delta L/2)$ and $70+(\Delta L/2)$, respectively. The levels within each Gaussian distribution were computed using an approach described by Lutfi (1989) in which the mean level of the distribution was converted to power and the desired standard deviation (s.d.) was expressed as a proportion of the mean. *Z*-score values limited to a range of ± 3 were then drawn at random during the stimulus-generation process and multiplied by the standard deviation in power units to arrive at a quantity that was added to the mean power and converted back to dB SPL. Note that the s.d. proportion cannot exceed 1/3 because any larger value will result in specification of negative power for $Z=-3.0$. To maintain mean values of the Gaussian power distributions that were equal in dB SPL to those used in the other sets of conditions, we converted those dB values to power and solved for the s.d. proportion necessary to satisfy the requirement that lower mean power $+2$ s.d.=upper mean power. Those proportions, which ranged from 1/3.03 for $\Delta L = 2.2$ dB to 1/85.86 for $\Delta L = 0.1$ dB, are given in Table I.

Unfortunately, due to a programming error, the requirement that $\sigma_e = \frac{1}{2}\Delta L$ and that $d'_{\text{ideal}} = 2$ was not maintained across Gaussian-power conditions. After discovering the error during data analysis, we corrected for it by computing the expected values of σ_{e} for each condition, given the error, and verified that these values were in fact correct by analysis of the levels presented on each trial, recorded with the trial-by-trial responses. We then used the expected values of σ_e to compute values of d'_{ideal} for each value of ΔL. These are also given in Table I.

For each of the three experimental conditions, ordered as ''constant level,'' ''Gaussian level,'' and ''Gaussian power,'' the subjects completed three blocks of practice trials, with decreasing values of ΔL: 2.2, 1.0, and 0.4 dB. The data were then collected with conditions tested in counterbalanced order, as follows. Starting with the constant level condition, the subjects completed two blocks of trials for $\Delta L = 2.2$ dB, then two blocks of trials for $\Delta = 1.9$ dB, and so on, dropping 0.3 dB for each successive pair of blocks. After completing $\Delta L = 0.1$ dB, the subjects started this process again for the Gaussian level conditions, and then for the Gaussian power conditions. This testing sequence was then reversed, starting with the Gaussian power condition and $\Delta L = 0.1$ dB, and finishing with the constant level condition and $\Delta L = 2.2$ dB. This resulted in data for eight-point functions for each of the three experimental conditions, constant level, Gaussian level, and Gaussian power, with 400 trials per point. For the constant level conditions, the resulting data provided standard psychometric functions.

D. Results

Values of d' and the criterion measure β were computed for each 100-trial block for each condition for each subject from the trial-by-trial data, and means and standard deviations of the four d' values per condition were computed for use in later analyses. No orderly criterion effects were expected in a two-interval, forced-choice task and none was observed.

Data for conditions without external variability were used to construct psychometric functions (PFs). Many alternative procedures and coordinates have been suggested for fitting PFs for intensity discrimination (e.g., Buus and Florentine, 1991; Dai, 1995; Moore *et al.*, 1999). Our primary goal was to use best-fitting PFs to obtain an estimate of the intensity difference required for $d'=1$, to serve as a measure of internal noise obtained in the absence of external noise. The coordinates of the PF, however, have some bearing on the choice of the optimum distribution of external noise. Because our eight-point PFs with 3200 trials per PF were better defined than those available in the literature, we considered several alternative coordinates suggested in the literature and chose the coordinates that permitted the best fit to the data. Moore *et al.* (1999) used a similar empirical approach. The PFs for individual subjects were fitted with straight lines in coordinates of d' vs ΔL, $\Delta I/I$, or $10\log(\Delta I/I)$, with slopes and intercepts chosen to minimize the squared deviation of the observed and predicted values of d'. The fits of d' against ΔL and $\Delta I/I$ resulted in intercepts close to zero and were repeated with functions constrained to pass through the origin. Both the two-parameter and the one-parameter fits were slightly better in ΔL than in $\Delta I/I$ coordinates, although the relation between these two coordinate systems is nearly linear over the range of intensity differences that we used. Fits were poorer in $10\log(\Delta I/I)$. We therefore used the one-parameter ΔL fits in later analyses. PFs for the seven subjects and for mean d' values across subjects are shown in Fig. 1. Values of the slope parameter and goodness of fit are given in Table II, along with DL_{I}, an estimate of ΔL required for $d'=1$ obtained by taking the reciprocal of the slope. Although the intensity DL frequently is referred to in the literature as ΔL, we will use DL_{I} here because ΔL is the clearest label for the mean difference between the sample distributions.

Data for conditions with external variability were used to obtain an independent estimate of the amount of internal noise, σ_i, by fitting Eq. (1) to the eight values of d', using

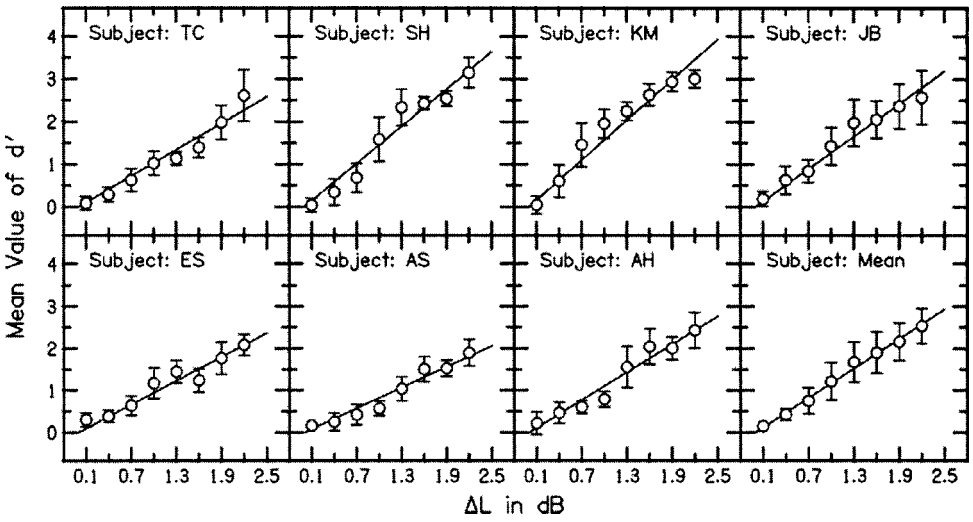


FIG. 1. Psychometric functions in experiment 1 for seven subjects and means across subjects. The error bars for individual subjects reflect variability in four estimates of d', each based on 100 trials. The error bars for the mean data reflect variability in the results for the seven subjects. Slopes and values of r^2 are given in Table II.

the value of Δ and σ_e for each condition and selecting the value of σ_i that minimized the squared deviation of the predicted d' values from the data. We initially fitted the data by assuming that the internal and external noise were additive in dB for the conditions where the external variability was Gaussian in dB and that they were additive in power for the conditions where the external variability was Gaussian in power. The shape of the external distribution does not, of course, determine the metric in which the internal and external noises are additive. We therefore considered two other alternatives, addition in power for the case where the external variability was Gaussian in dB and addition in dB for the case where the external variability was Gaussian in power. Other internal metrics are possible and we lack a method for selecting the metric that is optimum. Of the alternatives we considered, the fits were equally good for the condition in which the external variability was Gaussian in dB whether the internal and external noise were assumed to add in dB or in power. The fits were somewhat better for the condition in which the external variability was Gaussian in power when the internal and external noise were assumed to add in dB and the two sets of estimates were in good agreement with one another. We therefore report results for both external-variability conditions, assuming that the units in Eq. (1) are in dB. The data and best-fitting functions for the seven subjects and the mean d' values across subjects, for conditions in which the external variability was Gaussian in dB, are shown in Fig. 2. The corresponding data and functions for conditions in which the external variability was Gaussian in power are shown in Fig. 3. Values of σ_i for both sets of conditions are given in Table II, along with measures of goodness of fit.

Figure 4 shows a scatter plot for the seven subjects comparing estimates of internal noise, DL_I, obtained from the psychometric functions to two corresponding estimates of σ_i obtained by fitting Eq. (1). One estimate of σ_i was taken from Fig. 2, where external variability was Gaussian in dB, and the other from Fig. 3, where external variability was Gaussian in power. The Pearson product-moment correlation coefficient for DL_I vs $\sigma_{i\text{-dB}}$ is 0.70. For DL_I vs $\sigma_{i\text{-power}}$ it is 0.80. The correlation between the $\sigma_{i\text{-dB}}$ and $\sigma_{i\text{-power}}$ measures is 0.86. Because the $\sigma_{i\text{-dB}}$ and $\sigma_{i\text{-power}}$ measures were in general agreement with one another, and because we had no indication that estimates of internal noise were particularly sensitive to the distribution of external variability or to the metric in which the internal and external noise are assumed to add, we included only Gaussian-level conditions in the second experiment and based further analyses on the assumption that the effects of internal and external noise add in dB.

TABLE II. Summary of results for experiment 1. The first three columns describe the psychometric functions fitted to the data obtained in the constant-level condition. Values of r^2 show the proportion of variance in the data accounted for by the estimates. The DL_I is the reciprocal of the slope. The next two columns show the estimated internal noise and goodness of fit for that estimate, for the Gaussian-level condition. The final two columns show the estimated internal noise and goodness of fit for that estimate for the Gaussian-power condition. The effects on internal and external noise were assumed to combine in dB in both cases. The mean row contains values describing fits to the mean data.

Subject	PF slope	r^2	DL_I	$\sigma_{i\text{-dB}}$	r^2	$\sigma_{i\text{-power}}$	r^2
TC	1.03	0.95	0.97	0.99	0.97	0.98	0.94
SH	1.46	0.95	0.69	0.68	0.97	0.69	0.97
KM	1.57	0.93	0.64	0.62	0.81	0.52	0.91
JB	1.27	0.96	0.79	0.41	0.91	0.44	0.91
ES	0.95	0.92	1.06	0.90	0.91	1.23	0.98
AS	0.82	0.95	1.21	0.94	0.95	0.99	0.89
AH	1.11	0.95	0.90	0.95	0.96	0.85	0.95
Mean	1.17	0.99	0.85	0.77	0.97	0.78	0.98

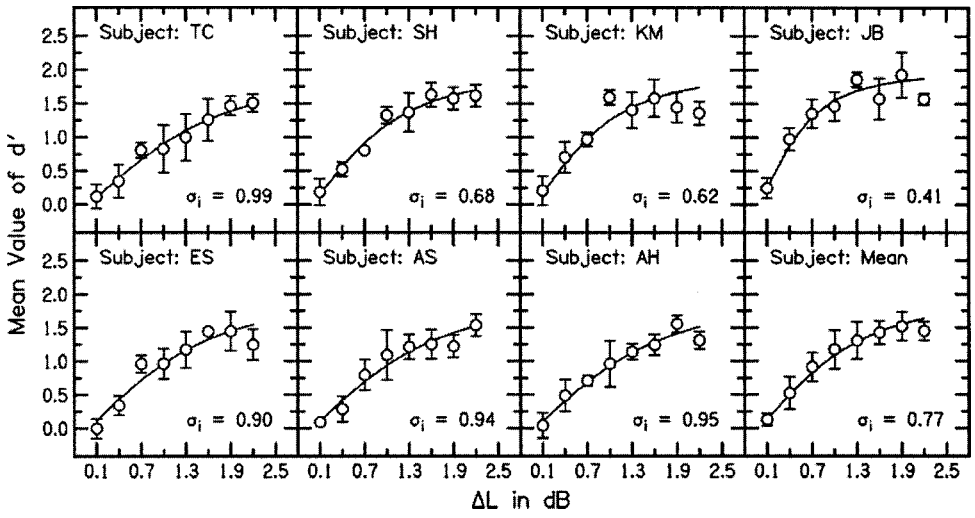


FIG. 2. Values of d' for seven subjects and means across subjects in conditions where external variability, Gaussian in dB, was added to each observation interval. The functions were generated from Eq. (1), with the value of σ_i chosen to maximize goodness of fit. Error bars here and in all later figures are as described for Fig. 1.

III. EXPERIMENT 2: THE EFFECT OF INTENSITY ON INTERNAL NOISE

The goal of the second experiment was to obtain an estimate of σ_i for intensity discrimination under conditions in which performance in the intensity discrimination task was expected to change.

A. Subjects, stimuli, apparatus, and procedure

The subjects, stimuli, apparatus, and procedure were the same as in experiment 1, with the following exceptions. The value of ΔL ranged from 0.4 to 2.2 dB in steps of 0.6 dB. The two distributions of levels were symmetric around 50 or 90 dB SPL. In conditions with external variability, the variability was Gaussian in dB. The data for conditions centered on 50 dB SPL were collected first. Training consisted of one practice block at $\Delta L = 2.2$ dB for the constant-level condition and one block at $\Delta L = 2.2$ dB for the Gaussian-level condition. Starting with the constant-level condition, the subjects then completed pairs of trial blocks for $\Delta L = 2.2$ dB, 1.6 dB, 1.0 dB, and finally 0.4 dB. This sequence was then repeated for the Gaussian-level condition. The conditions were then tested in reverse, starting with the Gaussian-level condition and $\Delta L = 0.4$ dB, and finishing with the constant-level condition and $\Delta L = 2.2$ dB. This testing order was then repeated to obtain data for levels centered on 90 dB SPL.

B. Results

Values of d' and the criterion measure β were computed for each 100-trial block for each condition for each subject from the trial-by-trial data and means and standard deviations of the four d' values per condition were computed for use in later analyses. The new data were combined with data for the same values of ΔL from experiment 1, for pedestal levels centered on 70 dB SPL, to create three-point functions describing performance and internal noise as a function of level. PFs were constructed by fitting straight lines through the origin in coordinates of d' vs ΔL, the coordinates found to be optimum in the analysis of data from experiment 1. The data from experiment 1, used here as the middle point of the three-point intensity function, were refitted using only the values of d' for $\Delta L = 0.4$, 1.0, 1.6, and 2.2 dB. PFs for the seven subjects and for mean d' values across subjects are shown in Fig. 5. Values of the slope parameter, the estimate of ΔL required for $d' = 1$, and goodness of fit are given in Table III. As expected, the PFs are steeper at higher levels, but performance was poorer than predicted for the largest

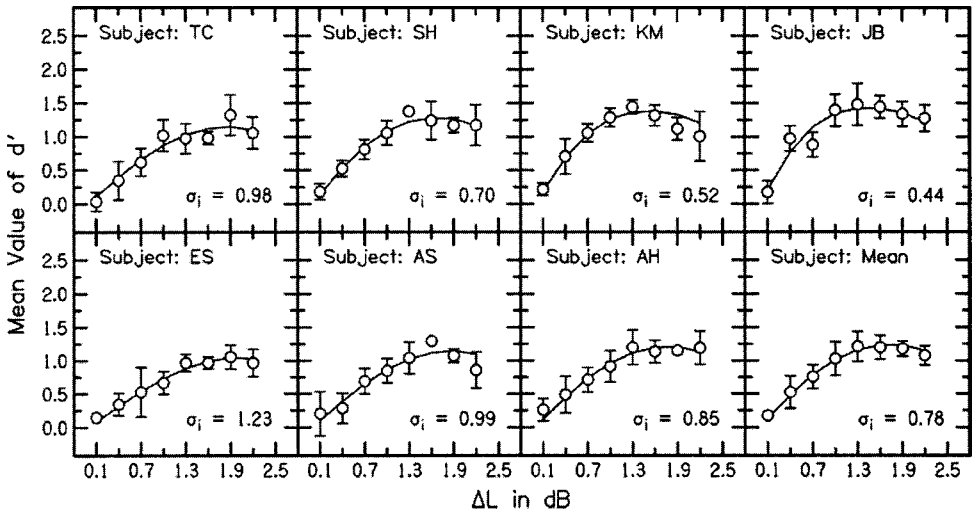


FIG. 3. Values of d' for seven subjects and means across subjects in conditions where external variability, Gaussian in power, was added to each observation interval. The functions were generated from Eq. (1), with the value of σ_i chosen to maximize goodness of fit. The shape of the function reflects a shift in d'_{deal} across values of ΔL as described in Table I.

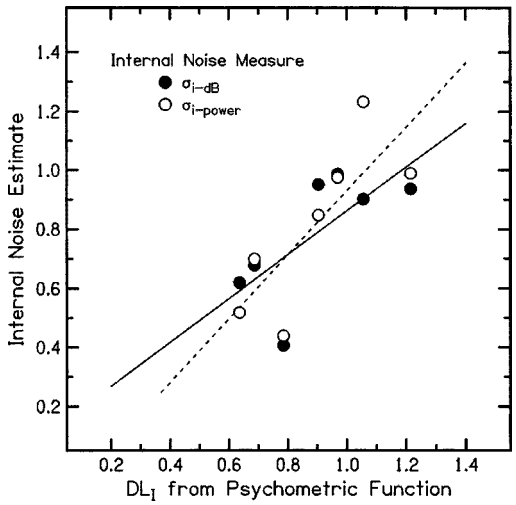


FIG. 4. A comparison of estimates of internal noise obtained in experiment 1 from the psychometric function in conditions with no external variability to estimates obtained in conditions where external variability was Gaussian in dB or in power. The solid line shows the regression function for the Gaussian dB conditions. The dotted line shows the function for the power conditions.

value of ΔL in the 90-dB SPL condition. Data for conditions with external variability were used to obtain an independent estimate of the amount of internal noise, σ_i, by fitting Eq. (1) to the four values of d' per level, using the value of ΔL and σ_e for each condition, as in experiment 1. The data from experiment 1 were again refitted using only the values of d' for $\Delta L = 0.4$, 1.0, 1.6, and 2.2 dB. The data and best-fitting functions for the seven subjects and for mean d' values across subjects are shown in Fig. 6. Values of σ_i are given in Table III, along with measures of goodness of fit.

Figure 7 shows a scatter plot for the seven subjects that compares estimates of internal noise, σ_i, obtained from the conditions with external variability, to corresponding estimates, DL_I, obtained from the psychometric functions. Data for the 70-dB conditions shown here are comparable to those in Fig. 4, except that the estimates in Fig. 7 are based on only four of the eight available data points, to maintain compatibility with the data for the 50- and 90-dB conditions. Although a majority of the subjects showed a reduction in both

TABLE III. Summary of results from experiment 2.

Subject	Pedestal level	Slope	r^2	DL_I	$\sigma_{i\text{-dB}}$	r^2
TC	50	0.54	0.98	1.87	2.14	0.96
	70	1.07	0.93	0.94	1.02	0.99
	90	1.56	0.93	0.64	0.68	0.96
SH	50	0.78	0.91	1.29	1.74	0.95
	70	1.47	0.98	0.68	0.63	0.96
	90	1.55	0.87	0.64	0.68	0.95
KM	50	0.81	0.97	1.23	1.44	0.70
	70	1.52	0.90	0.66	0.55	0.52
	90	1.55	0.92	0.65	0.56	0.95
JB	50	1.05	0.99	0.96	1.09	0.99
	70	1.24	0.96	0.81	0.48	0.45
	90	1.60	0.46	0.62	0.40	0.87
ES	50	0.26	0.93	3.80	3.35	0.89
	70	0.92	0.92	1.08	0.99	0.86
	90	1.24	0.94	0.81	1.14	0.61
AS	50	0.72	0.94	1.40	1.39	0.98
	70	0.85	0.94	1.18	0.92	0.96
	90	0.83	0.99	1.20	1.08	0.97
AH	50	0.99	0.98	1.01	1.15	0.97
	70	1.12	0.94	0.89	1.02	0.90
	90	1.39	0.90	0.72	0.78	0.94
Mean	50	0.73	0.96	1.65	1.76	0.92
	70	1.17	0.94	0.89	0.80	0.81
	90	1.39	0.86	0.75	0.76	0.89

estimates of internal noise as the intensity increased from 70 to 90 dB SPL, the effect of the increase from 50 to 70 dB SPL was greater. Because the data for the 70-dB condition were collected first, we were concerned that performance may have improved as subjects gained experience with the task. The pattern of results, however, does not suggest that learning was a factor. A learning effect would have increased the overlap of the data for the 50-dB SPL conditions with the data obtained earlier at 70 dB SPL and would have increased the separation between the data obtained at 70 and 90 dB SPL because improvement due to training would combine with the effect of higher presentation level. The observed pattern is just the opposite. When the effect of level is combined with the effect of individual differences, the Pearson product-moment correlation between DL_I and σ_i increases to $r = 0.94$.

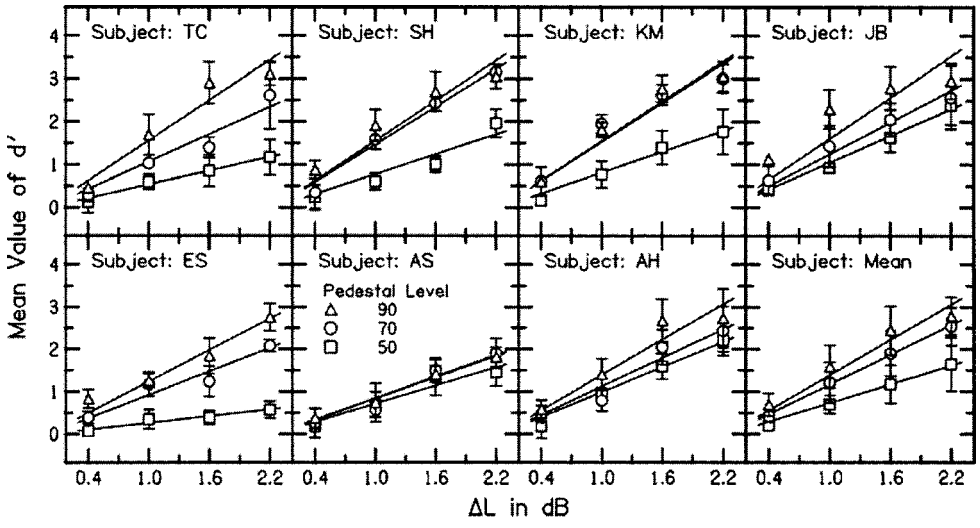


FIG. 5. Psychometric functions in experiment 2 for seven subjects and means across subjects. The data for the 70-dB condition are a subset of the data from Fig. 1.

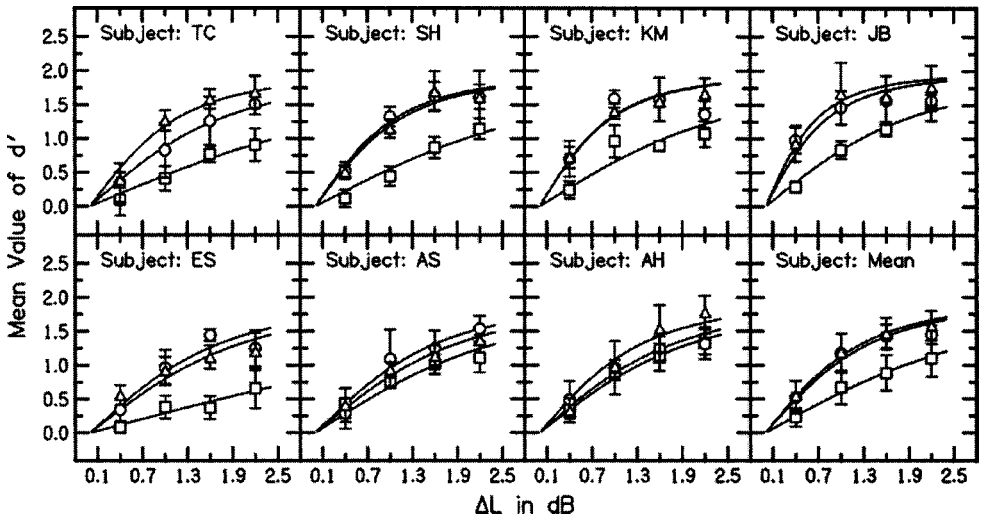


FIG. 6. Values of d' for seven subjects and means across subjects in conditions where external variability, Gaussian in dB, was added to each observation interval. The 90-dB condition is represented by triangles, 70 dB by circles, and 50 dB by squares. Values of internal noise estimates are given in Table III.

IV. DISCUSSION

The purpose of the experiments reported here was to determine whether manipulation of external variability in a sample discrimination framework would provide a valid measure of internal noise that could be used in studies of intensity discrimination and in studies that used sinusoidal maskers. To be valid, the new measure must be reliable and must yield results in general agreement with other accepted measures of internal noise. To be useful, it must also have advantages over other accepted measures. In the sections below, we address reliability, validity, and potential advantages, and then discuss a few additional measurement issues.

A. Reliability

Figures 4 and 7 provide a comparison of σ_i, the measure of internal noise obtained from Eq. (1), with DL_I, the measure based on the threshold for intensity discrimination. The threshold could have been estimated in a number of ways, but we used a measure based on the PF that used the same amount of data, collected in a similar way. If intensity discrimination is assumed to be limited only by internal noise, then DL_I provides a straightforward estimate of that noise that should be highly correlated with σ_i. The degree of agreement of these two measures is constrained by the reliability of each of the measures. We assessed reliability in several different ways. First, the data were divided into two equal parts, and internal noise estimates, σ_i and DL_I, were estimated for both halves. To counterbalance for any effect of learning or fatigue, we used the mean of the d' values for the first and fourth 100-trial blocks as the first half and the mean of the d' values for the second and third blocks as the second half. Equation (1) was then fitted to the data for Gaussian-level variation for each half, and PFs were fitted to the constant-level data for each half. The agreement between the first and second halves for the two internal noise measures for the seven subjects is shown in Fig. 8. The split-halves correlation coefficients for both measures are high ($r=0.96$ for σ_i, 0.91 for DL_I) and there is no significant difference between them ($z=0.296$; $p>0.05$). Note, however, that both regression lines have slopes less than 1.0 and

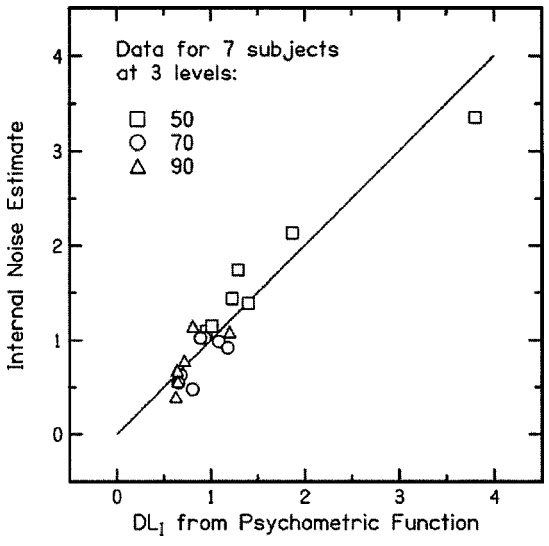


FIG. 7. A comparison of estimates of internal noise obtained in experiment 2 from the psychometric function, in conditions with no external variability, to estimates obtained in conditions where external variability was Gaussian in dB.

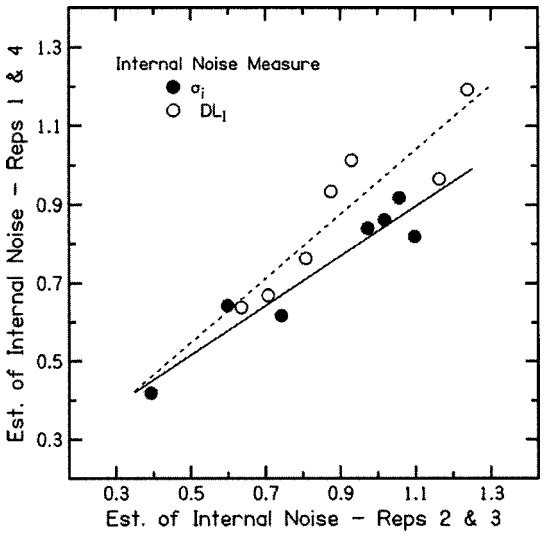


FIG. 8. The reliability of two of the internal noise measures obtained in experiment 1. The values of σ_i were obtained in Gaussian level conditions.

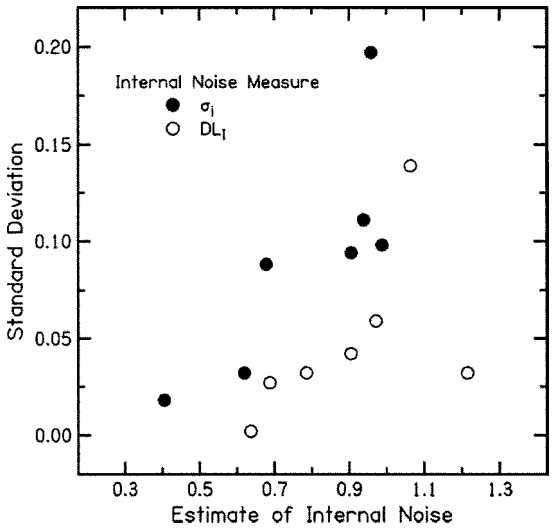


FIG. 9. The reliability of two of the internal noise measures obtained in experiment 1 expressed in terms of the standard deviation of the two estimates obtained for each measure for each subject in the split-halves analysis, plotted against the mean of those two estimates. Both estimates show greater variability in subjects who exhibit greater amounts of internal noise.

that the line for the σ_i measure is shallower than the line for DL_I. This indicates greater disagreement between the two replications in subjects with higher amounts of internal noise, and that this effect is somewhat greater for the σ_i measure. This can be seen in Fig. 9, in which the standard deviation across the two replications is plotted against the mean amount of internal noise for both measures. For the σ_i measure, the standard deviation is 10% of the mean, on average, while for the DL_I measure, it is only 5% of the mean.

The estimates of σ_i and DL_I for the 70-dB condition in Fig. 7 are based on data for four of the eight ΔL values in experiment 1, whereas the estimates in Fig. 4 are based on data for all eight ΔL values. A comparison of the estimates based on four and eight data points provides another measure of reliability and an indication of the amount of data required to obtain stable estimates. In this case, the correlation coefficient was 0.96 for the σ_i estimates and 0.99 for the DL_I estimates. This suggests that little is lost by going from estimates based on eight points to estimates based on four points. To test that hypothesis, split-half reliability coefficients were computed for the two internal noise measures for each of the three overall levels in experiment 2. The average correlation coefficient was 0.94 for the σ_i estimates and 0.98 for the DL_I estimates.

B. Validity

The σ_i measure of internal noise is reliable and is generally in good agreement with the DL_I measure. The best-fitting line has a slope less than 1. The fact that all of the points fall on or below the diagonal indicates that σ_i is always less than or equal to DL_I. This suggests that there is a common component in both measures, either a limitation on intensity discrimination that cannot be modeled as internal noise or a source of noise common to σ_i and DL_I. The fact that σ_i and DL_I decrease in the same way as a function of overall level suggests that the ''near-miss'' to Weber's law (McGill and Goldberg, 1968) can be modeled as a reduction in internal noise at higher levels. This result is consistent with the view that the near-miss results from the increase in neural information available at high levels through spread of excitation (e.g., Florentine and Buus, 1981; Viemeister, 1988).

C. Potential advantages

Although the σ_i measure may be somewhat less reliable than the DL_I measure, it is sufficiently stable to provide a meaningful measure of individual differences in an intensity discrimination task. We chose to explore use of the σ_i measure in the context of intensity discrimination because the DL_I provided a gold standard. Although the σ_i measure provides little new information in the context of intensity discrimination, it could be very useful in studies of masking, where factors other than internal noise play a significant role in determining threshold and the contribution of internal noise is either unknown or is estimated on the basis of a model. It is easy to obtain an estimate of the signal level required for $d'=1$ in masking tasks, but strong assumptions are required to convert that information into an estimate of internal noise.

In the Introduction, we made a distinction between correlation measures of internal noise that rely on the relation between stimulus waveform features and responses, and additivity measures, like the one used here, that are based on the change in performance associated with a change in external variability. Correlation measures of internal noise have a significant problem that is illustrated by recent studies that use a similar approach to derive perceptual weights (e.g., Dai *et al.*, 1996; Lutfi, 1995; Richards and Zhu, 1994). In those studies, low correlation is assumed to reflect a lack of attention to a given feature. If a low correlation is to be interpreted as a high level of internal noise, the stimulus feature on which subjects base their decisions must be well defined. The correlation measure of internal noise developed by Green (1964) and replicated by Spiegel and Green (1981) avoids this problem by using catch trials where no signal is presented and where the masker waveforms are identical to those on other catch trials. If the subject consistently selects the same masker waveform on these catch trials, internal noise is assumed to be low. This approach does not require the experimenter to know the feature of interest, but does require variability in the masker from interval to interval and from trial to trial. It could not be used in a masking task where the same sinusoidal masker was used in all intervals of all trials.

Use of an additivity measure of internal noise also requires an assumption concerning the stimulus feature on which subjects base their decisions. Interval-to-interval variation in the level of a sinusoidal forward masker or of the pedestal in an increment detection procedure, for example, has little effect on performance (Jesteadt and Schairer, 2003). We must either conclude that internal noise is very large in both tasks or that the decision process cannot be modeled as an energy detector. Other features in the data point to the latter interpretation of the results. The additivity measure

thus provides an additional test of the nature of the decision process.

The test of the additivity measure presented by Spiegel and Green (1981) was based on only two levels of external variability. Their noise maskers were either identical from interval to interval or they were independent. A clear advantage of the approach taken here is that Eq. (1) provides a framework for combining many independent d' estimates into a single estimate of σ_i.

The results presented here suggest that the additivity measure of internal noise is not sensitive to the form of the external noise distribution or to the metric in which the internal and external noises are assumed to add. The uniformity of the fits may be due to the use of conditions in which σ_e was set to $\frac{1}{2}\Delta L$. Use of a constant ratio minimized the difference between linear and logarithmic representations of the data. Given a larger set of conditions, it may be possible to obtain more information about the metric of additivity. The fits to the data shown in Fig. 2 and Fig. 3 indicate that performance is lower than predicted at the higher values of σ_e and/or ΔL. This suggests either that internal noise is actually increasing across these conditions or that the metric should be adjusted to correct for this systematic error. A larger set of conditions with different values of σ_e at each value of ΔL would be required to determine whether the systematic error was a function of σ_{e} or ΔL.

D. Additional measurement issues

The internal noise estimates reported here are based on fitting Eq. (1), with one free parameter, to sets of four or eight data points. Not all data points contribute equally to the estimate. Data obtained in conditions for which the external variability is large with respect to internal noise (e.g., Lutfi, 1989) do not provide information about differences in internal noise across subjects. Likewise, little is learned by introducing external variability that is too small to be detected by the subjects. The maximum change in d' occurs at the point where $\sigma_e = \sigma_i$. Use of a range of values of σ_e insures that the unknown point will be bracketed. Although conditions in these studies were defined in terms of ΔL and σ_e was set to $\frac{1}{2}\Delta L$, equally stable estimates of σ_i would probably have been obtained by holding ΔL constant and allowing d'_{ideal} to vary across conditions. We could, for example, have selected a value of ΔL that yielded a d' of approximately 2 when $\sigma_e = 0$, then varied σ_e over the range used here. Had we done that, given $\sigma_i = 0.77$ as estimated from the mean data in Fig. 2, we would have obtained values of d' in the range from 1.98 to 0.66.

The one-parameter fits presented here are based on the assumption that the mean difference between the decision distributions, Δ, is known. We selected stimuli and a task for which Δ could be represented by the physically observable quantity ΔL. In tasks such as detection of a tone in noise, where factors other than internal noise contribute to the limitation in performance, the mean difference between the decision distributions is more difficult to specify. The psychometric functions for such tasks are not plotted on coordinates where performance goes to chance as the difference between the signal and nonsignal intervals goes to zero. These are the cases of real interest, because internal noise in these cases cannot be estimated directly from the slope of the psychometric function or from the signal level required for $d' = 1$. To determine the feasibility of obtaining valid estimates of internal noise under these conditions, we estimated the ratio $\Delta L/\sigma_e$, as well as σ_i, from the data. It was necessary to estimate the ratio, rather than ΔL, because it was the ratio that we held constant at 2 across conditions. The geometric mean estimated ratio was 1.80, with a standard deviation of 10%. The mean of the internal-noise estimates for the seven subjects from the two-parameter fit was 0.65, compared to 0.78 for the one-parameter estimates. The correlation between the two sets of estimates was $r = 0.83$. These results indicate that this procedure could be used to obtain estimates of internal noise for conditions in which the mean difference between the decision distributions was unknown.

V. CONCLUSIONS

(1) A measure of internal noise based on the effect of varying the overall level of the stimuli in an intensity discrimination task from interval to interval is highly reliable.
(2) The new measure is valid, in that it provides estimates of internal noise that are in good agreement with estimates of the DL for intensity discrimination obtained from psychometric functions. Estimates obtained with the new measure decline as a function of overall intensity and are generally somewhat smaller than the DL, suggesting either a limitation on intensity discrimination that cannot be modeled as internal noise or a source of noise common to both the new measure and the DL.
(3) The new measure can be used to obtain estimates of internal noise for tasks with sinusoidal maskers, where there is no variability from interval to interval under normal testing conditions. It is not sensitive to the form of the external variability, but does require the assumption that the subject is attending to overall level, as an energy detector would.
(4) It should be possible to use the new measure in cases where the mean difference between the decision distributions is unknown.

ACKNOWLEDGMENTS

This work was supported by grants R01 DC00136 and T32 DC00013 from the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health and by the Nebraska Biomedical Research Development Fund. This work benefited from discussions with Robert Lutfi, Donna L. Neff, and Stephen Neely and from data collection and analysis software developed by Thomas Creutz and Hongyang Tan, as well as from helpful comments during the review process by Robert Lutfi and an anonymous reviewer.

Berg, B. G., and Robinson, D. E. (**1987**). "Multiple observations and internal noise," J. Acoust. Soc. Am. Suppl. 1 **81**, S33.
Buus, S., and Florentine, M. (**1991**). "Psychometric functions for level discrimination," J. Acoust. Soc. Am. **90**, 1371–1380.

Carney, L. H. (**1994**). ''Spatiotemporal encoding of sound level: Models for normal encoding and recruitment of loudness,'' Hear. Res. **76**, 31–44.
Colburn, H. S. (**1973**). ''Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination,'' J. Acoust. Soc. Am. **54**, 1458–1470.
Colburn, H. S. (**1977**). ''Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise,'' J. Acoust. Soc. Am. **61**, 525–533.
Dai, H. (**1995**). ''On measuring psychometric functions: a comparison of the constant-stimulus and adaptive up-down methods,'' J. Acoust. Soc. Am. **98**, 3135–3139.
Dai, H., Nguyen, Q., and Green, D. M. (**1996**). ''Decision rules of listeners in spectral-shape discrimination with or without signal-frequency uncertainty,'' J. Acoust. Soc. Am. **99**, 2298–2306.
deBoer, E. (**1966**). ''Intensity discrimination of fluctuating signals,'' J. Acoust. Soc. Am. **40**, 552–560.
Durlach, N. I., and Braida, L. D. (**1969**). ''Intensity perception. I. Preliminary theory of intensity resolution,'' J. Acoust. Soc. Am. **46**, 372–383.
Florentine, M., and Buus, S. (**1981**). ''An excitation-pattern model for intensity discrimination,'' J. Acoust. Soc. Am. **70**, 1646–1654.
Green, D. M. (**1960a**). ''Auditory detection of a noise signal,'' J. Acoust. Soc. Am. **32**, 121–131.
Green, D. M. (**1960b**). ''Psychoacoustics and detection theory,'' J. Acoust. Soc. Am. **32**, 1189–1203.
Green, D. M. (**1964**). ''Consistency of auditory detection judgments,'' Psychol. Rev. **71**, 392–407.
Javel, E., and Viemeister, N. F. (**2000**). ''Stochastic properties of cat auditory nerve responses to electric and acoustic stimuli and application to intensity discrimination,'' J. Acoust. Soc. Am. **107**, 908–921.
Jesteadt, W., and Schairer, K. S. (**2003**). ''Effect of masker variability on forward masking and increment detection,'' in *Abstracts of the Twenty-Sixth Midwinter Research Meeting of the Association for Research in Otolaryngology*, edited by P. A. Santi (Association for Research in Otolaryngology, Mt. Royal, NJ), pp. 221–222.
Lutfi, R. A. (**1989**). ''Informational processing of complex sound. Intensity discrimination,'' J. Acoust. Soc. Am. **86**, 934–944.
Lutfi, R. A. (**1995**). ''Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks,'' J. Acoust. Soc. Am. **97**, 1333–1334.
McGill, W. J., and Goldberg, J. P. (**1968**). ''A study of the near-miss involving Weber's law and pure-tone intensity discrimination,'' Percept. Psychophys. **4**, 105–109.
Moore, B. C., Peters, R. W., and Glasberg, B. R. (**1999**). ''Effects of frequency and duration on psychometric functions for detection of increments and decrements in sinusoids in noise,'' J. Acoust. Soc. Am. **106**, 3539–3552.
Oxenham, A. J., and Buus, S. (**2000**). ''Level discrimination of sinusoids as a function of duration and level for fixed-level, roving-level, and across-frequency conditions,'' J. Acoust. Soc. Am. **107**, 1605–1614.
Patterson, R. D. (**1976**). ''Auditory filter shapes derived with noise stimuli,'' J. Acoust. Soc. Am. **59**, 640–654.
Plack, C. J., and Oxenham, A. J. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**, 1598–1608.
Richards, V. M., and Zhu, S. (**1994**). ''Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients,'' J. Acoust. Soc. Am. **95**, 423–434.
Siebert, W. M. (**1968**). ''Stimulus transformation in the peripheral auditory system,'' in *Recognizing Patterns*, edited by P. Kolers and M. Eden (MIT Press, Cambridge, MA), pp. 104–133.
Siegel, R. A. (**1979**). ''Internal and external noise in auditory detection,'' Masters thesis, Massachusetts Institute of Technology.
Spiegel, M. F., and Green, D. M. (**1981**). ''Two procedures for estimating internal noise,'' J. Acoust. Soc. Am. **70**, 69–73.
Swets, J. A., Shipley, E. F., McKey, M. J., and Green, D. M. (**1959**). ''Multiple observations of signals in noise,'' J. Acoust. Soc. Am. **31**, 514–521.
Tanner, Jr., W. P., and Birdsall, T. G. (**1958**). ''Definitions of d' and η as psychophysical measures,'' J. Acoust. Soc. Am. **30**, 922–928.
Viemeister, N. F. (**1988**). ''Intensity coding and the dynamic range problem,'' Hear. Res. **34**, 267–274.
Watson, C. S. (**1962**). ''Signal detection and certain physical characteristics of the stimulus during the observation interval,'' Doctoral dissertation, University of Indiana.
Wier, C. C., Green, D. M., Hafter, E. R., and Burkhardt, S. (**1977**). ''Detection of a tone burst in continuous- and gated-noise maskers; Defects of signal frequency, duration, masker level,'' J. Acoust. Soc. Am. **61**, 1298–1300.
Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (**2001**). ''A phenomenological model for the responses of auditory-nerve fibers. I. Nonlinear tuning with compression and suppression,'' J. Acoust. Soc. Am. **109**, 648–670.

Phase effects in masking: Within- versus across-channel processes

José I. Alcántara,[a] Brian C. J. Moore, Brian R. Glasberg, and Alex J. K. Wilkinson
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

Urszula Jorasz
Institute of Acoustics, Adam Mickiewicz University, 85 Umultowska Street, 61-614 Poznan, Poland

(Received 15 August 2002; revised 23 May 2003; accepted 21 July 2003)

The effects of bandwidth and component phase on masking were investigated using 200-ms narrowband (1-ERB_N) and broadband (5-ERB_N) cosine-phase (CP) and random-phase (RP) harmonic complex maskers, centered at 1 or 6 kHz. A continuous notched-noise was used to restrict off-frequency listening. The masker fundamental frequency ($F0$) was 25 Hz. In experiment 1, thresholds were measured for sinusoidal signals at 1 and 6 kHz, gated with the maskers. Thresholds were lower in the CP than in the RP masker, for both bandwidths, but the effect was markedly greater for the wider bandwidth. For the CP maskers, thresholds were markedly lower for the 5-ERB_N than for the 1-ERB_N bandwidth; for the RP maskers, there was a small effect in the opposite direction. Experiment 2 used 1- and 6-kHz CP maskers. The masker components in the ERB_N around the signal frequency were presented to one ear, and the remaining components were presented contralaterally. Thresholds were much higher than when all components were presented to the same ear, and were higher than for the 1-ERB_N masker alone, suggesting that the low thresholds for broadband monaural presentation do not depend on ''high level'' across-channel comparisons. Simultaneous masked thresholds could be predicted well using a model based on a simulated auditory filter, a level-dependent compressive nonlinearity, and a sliding temporal integrator; it was not necessary to assume the involvement of across-channel processes or of selective listening in the masker dips. [DOI: 10.1121/1.1608959]

PACS numbers: 43.66.Nm, 43.66.Dc, 43.66.Mk [MRL]

I. INTRODUCTION

The effect of phase on masking has been well documented using harmonic tone complexes with identical power spectra, but with individual components added in different starting phases (Mehrgardt and Schroeder, 1983; Smith *et al.*, 1986; Kohlrausch and Sander, 1995; Carlyon and Datta, 1997; Summers and Leek, 1998; Lentz and Leek, 2001; Oxenham and Dau, 2001a; Gockel *et al.*, 2002). The majority of these studies have used ''Schroeder-phase'' complexes (Schroeder, 1970). A complex with a positive phase curvature is referred to as ''Schroeder positive'' (m_+). A complex with negative phase curvature is known as ''Schroeder negative'' (m_-). The m_+ complex produces masked thresholds up to 20 dB lower than for the m_- complex (Smith *et al.*, 1986; Kohlrausch and Sander, 1995; Carlyon and Datta, 1997; Summers and Leek, 1998; Oxenham and Dau, 2001a).

To account for the phase effects, it has been suggested that the m_+ complex produces a waveform on the basilar membrane with a high peak factor while the m_- complex produces a less peaky waveform. The detection of a tone in a masker with high peak factor might be easier than in a masker with a low peak factor because the listener selectively attends to the tone during the low-amplitude portions of the masker waveform, a process called ''dip listening'' (Mehrgardt and Schroeder, 1983; Mott and Feth, 1986; Alcántara and Moore, 1995; Kohlrausch and Sander, 1995). The phase curvature of the human auditory filter, expressed in rad/Hz^2, appears to decrease in absolute magnitude as the characteristic frequency increases (Lentz and Leek, 2001; Oxenham and Dau, 2001b). Thus, the phase response of a given place on the basilar membrane shows a negative curvature (Ruggero *et al.*, 1997), which counteracts the positive phase curvature of the m_+ complex, so that all components, at the output of the filter, have a similar starting phase, leading to a high peak factor on the basilar membrane (Recio and Rhode, 2000).

The extraction of information from the dips might be affected in two ways by the fast-acting compression that has been observed on the basilar membrane (Rhode and Robles, 1974; Robles *et al.*, 1986; Recio *et al.*, 1998). First, the high-amplitude peaks in the m_+ complex would be subjected to less gain than the dips, reducing any forward masking effect on the following dips (Gockel *et al.*, 2002). Second, the compression would result in a relative amplification of the low-level portions of the waveform, enhancing the salience of signal information appearing in the dips (Smith *et al.*, 1986; Horst *et al.*, 1990; Alcántara *et al.*, 1996).

It remains unclear whether the observed phase effects in masking depend entirely on within-channel processes, or whether they are influenced by across-channel processes. If the dips in the waveform evoked by a masker on the basilar membrane are almost synchronous in different frequency re-

[a]Electronic mail: jia10@cus.cam.ac.uk

 0001-4966/2003/114(4)/2158/9/$19.00

gions, then subjects might use information from frequency regions remote from the signal frequency to indicate the optimum times to listen for the signal, or to enhance detection of the signal via comparison of waveform envelopes across channels. Mechanisms of this type have been proposed to explain some forms of comodulation masking release (CMR) (Hall *et al.*, 1984; Buus, 1985; Moore, 1988, 1992; Hall *et al.*, 1995; Buus *et al.*, 1996).

Carlyon and Datta (1997) found that thresholds for detecting sinusoidal signals in m_+ complexes were reduced by the addition of masker components that fell well outside the passband of the auditory filter centered at the signal frequency (Glasberg and Moore, 1990). This might be taken as indicating a role for across-channel processes. However, based on analysis using a basilar-membrane model (Giguère and Woodland, 1994), Carlyon and Datta argued that the observed effects could be explained by a within-channel process; the addition of masker components outside the passband (defined in terms of the −3-dB points) of the auditory filter can result in a higher peak factor of the basilar-membrane filtered waveform at the place where the signal is detected. It should be noted, however, that the model they used had filters with greater bandwidths than are typically assumed for human listeners (Glasberg and Moore, 1990). Oxenham and Dau (2001a) pointed out that the results might not have been predicted so well by a model with more realistic frequency selectivity.

To clarify the role of within and across-channel processes in producing phase effects in masking, Oxenham and Dau (2001a) carried out a comprehensive study using m_+ and m_- complexes. They measured thresholds for 1- and 4-kHz tones presented simultaneously with m_+ and m_- complexes of various bandwidths. For their ''narrow-band'' stimuli, a small effect of masker phase was found for a center frequency of 4 kHz, but not for a center frequency of 1 kHz. Large effects of masker phase were found only when many components were used, covering a frequency range greater than the bandwidth of the auditory filter. The results from simulations using several different basilar-membrane models suggested that the observed phase effects could be accounted for with a single-channel model employing gammachirp filters (Irino and Patterson, 2001), which had a magnitude response consistent with accepted measures of frequency selectivity, but with a phase response altered so that it was exactly the same as that of the m_- stimulus used in the study.

Two aspects of the study of Oxenham and Dau (2001a) make the interpretation of their results unclear: (1) They used maskers with a sufficiently large number of components to produce highly modulated waveforms. This, in combination with their choice of $F0$ of 100 Hz, meant that even their ''narrow-band'' maskers had bandwidths greater than the auditory filter bandwidth at the signal frequency; the bandwidths were about 3 ERB_Ns at 1 kHz and 4 ERB_Ns at 4 kHz, where ERB_N stands for the average value of the equivalent rectangular bandwidth (ERB) of the auditory filter for normally hearing listeners at moderate sound levels (Moore, 2003); therefore they did not have a ''true'' within-channel condition. (2) When the masker bandwidth is relatively narrow, subjects may listen ''off frequency'' to detect the signal, especially when the signal is detected as a change in overall level; the change in excitation level on the high-frequency side of the excitation pattern is greater than the change in the center of the pattern (Zwicker, 1956). On the high-frequency side of the excitation pattern, both m_+ and m_- maskers are probably processed linearly (Ruggero, 1992), and the phase response of the basilar membrane has only a small curvature over the frequency range of a narrow-band masker. Both these factors might lead to similar masking effectiveness for the two masker phases. Such off-frequency listening is unlikely with broadband maskers, as the masker components remote from the signal frequency effectively prevent off-frequency listening. Thus, in the study of Oxenham and Dau, the signal might have been detected through different channels depending on the masker bandwidth.

In this study, we sought to clarify the relative influence of within- and across-channel mechanisms on the detection of signals in harmonic maskers with different component phases, using methods similar to those of Oxenham and Dau (2001a), but with modifications to the stimuli to address the problems discussed above. To produce maskers with bandwidths less than one ERB_N (Glasberg and Moore, 1990), but with enough components to produce highly modulated waveforms, we used an $F0$ of 25 Hz, a factor of 4 lower than used by Oxenham and Dau. To restrict off-frequency listening, a notched-noise centered at the signal frequency was presented together with the main masker.

Rather than using Schroeder-phase stimuli, we used cosine-phase (CP) and random-phase (RP) stimuli. Previous research suggests that CP produces waveforms with a high peak factor on the basilar membrane, while random phase RP produces lower peak factors (Duifhuis, 1970; Mehrgardt and Schroeder, 1983; Alcántara and Moore, 1995; Kohlrausch and Sander, 1995; Alcántara *et al.*, 1996; Gockel *et al.*, 2002).

II. GENERAL EXPERIMENTAL METHOD

A. Subjects

Four subjects took part in two experiments. Three were authors JIA, BRG, and AJKW and one was an undergraduate student, who was paid for his services. JIA and BRG were highly experienced in masking experiments. The remaining two subjects were not, and therefore they were each given practice until their performance was stable. The age of the subjects ranged from 21 to 57 years. All had audiometric thresholds less than 20 dB HL and air–bone gaps less than 10 dB HL, at frequencies between 250 and 8000 Hz.

B. Procedure

An adaptive three-interval forced-choice task with a ''3-down 1-up'' method was used to estimate the 79.4% correct point on the psychometric function (Levitt, 1971). The three masker bursts in a trial, accompanied by lights on a response box, were separated by 500 ms. The signal was presented during one of the bursts, selected at random. Subjects indicated the interval thought to contain the signal using a three-button response box; the buttons were positioned below the

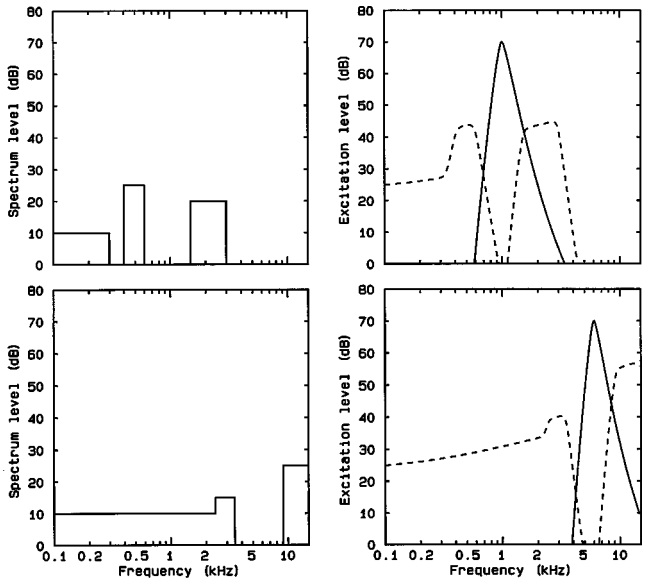


FIG. 1. Left-hand panels show the spectra of the noises used to restrict off-frequency listening and to mask combination tones, for signal frequencies of 1 kHz (top) and 6 kHz (bottom). Right panels show excitation patterns of the noises (dashed lines), calculated as described by Glasberg and Moore (1990), together with the excitation patterns evoked by 70 dB SPL signals (solid lines). Note that each noise produced negligible excitation at the signal frequency.

three LED lights. Feedback was provided after each trial by lighting the LED for the correct interval. The initial step size was 5 dB, and this was reduced to 2 dB after four reversals. Twelve reversals were obtained and threshold was defined as the mean of the signal levels at the last eight reversals. Three threshold estimates were obtained for each condition and each listener. Estimates for which the standard deviation of the levels at the last eight reversal points was greater than 5 dB were discarded and the test condition repeated.

C. Equipment

The complex tone maskers and sinusoidal signals were digitally generated using two separate digital-to-analog converters of a Tucker-Davies Technologies (TDT) System II, controlled by a PC. The sampling rate was 51.2 kHz. The masker and tone signals were individually attenuated by TDT PA4 programmable attenuators before being mixed together (TDT SM3) and passed to a headphone buffer (TDT HB6) and to the Sennheiser HD580 earphone. Levels in this paper are nominal levels at the input to the earphone, which has a ''diffuse field'' response. Levels at the eardrum are boosted relative to nominal levels for frequencies around 3 kHz. However, for the center frequencies used here, eardrum and nominal SPLs were similar.

A noise masker with a spectral notch at the signal frequency (1 or 6 kHz) was presented continuously in all experiments. The noise was designed to prevent off-frequency listening and the detection of combination tones, while not appreciably affecting signal-evoked excitation at characteristic frequencies (CFs) close to the signal frequency. It was presented to the same ear as the signal. The spectra of the noises used are shown in the left panels of Fig. 1. Details of the spectra are given in Table I. The right panels of Fig. 1 show excitation patterns of the noises (dashed lines), calculated as described by Glasberg and Moore (1990), together with the excitation patterns evoked by a 70 dB SPL signal (solid lines); measured thresholds were mostly below 70 dB SPL. The noise spectra were chosen so that the noise would produce (1) negligible excitation at the signal frequency, f_s; (2) relatively high excitation for frequencies $f_s \pm 0.3f_s$, so as to restrict off-frequency listening; and (3) a moderate amount of excitation for frequencies below $0.3f_s$, so as to mask combination tones.

TABLE I. Characteristics of the background noise maskers, each of which was composed of three bands.

Signal frequency	Band	Spectrum level	Frequency range (kHz)
1 kHz	1	10	0.02–0.3
	2	25	0.4–0.6
	3	20	1.5–3.0
6 kHz	1	10	0.02–2.3
	2	15	2.3–3.5
	3	25	9.0–15.0

The noise was specified in the frequency domain, with frequency components spaced by 1 Hz, and was converted to the digital domain using an inverse fast Fourier transform. This gave a 1-s sample, which was repeated seamlessly to give a continuous noise. The noise was played out from a Silicon Graphics Indy (16-bit resolution, 44.1-kHz sampling rate) and recorded onto CDR. During the experiment, the noise was played back continuously and added to the masker and signal via the TDT SM3 mixer. Although the noise was intended mainly to restrict off-frequency listening when a narrow-band masker was used, it was also presented continuously when a broadband masker was used. This was done to control for any side-effects of the noise and to ensure that off-frequency listening was equally restricted in all conditions. Pilot experiments showed that thresholds for signals presented in the broadband maskers (see below for details) were not affected by the presence or absence of the background noise.

III. EXPERIMENT 1. EFFECTS OF MASKER BANDWIDTH AND NUMBER OF COMPONENTS ON SIMULTANEOUS MASKING

A. Rationale

The aim of this experiment was to compare the magnitude of the phase effect in masking (RP versus CP masker) under conditions where all masker components were contained within 1 ERB_N and where they were spread over several ERB_Ns, and where off-frequency listening was restricted using a notched-noise. We used two center frequencies, 1 and 6 kHz. This allowed us to manipulate the number of components falling within one ERB_N, while keeping F0 constant.

B. Stimuli and conditions

The harmonic complex tone maskers had equal-amplitude components added in either CP or frozen RP; in the latter case, the same ''random'' starting phase was used throughout (we call this ''fixed frozen''). Possible conse-

quences of this are discussed later. All maskers had an overall duration of 200 ms (160-ms steady state, with 20-ms raised-cosine ramps) and had an F0 of 25 Hz. The maskers were centered at either 1 kHz, with a level per component of 63 dB SPL or at 6 kHz, with a level per component of 58 dB SPL. The total power in the ERB_N (Glasberg and Moore, 1990) around the center frequency was approximately the same for the 1- and 6-kHz maskers; the level per ERB_N was about 70 dB at 1 kHz and 72 dB at 6 kHz. The sinusoidal signal had a frequency equal to the center frequency of the masker, and it had the same duration as the masker and was gated synchronously with it. The signal had a starting phase of 90° relative to the central component of the CP or RP masker.

The maskers were 1-kHz narrow-band CP and RP (harmonic numbers 38–42; component frequencies 950–1050 Hz); 1-kHz broadband CP and RP (harmonic numbers 27–53; component frequencies 675–1325 Hz); 6-kHz narrow-band CP and RP (harmonic numbers 227–253; component frequencies 5675–6325 Hz); and 6-kHz broadband CP and RP (harmonic numbers 173–307; component frequencies 4325–7675 Hz). The narrow-band maskers spanned frequency ranges of 100 and 650 Hz at 1 and 6 kHz, respectively. These ranges are slightly less than the corresponding ERB_Ns of the auditory filters, which are approximately 130 and 670 Hz, respectively (Glasberg and Moore, 1990). Thus, on an ERB_N scale, the 6-kHz narrow-band masker had roughly the same bandwidth as the 1-kHz narrow-band masker, but contained a little more than five times as many components. In fact, the 6-kHz narrow-band masker had the same number of components as the 1-kHz broadband masker. The broadband stimuli spanned frequency ranges corresponding to five times the ERB_N at the signal frequency.

C. Results

The pattern of results was consistent across subjects, and mean results are shown by the four left-most open bars (CP) and shaded bars (RP) in Fig. 2. Thresholds are expressed as the signal level relative to the masker level in a 1-ERB_N-wide band around the signal frequency. Error bars denote ±1 standard error (SE) across subjects. The two right-most bars show the results of a later experiment. Asterisks show predictions of a model that is described later. Consider first the difference in thresholds for the CP and RP maskers. The former always led to lower thresholds than the latter, even for the narrow-band maskers. The difference between the two phase conditions was greatest for the broadband maskers (≈33 dB on average), and less for the narrow-band maskers (≈6 dB on average). The effect for the narrow-band masker at 1 kHz contrasts with the results of Oxenham and Dau (2001a). They found very little difference between m_+ and m_- maskers in their ''narrow-band'' condition at 1 kHz, although they did find an effect for a center frequency of 4 kHz. Their failure to find a phase effect at 1 kHz may have been due to the lack of any noise to prevent off-frequency listening (on the high-frequency side of the excitation pattern) in their experiment.

Consider now the effect of bandwidth. For the CP masker, thresholds dropped dramatically when the masker bandwidth was increased from 1 to 5 ERB_Ns. The decrease was similar for the two center frequencies. This effect is similar to what has been observed previously for m_+ maskers (Carlyon and Datta, 1997; Oxenham and Dau, 2001a). The components outside the ERB_N centered at the signal frequency might have produced a release from masking either because they led to a waveform with a higher peak factor at the output of the auditory filter centered at the signal frequency (more components lead to a more peaky waveform), or because of an across-channel effect of some kind. This point is discussed in more detail later.

For the RP masker, thresholds increased somewhat when the bandwidth was increased from 1 to 5 ERB_Ns. This effect is somewhat surprising, as one might expect the RP masker to have effects similar to a random noise masker, and for the latter thresholds hardly change when the masker bandwidth is increased beyond one ERB_N keeping the spectral level constant (Fletcher, 1940). Part of the effect of bandwidth can be attributed to the fact that the narrow-band masker actually had a bandwidth slightly below 1 ERB_N, so increasing the bandwidth would have produced a small increase in the output of the auditory filter centered at the signal frequency. The effect might also have been related to our use of a fixed (frozen) random phase (Hanna and Robinson, 1985). To check on this, three subjects were rerun using the 1-kHz signal. In one condition (variable frozen), the masker component phase was chosen randomly and independently for each run, but was fixed within a run. In another condition (random), the masker component phase was chosen ran-

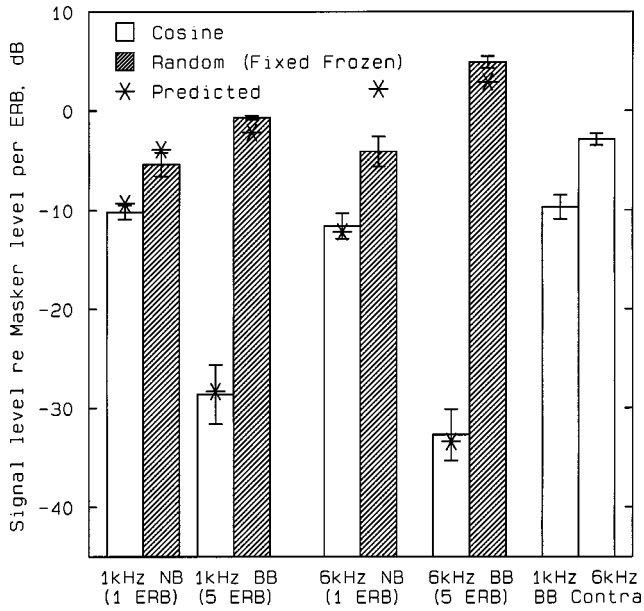


FIG. 2. Mean results of four subjects. The four left-most pairs of bars show results of experiment 1. The masker center frequency and bandwidth are shown at the bottom. Open and shaded bars indicate that the masker components were added in cosine phase (CP) or (fixed frozen) random phase (RP), respectively. The signal level at threshold is expressed relative to the masker level per ERB_N. The two right-most bars show the results of experiment 2, when the flanking components of the 5-ERB_N-wide CP masker were presented to the ear contralateral to that receiving the signal and on-frequency band. In all cases, error bars indicate±one standard error (SE) across subjects. Asterisks show thresholds predicted using the model described in the text.

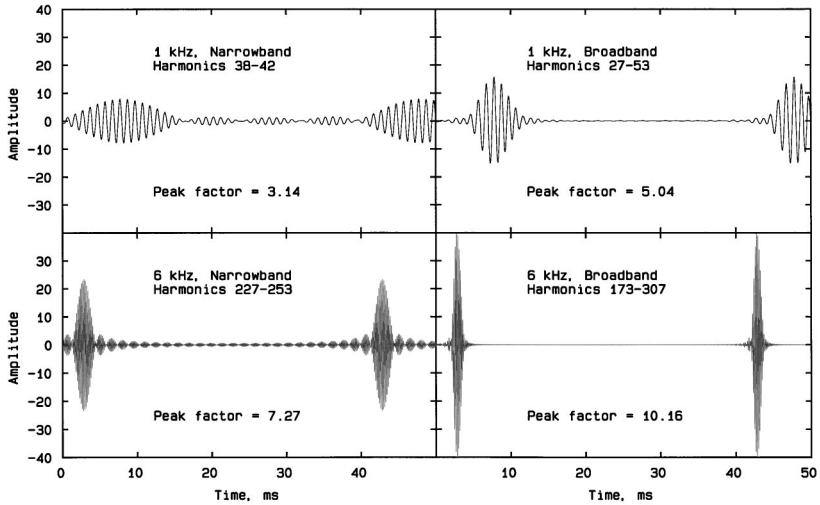


FIG. 3. Simulation of the outputs of auditory filters centered at 1 kHz (top) or 6 kHz (bottom) in response to narrow-band (left) or broadband (right) CP maskers. The simulation did not take basilar-membrane compression into account.

domly for each and every masker presentation. The results showed that the effect of increasing the noise bandwidth was similar for the fixed-frozen and variable-frozen conditions, and was slightly smaller for the random condition.

To assess the statistical significance of the effects described for the main experiment, a within-subjects three-way analysis of variance (ANOVA) was conducted, with factors center frequency, bandwidth (1 ERB_N versus 5 ERB_N) and phase. In what follows, only effects significant at $p<0.05$ will be discussed. The main effects of phase and bandwidth were highly significant [$F(1,3)=1150.0$, $p<0.001$ and $F(1,3)=93.6$, $p<0.002$, respectively], as was the interaction of bandwidth and phase [$F(1,3)=124.33$, $p<0.002$]. *Post-hoc* tests, based on the least-significant differences test, showed that thresholds were significantly higher in the RP than in the CP phase masker for both the 1-ERB_N wide maskers ($p<0.01$) and the 5-ERB_N wide maskers ($p<0.001$). The decrease in threshold with increasing bandwidth for the CP masker was highly significant ($p<0.001$). The increase in threshold with increasing bandwidth for the RP masker was also significant ($p<0.01$).

D. Discussion and model analysis

As mentioned earlier, the components outside the ERB_N centered at the signal frequency might have produced a release from masking because they led to a waveform with a higher peak factor at the output of the auditory filter centered at the signal frequency. This effect is illustrated in Fig. 3, which shows calculated waveforms at the outputs of auditory filters centered at 1 or 6 kHz in response to the CP maskers (without any signal). The calculations assumed that the amplitude characteristic of each side of the auditory filter can be described by a rounded-exponential function (Patterson *et al.*, 1982):

$$W(g)=[(1+pg)\exp(-pg)]^{0.5},$$

where g is the deviation from the center frequency of the filter divided by the center frequency, and p is a parameter defining the sharpness of the filter. The exponent of 0.5 converts the response from the intensity domain to the amplitude domain. The values of p for the lower and upper halves of the filter, respectively, were taken as 23.0 and 30.2 at 1 kHz, and 27.2 and 35.7 at 6 kHz. These value correspond to those estimated by Glasberg and Moore (1990) for the levels used in our experiment. The phase curvature of the response of the auditory filter was taken as -5×10^{-5} rad/Hz^2 at 1 kHz and -3×10^{-6} rad/Hz^2 at 6 kHz (Lentz and Leek, 2001; Oxenham and Dau, 2001b). The phase response was assumed to approach a lead of 90° at low frequencies (Robles and Ruggero, 2001). The waveforms at the outputs of the simulated filters were calculated by summing sinusoidal components with the appropriate amplitudes and phases. Note that the calculated waveforms do not include the effects of fast-acting compression in the cochlea, which would have the effect of amplifying low-level portions of the waveforms relative to higher-level portions.

For the narrow-band masker centered at 1 kHz (top-left panel) the peak factor, defined as the ratio of the maximum amplitude to the root-mean-square (rms) amplitude, was 3.14. A distinct oscillation is apparent in the valley between major peaks. The magnitude of the low-amplitude portion was estimated by calculating its rms value over a 10-ms interval centered between the major waveform peaks. We chose to average over a 10-ms interval, as this corresponds roughly to the equivalent rectangular duration of the auditory ''temporal window'' (Moore *et al.*, 1988; Plack and Moore, 1990; Oxenham and Moore, 1994). The resulting value, expressed relative to the rms value of the whole waveform, was -11.9 dB. Hereafter, this relative value is referred to as L_{valley}. For the broadband masker centered at 1 kHz (top-right panel), the peak factor increased to 5.04, and L_{valley} decreased to -37.0 dB. This illustrates how adding extra components can produce a large decrease in the level in the valleys. For the narrow-band masker centered at 6 kHz (bottom-left panel), the peak factor was 7.27 and the value of L_{valley} was -20.6 dB. The peak factor is higher and the value of L_{valley} is lower than for the 1-kHz narrow-band masker. This illustrates the effect of having a larger number of components within the passband of the auditory filter. Finally, for the broadband masker centered at 6 kHz, the peak factor was 10.16 and the value of L_{valley} was -44.0 dB. Again, adding

components outside the passband of the filter had a large effect on L_{valley}.

It is of interest to compare the results for the 1-kHz broadband masker and the 6-kHz narrow-band masker. These maskers both contained 27 components. However, for the 1-kHz masker the components were spread over a range greater than 1 ERB_N, while for the 6-kHz masker they were all contained within 1 ERB_N. The peak factor of the waveform at the output of the simulated auditory filter at the masker center frequency was somewhat greater for the latter than for the former (7.27 versus 5.04). However, the obtained thresholds were much higher for the latter than for the former. This suggests that the peak factor is not critical in determining thresholds. In contrast, the value of L_{valley} was markedly lower for the 1-kHz broadband masker (−37.0 dB) than for the 6-kHz narrow-band masker (−20.6 dB), and this pattern is consistent with the obtained thresholds.

To assess more rigorously whether the thresholds obtained for the CP maskers in simultaneous masking could be explained in terms of a within-channel process, we used a model with four stages:

(1) A simulated auditory filter was used to calculate a filtered waveform as described above.

(2) The Hilbert envelope of the filter output was determined (Bracewell, 1986).

(3) The envelope was subjected to a level-dependent compressive nonlinearity, defined in terms of the gain, G (output level minus input level), which was

$$G = 55/[1 + \exp(0.31x - 0.00035x^2 + 0.000015x^3)],$$

where x is equal to the output level minus 82.2. This nonlinearity was based on basilar-membrane input–output functions determined in animals. Its form is illustrated in Fig. 1.18 of Moore (2003).

(4) The compressed envelope was passed through a sliding temporal integrator with characteristics as described by Oxenham and Moore (1994). The equivalent rectangular duration of the integrator was 7 ms.

The model output was determined for each CP masker alone (1 and 6 kHz, broadband and narrow band), and for that masker plus a sinusoidal signal at the appropriate center frequency. The signal had a phase of 90° relative to the masker component at the same frequency, as was also the case in our experiment. The ratio of outputs for these two cases, $R(t)$, was calculated. We compared results for two cases. One case was based on the concept of dip-listening, assuming that listeners are able selectively to extract information about the signal during the dips in the masker waveform. For this case, the 10-ms interval during the masker period for which $R(t)$ was highest was determined; this interval always occurred between the main peaks in the masker waveform. The mean value of $R(t)$ over the 10-ms interval, $R(t)_{10}$, was used as a decision variable. The second case was based on the assumption that listeners do not listen selectively during the masker dips, but combine information over the whole masker period. For this case, $R(t)$ was averaged over the whole of one 40-ms masker period (centered in a longer segment, so as to exclude effects of the onset and offset ramps). The resulting quantity, $R(t)_{40}$, was used as a decision variable.

To predict thresholds, a starting value of $R(t)_{10}$ or $R(t)_{40}$ was chosen, and the input signal level was adjusted for each condition (two signal frequencies and two masker bandwidths) so as to achieve that value of $R(t)_{10}$ or $R(t)_{40}$. This gave a set of predicted thresholds. An iterative procedure was used to find the value of $R(t)_{10}$ or $R(t)_{40}$ for which the mean of the predicted thresholds was equal to the mean of the obtained thresholds. Both decision variables led to reasonably accurate predictions, but errors were smaller for $R(t)_{40}$. The mean-squared deviation of the predicted values from the obtained values was 4.4 dB^2 for $R(t)_{10}$ and 0.49 dB^2 for $R(t)_{40}$. The predicted thresholds based on $R(t)_{40}$ are shown as asterisks in Fig. 2.

Predictions based on $R(t)_{40}$ are also shown for the RP maskers. To generate these predictions, ten different samples of each masker (with different selections of random phases) were used and predicted thresholds were averaged across the ten. The predicted thresholds for the RP maskers are close to the obtained thresholds, except for the narrow-band masker centered at 6 kHz, where the predicted threshold is about 5 dB higher than the obtained threshold. We conclude that a single-channel model is able to account well for the thresholds found for the CP maskers. It does not appear necessary to assume selective listening during the dips in the masker envelope. The same model accounts reasonably well for the thresholds obtained with the RP maskers.

IV. EXPERIMENT 2. EFFECT OF CONTRALATERAL PRESENTATION OF THE FLANKING COMPONENTS

A. Rationale

Although the analysis presented above suggests that the results of experiment 1 could be predicted reasonably well based on a within-channel process, it is desirable to conduct an independent test of a role for across-channel processes. To assess whether the lower thresholds in the broadband maskers occurred partly because of across-channel comparisons of the type that are thought to be involved in some forms of comodulation masking release (CMR), the signal and the masker components in the ERB_N around the signal frequency (the on-frequency band) were presented to one ear, and the remaining (flanking) components were presented to the other ear. It is known that CMR can occur when the on-frequency masking band is presented to one ear, and the flanking bands are presented to the other ear (Schooneveldt and Moore, 1987; Hall *et al.*, 1988; Moore and Shailer, 1991; Hall *et al.*, 1995). If the effects observed in experiment 1 involved across-channel comparisons comparable to those occurring for CMR, then one would expect that contralateral presentation of the flanking components would result in lower thresholds than obtained with the on-frequency band alone. A similar experiment, with a similar rationale, was conducted by Oxenham and Dau (2001a). Using the m_+ masker, they found that the contralateral components did not produce a release from masking. However, this lack of effect may have occurred because they did not take precautions to

prevent subjects from using information from the high-frequency side of the excitation pattern for the narrowband maskers.

B. Stimuli and test conditions

The maskers were the same as the 5-ERB_N wide CP maskers used in experiment 1, except that the components were split between the two ears. For the 1-kHz signal frequency, the on-frequency band contained harmonics 38–42, and the flanking contralateral band contained harmonics 27–37 and 43–53. For the 6-kHz signal frequency, the on-frequency band contained harmonics 227–253, and the flanking contralateral band contained harmonics 173–226 and 254–307.

All aspects of signal generation were as described in experiment 1, except that two rather than one transducers of the Sennheiser HD 580 headset were used. The notched noise was played continuously to the ear containing the signal and the on-frequency band.

C. Results

Once again, results were very similar across the four subjects, and mean thresholds are shown in Fig. 2, by the unfilled right-hand bars. For the 1-kHz signal, thresholds for the narrowband masker with contralateral flanking components were almost the same as for the narrowband masker alone (left-most open column). Thus, there was no release from masking comparable to that observed in CMR. For the 6-kHz signal, the contralateral flanking components actually led to a threshold increase of about 9 dB, relative to threshold observed for the on-signal band alone. Thus, far from producing a release from masking, the flanking component produced an interference effect. For both the 1- and 6-kHz signals, thresholds in the presence of the contralateral flanking components were much higher than when all components were presented to the same ear (broadband condition of experiment 1).

To assess the statistical significance of these effects, an ANOVA was conducted with factors center frequency and type of flanking components (absent, present in ipsilateral ear, or present in contralateral ear). The main effect of frequency was not significant. The main effect of type of flanking component was significant: $F(2,6)=140.29$, $p<0.001$. *Posthoc* tests indicated that the mean threshold was significantly lower when the flanking components were in the same ear as the on-frequency components than when they were in the opposite ear ($p<0.001$). Also, the mean threshold was significantly higher when the flanking components were in the opposite ear to the on-frequency components than when there were no flanking components ($p<0.05$). Although the latter effect appeared to be stronger at 6 kHz than at 1 kHz, the interaction of frequency and type of flanking component was not significant: $F(2,6)=2.8$, $p=0.138$.

We conclude that the reduction in threshold produced by increasing the bandwidth of the CP masker from 1 ERB_N to 5 ERB_Ns, as observed in experiment 1, was not the result of ''high level'' across-channel comparisons of the type thought to occur for some forms of CMR. If across-channel processes are involved, then they depend on the flanking components being presented to the same ear as the on-frequency components.

V. GENERAL DISCUSSION

A. Within-channel processes

Our results show that, in simultaneous masking, CP maskers lead to lower thresholds than RP maskers even when the masker bandwidth is less than 1 ERB_N, provided that precautions are taken to restrict off-frequency listening. This indicates a within-channel effect of masker phase. The effect may arise partly because of a process of dip listening, and/or partly because of the effects of fast-acting compression on the basilar membrane. Previous failures to find effects of masker phase for narrow-band maskers (Carlyon and Datta, 1997; Buss *et al.*, 1998; Oxenham and Dau, 2001a) presumably resulted from a failure to restrict off-frequency listening and/or from an inappropriate choice of component phases. Oxenham and Dau (2001a) did find a phase effect in simultaneous masking for their 4-kHz ''narrow-band'' masker, but, as noted earlier, this masker actually had a bandwidth of nearly 4 ERB_Ns.

B. The role of dip-listening

The modeling described in Sec. III D showed that the pattern of results from experiment 1 could be accounted for well without assuming selective listening in the dips of the masker waveform evoked on the basilar membrane. Although information about the signal appeared primarily in the dips of the CP maskers, the effect of the simulated basilar-membrane compression was to increase the salience of the signal in the dips ''automatically.'' The term dip-listening is often taken to imply a process of selective listening in time, using ''glimpses'' of the signal to make inferences about the presence of the signal (Moore, 2002). Our analysis suggests that such selective listening may not be necessary, at least for the masking conditions studied here. Dip listening may well play a role in other situations, for example when listening for a target talker in the presence of background speech (Duquesnoy, 1983; Peters *et al.*, 1998) or a fluctuating noise (Howard-Jones and Rosen, 1993).

C. Across-channel processes

Our analysis using the model described earlier suggested that the results of experiment 1 could be accounted for using a single-channel model including a simulation of basilar-membrane filtering and compression. Oxenham and Dau (2001a) also found that phase and bandwidth effects obtained using m_+ and m_- maskers could be accounted for using a single-channel model; they used a variety of initial filters in combination with the model of Dau *et al.* (1997) or combined with a temporal window model similar to that used here.

The results of experiment 2 indicated that, if across-channel processes are involved, they do not behave like those that influence some forms of CMR. The presentation of the flanking components to the opposite ear to the on-signal

band did not produce a release from masking. A similar finding was reported by Oxenham and Dau (2001a) for the m_+ masker, although, as noted earlier, their failure to find an effect could have been due to the lack of precautions to prevent off-frequency listening for the narrow-band masker. Our results suggest that the masking release produced by flanking components presented ipsilaterally to the ear receiving the signal and on-frequency components does not result from a ''high-level'' process of across-channel comparison.

VI. CONCLUSIONS

We summarize here our main findings and the conclusions that can be drawn from them.

(1) In simultaneous masking, CP maskers led to lower thresholds than RP maskers even when the masker bandwidth was less than 1 ERB_N, when a notched-noise masker was used to restrict off-frequency listening. Thus, there is a within-channel effect of masker component phase. Previous failures to find an effect of masker component phase for narrow-band maskers can be attributed to the absence of an additional masker to restrict off-frequency listening and/or an inappropriate choice of component phases.

(2) In simultaneous masking, thresholds in the CP masker dropped dramatically when the masker bandwidth was increased from 1 ERB_N to 5 ERB_Ns, while thresholds in the RP masker increased slightly. The drop for the CP masker can probably be attributed to the flanking masker components leading to a waveform with deeper ''valleys'' at the place on the basilar membrane where the signal is detected.

(3) For the 5-ERB_N wide CP maskers, presentation of the flanking components (outside the ERB_N centered at the signal frequency) to the ear contralateral to that receiving the signal and on-frequency band did not result in less masking than obtained using the on-frequency band alone. This suggests that the masking release observed when all components are presented to the same ear does not depend on ''high-level'' across-channel comparisons of the type involved in some forms of CMR.

(4) The simultaneous masking results of experiment 1 could be accounted for reasonably well using a single-channel model including a simulation of basilar-membrane filtering and compression. It does not appear necessary to assume temporally selective listening during the dips in the waveform evoked by the masker on the basilar membrane.

ACKNOWLEDGMENTS

This work was supported by the Medical Research Council (UK). Author JIA was supported by the RNID (UK). We thank Michael Stone for help in calculating the waveforms at the outputs of the auditory filters, and Marjorie Leek and two anonymous reviewers for helpful comments.

Alcántara, J. I., and Moore, B. C. J. (**1995**). ''The identification of vowel-like harmonic complexes: Effects of component phase, level, and fundamental frequency,'' J. Acoust. Soc. Am. **97**, 3813–3824.

Alcántara, J. I., Holube, I., and Moore, B. C. J. (**1996**). ''Effects of phase and level on vowel identification: Data and predictions based on a nonlinear basilar-membrane model,'' J. Acoust. Soc. Am. **100**, 2382–2392.

Bracewell, R. N. (**1986**). *The Fourier Transform and its Applications* (McGraw–Hill, New York).

Buss, E., Grose, J. H., and Hall, J. W. (**1998**). ''Fast-acting compression of auditory stimuli: A test with narrowband stimuli,'' Assoc. Res. Otolaryngol., Abs. **21**, 798.

Buus, S. (**1985**). ''Release from masking caused by envelope fluctuations,'' J. Acoust. Soc. Am. **78**, 1958–1965.

Buus, S., Zhang, L., and Florentine, M. (**1996**). ''Stimulus-driven, time-varying weights for comodulation masking release,'' J. Acoust. Soc. Am. **99**, 2288–2297.

Carlyon, R. P., and Datta, A. J. (**1997**). ''Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system,'' J. Acoust. Soc. Am. **101**, 3636–3647.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). ''Modeling auditory processing of amplitude modulation: I. Detection and masking with narrowband carriers,'' J. Acoust. Soc. Am. **102**, 2892–2905.

Duifhuis, H. (**1970**). ''Audibility of high harmonics in a periodic pulse,'' J. Acoust. Soc. Am. **48**, 888–893.

Duquesnoy, A. J. (**1983**). ''Effect of a single interfering noise or speech source on the binaural sentence intelligibility of aged persons,'' J. Acoust. Soc. Am. **74**, 739–743.

Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–65.

Giguère, C., and Woodland, P. C. (**1994**). ''A computational model of the auditory periphery for speech and hearing research. I. Ascending path,'' J. Acoust. Soc. Am. **95**, 331–342.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

Gockel, H., Moore, B. C. J., and Patterson, R. D. (**2002**). ''Asymmetry of masking between complex tones and noise: The role of temporal structure and peripheral compression,'' J. Acoust. Soc. Am. **111**, 2759–2770.

Hall, J. W., Cokely, J., and Grose, J. H. (**1988**). ''Combined monaural and binaural masking release,'' J. Acoust. Soc. Am. **83**, 1839–1845.

Hall, J. W., Grose, J. H., and Mendoza, L. (**1995**). ''Across-channel processes in masking,'' in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). ''Detection in noise by spectro-temporal pattern analysis,'' J. Acoust. Soc. Am. **76**, 50–56.

Hanna, T. E., and Robinson, D. E. (**1985**). ''Phase effects for a sine wave masked by reproducible noise,'' J. Acoust. Soc. Am. **77**, 1129–1140.

Horst, J. W., Javel, E., and Farley, G. R. (**1990**). ''Coding of spectral fine structure in the auditory nerve. II. Level-dependent nonlinear responses,'' J. Acoust. Soc. Am. **88**, 2656–2681.

Howard-Jones, P. A., and Rosen, S. (**1993**). ''Uncomodulated glimpsing in 'checkerboard' noise,'' J. Acoust. Soc. Am. **93**, 2915–2922.

Irino, T., and Patterson, R. D. (**2001**). ''A compressive gammachirp auditory filter for both physiological and psychophysical data,'' J. Acoust. Soc. Am. **109**, 2008–2022.

Kohlrausch, A., and Sander, A. (**1995**). ''Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets,'' J. Acoust. Soc. Am. **97**, 1817–1829.

Lentz, J. J., and Leek, M. R. (**2001**). ''Psychophysical estimates of cochlear phase response: masking by harmonic complexes,'' J. Assoc. Res. Otolaryngol. **2**, 408–422.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Mehrgardt, S., and Schroeder, M. R. (**1983**). ''Monaural phase effects in masking with multicomponent signals,'' in *Hearing-Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer-Verlag, Berlin).

Moore, B. C. J. (**1988**). ''Dynamic aspects of auditory masking,'' in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. Edelman, W. Gall, and W. Cowan (Wiley, New York).

Moore, B. C. J. (**1992**). ''Across-channel processes in auditory masking,'' J. Acoust. Soc. Jpn. (E) **13**, 25–37.

Moore, B. C. J. (**2002**). ''Interference effects and phase sensitivity in hearing,'' Philos. Trans. R. Soc. London, Ser. A **360**, 833–858.

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego).

Moore, B. C. J., and Shailer, M. J. (**1991**). ''Comodulation masking release as a function of level,'' J. Acoust. Soc. Am. **90**, 829–835.

Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (**1988**). ''The shape of the ear's temporal window,'' J. Acoust. Soc. Am. **83**, 1102–1116.
Mott, J. B., and Feth, L. L. (**1986**). ''Effects of the temporal properties of a masker upon simultaneous masking patterns,'' in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson (Plenum, New York).
Oxenham, A. J., and Dau, T. (**2001a**). ''Reconciling frequency selectivity and phase effects in masking,'' J. Acoust. Soc. Am. **110**, 1525–1538.
Oxenham, A. J., and Dau, T. (**2001b**). ''Towards a measure of auditory-filter phase response,'' J. Acoust. Soc. Am. **110**, 3169–3178.
Oxenham, A. J., and Moore, B. C. J. (**1994**). ''Modeling the additivity of nonsimultaneous masking,'' Hear. Res. **80**, 105–118.
Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). ''The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold,'' J. Acoust. Soc. Am. **72**, 1788–1803.
Peters, R. W., Moore, B. C. J., and Baer, T. (**1998**). ''Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people,'' J. Acoust. Soc. Am. **103**, 577–587.
Plack, C. J., and Moore, B. C. J. (**1990**). ''Temporal window shape as a function of frequency and level,'' J. Acoust. Soc. Am. **87**, 2178–2187.
Recio, A., and Rhode, W. S. (**2000**). ''Basilar membrane responses to broadband stimuli,'' J. Acoust. Soc. Am. **108**, 2281–2298.
Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). ''Basilar-membrane responses to clicks at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **103**, 1972–1989.
Rhode, W. S., and Robles, L. (**1974**). ''Evidence from Mössbauer experiments for nonlinear vibration in the cochlea,'' J. Acoust. Soc. Am. **55**, 588–596.
Robles, L., and Ruggero, M. A. (**2001**). ''Mechanics of the mammalian cochlea,'' Physiol. Rev. **81**, 1305–1352.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). ''Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and response phases,'' J. Acoust. Soc. Am. **80**, 1364–1374.
Ruggero, M. A. (**1992**). ''Responses to sound of the basilar membrane of the mammalian cochlea,'' Curr. Opin. Neurobiol. **2**, 449–456.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Schooneveldt, G. P., and Moore, B. C. J. (**1987**). ''Comodulation masking release (CMR): effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band,'' J. Acoust. Soc. Am. **82**, 1944–1956.
Schroeder, M. R. (**1970**). ''Synthesis of low peak-factor signals and binary sequences with low autocorrelation,'' IEEE Trans. Inf. Theory **IT-16**, 85–89.
Smith, B. K., Sieben, U. K., Kohlrausch, A., and Schroeder, M. R. (**1986**). ''Phase effects in masking related to dispersion in the inner ear,'' J. Acoust. Soc. Am. **80**, 1631–1637.
Summers, V., and Leek, M. R. (**1998**). ''Masking of tones and speech by Schroeder-phase harmonic complexes in normally hearing and hearing-impaired listeners,'' Hear. Res. **118**, 139–150.
Zwicker, E. (**1956**). ''The foundations for the determination of the information capacity of the auditory system''. ''Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs,'' Acustica **6**, 356–381.

Threshold differences for interaural time delays carried by double vowels[a)]

Michael A. Akeroyd[b)]
Department of Neuroscience, University of Connecticut Health Center, Farmington, Connecticut 06030 and Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, United Kingdom

(Received 7 March 2002; revised 25 July 2003; accepted 25 July 2003)

Experimental measurements were made of threshold interaural time differences (ITDs) for a ''target'' vowel presented simultaneously with a fixed-ITD ''distracter'' vowel. Three double-vowel pairs were used, comprising an ''er'' (/ə/) together with either an ''ai,'' ''ar,'' or ''oo'' (respectively, /e/, /ɔ/, and /u/). Threshold ITDs were found to be larger for the target vowel when it was part of a double-vowel pair than in control conditions in which it was presented alone. The effect size depended upon the choice of target vowel and distracter vowel, the level of the target relative to the distracter, and whether the two vowels had the same or different fundamental frequencies. The experiment was analyzed using a multichannel modification of Heller and Trahiotis' [J. Acoust. Soc. Am. **99**, 3632–3637 (1996)] model, which used a weighted combination of the detectabilities of the ITD of the target and the distracter. It gave predictions consistent with the observed effects of level and with some of the effects of the choice of target vowel, but it could not describe the effect of the target–distracter differences in fundamental frequency. It was found that a single-channel version of the model, in which the chosen channel was allowed to depend upon fundamental frequency (which could be derived using a monaural autocorrelation model) did give a set of predictions in qualitative accord with the data. [DOI: 10.1121/1.1611884]

PACS numbers: 43.66.Pn, 43.66.Ba, 43.66.Hg [LRB]

I. INTRODUCTION

A common task performed by the auditory system is the separation of two simultaneous voices, each coming from a different spatial location, such as in the famous ''cocktail-party'' situation. Because the spectra of two voices overlap, giving harmonics and formants that are sometimes shared, the cues to the spatial location of the two voices are intermingled. How does the auditory system determine the spatial location of each voice? The experiment here investigates this question by studying the detectability of changes in the interaural time delay (ITD) of a short, synthetic, vowel presented simultaneously with a second synthetic vowel.

McFadden and Pasanen (1976) studied binaural performance with spectrally nonoverlapping sounds, sufficiently far apart in frequency such that any interactive effects could be attributed to across-channel effects. They found that the threshold difference for ITD for one narrow band of noise (the ''target''), centered at 4000 Hz, was substantially increased by adding a second narrow band of noise (the ''distracter''), centered at 500 Hz and of fixed ITD. The mere presence of a distracter sound was not, however, sufficient to give this effect: no increase in threshold ITD was observed when the distracter was a continuous, interaurally uncorrelated noise, low-pass filtered at 650 Hz, or when the target was a narrow-band noise centered at 500 Hz and the distracter was a narrow-band noise centered at 4000 Hz. These observations of ''binaural interference'' have subsequently been confirmed and extended (e.g., Bernstein and Trahiotis, 1995; Heller and Trahiotis, 1995, 1996; Smith-Olinde *et al.*, 1998).

Increases in threshold ITD for a target have also been observed if it and the distracter are pure or complex tones below 1500 Hz. Buell and Hafter (1991) used pure tones between 354 and 800 Hz as targets and distracters, but only observed a change in threshold ITD when the frequencies of the two tones were harmonically related, although Stellmack and Dye (1993) did observe an effect with inharmonic complexes. Woods and Colburn (1992) observed an increase in threshold ITD for a 600-Hz pure-tone target presented against a fixed-ITD 400+800-Hz two-tone distracter. The effect remained if the distracter began 150 ms before the target, even though in this condition the listeners reported perceiving the distracter and the target as separate objects [Stellmack and Dye (1993) reported similar results]. Stellmack (1994) observed that thresholds recovered if the portion of the distracter that was simultaneous with the target was removed. Stellmack and Dye (1993) used a pure-tone target of 753 Hz amidst fixed-ITD complex tone distracters with frequency spacings from 10 to 100 Hz, and noted that the increase in the threshold ITD of the target was largest for intermediate spacings. They, and Dye (1990), who used three-component or five-component complexes as stimuli with components between 300 and 1200 Hz, found that the

[a)]Parts of this research were presented at the 141st meeting of the Acoustical Society of America [M. A. Akeroyd, J. Acoust. Soc. Am. **109**, 2375(A) (2001)] and at the British Society of Audiology Short Papers Meeting on Experimental Studies of Hearing and Deafness [M. A. Akeroyd, Int. J. Audiol. **41**, 253–254 (2002)].

[b)]Present address: MRC Institute of Hearing Research (Scottish Section), Glasgow Royal Infirmary, Alexandra Parade, Glasgow G31 2ER, United Kingdom. Electronic mail: maa@ihr.gla.ac.uk.

size of the effect increased in proportion to the number of components in the distracter.

Dye (1990) accounted for the dependence of threshold ITD on the number of distracter components by assuming a process wherein the detectability (d') of the changes in ITD of the target and distracters was averaged. Because the distracters were fixed in ITD, they contributed a d' of zero to the average. Thus, the greater the number of distracters, the smaller would be the d' expected for any target ITD. Buell and Hafter (1991) showed that a more-detailed d' model, in which the detectabilities were weighted in inverse proportion to the corresponding threshold ITDs, could describe their own results and those of McFadden and Pasanen (1976). This model was extended by Heller and Trahiotis (1996), incorporating into the weightings measurements of the ITD-based extent of laterality of the individual components of the target.

These studies have all shown that the fidelity of ITD processing of one sound can be affected by the presence of a second sound. They have used relatively simple stimuli: bandlimited noises, pure tones, or complex tones with, at most, a small number of components. The present experiment uses short, synthetic vowels with overlapping harmonics and overlapping lower formants. These ''double-vowel'' stimuli have formed the basis of many studies of the separation of sounds. Listeners find it difficult to identify a target vowel when it is presented simultaneously on the same fundamental frequency (F_0) as a distracter, but they find it increasingly easier to do so if a small difference in F_0 is introduced (e.g., Scheffers, 1983; Assmann and Summerfield, 1990, 1994; de Cheveigné *et al.*, 1997). Performance improves only marginally if the target and distracter are presented with different ITDs (Shackleton and Meddis, 1992), suggesting that monaural cues like F_0, but not binaural cues like ITD, are primary in guiding the separation of simultaneous short sounds (Culling and Summerfield, 1995, but see Hukin and Darwin, 1995, for a contrasting result with sentences).

Two questions are considered here: (1) Is an increase in threshold ITD as a result of ''distracting'' stimulation also observed with vowel stimuli? (2) Can the size of any increase be predicted by theoretical models, such as Heller and Trahiotis' (1996) d' model? Question 1 is studied in Sec. II, which describes an experiment measuring the threshold ITD for a target vowel in the presence of a fixed-ITD distracter vowel, as a function of the choice of target and distracter vowels, whether they had the same or different fundamental frequencies, and their relative level. The results indicated that an increase in threshold ITD was indeed observed with vowel stimuli, but that its size depended upon all three factors. Question 2 is studied in Sec. III. A variety of models based on Heller and Trahiotis' (1996) d' model is considered, some incorporating a compulsory across-frequency combination of information across all the channels of an auditory filterbank, others utilizing only a single channel. All the models gave predictions consistent with the effects of relative level. The multichannel model offered reasonable predictions of the data when there was no difference in F_0 between the vowels, but it could not describe the effects observed when there was a 25-Hz difference in F_0. The single-channel d' model, in which the choice of frequency channel was made dependent on F_0, however, gave a set of predictions that *did* qualitatively reflect F_0 differences. Moreover, it is shown that a monaural, autocorrelation analysis of F_0 (Meddis and Hewitt, 1992) could support the choice of channel. This model employs monaural analysis in order to guide source segregation followed by subsequent binaural analysis (e.g., Hill and Darwin, 1996; Shackleton, 1998, Darwin and Hukin, 1999).

II. EXPERIMENT: THRESHOLD ITDS FOR A TARGET VOWEL

A. Method

1. Design

The effect of a distracter vowel on the threshold ITD of a target vowel was studied. In a set of ''double-vowel'' conditions, threshold ITDs were measured for a target vowel when presented simultaneously with a fixed-ITD distracter vowel. Three values of target level *re*: distracter level were tested: -6, 0, and $+6$ dB. The double vowels were formed by adding the neutral vowel /ə/ (''herd'') to one of /e/, /ɔ/, or /u/ (''hair,'' ''hard,'' or ''who'd,'' respectively). In half the conditions, the /ə/ was the target and the /e/, /ɔ/, or /u/ was the distracter; in the other half of the conditions, these relations were swapped. In a set of control ''single-vowel'' conditions, threshold ITDs were measured for each of /ə/, /e/, /ɔ/, and /u/ when presented alone. The ''threshold ratio'' was then defined as the ratio of the threshold ITD, in μs, for a given double-vowel condition and listener, to the single-vowel threshold ITD, in μs, for that listener. It is a measure of the degree of interference caused by the distracter vowel.

The vowel pairs were crafted to give three different relations between the first formant ($F1$) frequency of the constituents. One aim was to study the effects for frequencies around 600 Hz, which have been found to dominate many aspects of binaural performance, including the sensitivity to ITD and the extent of ITD-based lateralization of noises or pure tones (e.g., Raatgever, 1980; Buell and Hafter, 1991). First, the lower three formant frequencies $F1$, $F2$, and $F3$ of the neutral vowel /ə/ were set to 500, 1500, and 2500 Hz, respectively (Lieberman and Blumstein, 1988, p. 46). Next, the formant frequencies of the other vowels were derived from the values measured by Hillenbrand *et al.* (1995, Table V) for American English. For /e/, /ɔ/, and /u/ they reported $F1$ frequencies of 476, 652, and 378 Hz, respectively. When paired with /ə/ these values would give $F1$ relations of $F1_{/ə/} \approx F1_{/e/}$, $F1_{/ə/} < F1_{/ɔ/}$, and $F1_{/ə/} > F1_{/u/}$. Accordingly, these $F1$ frequencies were rounded to 500, 625, and 375 Hz, so giving relations of $F1_{/ə/} = F1_{/e/}$, $F1_{/ə/} < F1_{/ɔ/}$, and $F1_{/ə/} > F1_{/u/}$. A secondary aim of the design was to remove formant-frequency differences irrelevant to the present design, so that the $F2$ frequencies were set to be multiples of 500 Hz, the $F3$, $F4$, and $F5$ frequencies were set to be the same for all vowels, and the $F1-F5$ formant frequencies were set to be multiples of the F_0 of their vowel.

In the first set of conditions the /ə/ vowel was synthesized on a fundamental frequency (F_0) of 100 Hz and the

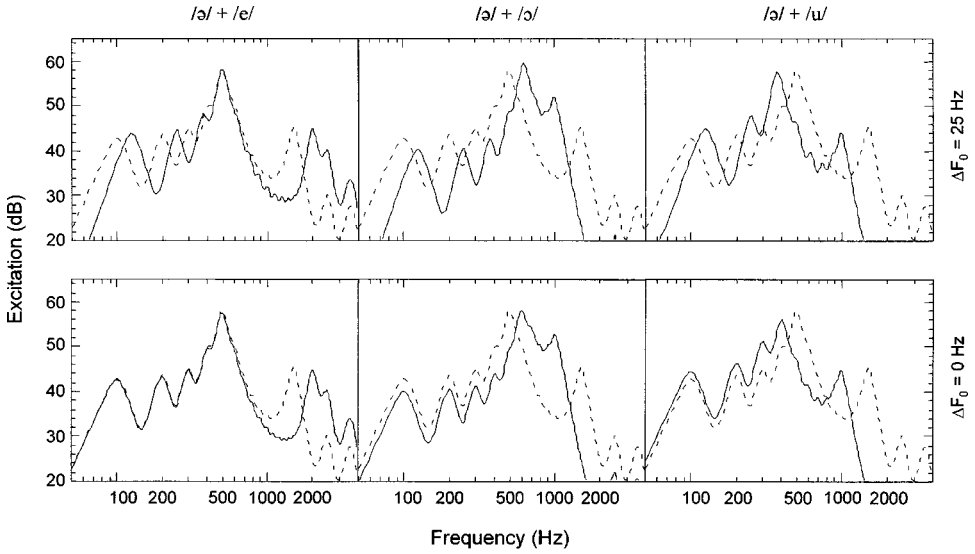


FIG. 1. Excitation patterns of the vowel stimuli. The three columns show the excitation patterns of /ə/ +/e/, /ə/+/ɔ/, and /ə/+/u/. For the top row, the /ə/ (dashed lines) had an F_0 of 100 Hz and the /e/, /ɔ/, and /u/ (solid lines) had an F_0 of 125 Hz, giving a ΔF_0 of 25 Hz. For the bottom row, all the vowels had an F_0 of 100 Hz, giving a ΔF_0 of 0 Hz. The excitation patterns were calculated using Moore *et al.*'s (1997) model and incorporate the transmission characteristics of the ear and a TDH-39 headset.

other vowels on a F_0 of 125 Hz, so giving a ΔF_0 of 25 Hz or approximately 4 semitones. The target vowel's ITD was always $\geqslant 0$ μs, but the distracter vowel's ITD was fixed at -400 μs (a positive ITD corresponded to the left ear leading). These differences in F_0 and ITD were chosen in order to lead to clear, separate percepts of the individual vowels. For similar ITD differences but a ΔF_0 of only 1 semitone (6 Hz), Shackleton and Meddis (1992) noted that listeners reported two separate vowel percepts at clearly different lateralizations. Many experiments on the identification of the constituent vowels of double vowels have shown that performance depends on ΔF_0, increasing by as much as 30%–40% from 0 to 1 or 2 semitones (see de Cheveigné *et al.*, 1997, Fig. 1). Shackleton *et al.* (1994) reported that changing the ITD of one vowel from 0 to 400 μs led to gain in identifiability of just 0.1% (at 0 semitones) or 12% (at 1 semitone), while in contrast a change in ΔF_0 from 0 to 1 semitones led to a gain of 36%. The author's own experience was that the two vowels were, indeed, lateralized in noticeably different places: the target was somewhere on the left side of the head, depending on its actual ITD, but the distracter was on the far right side of the head.

In a second set of conditions, run after all the $\Delta F_0 = 25$-Hz conditions were completed, all the vowels were synthesized with a fundamental frequency of 100 Hz. Thus, there was no difference in F_0 to help in perceptually separating the target vowel from the distracter vowel in these conditions.

2. Stimuli synthesis and apparatus

The waveforms for each vowel were digitally presynthesized and stored on computer disk, using a MATLAB implementation of Klatt's cascade synthesizer (Klatt, 1980) at a sampling rate of 20 000 samples per second. As noted above, the formant frequencies were based on those measured by Hillenbrand *et al.* (1995) for American English vowels but rounded and simplified.[1] After synthesis, the vowels were scaled so that the overall power of each vowel was equal. The top row of Fig. 1 shows the excitation patterns of the three pairs /ə/+/e/, /ə/+/ɔ/, and /ə/+/u/ for a target level *re*: distracter level of 0 dB and for a ΔF_0 of 25 Hz. The bottom row shows the corresponding excitation patterns for a ΔF_0 of 0 Hz. These patterns were calculated by Moore *et al.*'s (1997) "LOUDAES" model, and incorporate the transmission characteristics of a TDH-39 headphone and the outer and middle ear.

At the beginning of each experimental run the waveforms of the relevant vowels were loaded from computer disk. The ITD required on each trial was applied via appropriate linear phase shifts to a 4096-point FFT of the left-ear waveform. The stimuli destined for the left and right ears were gated coincidentally using 10-ms duration (measured across the 10%–90% points) cosine-squared onset and offset ramps, yielding stimuli of 204.8-ms overall duration. This process led to an ongoing ITD without any onset ITD. The waveforms of the two vowels were then digitally mixed, with the appropriate amplitude gains for the required target *re*: distracter level. All of these computations were carried out on a TDT AP2. After synthesis, the sounds were converted to analog waveforms (TDT PowerDac), low-pass filtered at 8.5 kHz (TDT FLT2), and then presented over headphones (TDH-39, mounted in MX/41-AR cushions) to listeners seated in individual, single-walled, IAC booths, at an overall level of 64 dB SPL. The experiments were conducted at the University of Connecticut Health Center.

3. Procedures

For each listener, individual thresholds were obtained from three successive adaptive runs for each of the single-vowel and $\Delta F_0 = 25$-Hz double-vowel conditions, intermingled and visited in a random order. Then, three more thresholds were obtained at each condition, visited in the reverse order, and then three more at each in the original order, and so on. The values of threshold ITD reported below are the average of the individual thresholds from the last six adaptive runs taken from a set of between 12 and 18 adaptive runs. The design was then repeated for the $\Delta F_0 = 0$-Hz double-vowel conditions, intermingled with a second set of

single-vowel conditions. The experiment was conducted across many sessions; listeners typically completed about nine adaptive runs in each session.

A four-interval, two-alternative forced-choice adaptive tracking procedure was used (e.g., Bernstein and Trahiotis, 1982). Intervals 1 and 4 served as ''cue'' intervals in which the target vowel was presented with an ITD of 0 μs. The target vowel was presented with an ITD of τ μs (where $\tau > 0$) in intervals 2 or 3 with equal *a priori* probability. In the remaining interval the target vowel was again presented with an ITD of 0 μs. The listener's task was to determine which of intervals 2 or 3 contained the τ-μs ITD target. In the double-vowel conditions, the distracter vowel was presented simultaneously in all of the intervals with an ITD fixed at -400 μs. In the single-vowel conditions the distracter vowel was not presented. The ITD of the target vowel was varied adaptively, using a two-down, one-up rule asymptoting at the 70.7%-correct point on the psychometric function (Levitt, 1971). The adaptive track used a logarithmic scale of ITD. The initial step size for the adaptive track corresponded to a factor of 2.51 (equivalent to $0.4 \log_{10}$ units) for trials up to and including the second change of direction of the track, then was reduced to a factor of 1.58 ($0.2 \log_{10}$ units) up to and including the next two changes of direction, and was then fixed at a factor of 1.26 ($0.1 \log_{10}$ units) for the remaining 12 changes of direction (Saberi, 1995). Threshold ITD for the individual run was determined from the arithmetic mean of the logarithmic value at each of the last ten reversals. Average thresholds and the statistical analyses were also calculated using threshold ITDs in log units.[2] The ANOVAs reported below were calculated using the univariate general linear model of SPSS version 11. They had listener as a random factor, the three experimental manipulations of relative level, vowel pair, and target vowel as fixed factors, and used six values of threshold ITD (in log units) per cell.

4. Listeners

Four listeners, aged 23–26, participated in the experiment. Three of the listeners (A, B, and D) had participated in earlier experiments at the University of Connecticut Health Center. All of the listeners had audiometric thresholds lower than 15 dB at octave frequencies from 250 to 8000 Hz inclusive.

B. Results

1. Fundamental-frequency difference=25 Hz

For the first set of single-vowel conditions, the means of the threshold ITDs for /ə/, /e/, /ɔ/, and /u/ were, respectively, 34, 45, 36, and 42 μs. An ANOVA showed that the threshold ITDs did not differ significantly across the four vowels [$F(3,9)=0.74$]. The overall means for each of the four listeners A, B, C, and D were, respectively, 91, 14, 45, and 42 μs; these differences were significant [$F(3,9)=25.5$; $p<0.001$].

The left panel of Fig. 2 shows the threshold ratios for a ΔF_0 of 25 Hz and for the four listeners individually but averaged across vowel pair and target, as a function of the relative level of the target vowel. As the ratios were generally larger than 1.0, the double-vowel threshold ITDs were generally larger than the mean single-vowel threshold ITD. The ratios decreased as the relative level was increased, although the ratios derived for the four listeners did differ somewhat. The ANOVA showed that both these main effects were significant (relative level: $F(2,6)=13.7$, $p=0.006$; listener: $F(3,7)=7.5$, $p=0.013$). None of the interactions of main effect with listener was found to be significant.

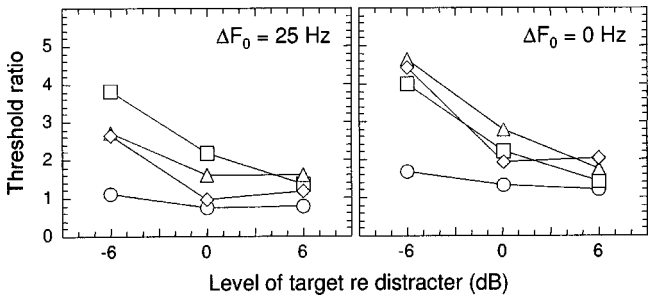


FIG. 2. The mean threshold ratios for the four listeners as a function of relative level, averaged across vowel. The two panels are for the $\Delta F_0 = 25$-Hz conditions (left) and $\Delta F_0 = 0$-Hz conditions (right). The four listeners are: A, circles; B: squares; C, triangles; and D, diamonds.

The top panel of Fig. 3 shows the threshold ratio, averaged across listeners, for each of the vowel pairs as a function of the relative level of the target. The three plots show the results from the three pairs, both when the target vowel was /ə/ (open symbols) and when the distracter was /ə/ and therefore the target was either /e/, /ɔ/, or /u/ (filled symbols). The general effect of level is clear. The threshold ratios also depended upon the choice of target vowel within each pair. An ANOVA showed that this interaction of target×pair was significant [$F(2,6)=5.4; p=0.05$] but that of level×target ×pair was insignificant [$F(4,12)=0.62$]: the threshold ratio

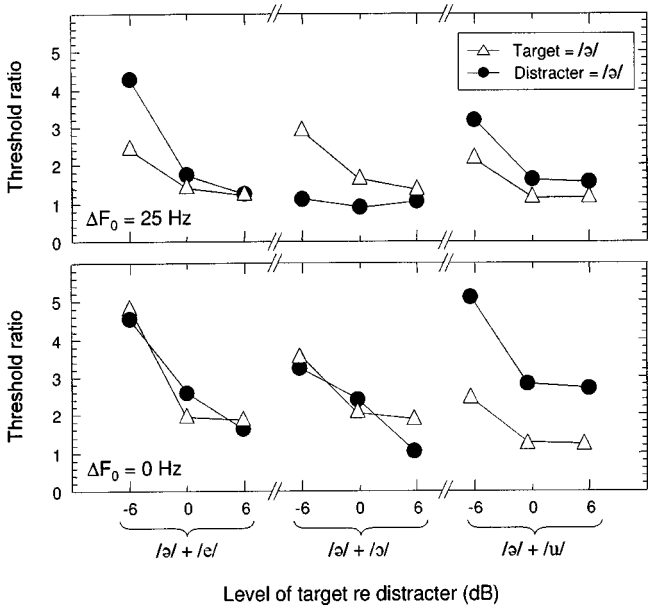


FIG. 3. The mean threshold ratios for each of the target vowels and vowel pairs, as a function of the level of the target relative to the distracter. The two panels are for the $\Delta F_0 = 25$-Hz conditions (top) and $\Delta F_0 = 0$-Hz conditions (bottom). Within each plot, the two lines show the results when the target vowel was /ə/ (open triangles) or when the distracter was /ə/ and therefore the target was either /ɛ/, /ɔ/, or /u/ (closed circles).

for the target /ə/ was generally *less* than that for the other vowel in two pairs (/ə/+/e/;/ə/+/u/) but *greater* in the middle pair (/ə/+/ɔ/). The mean ratios for each level, averaged across vowel target and pair, were 2.5, 1.4, and 1.3 for −6, 0, and +6 dB, respectively, giving a grand ratio of 1.5.

2. Fundamental-frequency difference=0 Hz

For the repeats of the single-vowel conditions, the mean threshold ITDs for /ə/, /e/, /ɔ/, and /u/, were, respectively, 36, 30, 32, and 38 μs, and so were within 15 μs of those measured earlier. The means for each of the four listeners A, B, C, and D were, respectively, 75, 16, 35, and 33, μs, respectively, and so were within 30 μs of those measured earlier. An ANOVA showed that the single-vowel threshold ITDs did not differ significantly across the sets of runs.

The right panel of Fig. 2 shows the average threshold ratio for the four listeners for the $\Delta F_0=0$-Hz conditions. As before, the four listeners generally differed amongst themselves, but the threshold ratios were always larger than 1.0 and showed a similar dependence on level [relative level: $F(2,6)=16.8$, $p=0.003$; listener: $F(3,1.6)=17.8$, $p=0.08$]. As before, none of the interactions of main effect with listener was found to be significant.

The bottom panel of Fig. 3 shows the threshold ratio for each of the vowel pairs as a function of the relative level of the target. Again, there was an effect of the choice of vowel pair and target: the ANOVA showed that the interaction of target×pair was significant [$F(2,6)=13.3$; $p=0.006$] but that of level×target×pair was insignificant [$F(4,12)=0.79$]. The effect of the choice of target vowel within each pair was slightly different than in the $\Delta F_0=25$-Hz conditions (top panel), however, in that only in one pair (/ə/+/u/) was there a reliable difference between the targets, /ə/ giving a considerably lower threshold ratio than /u/. In the /ə/+/ɔ/ pair, there was some evidence that the ratio for /ə/ was larger than for /ɔ/, but for the /ə/+/e/ pair, those differences were inconsistent across level. The mean ratios for each level, averaged across vowel target and pair, were 3.9, 2.1, and 1.7 for −6, 0, and +6 dB, respectively, giving a grand ratio of 2.2.

C. Discussion

Overall, the experiment indicated that the presence of a simultaneously presented distracting vowel led to a general increase in threshold ITD for a target vowel. The size of the increase was dependent upon the relative levels of the two vowels, a difference in fundamental frequency, and the choice of target and distracter themselves. The results extend earlier observations (e.g., McFadden and Pasanen, 1976; Dye, 1990; Buell and Hafter, 1991), which used relatively simple stimuli such as narrow-band noises or pure tones, and demonstrate that such interference is observed with spectrally complex sounds such as vowels.

A quantitative analysis of the results will be presented in the next section; here, the discussion is limited to a qualitative analysis of the interaction between the binaural dominance region around 600 Hz (e.g., Bilsen and Raatgever, 1973; Raatgever, 1980), and the vowel excitation patterns. For simplicity, the argument is limited to a target level *re*: distracter level of 0 dB.

First, consider the /ə/+/u/ pair (Fig. 1, right column). Here, the /ə/ vowel (dashed line) generates *more* excitation than the /u/ vowel (solid line) around 600 Hz. If the 600-Hz region dominates performance, then the threshold ratio when the target was /ə/ would be expected to be *lower* than when the target was /u/. The experimental data support this expectation; the mean threshold ratios for targets of /ə/ vs /u/ were, respectively 1.2 vs 1.6 ($\Delta F_0=25$ Hz) and 1.3 vs 2.8 ($\Delta F_0=0$ Hz). Second, consider the /ə/+/ɔ/ pair (Fig. 1, middle column). Here, /ə/ generates *less* excitation than /ɔ/ at 600 Hz, and so the threshold ratio for the /ə/ target would be expected to be *greater* than that for /ɔ/ target. The experimental data support this expectation for $\Delta F_0=25$ Hz but not for 0 Hz, the mean threshold ratios for targets of /ə/ vs /ɔ/ being, respectively, 1.6 vs 0.9 ($\Delta F_0=25$ Hz) and 2.0 vs 2.4 ($\Delta F_0=0$ Hz). Finally, consider the /ə/+/e/ pair (Fig. 1, left column). Here, the two vowels generate approximately equal excitation at 600 Hz, and so the threshold ratios would be expected to be the same, no matter which one was the target. The experimental data, however, do not bear out this expectation, as there was an advantage for /ə/: the mean threshold ratios for targets of /ə/ vs /e/ were, respectively, 1.5 vs 1.8 ($\Delta F_0=25$ Hz) and 2.0 vs 2.6 ($\Delta F_0=0$ Hz). Overall, the effects of the choice of target vowel and vowel pair are partly consistent with a binaural dominance region around 600 Hz, being more consistent for the $\Delta F_0=25$-Hz conditions than the $\Delta F_0=0$-Hz conditions, with the exception of the /ə/ vs /e/ pair.

III. QUANTITATIVE MODELS BASED ON *d′*

This section considers the application of Heller and Trahiotis' (1996) *d′* model to double-vowel stimuli. Heller and Trahiotis measured the threshold ITD for a complex-tone target in the presence of a complex-tone distracter. They interpreted their results using an extension of a two-channel *d′* model originally developed by Buell and Hafter (1991). In the model, the target and distracter were assumed to be represented in two separate processing channels. Performance was based on a weighted combination of the detectabilities (*d′*) of the ITD information present in both channels. Heller and Trahiotis demonstrated that the model was reasonably accurate at accounting for the threshold ratios of sinusoidally amplitude-modulated tones, and Buell and Hafter (1991) had earlier demonstrated that the original model could account for the threshold ratios of pure tones and two-tone complexes.

In Heller and Trahiotis' (1996) and Buell and Hafter's (1991) experiments, the target and distracter sounds occupied different frequency regions. It is reasonable to assume that each sound was indeed represented in a separate processing channel. For the double-vowel stimuli employed here, however, this assumption seems invalid: the vowels were broadband, overlapping in frequency and sharing some harmonics. In order to allow the prediction of threshold ratios for double vowels, it was therefore decided to incorporate directly the

fact that each channel will represent, to a greater or lesser extent, *both* vowels. The question of the bandwidth of the stimuli was addressed in two different ways. In the first (Secs. III A and III B), it was decided to use a multichannel approach, combining d' across a large number of frequency channels covering 50–1500 Hz (the approximate range over which listeners effectively utilize fine-structure-based ITD information). In the second (Sec. III C), it was decided to use a single-channel approach, using d' from just one channel placed at, or near, the first-formant frequency of the target vowel.

A. Description of the multichannel model

Heller and Trahiotis (1996) assumed that the detectability, d'_T, of any change in the physical ITD of the target, *when presented alone*, was given by $d'_T = \Delta x_T / \sigma_T$, where σ_T was the standard deviation associated with processing ITDs, and where Δx_T was the change in perceived lateralization caused by a change $\Delta\tau_T$ in the ITD of the target

$$\Delta x_T = s_T \Delta\tau_T, \tag{1}$$

where s_T was an empirically measured slope. The processing standard deviation σ_T was assumed to be proportional to the empirically measured *threshold* ($d'=1$) ITD for the target when presented alone, ΔT_T

$$\sigma_T = s_T \Delta T_T. \tag{2}$$

A parallel set of assumptions was made for the detectability d'_D of any change in ITD in the distracter *when presented alone*, with variables Δx_D, σ_D, and $\Delta\tau_D$, together with empirical measurements s_D and ΔT_D. Overall, the detectability of a change in ITD of the target, *when presented simultaneously with the distracter*, was then determined by a weighted sum of the individual detectabilities d'_T and d'_D, giving [Heller and Trahiotis, Eq. (A4)]

$$d' = (w_T \Delta x_T + w_D \Delta x_D) / \sqrt{w_T^2 \sigma_T^2 + w_D^2 \sigma_D^2}, \tag{3}$$

where w_T and w_D are the weights applied to the target and the distracter. Heller and Trahiotis assumed that these weights were inversely proportional to the ITD-processing variances σ_T^2 and σ_D^2

$$w_T \propto 1/\sigma_T^2, \quad w_D \propto 1/\sigma_D^2. \tag{4}$$

Equation (1) can therefore be rearranged to give

$$d' = \left(\frac{\Delta\tau_T}{s_T \Delta T_T^2} + \frac{\Delta\tau_D}{s_D \Delta T_D^2} \right) \Bigg/ \sqrt{\frac{1}{s_T^2 \Delta T_T^2} + \frac{1}{s_D^2 \Delta T_D^2}}. \tag{5}$$

The threshold ITD $\Delta\tau_T$ for the target when presented simultaneously with the distracter could then be found as all the other variables are known: $d'=1$ at threshold; ΔT_T, ΔT_D, s_T, and s_D were measured, and $\Delta\tau_D = 0$ as the ITD of the distracter vowel was fixed.

In the present conditions the target and distracter vowels overlapped. It was therefore decided to combine the ITD information present in *all* the channels of a multichannel auditory frequency analysis. Each channel corresponds to a separate auditory filter instead of to a separate sound: there is no explicit segregation of the target from the distracter. The starting point for the present models is therefore an *N*-channel extension of Eq. (5) [Buell and Hafter, 1991, Eq. (1); Woods and Colburn, 1992, Eq. (3)]

$$d' = \left(\sum_{n=1}^{N} \frac{\Delta t_n}{s_n \Delta T_n^2} \right) \Bigg/ \sqrt{\sum_{n=1}^{N} \frac{1}{s_n^2 \Delta T_n^2}}. \tag{6}$$

Here, detectability is summed over *frequency* channels, and so (1) Δt_n represents the change in ITD in channel *n* resulting from a change in the ITD of the target vowel $\Delta\tau_T$ (the change from τ to t is to indicate that the value is the ITD in a frequency channel, not the ITD of the sound itself); (2) s_n is the empirical slope of the function linking a change in lateralization in channel *n* to a change in ITD in channel *n*; and (3) ΔT_n is the empirical value for the threshold ITD in channel *n* for a sound presented alone. Values of ΔT_n and s_n were not measured in the present experiment. It was, therefore, decided to assume that ΔT_n was proportional to the threshold ITD of a pure tone as a function of frequency, for which Buell and Hafter's (1991, p. 1896) function was used

$$\Delta T_n \propto 0.000\,003\,45 f_n^2 - 0.004\,28 f_n + 2.124\,52, \tag{7}$$

where f_n is the center frequency of the *n*th channel, in Hz. It was also decided to assume that s_n was given by Schiano *et al.*'s (1986) data on the lateralization of a pure tone as a function of frequency, for which a cubic polynomial was fitted to Stern and Shear's (1996, Fig. 2) normalization of those data

$$s_n = -1.015 \times 10^{-9} f_n^3 + 1.530 \times 10^{-6} f_n^2 - 1.251 \times 10^{-4} f_n + 0.558. \tag{8}$$

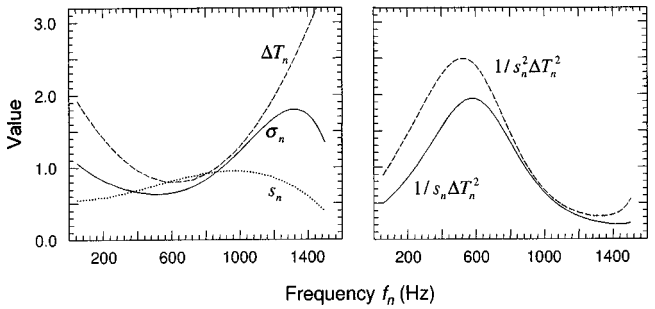


FIG. 4. Left panel: the frequency (f_n) dependence of the threshold ITD of a pure tone [ΔT_n, Eq. (7), with a proportionality constant of 1.0], the slope linking ITD to lateralization [s_n, Eq. (8)], and their product, the standard deviation associated with processing ITDs [σ_n, Eq. (2)]. Right panel: the frequency dependence of the factors in the numerator ($1/s_n\Delta T_n^2$) and denominator ($1/s_n^2\Delta T_n^2$) in the d' summation [Eq. (6)].

The left panel of Fig. 4 illustrates these functions: ΔT_n (dashed line), s_n (dotted line), and their product σ_n (solid line). The right panel shows the factors in Eq. (6): $1/s_n\Delta T_n^2$ (solid line) and $1/s_n^2\Delta T_n^2$ (dashed line). Note that both the latter two functions give maxima in the region of 500 Hz, and so those frequencies dominate the summations in Eq. (6). Frequencies lower than about 200 Hz or higher than about 1000 Hz play a much-reduced role.

The crucial part of the model is that Δt_n—being the change in ITD in channel *n* resulting from a change in the ITD of the target vowel $\Delta\tau_T$—is *not* necessarily equal to

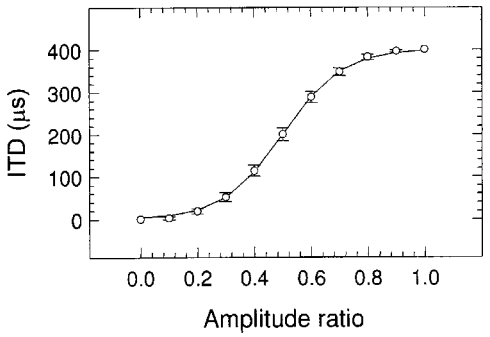


FIG. 5. The symbols mark the ITD measured at the output of a gammatone filter for the combination of a 400-μs ITD ''target'' noise and a 0-μs ''distracter'' noise. The abscissa is the amplitude ratio of the target [Eq. (9)]. The error bars are the standard deviation across 250 computational trials. The solid line is the best-fitting logistic function [Eq. (10)].

$\Delta\tau_T$. For instance, if none of the excitation in a channel is due to the target, then it will make no difference to Δt_n however much the target ITD $\Delta\tau_T$ is changed. If instead all the excitation in the channel is attributable to the target vowel, then Δt_n will equal $\Delta\tau_T$, but if the target and distracter contribute equally to the excitation, then Δt_n will be expected to equal $0.5\Delta\tau_T$. A computer simulation was run to find a numerical relation for the full range of target and distracter levels. In particular, a 1-second, 1–1000-Hz target noise of $\Delta\tau_T = 400$-μs ITD was generated at a sampling rate of 50 kHz, a similar distracter noise of 0-μs ITD generated, the two added, then filtered using a 500-Hz gammatone auditory filter (Patterson *et al.*, 1995), then the cross-correlation function measured. The position of the ''best'' peak—that one closest to 0 μs—was measured to a resolution of 1 μs using a 6-order polynomial fitted to the cross-correlation points within ± 250 μs of the largest one. The process was repeated 250 times for each of a range of levels of the target value. The results are shown in Fig. 5 as a function of the ''amplitude ratio''

$$r_n = \frac{a_{T,n}}{a_{T,n} + a_{D,n}}, \tag{9}$$

where $a_{T,n}$ and $a_{D,n}$ are the mean amplitudes of the target and the distracter, respectively, measured in the nth channel. A logistic function was then fitted to the data, giving

$$\Delta t_n = \frac{\Delta\tau_T}{1 + \exp((0.5 - r_n)/0.1035)}. \tag{10}$$

This function deviates from the computational data by, at most, 5 μs. It defines the degree to which the ITD, measured at the output of a frequency channel, stemming from the target is reduced in the presence of a distracter. It is this part of the model that incorporates the fact that each channel will represent, to a greater or lesser extent, both vowels: the relative excitations they generate in a given channel will affect the amplitude ratio r_n and therefore the values of Δt_n, so giving a dependence on which vowel pair is tested and which vowel of the pair is the target. This assumption is the *only* part of the multichannel model that is affected by the choice of vowel: there is no other segregation or grouping, direct or indirect, of the acoustic information into ''target'' or ''distracter.''

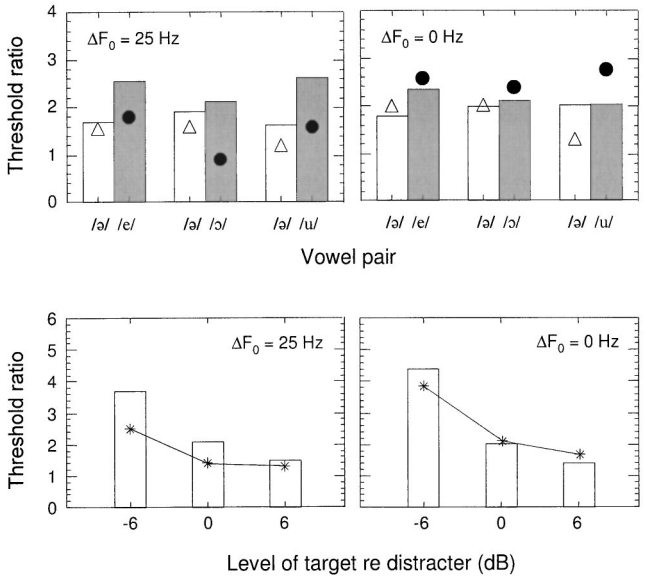


FIG. 6. Top row: Predictions from the multichannel d' model for $\Delta F_0 = 25$ Hz (left) and for $\Delta F_0 = 0$ Hz (right). In each panel, the symbols show the experimental data and the histogram bars show the corresponding predictions (triangles/open bars: target=/ə/; circles/filled bars: distracter =/ə/). Both the data and predictions are for a target level of 0 dB relative to the distracter. Bottom row: predictions from the same model (bars) compared to the experimental data (asterisks) as a function of target level, averaged across vowel pair and target.

Predictions of threshold ratio were found by first (1) for all the *single* vowels, setting r_n to be 1.0 for each channel; (2) setting the ITD of the target to be 40 μs (being close to the mean single-vowel threshold); (3) calculating Δt_n, ΔT_n, and s_n for each channel; (4) entering those values in Eq. (6); and (5) then finding the value of the proportionality constant in Eq. (7) that led to a mean d' of 1.0 across all the single vowels. Next, for each of the *double* vowels (6) the values of r_n were found by measuring the excitations of the vowels using Moore *et al.*'s (1997) ''LOUDAES'' model (see Fig. 1); (7) for a range of target ITDs $\Delta\tau_T$ from 0 to 1000 μs, calculating Δt_n, ΔT_n, and s_n for each channel; (8) entering those values in Eq. (6); and (9) finding whichever target ITD that gave a d' of 1.0 for the previously measured proportionality constant. Finally (10) the threshold ratio was found by taking the ratio of the double-vowel target ITD to the assumed single-vowel threshold ITD of 40 μs. The choice of channel frequencies was based on the set available in the LOUDAES model: 18 were used, with center frequencies placed at 1-ERB intervals, being 49, 81, 116, 155, 248, 302, 362, 429, 504, 588, 681, 781, 899, 1028, 1171, 1330, and 1507 Hz.

B. Predictions of the multichannel model

The predictions of the multichannel model for a target *re*: distracter level of 0 dB are illustrated in the top row of Fig. 6. The symbols mark the experimental threshold ratios from the $\Delta F_0 = 25$-Hz (top-left panel) and $\Delta F_0 = 0$-Hz (top-right panel) conditions. The histogram bars mark the corresponding predictions. The model has done a reasonable job

of predicting ratios obtained with two of the three vowel pairs for the $\Delta F_0 = 0$-Hz conditions, and has predicted the order of threshold ratios for two of the three vowel pairs for the $\Delta F_0 = 25$-Hz conditions. This qualified success is likely to be due to the dominance of frequencies around 500–600 Hz in the model (see Fig. 4). In contrast, the model predicts no major effect of changing ΔF_0: considered across all three target levels of +6, 0, and −6 dB, it predicted the $\Delta F_0 = 0$-Hz threshold ratios to be 1.03× larger than the $\Delta F_0 = 25$-Hz threshold ratios, whereas the experimental results showed the $\Delta F_0 = 0$-Hz ratios to in fact be 1.4× larger. The model also gave reasonable predictions of the effect of varying the level of the target vowel. The plots in the bottom row of Fig. 6 illustrate the experimental data (symbols), averaged across vowel target and pair and their corresponding predictions (histogram bars). The overall effect of level is clear; it is believed that the source of this qualitatively successful prediction is the relation between a change in target ITD $\Delta\tau_T$ and the resulting change in ITD in any channel Δt_n [see Eqs. (9) and (10)]. Indeed, if all the frequency-dependent complications are removed, by assuming a single-channel model in which r is 0.33, 0.5, and 0.66 for the three levels of −6, 0, and +6 dB, respectively, then the predicted ratios become 6.0, 2.0, and 1.2.[3]

The source of the ΔF_0 problem is a general overprediction of the threshold ratio for all the $\Delta F_0 = 25$-Hz conditions. A small selection of variations was tested in order to see if any could improve the predictions of this data. They included (1) incorporating an additional weighting so that channels with more excitation—especially those with the vowel formants exciting them—were weighted preferentially to channels with less excitation, and then implementing this weighting using decibels or linear units (the latter ''stretching'' the weighting considerably); or (2) using two different functions describing threshold ITD at each frequency for Eq. (7), namely, Stern *et al.*'s (1988) fit to Raatgever's (1980) ''dominance'' data, and a polynomial fit to Zwislocki and Feldman's (1956) data on the threshold ITD of a pure tone as a function of frequency.[4] Although these modifications changed the details of the predictions for each vowel target—for instance, Stern *et al.*'s function, for a ΔF_0 of 0 Hz, led to a fairly accurate difference between /ə/ and /u/ but also to a fairly inaccurate difference between /ə/ and /ɔ/—none offered any overall effect of changing ΔF_0. It seems clear that the multichannel d' model is unable to predict the effect of ΔF_0 on the threshold ratio of double vowels. One reason underlying this effect may be that a separation of F_0 leads to a stretching of the spectral pattern of the resolved harmonics. The result is that some channels will have more information about the target, but an almost equal number will have less, so, on average, the total amount of information would be expected to remain constant.

To summarize, Buell and Hafter (1991) and Heller and Trahiotis (1996) developed the d' model for nonoverlapping pure- or complex-tone stimuli. The analyses here have shown how it can be extended to double-vowel stimuli with overlapping formants and harmonics. Although this multichannel d' model could offer a good description of the relative level of the target vowel and of some details of the choice of target and vowel pair, it offered no account of the effect of ΔF_0. It seems that the exceptional case was when ΔF_0 was 25 Hz, for which the model did not describe the experimental observation that the threshold ratios are generally smaller than when ΔF_0 was 0 Hz. The possibility that this effect is due to some selective choosing of information is explored in the next section.

C. A single-channel model

The multichannel d' model has no explicit segregation of information into ''target'' or ''distracter,'' nor is there an explicit analysis of F_0. The one part of the model that can depend on differences in F_0 is Eq. (9), defining the amplitude ratio r, and then only if a change leads to a significant change in the excitation patterns and hence the values of r. These changes were small, however, and so the model showed only minor effects of varying ΔF_0. Here, a different approach was taken: it was decided to try a single-channel model, using d' information from just *one* channel.

It was found that, if the choice of channel depended upon ΔF_0, then the predictions varied with changes in ΔF_0. This discussion concentrates upon one pair of strategies for making the choice: for $\Delta F_0 = 25$ Hz, the chosen channel was that one placed at the first-formant ($F1$) frequency of the target vowel, but for $\Delta F_0 = 0$ Hz, the chosen channel was that one which gave the most excitation. That is, when there was a difference in F_0, suggesting that two sound sources were present, the ''best'' channel from those excited by the target was selected, but when there was no difference in F_0, suggesting only one sound source, the ''best'' overall channel was selected. This strategy is appealing because the choice of channel can be made automatically from the *monaural* information regarding fundamental frequency. To test this approach, the double-vowel stimuli were analyzed by an autocorrelation model based on that of Meddis and Hewitt (1992). In particular (1) the double-vowel waveforms were passed through a multichannel gammatone filterbank, with filters spaced at 0.1-ERB intervals from 50 to 4000 Hz; (2) the filter outputs were half-wave rectified and compressed using a power-law function with an exponent of 0.4 (Oxenham and Moore, 1995); (3) within-channel autocorrelation functions were calculated on those compressed outputs, using an exponential window of 10-ms time constant and of 30-ms overall duration [Meddis and Hewitt, 1992, Eq. (1)]; (4) finally, the functions were averaged across frequency. As expected from many studies of the identification of double vowels (e.g., Assmann and Summerfield, 1990; Meddis and Hewitt, 1992), the average functions showed two clear peaks, at 8 ms (=1/125 Hz) and 10 ms (=1/100 Hz) for the $\Delta F_0 = 25$-Hz stimuli, but only one peak, at 10 ms, for the $\Delta F_0 = 0$-Hz stimuli.[5] The crucial decision regarding whether to employ the most-excited channel or the channel corresponding to the $F1$ of the target is then made thus: if there are two peaks, pointing towards two fundamentals and two sound sources, then an explicit segregation of target from distracter can be done and the $F1$-channel strategy can be used, but if there is only one peak, so there is no evidence for more than one source, then the most-excited-channel strategy is used as a default.

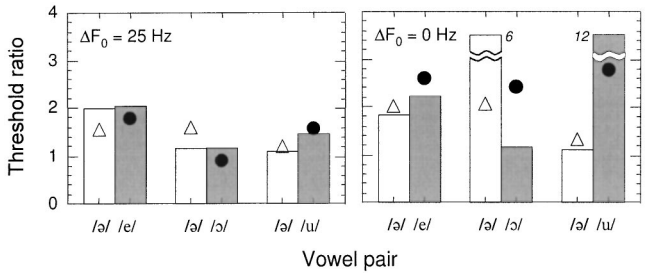


FIG. 7. Predictions from the single-channel d' model, using the $F1$-channel strategy (left panel, $\Delta F_0 = 25$ Hz) and the most-excited-channel strategy (right panel, $\Delta F_0 = 0$ Hz). The format is the same as that used in the top row of Fig. 6. Note that two of the predictions in the right panel were too extreme to be plotted and are marked numerically instead.

The histogram bars in the right panel of Fig. 7 show the predictions of the most-excited channel strategy for the $\Delta F_0 = 0$-Hz, equal-level stimuli. These channels were, for /ə/+/e/, at 504 Hz, for /ə/+/ɔ/ at 587 Hz, and for /ə/+/u/ at 504 Hz again. These predictions were formed by running the earlier d' model but with just this one channel entered into Eq. (6) (for the single vowels, the 504-Hz channel was used for all, as that was, on average, the most-excited channel of all the single vowels; the average autocorrelation functions will show only one clear peak so the choice of most-excited channel is consistent with the model). There was a clear effect of choice of vowel pair, but the predictions for /ɔ/ and /u/ are unfortunately extreme (this effect is believed to be due to both these vowels generating little excitation at the chosen frequency, so giving a small value of r [Eq. (9)] and so requiring a considerable ITD to give a d' of 1.0). Averaged across vowel, the mean ratio was 2.7.

The left panel of Fig. 7 shows the corresponding predictions of the $F1$-channel strategy for the $\Delta F_0 = 25$-Hz, equal-level stimuli. For simplicity, the experimenter's knowledge of the correct $F1$ was used: that is, when the target was /ə/, the channel was 500 Hz no matter what the distracter vowel was, when the target was /e/, it was 500 Hz, for /ɔ/, it was 625 Hz, and for /u/, it was 375 Hz.[6] The predictions were formed in the same manner as for the earlier single-channel strategy. It can be seen from the figure that the large differences across target observed in the previous single-channel strategy have been removed. Moreover, the mean threshold ratio was now 1.4, encouragingly close to the experimental value (1.4) and smaller than the value for $\Delta F_0 = 0$ Hz found by the other strategy.

The use of two different strategies was essential as it was found that, like with the earlier multichannel model, applying the same strategy to both the $\Delta F_0 = 25$-Hz and 0-Hz cases did not lead to differential predictions: it was only when each strategy was applied to separate cases that a reliable effect was found.[7] Both strategies were found to give qualitatively similar effects of overall level to those obtained from the multichannel model, as they use the same dependence of ITD on level [see Eqs. (9) and (10)]. It would seem that this segregation-of-information approach is a reasonable one. Moreover, the choice of formant frequency can also be determined from the autocorrelation functions. When the target vowel was /ɔ/ in /ə/+/ɔ/, the 625-Hz channel was dominated by a F_0 of 125 Hz, but when the target vowel was /u/ in /ə/+/u/, the 375-Hz channel was dominated by that F_0. If the rule was to choose the channel dominated by the target F_0 (as determined by the across-frequency average of the autocorrelation functions), then the choices made would, indeed, be those required for the $F1$ strategy. This rule is somewhat complicated, however, by the case of /ə/. Its $F1$ was at 500 Hz, which is a multiple of both 125 and 100 Hz. Thus, the autocorrelation offers no guidance as to its correct fundamental. To avoid this problem, it is possible that such a formant is always allocated to the target unless a better one (i.e., 625 or 375 Hz) is used, but it is clear that is a somewhat *ad hoc* exception. A general solution has not yet been found.

Overall, a single-channel model can give a distinction between the $\Delta F_0 = 25$- and 0-Hz cases, but only by varying the strategy for choosing the channel selected. Monaural autocorrelation analysis can guide the choice of channel. In essence, this strategy reflects monaural grouping followed by binaural processing (e.g., Hill and Darwin, 1996; Shackleton, 1998; Darwin and Hukin, 1999). This model, like the others, can qualitatively predict the effect of varying the level of the target vowel. Certain problems remain, however, especially the extreme predictions for some vowels and the degree to which they depend on the precise form of the ITD at the output of a channel.

The defining aspect of this model is that just one channel is used. It is quite possible that a ''middle-way'' approach, in which some subset of channels is used, taking information from other, less-good channels in addition to the best, offers a better model of the auditory processing. This approach has been shown to be fruitful for the problem of the recognition of double vowels (e.g., Meddis and Hewitt's, 1992, autocorrelation model, in which individual channels were segregated into groups). Furthermore, there are numerous studies showing across-frequency integration of binaural information (e.g., Jeffress, 1972; Trahiotis and Stern, 1989; Hill and Darwin, 1996; Trahiotis *et al.*, 2001). The questions of the factors that determine which subsets of channels supply the binaural information in simultaneous speech, and the degree to which the resulting groups are integrated across frequency, require further study.

IV. SUMMARY

Threshold differences for ITD were measured for the ''target'' vowel of three double-vowel pairs. The formant frequencies were crafted so that the frequency of the first formant of the target was either equal to, less than, or greater than that of the distracter. The intensity of the target vowel was either -6, 0, or $+6$ dB relative to the distracter. The fundamental frequency (F_0) of both vowels was either the same (i.e., $\Delta F_0 = 0$ Hz) or different ($\Delta F_0 = 25$ Hz).

The results showed that, in general, threshold ITDs were larger than when the vowels were presented alone. This effect mirrors that of ''binaural interference'' (e.g., Buell and Hafter, 1991). The size of the effect was dependent upon the choice of vowel pair, the choice of target vowel, the relative level of the target, and the difference in F_0. A qualitative

analysis based on the binaural ''dominance'' region around 600 Hz was considered and found to be in partial agreement with the data.

A multichannel version of Heller and Trahiotis' (1996) d' model was developed, incorporating the degree to which each frequency channel represented both the target and distracter vowels. It was found that the model was consistent with the overall effect of level and some of the detail of the choice of vowel pair or target, but it was unable to describe the effect of introducing an F_0 difference. A single-channel version of the model was explored. This model was qualitatively consistent with the F_0 effect, but only if the choice of channel was allowed to vary with changes in F_0. It was noted that this selection of channel could be informed by an autocorrelation analysis of monaural information. This model could describe other aspects of the effect of choice of vowel—although some values were extremely overpredicted—and some of the effects of differences in F_0. These two models represent opposite extremes of across-frequency analysis: compulsory combination of all channels vs selection of just one channel. It was speculated that the full solution will be found by considering some form of partial across-frequency analysis.

ACKNOWLEDGMENTS

This study was supported by an MRC Career Development Award and was conducted while the author was visiting the University of Connecticut Health Center. I thank Ron van Ness-Otunnu (supported by research grant numbers NIH DC-04073 and NIH DC-00234 from the NIDCD, NIH) for collecting the experimental data. The experimental code used the ''SIGLIB'' software package written by Dr. M. van der Heijden and the binaural modeling used the author's own software (Akeroyd, 2001). I also thank Les Bernstein, Chris Darwin, Klaus Hartung, Quentin Summerfield, and Tino Trahiotis for useful discussions and comments, and Trevor Shackleton and an anonymous reviewer for their insightful comments and suggestions during review.

[1]The $F1$ and $F2$ frequencies were (in the order /ə/, /e/, /ɔ/, /u/): $F1=500$, 500, 625, 375 Hz; $F2=1500$, 2000, 1000, and 1000 Hz. $F3$ was always 2500 Hz, $F4$ was always 3500 Hz, and $F5$ was always 3750 Hz. The /e/, /ɔ/, and /u/ vowels were based on Hillenbrand *et al.*'s (1995, Table V) values for /e/, /ɔ/, and /u/: $F1=476$, 652, 378 Hz; $F2=2089$, 997, 997 Hz; $F3=2691$, 2538, 2343; $F4=3649$; 3486; 3357 Hz. The values of $F1$ and $F2$ used here place each of the vowels near the lower-left corner of the ellipses plotted in Fig. 3 of Hillenbrand *et al.*

[2]All the calculations and means reported in this paper are arithmetic means of threshold differences or ITDs in $\log_{10}$ units, which are equivalent to geometric means of ratios or ITDs in microseconds.

[3]Taking the −6-dB condition as an example, the calculation proceeds thus: the overall level of the target and distracter were 58 and 64 dB, respectively, and so a_T and a_D were 794 and 1584 units, giving an amplitude ratio r of $794/(794+1584)=0.33$, and therefore a Δt_n of 1/6 (for a unitary value of ΔT_T), producing a final threshold ratio of 6.0.

[4]The following function was found to give a good fit to Zwislocki and Feldman's (1956, Fig. 8) data: $\Delta T_n = 0.000\,047\,47 f_n^2 - 0.0823 f_n + 47.688$.

[5]As a ΔF_0 of 25 Hz—approximately 4 semitones—is large by the standards of most studies on double-vowel *segregation*, the performance of both listeners and models on segregation will be expected to be effectively at ceiling (e.g., Meddis and Hewitt, 1992, Fig. 9; de Cheveigné *et al.*, 1997, Fig. 1).

[6]As the LOUDAES (Moore *et al.*, 1997) filterbank had channels spaced at 1-ERB intervals, there were no channels placed exactly at the three $F1$ frequencies of 375, 625, and 500 Hz. Instead, the d's at 375 Hz and 625 Hz were found by a linear interpolation of d's across the 362+429-Hz and 588+681-Hz channels, respectively. The 504-Hz channel was used directly for the 500-Hz formant, being sufficiently close to require no interpolation.

[7]A selection strategy based on *binaural* information was also considered. It chose the channel that gave the ''best'' ITD: that is, the channel that gave the largest (most-positive) ITD of the target. As the ITD at the output of a channel was assumed to be a monotonic function of the amplitude ratio [see Eq. (10) and Fig. 5], this rule was equivalent to choosing the channel with the largest amplitude ratio. The range of channels was limited to 360 to 900 Hz in order to avoid choosing extreme-frequency channels in which the binaural system is particularly insensitive to ITDs (see the ΔT_n function plotted in Fig. 4). When applied to both sets of conditions, the results showed, however, only a minor effect of varying ΔF_0, as the mean threshold ratios were 1.4 ($\Delta F_0=25$ Hz) and 1.5 ($\Delta F_0=0$ Hz).

Akeroyd, M. A. (**2001**). ''A binaural cross-correlogram toolbox for MATLAB,'' http://www.biols.susx.ac.uk/home/Michael_Akeroyd/download2.html (software package).

Assmann, P. F., and Summerfield, Q. (**1990**). ''Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **88**, 680–697.

Assmann, P. F., and Summerfield, Q. (**1994**). ''The contribution of waveform interactions to the perception of concurrent vowels,'' J. Acoust. Soc. Am. **95**, 471–484.

Bernstein, L. R., and Trahiotis, C. (**1982**). ''Detection of interaural delay in high-frequency noise,'' J. Acoust. Soc. Am. **71**, 147–152.

Bernstein, L. R., and Trahiotis, C. (**1995**). ''Binaural interference effects measured with masking-level difference and with ITD-and IID-discrimination paradigms,'' J. Acoust. Soc. Am. **98**, 155–163.

Bilsen, F. A., and Raatgever, J. (**1973**). ''Spectral dominance in binaural lateralization,'' Acustica **28**, 131–132.

Buell, T. N., and Hafter, E. R. (**1991**). ''Combination of binaural information across frequency bands,'' J. Acoust. Soc. Am. **90**, 1894–1900.

Culling, J. F., and Darwin, C. J. (**1994**). ''Perceptual and computational separation of simultaneous vowels: Cues arising from low-frequency beating,'' J. Acoust. Soc. Am. **95**, 1559–1569.

Culling, J. F., and Summerfield, Q. (**1995**). ''Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay,'' J. Acoust. Soc. Am. **98**, 785–797.

Darwin, C. J., and Hukin, R. W. (**1999**). ''Auditory objects of attention: The role of interaural time differences,'' J. Exp. Psychol. Hum. Percept. Perform. **25**, 617–629.

de Cheveigné, A., Kawahara, H., Tsuzaki, M., and Aikawa, K. (**1997**). ''Concurrent vowel identification. I. Effects of relative amplitude and F_0 difference,'' J. Acoust. Soc. Am. **101**, 2839–2847.

Dye, R. H. (**1990**). ''The combination of interaural information across frequencies: Lateralization on the basis of interaural delay,'' J. Acoust. Soc. Am. **88**, 2159–2170.

Heller, L. M., and Trahiotis, C. (**1995**). ''Interference in detection of interaural delay in a sinusoidally amplitude-modulated tone produced by a second, spectrally remote sinusoidally amplitude-modulated tone,'' J. Acoust. Soc. Am. **97**, 1808–1816.

Heller, L. M., and Trahiotis, C. (**1996**). ''Extents of laterality and binaural interference effects,'' J. Acoust. Soc. Am. **99**, 3632–3637.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

Hill, N. I., and Darwin, C. J. (**1996**). ''Lateralization of a perturbed harmonic: Effects of onset asynchrony and mistuning,'' J. Acoust. Soc. Am. **100**, 2352–2364.

Hukin, R. W., and Darwin, C. J. (**1995**). ''Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel,'' J. Acoust. Soc. Am. **98**, 1380–1387.

Jeffress, L. A. (**1972**). ''Binaural signal detection: Vector theory,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. II.

Klatt, D. H. (**1980**). ''Software for a cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 971–995.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Lieberman, P., and Blumstein, S. E. (**1988**). *Speech Physiology, Speech Perception, and Acoustic Phonetics* (Cambridge University Press, Cambridge).
McFadden, D., and Pasanen, E. G. (**1976**). ''Lateralization at high frequencies based on interaural time differences,'' J. Acoust. Soc. Am. **59**, 634–639.
Meddis, R., and Hewitt, M. J. (**1992**). ''Modeling the identification of concurrent vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **91**, 233–245.
Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). ''A model for the prediction of thresholds, loudness and partial loudness,'' J. Audio Eng. Soc. **45**, 224–240.
Oxenham, A. J., and Moore, B. C. J. (**1995**). ''Additivity of masking in normally hearing and hearing-impaired subjects,'' J. Acoust. Soc. Am. **98**, 1921–1934.
Patterson, R. D., Allerhand, M. H., and Giguere, C. (**1995**). ''Time-domain modeling of peripheral auditory processing: A model architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.
Raatgever, J. (**1980**). ''On the binaural processing of stimuli with different interaural phase relations,'' Doctoral dissertation, Delft University.
Saberi, K. (**1995**). ''Some considerations on the use of adaptive methods for estimating interaural delay thresholds,'' J. Acoust. Soc. Am. **98**, 1803–1806.
Scheffers, M. T. M. (**1983**). ''Sifting vowels: Auditory pitch analysis and sound segregation,'' Doctoral dissertation, Groningen.
Schiano, J. L., Trahiotis, C., and Bernstein, L. R. (**1986**). ''Lateralization of low-frequency tones and narrow bands of noise,'' J. Acoust. Soc. Am. **79**, 1563–1570.
Shackleton, T. M. (**1998**). Comment on ''Binaural mechanisms that emphasize consistent interaural timing information over frequency,'' by R. M. Stern and C. Trahiotis, *in Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), p. 394.
Shackleton, T. M., and Meddis, R. (**1992**). ''The role of interaural time difference and fundamental frequency difference in the identification of concurrent vowel pairs,'' J. Acoust. Soc. Am. **91**, 3579–3581.
Shackleton, T. M., Meddis, R., and Hewitt, M. J. (**1994**). ''The role of binaural and fundamental frequency difference cues in the identification of concurrently presented vowels,'' Q. J. Exp. Psychol. A **47**, 545–563.
Smith-Olinde, L., Koehnke, J., and Besing, J. (**1998**). ''Effects of sensorineural hearing loss on interaural discrimination and virtual localization,'' J. Acoust. Soc. Am. **103**, 2084–2099.
Stellmack, M. A. (**1994**). ''The reduction of binaural interference by the temporal nonoverlap of components,'' J. Acoust. Soc. Am. **96**, 1465–1470.
Stellmack, M. A., and Dye, R. H. (**1993**). ''The combination of interaural information across frequencies: The effects of number and spacing of components, onset asynchrony, and harmonicity,'' J. Acoust. Soc. Am. **93**, 2933–2947.
Stern, R. M., and Shear, G. D. (**1996**). ''Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay,'' J. Acoust. Soc. Am. **100**, 2278–2288.
Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (**1988**). ''Lateralization of complex binaural stimuli: A weighted-image model,'' J. Acoust. Soc. Am. **84**, 156–165.
Trahiotis, C., and Stern, R. M. (**1989**). ''Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and phase,'' J. Acoust. Soc. Am. **86**, 1285–1293.
Trahiotis, C., Bernstein, L. R., and Akeroyd, M. A. (**2001**). ''Manipulating the 'straightness' and 'curvature' of patterns of interaural cross correlation affects listeners' sensitivity to changes in interaural delay,'' J. Acoust. Soc. Am. **109**, 321–330.
Woods, W. S., and Colburn, H. S. (**1992**). ''Test of a model of auditory object formation using intensity and interaural time difference discrimination,'' J. Acoust. Soc. Am. **91**, 2894–2902.
Zwislocki, J., and Feldman, R. S. (**1956**). ''Just noticeable differences in dichotic phase,'' J. Acoust. Soc. Am. **28**, 860–864.

Perceptual segregation of competing speech sounds: The role of spatial location[a)]

Ward R. Drennan,[b)] Stuart Gatehouse, and Catherine Lever
MRC Institute of Hearing Research, Scottish Section, Glasgow Royal Infirmary—Queen Elizabeth Building, 16 Alexandra Parade, Glasgow G31 2ER, Scotland

(Received 25 August 2001; revised 21 July 2003; accepted 22 July 2003)

Culling and Summerfield [J. Acoust Soc. Am. **92**, 785–797 (1995)] showed that listeners could not use ongoing interaural time differences (ITDs) to achieve source segregation. The present experiments tested a free-field analog of their experiment. The stimuli consisted of narrow bands of noise, pairs of which represented the first and second formants of the whispered vowels ''ar,'' ''ee,'' ''er,'' and ''oo.'' A target noise-band pair (vowel) was presented at various angles on the listeners' left while a complementary distracter was presented on the listeners' right. Listeners correctly identified the target vowel in the free-field well above chance. Performance remained well above chance in headphone experiments that retained spatial cues but eliminated reverberations and head movements. The full range of cues that normally determine perceived spatial location provided sufficient information for segregation. Further experiments, which systematically evaluated the contribution of these cues in isolation and in combination, showed that some listeners, following training, exhibited the ability to segregate based on ongoing ITDs alone. Substantial individual differences were observed. The results show that listeners can use spatial cues to segregate simultaneous sound sources. [DOI: 10.1121/1.1609994]

PACS numbers: 43.66.Pn, 43.66.Rq [LRB]

I. INTRODUCTION

In a typical listening environment, there are multiple sounds which listeners perceive as independent acoustic entities. The acoustic energy emitted from the sources and their corresponding neural representations overlap in time and frequency. The auditory system must then use the available information to group the components into discrete objects. Successful grouping of overlapping auditory stimuli into discrete objects is known as perceptual segregation. Understanding how the auditory system segregates objects is important in addressing the problems of the hearing impaired and users of auditory prostheses who, for various reasons, have difficulty understanding speech in noisy environments. There are a number of cues known to contribute to segregation including harmonicity, across-frequency comodulation, and onset asynchrony. Two reviews that have addressed these factors are Yost and Sheft (1993) and Moore (1997, pp. 259–269).

Harmonicity is perhaps the most powerful cue for source segregation. Tones with a common harmonic relation are readily heard as a group. Broadbent and Ladefoged (1957) observed that a set of simultaneous components with two fundamental frequencies is heard as two objects. Moore *et al.* (1986) observed that mistuned harmonics stand out as separate acoustic entities. Darwin and Gardner (1986) noted that a harmonic with a frequency that is an exact multiple of the fundamental frequency influences the timbre of a vowel significantly more than does a mistuned harmonic.

Synchronized onsets of multiple components in a complex sound are also important in object formation. Bregman and Pinker (1978) have shown that nonsimultaneous onsets result in the segregation of specific components which have an influence on timbre. Additionally, a number of studies have shown that nonsimultaneous onsets play a role in sound source segregation (e.g., Rasch, 1978; Darwin, 1984; Roberts and Moore, 1991; Darwin and Ciocca, 1992; Hukin and Darwin, 1995). Coherent amplitude modulation is a third cue that has been shown to contribute to source segregation (Kubovy and Daniel, 1983; Hall *et al.*, 1984; Hall and Grose, 1990).

Harmonicity, simultaneous onsets, and comodulation operate monaurally. Humans have a strong subjective sense that sounds emanating from different locations are different objects, and people perceive that they are attending selectively to different spatial locations. Based upon introspection, this suggests that spatial cues might be used to segregate sources. The hypothesis tested in the following study is that the cues that accompany spatial location can be grouped across frequency. These cues will then facilitate source segregation. The cues that accompany spatial position include interaural intensity differences (IIDs), interaural timing differences (ITDs), and spectral-shape cues, predominately above 4 kHz resulting from the acoustics of the pinna. Middlebrooks and Green (1991) provide an excellent review of the cues that accompany spatial location.

Previous research has shown that infinite IIDs contribute to simultaneous source segregation, but ITDs do not. Culling and Summerfield (1995) showed that infinite IIDs lead to segregation. Infinite IIDs, however, are monaural presentations, which differ categorically from naturally occurring IIDs. Natural IIDs are at most 25 dB for 90° azimuth at 6.4 and 12 kHz and less for smaller angles and lower frequencies. IIDs for frequencies at or below 3 kHz do not exceed 15

[a)]Data from this study were presented at the 138th meeting of the Acoustical Society of America, November 3, 1999 in Columbus, OH.

[b)]Present address: Kresge Hearing Research Institute, Ann Arbor, MI. Electronic mail: warddrennan@yahoo.com

 0001-4966/2003/114(4)/2178/12/$19.00

dB for distant sources (Shaw, 1974). Culling and Summerfield further observed that ongoing ITDs do not contribute to segregation. Ongoing ITDs consist of interaural time differences in the fine-structure and envelope of acoustic waves in the absence of onset ITDs. Onset ITDs always accompany ongoing ITDs in the natural environment.

Hukin and Darwin (1995) demonstrated that ITDs (inclusive of onset and ongoing ITDs), when included in a single tone among a complex of tones representing a vowel, do not contribute to the vowel's timbre and hence appear insignificant in source segregation. Dye (1990), Hill and Darwin (1996), and Darwin and Hukin (1997) also showed that ITDs, on their own, do not contribute to segregation.

ITDs, however, are known to be the dominant cue for determining the azimuth of a sound source when low-frequency information is present (Wightman and Kistler, 1992; Culling *et al.*, 1994). Given that ITDs are critical for localization and given that people have a strong subjective sense that spatial information contributes to source segregation, the fact that ITDs do not appear to play a role in segregation is perplexing.

A model introduced by Woods and Colburn (1992) can be used to explain this observation. The model assumes that object formation occurs prior to the assignment of perceptual attributes such as lateral position. If object formation occurs prior to the assignment of azimuth, then object formation and localization can be independent. ITDs need not necessarily play a role in segregation even if they were dominant for localization. Nevertheless, if spatial cues contribute to source segregation, as introspection might suggest, then object formation and the assignment of lateral position could be concurrent and parallel processes.

The role of naturally occurring combinations of ITDs and IIDs in segregation was not addressed in the aforementioned studies. Previous works addressed the effects of limited cues such as ITDs, ongoing ITDs, or infinite IIDs. Gaik (1993) has shown that psychophysical response patterns differ between natural and unnatural combinations of ITDs and IIDs. Therefore, the use of naturally occurring combinations of ITDs and IIDs might yield different results.

The present experiments investigate the role of realistic spatial cues in across-frequency grouping and subsequent source segregation in the absence of other strong monaural grouping cues. The eliminated cues include harmonicity, comodulation, and onset asynchronies. The monaural onset asynchronies eliminated in these experiments are on the order of 100 ms, several orders of magnitude larger than the onset asynchronies observed with spatial separation (hundreds of microseconds). The paradigm of Culling and Summerfield (1995) is adopted first in a free-field setting and then in a headphone setting in which the roles of IIDs and of onset and ongoing ITDs are investigated independently.

II. EXPERIMENT 1: FREE FIELD

A. Purpose

The objective of this experiment was to determine if the cues that accompany spatial location in free-field listening can be used to segregate sound sources in the absence of other strong cues for segregation.

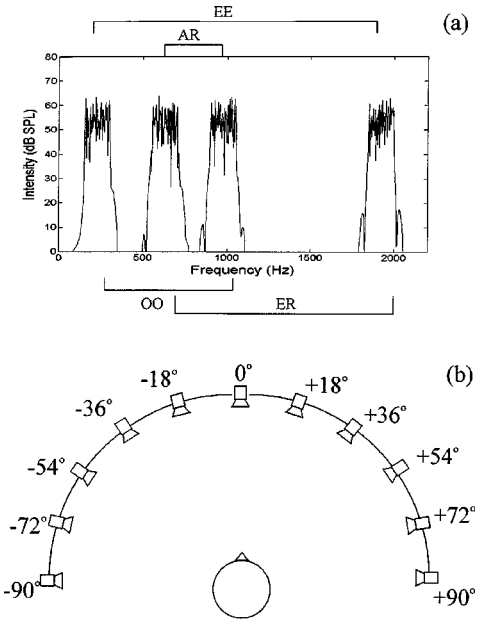


FIG. 1. (a) Spectrum of four narrow bands of noise, 150 Hz wide. The bands were combined in pairs to create the ''vowels.'' (b) Schematic of the dome for free-field listening. The stimuli were presented over speakers mounted within the dome. The listener sat at the center of the dome facing the middle speaker as shown.

B. Listeners

Nine normal-hearing listeners with pure-tone sensitivity better than 20 dB HL between 250 and 8000 Hz completed the study. Their ages ranged from 18 to 49 years. They were paid to participate.

C. Method

The stimuli were identical to those used by Culling and Summerfield (1995). Four ''vowels'' were created. Each ''vowel'' consisted of a pair of narrow-band noises selected from a set of four. The noises represented the first and second formants of the vowels /ɑ/, /i/, /ɛr/, and /u/. These stimuli were not vowels as heard in normal daily discourse but a representation of whispered, voiceless vowels which could be labeled appropriately.

Figure 1(a) shows how the four narrow-band noises were combined to create the whispered ''vowels.'' The narrow-band noises were 150 Hz wide and centered on 225, 625, 975, and 1925 Hz. The duration of the stimuli was 400 ms including 10-ms linear rise/fall ramps. The noise bands were presented as complementary pairs such that all four bands were played simultaneously: two bands from the listeners' left and two bands from the listeners' right. The listeners were asked to identify the ''vowel'' heard on the left. Ten different versions of each vowel were created using ten sets of independent narrow-band noises. Each vowel was selected randomly from ten exemplars such that the vowel on the right did not have a consistent match with the vowel on the left. These ten exemplars were used to reduce the probability that listeners could use fine-structure detail, unique envelope characteristics, or within-band spectral detail to discriminate the vowels. The results of Siegel and Colburn (1989) and Isabelle and Colburn (1995) suggest that ten ex-

emplars are sufficient so that subtle details of frozen-noise waveforms are not learned and do not materially affect results in binaural detection tasks.

The stimuli were presented from speakers mounted in a 1.15-m-radius hemisphere as shown in Fig. 1(b). The hemisphere was covered with sound-absorbing foam and placed against the back wall of a sound-treated chamber approximately 2.5 m long by 2.6 m wide by 2.5 m high. Reverberation times ranged from 0.16 to 0.26 s, depending on the speaker. IIDs measured with a head and torso simulator were within ±3 dB of the averaged data of Shaw (1974) and ITDs were within 90 μs of the data reported by Kuhn (1977). The details of the reverberation characteristics and the methods used to determine them were reported by Lorenzi *et al.* (1999). The environment has similar reverberation characteristics to rooms in which people normally listen.

The speakers were at 0° elevation and 18° apart at azimuths ranging from −90° to +90°. Measured at the center of the hemisphere, the level of the stimuli was 70 dB SPL, and the frequency response of the speakers was flat within 1.5 dB SPL from 150 Hz to 20 kHz. For white noise at 70 dB SPL, the intensity across speakers was constant within 0.5 dB.

The listeners sat at the center of the hemisphere equidistant from all of the speakers. They were asked not to move their heads and to continuously face the speaker at 0°. The listeners responded using a touch sensitive computer screen located in front of them just above waist level. They chose one of the four ''vowels'' presented on the screen as ''ar,'' ''ee,'' ''er,'' and ''oo.'' The task was difficult. The vowel representations were not natural, and correct identification was learned during a training period in which listeners were presented example stimuli without distracters and had knowledge of their identity.

There were two phases of training. During the first phase of training, the four single ''vowels'' (example stimuli) were presented sequentially without distracters from the left side at −90° azimuth. The four example stimuli were presented in order (''ar,'' ''ee,'' ''er,'' ''oo'') and repeated five times. All ten exemplars were used and randomly selected. With knowledge of the order of presentation, the listeners learned which noise-band pair corresponded to which single vowel. No responses were given during the presentation of the example stimuli. To test their performance, listeners were then given the four-alternative forced-choice task with 80 trials. On each trial, a single vowel was presented on the left (−90°) *without* distracters. No feedback was given. The sequence of 20 examples followed by an 80-trial-block, four-alternative forced-choice task was repeated for at least an hour and until each listener achieved at least 75% correct. The total time required to achieve 75% correct ranged from 1 to 3 h. The total number of examples heard ranged from about 180 to 540, and the total number of forced-choice trials ranged from about 720 to 2160, depending on the listener. Some of the better listeners achieved more than 95% correct in 1 h. Some listeners confused specific pairs of ''vowels.'' To assist them in learning the difference, five repetitions of the confused single vowels were presented in alternation without distracters. Those listeners were then given a two-alternative forced-choice task using only the confused pair. The procedure was repeated (five examples followed by a forced-choice task) until they achieved 75% correct or more with just the pair. The total number of trials necessary ranged from 40 to several hundred, depending on the listener. After learning to identify the pair, testing returned to the sequential presentation of all four examples five times and the four-alternative forced choice task.

In the second phase of training, distracters were added. The listeners heard the example stimuli presented once in order (''ar,'' ''ee,'' ''er,'' ''oo''). For these example stimuli, the target vowel was presented at −90° accompanied by its complement at +90°. Then, the listeners were given a four-alternative forced-choice task with distracters and no feedback. Each trial block consisted of 80 trials. The listeners were asked to focus their attention on the ''vowel'' heard on the left and therefore were aware that the vowel at −90° azimuth was the target. They had the option of hearing the four example stimuli played once prior to each trial block. Most (the better) listeners requested them before only a few trial blocks. A few of the poorer listeners requested them before several trial blocks. A few listeners had difficulty with specific pairs of vowels. They were given a two-alternative, forced-choice task in which they discriminated only those two vowels until their scores appeared to be near asymptotic levels. No feedback was provided. They then continued with the four-alternative-forced choice task to complete the second phase of training. Overall, the number of examples presented in the first training phase amounted to hundreds, while in the second phase only tens.

After training, testing proceeded in 200-trial blocks. In one day of data collection, listeners heard every target presented at every angle 50 times totaling 1200 trials per day. No trial-to-trial feedback was given and no examples were presented during the final testing. The procedure was repeated on a second day so that, in total, listeners heard each ''vowel'' at each angle 100 times. Typically, the listeners participated 1 day per week and required 3–4 days of testing. The entire testing period including training spanned 3–4 weeks.

D. Results

Figure 2 shows the results of all listeners. Test-retest reliability was good (within 5%–10%); therefore, performance is averaged over the two repetitions for each listener. Each point includes data from 400 trials (100 presentations of each vowel). The dotted line indicates the upper bound of the 95% confidence interval for chance performance. The calculation of the confidence interval for chance levels is based on a one-tailed binomial probability distribution.

Although performance varies among listeners, it is clear that all nine listeners have the ability to segregate speech sounds when the cues that accompany spatial location are present. A repeated-measures analysis of the within-listeners effect of angle showed that the ability to segregate declines as the angle between sources decreases ($F_5 = 25.275$, $p < 0.0005$).

L7's performance in the 0° condition was better than chance. This, in theory, should not be possible, so some information that the authors did not intend to provide must

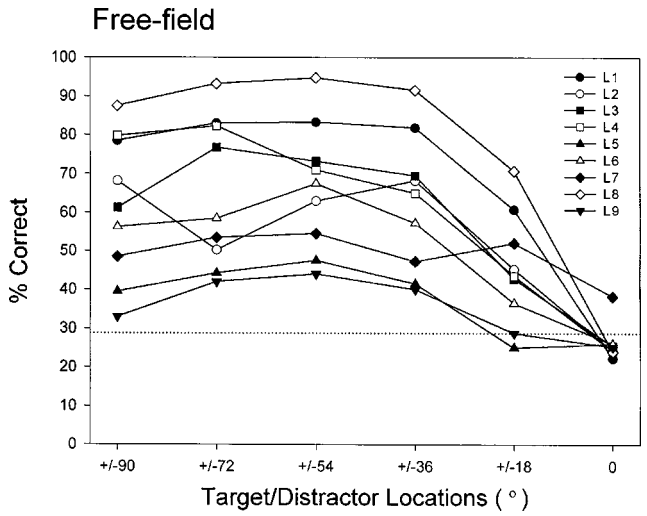


FIG. 2. Combined data for free-field listening, experiment 1. Each data point represents data from 400 trials including the four ''vowels'' presented 100 times over 2 days. Differences in percent correct equal to or greater than 10% are significantly different at the 5% level. The dotted line shows the upper bound of the 95% confidence interval for chance performance.

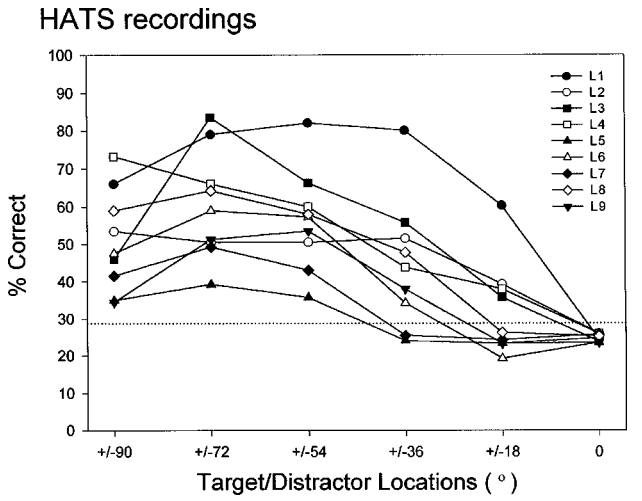


FIG. 3. Data for the HATS recordings, experiment 2. The dotted line shows the upper bound of the 95% confidence interval for chance performance.

have been available. The reason for the aberration is not known. Perhaps ten exemplars were not sufficient to eliminate this listener's ability to identify stimuli using within-channel spectral details. The aberration was observed for one listener only, and only in the free field. In the numerous subsequent headphone experiments, there were no aberrations at 0°.

Experiment 1 showed clearly that listeners can use the cues that accompany spatial position to achieve segregation.

III. EXPERIMENT 2: HEAD AND TORSO SIMULATOR RECORDINGS

A. Purpose

The listeners in experiment 1 were instructed to face forward towards the speaker at 0° and asked not to move their heads; however, no restraints were placed on their heads and no formal monitoring was employed. Listeners could have used head movements to facilitate segregation. The purpose of the second experiment was to determine if listeners could segregate without access to cues that may have been provided by head movements.

B. Method

All of the stimuli were recorded digitally using a Brüel and Kjær head and torso simulator type 4100 D (HATS) positioned at the center of the hemisphere, equidistant from all 11 speakers, just as if it were a human listener in experiment 1. The HATS contained Brüel and Kjær microphones type 4189 connected to a Brüel and Kjær preamplifier type 2671. The preamplifier was connected in series with a Brüel and Kjær sound quality conditioning amplifier type 2672, a 40-Hz cutoff high-pass filter, and a DASCAM DA-20 MK II DAT recorder. The recordings were made at a sampling rate of 44.1 kHz. The digitized waveform files were presented over Beyerdynamic DT 48 headphones using a TDT DA3-4 digital to analog converter. The stimuli were presented at 70 dB SPL. The listeners repeated all conditions (2400 trials), identifying the ''vowel'' heard on the left.

The head-related transfer function (HRTF) of the HATS differs from that of the listeners in the study. The exact replication of the HRTFs is not critical, because the ongoing and onset ITDs and IIDs of the HATS are sufficient to elicit lateralization of the stimuli. The spectral details of the HRTFs will have little or no influence in this experiment because there are ten independent exemplars of the noise which are all narrow band. High-frequency (>4 kHz) pinna cues are not available because the narrow-band noises did not contain energy above 2000 Hz.

C. Results

Comparison of Figs. 2 and 3 shows that on average the nine listeners performed slightly worse in the headphone condition than in the corresponding condition in experiment 1, particularly at ±90°. Performance with the HATS was significantly worse ($t_8 < 2.525$, $p < 0.036$) than in the free field at each angle except 72° and 0°. Poorer performance in the HATS condition might be a result of removing head movements, a result of differences in HRTFs between the listener and the HATS or a combination of the two. Clearly, however, performance in the HATS-recordings condition is well above chance, and, therefore, the listeners are successfully segregating without head movements.

IV. EXPERIMENT 3: VIRTUAL STIMULI

A. Purpose

Numerous spatial cues could contribute to listeners' abilities to segregate the vowels. These include ongoing ITDs, the corresponding onset ITDs, and IIDs. These spatial cues are frequency dependent (Kuhn, 1987; Shaw, 1974). Culling and Summerfield (1995) investigated the role of limited sets of frequency-dependent spatial cues in segregation which were not wholly representative of what a listener might actually experience. One purpose of these experiments was to investigate the ability of listeners to segregate the

TABLE I. (a) The delays in microseconds applied to each noise band for each angle. The delays were applied to pairs of noise bands, leading or lagging, depending on the desired direction of the source. (b) The attenuation in dB applied to each noise band for each angle. The attenuation was applied to pairs of noise bands independently simulating head-shadow effects in the free field.

	Angle				
Frequency	18	36	54	72	90
	(a) Delays in μs				
225	250	480	640	760	800
625	230	410	590	680	710
975	170	320	560	640	680
1925	180	280	420	600	650
	(b) Attenuation in dB				
225	1	2	2.5	3	3.5
625	3	6	7	7	7
975	4	7	9	8	7
1925	4	7	9	10	9

vowels using more natural spatial cues. The primary purpose of experiment 3 was to disassociate the cues, evaluating the contribution of each cue to source segregation and determining the manner or manners in which the cues interact. Also, previous studies (e.g., Culling and Summerfield, 1995; Darwin and Hukin, 1998) had employed ongoing ITDs with the other cues, e.g., onset ITDs and IIDs, set to 0. The implementation of similar conditions in experiment 3 would foster valid comparisons among the studies.

Additionally, reverberations in the experimental room might have negatively affected the ability of listeners to segregate in experiment 2. The reverberation characteristics of the room were described in detail by Lorenzi *et al.* (1999). These reverberations result in multiple ITDs and IIDs from the source and reflections. Thus, the listeners might have been influenced in experiment 2 with patterns of ITDs and IIDs that were ambiguous with respect to the actual spatial positions of the noise bands. An additional purpose of experiment 3 was to eliminate the complexities arising from reverberations.

B. Method

Six conditions were executed using virtual stimuli. Each examined a set of one or more cues which contributes to the perception of spatial location. IIDs and ITDs were carefully estimated from previously published data collected from KEMAR recordings in an anechoic chamber. The estimates of IID and ITD for each angle were read off figures published by Kuhn (1987) for ITDs and by Shaw (1974) for IIDs. This approach yielded realistic, frequency-dependent values of IIDs and ITDs, which were applied to the noise bands, recreating a free-field experiment without the reverberation effects observed in the HATS recordings. The IIDs and ITDs used are shown in Tables I(a) and I(b), respectively. The ITDs used here are different from the Culling and Summerfield (1995) stimuli in three ways: (1) onset ITDs were present in some conditions, (2) some ITDs for angles of 72° or more were larger than those used by Culling and Summerfield, and (3) the ITDs are frequency dependent in a manner consistent with natural spatial cues. Subsequent informal listening by one of the authors (L4) confirmed that the noise bands (either in isolation or in appropriate pairs) were perceived at the expected lateral position.

The experiment was executed in the same manner as the previous experiment. Listeners heard 100 trials of each ''vowel'' at each angle. No feedback was provided. Listeners repeated the tasks as in experiments 1 and 2, hearing a total of 2400 trials (100 trials×4 vowels×6 angles) for each condition.

The following conditions were executed:

(1) Condition A (IIDs and ITDs): The appropriate attenuation and delays were applied to the noise bands for each ear and for each vowel separately (Table I). The procedure created stimuli for a headphone experiment containing ITDs and IIDs similar to those that were available in experiment 2 but devoid of reverberations.
(2) Condition B (ITDs only): ITDs were applied to the stimuli as in condition A, but IIDs were set to 0 dB.
(3) Condition C (IIDs only): IIDs were applied to the stimuli as in condition A, but no interaural delays were incorporated into the stimuli. All ITDs were 0 μs.
(4) Condition D (ongoing ITDs only): The interaural delays were administered as in condition A, but the envelopes were adjusted so that onset ITDs were 0 μs. An envelope with 10-ms linear rise/fall ramps and a 380-ms flat period (with magnitude equal to 1) was multiplied by a sample of the narrow-band noise with a duration greater than 400 ms. For the contralateral ear, the noise was shifted according to the ITDs reported by Kuhn (1987) and then multiplied by the same envelope. Onset ITDs were all 0 μs and IIDs were 0 dB.
(5) Condition E (onset ITDs only): Ongoing ITDs and IIDs were set to 0 and the only usable cues were onset ITDs. The same multiplication technique was used as in condition D, except that, for the contralateral ear, the envelope described above was shifted in time while the temporal fine structure was not shifted.
(6) Condition F (Culling and Summerfield): the same stimuli that were used by Culling and Summerfield (1995) were employed. These stimuli had ongoing ITDs of 390 μs. Thus, they were similar to the stimuli used in condition D, except that the ongoing ITDs were not dependent upon a specific angle or frequency. The listeners heard the ''vowels'' 300 times each, totaling 1200 trials on a single day. The value of 300 was chosen so that the duration of testing in the last condition matched those of the earlier conditions. Larger *N*'s also improve statistical power.

Experiment 3 required 11 days to complete. Listeners participated for 1.5 to 2 h each day including 30 min of break time.

C. Results

1. IIDs and ITDs

The results of condition A (IIDs and ITDs) are shown in Fig. 4(a). All nine listeners use the cues that accompany spa-

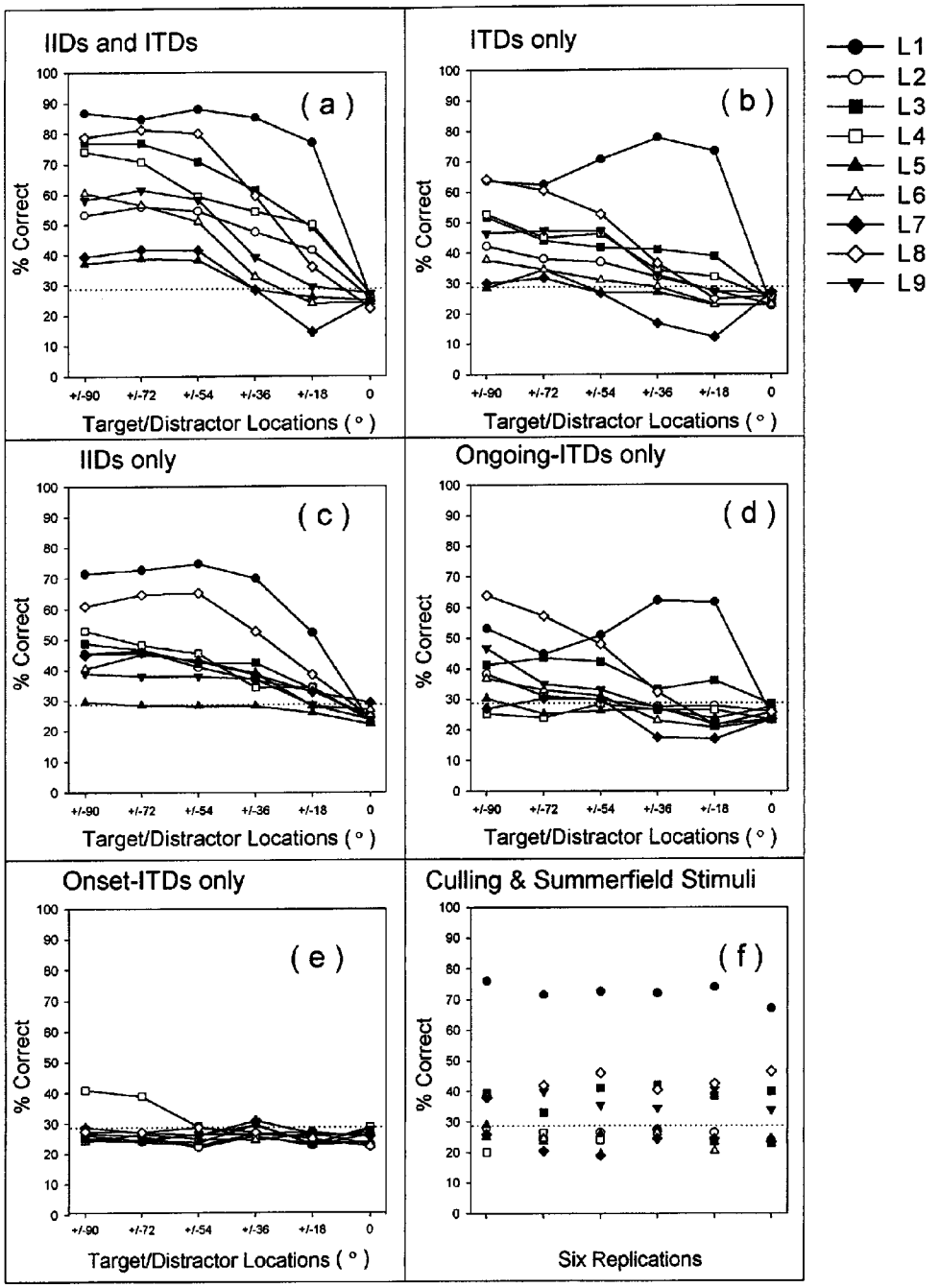


FIG. 4. Data from experiment 3. Panel (a) IIDs and ITDs, virtual stimuli recreating spatial cues in an anechoic environment based on data from Kuhn (1987) and Shaw (1974). Panel (b) ITDs with no IIDs. Panel (c) IIDs with no ITDs. Panel (d) Fine-structure ITDs with no onset asynchrony and no IIDs. Panel (e) Onset asynchrony only. Panel (f) Six repetitions with Culling and Summerfield's stimuli. The dotted lines show the upper bound of the 95% confidence intervals for chance performance.

tial location to segregate the stimuli. A repeated-measures ANOVA showed a significant effect of angle ($F_{5,40} = 28.768$, $p < 0.0005$). Large individual differences are evident in these data. When IIDs and ITDs were presented, performance ranged from 37.25% to 86.75% correct when the stimuli were at $\pm 90°$. Ninety-five percent confidence intervals for these percent correct values, based on binomial probabilities, were always less than $\pm 5\%$.

A paired *t*-test showed that performance at 90° in this experiment was superior to performance at 90° in experiment 2 ($t_8 = 2.961$, $p = 0.0182$). The difference is probably a result of acoustic reflections within the dome. At 90°, a strong first reflection was likely, either off the hemispheric dome, treated with foam that was marginally effective, or off the walls, which were treated with a thick curtain, also marginally effective.

2. Reduced-cue conditions

Figures 4(b) and (c) show the results of conditions B (ITDs only) and C (IIDs only), respectively. Eight out of nine listeners performed above chance levels in the IIDs-only condition, and seven out of nine listeners performed above chance levels in the ITDs-only condition. Chance levels were

determined according to binomial probabilities given 400 independent forced choice trials in each condition. The results show that the majority of listeners were able to use IIDs alone or ITDs alone to segregate. Performance, however, was worse in these conditions compared with condition A (IIDs and ITDs). Averaging over the nonzero angles, performance was 10.7% and 14.6% worse for the IIDs-only and ITDs-only conditions, respectively. A two-way, repeated measures ANOVA (3 conditions: IIDs and ITDs, ITDs only, and IIDs only by six angles) showed a significant effect of condition ($F_{2,16}=17.875$, $p<0.0005$) and of angle ($F_{5,40}=23.132$, $p<0.0005$).

Performance for the ongoing-ITDs condition (condition D) is shown in Fig. 4(d). Performance in this condition averaged 5.9% worse than in the ITDs condition (condition B): 40.1% vs. 34.2% over all nonzero angles. A two-way, repeated measures ANOVA (2 conditions: ITDs-only and ongoing-ITDs only and six angles) showed significant differences between the two conditions ($F_{1,8}=8.429$, $p=0.020$) and among angles ($F_{5,40}=8.032$, $p<0.0005$). Six listeners (L1, L2, L3, L6, L8, and L9) were segregating (above chance) at the three largest angles with ongoing ITDs only.

Figure 4(f) (condition F) shows data from the Culling and Summerfield condition. Four listeners out of nine could segregate above chance levels, one to a great extent (L1) and three others to a lesser extent (L3, L8, and L9). The results for L1, L3, and L8 match closely between condition D at 36° (the condition at which the ongoing-ITDs are nearest to the Culling and Summerfield stimuli) and condition F. L9's performance is chance at 36° for the ongoing-ITDs condition but above chance, although variable, in the Culling and Summerfield condition. Performance is at chance levels for the onset-ITDs-only condition [Fig. 4(e), condition E] except for L4 who is marginally above chance at the two largest angles.

L1 (the third author) displayed a strange pattern of results. In the ITDs-only condition and the ongoing-ITDs condition, this listener's performance improved as the angle decreased. For all other listeners, performance decreased as a function of decreasing angle. This listener also had excellent performance in the Culling and Summerfield experiment. The Culling and Summerfield ongoing ITD of 390 μs corresponds to an angle of 18°–36°, depending on frequency. This is the same range of angles over which L1 has the best performance in the condition with ongoing ITDs only. L1 had several days of experience listening to the Culling and Summerfield stimuli prior to commencing this study. This listener might have learned to tune to the specific ITD, which was reflected in the data. Another possibility is that this listener was mislateralizing some bands (possibly the 625- or 975-Hz band) at the larger angles because of ambiguity caused by slipped cycles. With narrow-band continuous stimuli (especially sinusoids) the ear from which the signal leads is unclear. For example, a continuous narrow-band acoustic stimulus with a 1-ms period leading from the right by $\frac{1}{4}$ cycle (consistent with a source at a moderate azimuth angle on the right) would be indistinguishable from one lagging in the right ear by $\frac{3}{4}$ cycles (consistent with a source at a large azimuth angle on the left side). The ambiguity decreases with increasing bandwidth because slipped cycles occur at different delays dependent on frequency while the actual delays (nonslipped cycles) are independent of frequency. The 150-Hz-wide stimuli in these experiments are likely sufficiently wide to avoid such ambiguity.

TABLE II. (a) Confusion matrix for the IIDs-only condition, experiment 3B. The data show the percentage of each response. The data were combined over nine listeners and three angles, ±90°, 72°, and 54°. (b) Confusion matrix for the ITDs-only condition, experiment 3C. (c) Confusion matrix for the IIDs and ITDs condition, experiment 3A.

	Target				
Response	ar	ee	er	oo	Total
		(a) IID only			
ar	49.4	9.4	9.3	37.8	26.5
ee	6.9	53.3	36.4	6.1	25.7
er	13.2	27.1	44.7	12.0	24.2
oo	30.6	10.3	9.6	44.0	23.6
		(b) ITD only			
ar	44.2	8.5	24.7	12.1	22.4
ee	10.2	51.3	18.2	31.4	27.8
er	28.3	20.1	43.2	17.9	27.4
oo	17.3	20.1	13.9	38.6	22.4
		(c) IIDs and ITDs			
ar	62.0	4.1	9.7	14.4	22.6
ee	4.3	70.9	24.2	14.1	28.4
er	11.4	17.4	54.3	11.0	23.5
oo	22.2	7.6	11.8	60.5	25.5

The data for L7 were also often atypical. In the IIDs and ITDs, ITDs-only, and ongoing-ITDs-only conditions, this listener had above chance performance at large angles, but significantly below chance performance at small angles, except for 0 degrees where performance was at chance levels. This reflected an angle-specific bias, which only appears when ITDs are present. The result indicates this listener is sensitive to ITDs but is mislabeling the vowels when only relatively small ITD cues are present.

D. Confusions

1. Overall pattern of confusions

Distinct patterns of confusions were present in the data, especially in the reduced-cue conditions. These patterns are consistent with the duplex theory of Lord Rayleigh (1907). The work of Lord Rayleigh and numerous subsequent studies has shown that the effects of ITDs and IIDs on the lateral displacement of sound sources are frequency dependent. Given realistic ITD cues with IIDs set to 0, for example, the low-frequency stimuli (<1.5 kHz) are perceptually displaced more from the midline than high-frequency stimuli (>1.5 kHz). Given realistic IIDs with ITDs set to 0, high-frequency stimuli are perceptually displaced more from the midline than low-frequency stimuli. Blauert (1997, pp. 137–177) provides a review.

In the present experiments, successful determination of the side (left or right) of the 225- and 625-Hz bands is necessary to distinguish ''ee'' from ''er'' and ''ar'' from ''oo.'' In the IID-only condition, given that ITDs were 0 μs and ITDs dominate in the low-frequency region (Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002), the 225-,

TABLE III. Ratio of the probability of type 1 or type 2 confusions and percent correct. $P(\text{type1})/P(\text{C})$ was calculated for the IIDs and ITDs and IIDs-only condition and $P(\text{type2})/P(\text{C})$ was calculated in the other conditions. The probabilities are based on the observations at ±54°, 72°, and 90°. A ''*'' indicates overall performance was greater than 50%. A probability ratio with ''*'' indicates performance was between 25% and 50% correct and listeners were answering correctly more often than making the expected confusion. A probability ratio and ''-'' indicates that performance was between 25% and 50% and the probability of a correct response was not significantly different than the probability of the expected confusion. A ''-'' alone indicates chance performance.

	Listener								
Condition	L1	L2	L3	L4	L5	L6	L7	L8	L9
IIDs and ITDs	*	*	*	*	0.37*	*	0.84*	*	*
ITDs only	*	0.62*	0.64*	0.38*	1.06-	0.73*	0.77*	*	0.82*
ITDs only	*	1.03-	0.69*	0.35*	-	0.96-	1.06-	*	0.98-
Ongoing ITDs	*	0.94-	0.71*	-	-	0.64*	-	*	0.91-
Culling and Summerfield	*	-	0.87*	-	-	-	-	0.71*	0.86*

625-, and 975-Hz bands were probably perceived near the midline. The intracranial position of the 1925-Hz band would be expected to be further from the center given the magnitude of IIDs in that frequency region. Therefore, discrimination of the position of the 225- and 625-Hz bands is expected to be difficult, and ee/er and ar/oo (type 1) confusions were expected to be more extensive in the IID-only condition.

Table II(a) shows the confusion matrix for the IID-only condition. Performance is summed over all listeners and over angles ±54°, 72°, and 90°. These angles were chosen because performance was generally consistent and best for the larger angles. Each number shows the proportion of times the listener gave a particular answer for a particular target. As expected, in the IID-only case, the ee/er and ar/oo (type 1) confusions were more frequent than the ee/oo and ar/er (type 2) confusions (33.0% vs 9.7%).

Similarly, successful determination of the side of the 975- and 1925-Hz bands must be determined in order to distinguish ''ee'' from ''oo'' and ''ar'' from ''er.'' In the ITD-only condition, given that IIDs were 0 dB and that sensitivity to ongoing ITDs is considerably less in the frequency range of the 1925-Hz band (Zwisloski and Feldman, 1956), the intracranial position of the 1925-Hz band is expected to be nearer the midline. Therefore, discrimination of the position of the 1925-Hz band is expected to be difficult and ee/oo and ar/er (type 2) confusions are expected to be more extensive in the ITD-only condition. Table II(b) shows the confusions for the ITD-only condition were as expected (26.1% vs 16.9%).

The confusions for the IIDs and ITDs condition are shown in Table II(c). There was a trend in the IIDs and ITDs condition that the type 1 confusions were more common than the type 2 confusions (19.6% vs 10.7%). Type 1 confusions (ee/er and ar/oo) arise when the sides of the high frequency bands (975 and 1925 Hz) are discriminable, but the sides of the lower frequency bands are not. Because ITD cues are most important in the low-frequency region, the predominance of type 1 confusions when IIDs and ITDs are present suggest that ITDs play a relatively weak role in the segregation task. Thus, the observation suggests that IIDs dominate segregation ability when IIDs and ITDs are present. Zurek (1993) showed that in binaural intelligibility level difference (BILD) experiments, when stimuli are on opposite sides of the head, the S/N ratio at the better ear has a larger positive impact on performance than binaural interactions. In the present experiments, a S/N advantage exists at the left ear in all conditions containing nonzero IIDs. Thus, the IIDs in these experiments result in a binaural cue *and* a monaural cue—a S/N advantage at the left ear. This S/N ratio averages as much as 8 dB for pairs of noise bands. The only discernable cues ITDs yield are binaural. The apparent dominance of IIDs in segregation is probably due to the S/N advantage at the left ear.

The statistical significance of these confusion rates was analyzed using a multinomial logistic regression described by Agresti (1990, pp. 313–317) and by Long (1997). The regression was used to determine the ratio of confusion probabilities in each condition and their respective 95% confidence intervals (CIs). The three conditions discussed above were included in the multinomial logistic regression: ITD and IIDs, IID-only, and ITD-only. Using the methods described in the Appendix, the following conclusions were drawn:

(1) In the IID-only condition, type 1 confusions are 3.39 times more probable than type 2 confusions (95% CI: 3.16–3.63).
(2) In the ITD-only condition, type 2 confusions are 1.54 times more probable than type 1 confusions (95% CI: 1.46–1.64).
(3) In the IIDs and ITDs conditions, type 1 confusions are 1.83 times more probable than type 2 confusions (95% CI: 1.70–1.96).

2. Individual differences in confusions

It is possible that individual listeners could achieve 50% correct performance simply by lateralizing a single band. If this were the case, the listener would not be grouping across frequency, and a distinct pattern of confusions would emerge. His or her best possible performance would be 50% and the confusion pattern would indicate either that type 1 confusions were as common as correct responses, or that type 2 confusions were as common as correct responses, de-

pending on which band was lateralized. For example, in the IID-only condition, as discussed above, it is expected that the perceived intracranial position of the 1925-Hz band is further from the center than the lower bands. The position of the low bands in the IIDs-only case might be indeterminate to the listener. If a listener were sensitive only to the lateral position of the high band, ee/er and ar/oo (type 1) confusions would be as common as a correct response. If type 1 confusions were less common than a correct response, then the listener must be integrating information across frequency at least some of the time. Similarly, in the ITD-only condition, the perceived intracranial position of the low-frequency bands is expected to be further from the center than the high-frequency bands. The high-band position might be indeterminate. If listeners were only sensitive to the lateral position of the low-frequency bands, ee/oo and ar/er (type 2) confusions would be as common as correct responses.

Multinomial logistic regressions (described in the Appendix) were applied to individual data to determine whether the expected confusion was as common as a correct response. When performance was in the marginal region between chance levels and 50% correct, the logistic regressions were used to determine if the probability ratio of type 1 or type 2 confusions to percent correct is significantly different from one. In conditions A (IIDs and ITDs) and C (IID only), type 1 confusions were most prominent; therefore, $P(\text{Type1})/P(\text{C})$ was calculated. In all other conditions, type 2 confusions were more prominent; therefore, $P(\text{Type2})/P(\text{C})$ was calculated. Table III shows the probability ratio [$P(\text{Type1})/P(\text{C})$ or $P(\text{Type2})/P(\text{C})$] for each listener in each condition. A ''*'' alone indicates that correct responses occurred significantly more frequently than the expected type of confusion. A ''-'' alone indicates performance was at chance levels. For cases in which performance was in the marginal region of 25%–50%, the probability ratio is given with either a ''*'' or ''-''. A ''*'' indicates the confusions did not occur as frequently as correct responses; i.e., the null hypothesis that the probability ratio is equal to 1 could be rejected with at least 95% confidence. In this case, the listener was likely to be grouping the spatial information across frequencies to achieve segregation at least some of the time. A ''-'' indicates the expected confusions were as common as a correct response and the listener might have been (but was not necessarily) basing a decision on the lateral position of a single narrow-band noise.

Generally, there was not strong evidence that listeners were lateralizing a single band, except in the IID-only condition. Note that in the 45 observations (9 listeners$\times$5 conditions) of Table III, there were seven cases with evidence of the possible use of the strategy. Four of the seven cases are in the IID-only condition. The confusion patterns indicate that, in the IID-only condition, about half of the listeners could only glean segregation information from the 1925- or 975-Hz bands. In the ITD conditions, there was a trend to make confusions consistent with lateralizing the low-frequency band. All but one listener answered correctly more often then they made the expected confusion. Therefore, nearly all listeners integrated ITD information from multiple frequency regions.

V. DISCUSSION

This study used the paradigm of Culling and Summerfield (1995) to investigate the role of realistic spatial cues in source segregation. Culling and Summerfield showed that vowels with infinitely large IIDs could be segregated, but vowels with only ongoing ITDs could not be segregated. Culling and Summerfield also demonstrated that interaural decorrelation of the noise bands readily led to segregation. It appeared, though, that IIDs played a significant role in segregation and that ITDs did not. The present study has demonstrated that, with sufficient training, both IIDs and ITDs lead to segregation. This is true for realistic IIDs and ITDs presented in combination and separately. All listeners could segregate using the cues in combination. Eight out of nine listeners could segregate successfully when both ongoing and onset ITDs were present (Table III). Half of these listeners could use ongoing ITDs alone to segregate, and half could use IIDs alone to segregate.

The present results show that ITDs can play a role in segregation, but it appears that IIDs have a bigger impact. As noted in the results, IID (Type 1) confusions are more common when IIDs and ITDs are present, suggesting IID dominance. This observation bears some similarity to the results of Bronkhorst and Plomp (1988). They demonstrated that when speech was presented in front of the listener with the noise at azimuth angles from 60° to 120°, the difference in the speech reception threshold (SRT) of sentences in noise with spatial separation was 1–3 dB greater for IIDs only than for ITDs only. They further observed that the beneficial effect of ITDs is diminished in the presence of IIDs. These observations are consistent with the apparent dominance of IIDs in the present segregation experiments.

IIDs appear to dominate segregation using spatial cues, but the present studies have demonstrated that ITDs can contribute. Previous studies by Culling and Summerfield (1995), Darwin and Hukin (1997), and Darwin and Hukin (1999), however, showed that ITDs did not appear to play a role in segregation. The present stimuli were similar to, and in one case, the same as Culling and Summerfield's stimuli, so the question remains as to why there is a difference in the results. Four possible reasons for the differences are as follows:

(1) Naturally occurring, frequency-dependent *onset* ITDs were used in the present experiments but not in the previous experiments.
(2) The present stimuli had more realistic, frequency-dependent *ongoing* ITDs. A few of the ITDs at 72° and 90° were also larger than those used by Culling and Summerfield (maximum 700 μs).
(3) Listeners had significantly more experience listening to the stimuli (30 h) in these experiments than in previous experiments.
(4) Individual sensitivities to ITDs were different among listeners from the different studies.

The data in the ongoing ITD and onset and ongoing ITD conditions are consistent with the ''central spectrum'' model presented by Raatgever and Bilsen (1986). The central spec-

TABLE IV. Results from multinomial logistic regression for confusion matrices including data from all listeners and angles ±54°, 72°, and 90°.

Condition	B	StdErr	Wald	df	Sig.	e^B	Low	High
IID (B1)	−1.221	0.035	1209.4	1	<0.0005	0.294	0.275	0.316
ITD (B2)	1.657	0.046	1285.2	1	<0.0005	5.241	4.787	5.738
All (B3)	0.619	0.051	149.2	1	<0.0005	1.858	1.682	2.052

trum model is based upon the binaural coincidence detector first proposed by Jeffress (1948). In the central spectrum model, power is represented as a function of frequency and interaural delay in a three-dimensional plot showing peaks or ridges matching the ITD of an acoustic object. For example, if a Culling and Summerfield ''ee'' with its distracter ''ar'' were presented to the model, four short ridges would appear in the central spectrum. Two ridges would lie along a line representing an ITD of 390 μs at the frequencies for the 150-Hz-wide noise bands centered on 225 and 1925 Hz. Another two short ridges would lie along a line at −390 μs centered on 625 and 975 Hz. The auditory system, given the central spectrum, could ''look'' across lines of equal ITD and use this information to segregate.

The central spectrum model assumes a high degree of frequency selectivity and ITD selectivity. The auditory system can extract ITD information across the spectrum (McFadden and Pasanen, 1976; Bernstein and Trahiotis, 1982), but the degree of sensitivity is not consistent across the spectrum. Thresholds for ITD sensitivity in the low-frequency region (Klumpp and Eady, 1956; Zwisloski and Feldman, 1956) are higher than thresholds in the high-frequency region (Henning, 1974; McFadden and Pasanen, 1976; Nuetzel and Hafter, 1981; Bernstein and Trahiotis, 1994). In the low-frequency region (<1.5 kHz), fine-structure, ongoing ITDs are easily detectable. In the high-frequency region (>1.5 kHz), envelope ITDs have less potency. Additionally, the degree of lateralization invoked by ITDs in narrow-band noises 1000 Hz and below is considerably greater than the degree of lateralization invoked at 2000 Hz and above (Trahiotis and Bernstein, 1986). These data suggest that the degree of perceptual separation perceived with ITDs only for the 1925-Hz band (150 Hz wide) will be less than the perceptual separation of the three lower-frequency bands. These observations are consistent with the patterns of confusions observed in the present experiments. The integration of ITD information across low and high frequencies might be compromised because of the variable perceptual potency of ongoing and envelope ITDs across frequency. This could contribute to the weakness of ITDs in source segregation.

The results presented here demonstrate that ITDs can be grouped across frequency. Some listeners appeared to successfully segregate the vowels with ongoing ITD cues (4/9) and with ongoing and onset ITDs (8/9). It should be noted that it is possible that this task could be done without segregating the vowels but by learning a spatial-frequency image unique to each possible vowel plus distracter combination. Above chance performance could be accomplished by separating the individual narrow-band noises rather than by segregating a learned vowel. If such a strategy were used in the ongoing ITD condition, across-frequency grouping (AFG) of ITDs might not be taking place. This approach is unlikely primarily because of the training procedure. In the first phase of training, the listeners heard hundreds of examples of single vowels *without* distracters and, after achieving 75% correct and sometimes up to 95% correct, they heard only a handful of examples *with* distracters. The listeners would have to relearn an already well-learned mapping with only a handful of examples and a limited amount of time. Additionally, testing took place over several weeks, so the listeners would have to relearn and remember a new mapping over several weeks without feedback. The listeners would also have to ignore the instructions to listen to the left. Therefore, it is likely that the listeners who were above chance in the ongoing ITD condition achieved that level of performance by grouping ITDs across frequency.

Object formation, in the experiments presented here, is dependent on binaural cues. Therefore, the results are not consistent with the theory that object formation occurs prior to the determination of location. This theory was incorporated into a model by Woods and Colburn (1992) and discussed by Buell and Hafter (1991). The monaural cues for segregation of natural, harmonic, and comodulated stimuli (e.g., speech) are probably sufficiently strong that segregation can be achieved easily with only monaural cues. If this is the case, binaural cues might be unnecessary or even irrelevant. The auditory system, however, tends to make use of all the cues available to it. Thus, binaural cues might enhance monaural cues to segregation, contributing concurrently to object formation. A parallel process cannot be ruled out.

ACKNOWLEDGMENTS

The Scottish Executive and the United Kingdom's Medical Research Council supported this work. The authors are grateful for the assistance of software engineer Tim Folkard, hardware engineer Scott Morrison, and statistician Brady West. Les Bernstein, Chris Darwin, Ewan Macpherson, Christopher Stecker, Steve Colburn, and an anonymous reviewer provided helpful comments on previous versions of this manuscript.

APPENDIX: MULTINOMIAL LOGISTIC REGRESSION

A multinomial logistic regression described by Agresti (1990, pp. 313–317) and Long (1997) was employed to predict a given ratio of confusion probabilities using dummy variables representing the condition types as independent variables. This method allows, for example, determination of which conditions significantly influence the ratio of probability of type 1 errors to the probability of type 2 errors (or vice versa). The procedure was used to determine the ratio of type 2 to type 1 errors (or vice versa) in the IID and ITD, IID-

only, and ITD-only conditions. The procedure was also used to determine the ratio of type 1 or type 2 errors to correct responses for individual listeners. SPSS Version 10.0 (1999) was used to execute the analysis. A description of the process is included in the SPSS reference manual. The ratio of probability of type 2 confusions to the probability of type 1 confusions was determined using the following equation:

$$\ln\left(\frac{P(\text{Type2})}{P(\text{Type1})}\right) = B_1 + B_2(\text{ITDonly}) + B_3(\text{IIDITD}) \quad \text{(A1)}$$

in which B_1, B_2, and B_3 are model parameters which are "fit." The numbers 1 or 0 are substituted for the dummy variables (ITDonly and IIDITD) to calculate the appropriate ratios for these conditions using the determined values of B_1, B_2, and B_3. For example, consider the output of the regression shown in Table IV. The first and second columns show B_1, B_2, and B_3 and their respective standard errors. The third column shows the "Wald" statistic, which is similar to the F statistic in an ANOVA: larger Wald values indicate large significant differences among B values. In order to calculate the probability ratio for the IID-only condition [the reference condition for Eq. (A1)], ITDonly and IIDITD are set to 0 and values from the table are plugged into Eq. (A1). The ratio of the probability of type 2 errors to the probability of type 1 errors is then e^{B1}. For the ITD-only condition, ITDonly is 1 and IIDITD is 0 and for the IIDs and ITDs condition, IIDITD is 1 and ITDonly is zero. The ratios are then $e^{(B1+B2)}$ and $e^{(B1+B3)}$ for the ITD-only condition and IIDs and ITDs conditions, respectively.

The p values, all less than 0.0005, show that the difference in the ratios of the probabilities of type 2 and type 1 confusions among the three conditions are highly significant. The confidence interval for the ratio $P(\text{Type2})/P(\text{Type1})$ in the IID-only condition is determined by calculating $e^{(B1\pm 1.96*\text{SE})}$ where SE is the standard error for B_1.

To easily calculate the confidence intervals for the ITD-only and IID and ITD conditions (IIDITD), the regressions were redone using the appropriate reference conditions as shown in Eqs. (A2) and (A3). The probability ratios for each condition were equal, independent of which logistic equation was used. The standard error for B_1 in the equations below was used to determine the 95% confidence interval for their respective reference conditions: ITD only in Eq. (A2) and IID and ITD in Eq. (A3).

$$\ln\left(\frac{P(\text{Type2})}{P(\text{Type1})}\right) = B_1 + B_2(\text{IIDonly}) + B_3(\text{IIDITD}), \quad \text{(A2)}$$

$$\ln\left(\frac{P(\text{Type2})}{P(\text{Type1})}\right) = B_1 + B_2(\text{IIDonly}) + B_3(\text{ITDonly}). \quad \text{(A3)}$$

As above, the standard error for each condition is determined by calculating $e^{(B1\pm 1.96*\text{SE})}$ where SE is the standard error for B_1.

For analyzing confusions among individual listeners, the following equation was used:

$$\ln\left(\frac{P(\text{Type}X)}{P(C)}\right) = B_1 + B_2(B) + B_3(C) + B_4(D) + B_5(F) \quad \text{(A4)}$$

in which B, C, D, and F are dummy variables set to 1 or 0 depending on the condition. In order to calculate the appropriate ratio for the reference condition A (IIDs and ITDs), for example, the equation is solved for the ratio, setting the values B, C, D, and F to 0. These letters are dummy variables for the ITDs-only, IIDs-only, ongoing-ITDs only and Culling and Summerfield conditions, respectively. To calculate the appropriate ratio in condition B (ITDs only), B is set to one, and C, D, and F are set to 0. The equation is then solved for the ratio. For the other conditions, the value of the appropriate dummy variable is set to 1 and the remaining dummy variables are set to 0. In this way, the expected probability ratios were calculated. The ratio of the probability of the expected confusion to the probability of a correct answer was calculated separately for each listener when performance was between chance levels and 50% correct. The associated 95% confidence intervals for the expected confusion were also calculated by running the regression six times, making each condition the reference and calculating the confidence interval, CI$=e^{(B1\pm 1.96*\text{SE})}$ as above.

Agresti, A. (**1990**). *Categorical Data Analysis* (Wiley, New York).

Bernstein, L. R., and Trahiotis, C. (**1982**). "Detection of interaural delay in high-frequency noise," J. Acoust. Soc. Am. **71**, 147–152.

Bernstein, L. R., and Trahiotis, C. (**1994**). "Detection of interaural delay in high-frequency sinusoidally amplitude-modulated tones, two-tone complexes and bands of noise," J. Acoust. Soc. Am. **95**, 3561–3567.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

Bregman, A. S., and Pinker, S. (**1978**). "Auditory streaming and the building of timbre," Can. J. Psychol. **32**, 19–31.

Broadbent, D. E., and Ladefoged, P. (**1957**). "On the fusion of sounds reaching different sense organs," J. Acoust. Soc. Am. **29**, 708–710.

Bronkhorst, A. W., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level difference on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508–1516.

Buell, T. N., and Hafter, E. R. (**1991**). "Combination of binaural information across frequency bands," J. Acoust. Soc. Am. **90**, 1894–1900.

Culling, J., and Summerfield, Q. (**1995**). "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay," J. Acoust. Soc. Am. **98**, 785–797.

Culling, J., Summerfield, Q., and Marshall, D. (**1994**). "Effects of simulated reverberation on the use of binaural cues and fundamental-frequency differences of separating concurrent vowels," Speech Commun. **14**, 71–95.

Darwin, C. J. (**1984**). "Perceiving vowels in the presence of another sound: Constraints on formant perception," J. Acoust. Soc. Am. **76**, 1636–1647.

Darwin, C. J., and Ciocca, V. (**1992**). "Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component," J. Acoust. Soc. Am. **91**, 3381–3390.

Darwin, C. J., and Gardner, R. B. (**1986**). "Mistuning a harmonic of a vowel: Grouping and phase effects on vowel quality," J. Acoust. Soc. Am. **79**, 838–845.

Darwin, C. J., and Hukin, R. W. (**1997**). "Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity," J. Acoust. Soc. Am. **102**, 2316–2324.

Darwin, C. J., and Hukin, R. W. (**1998**). "Perceptual segregation of a harmonic from a vowel by interaural time difference in conjunction with mistuning and onset asynchrony," J. Acoust. Soc. Am. **103**, 1080–1084.

Darwin, C. J., and Hukin, R. W. (**1999**). "Auditory objects of attention: The role of interaural time differences," J. Exp. Psychol. **25**, 617–629.

Dye, R. H. (**1990**). "The combination of interaural information across frequencies: Lateralization on the basis of interaural delay," J. Acoust. Soc. Am. **88**, 2159–2170.

Gaik, W. (**1993**). ''Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling,'' J. Acoust. Soc. Am. **94**, 98–110.
Hall, J. W., and Grose, J. H. (**1990**). ''Comodulation masking release and auditory grouping,'' J. Acoust. Soc. Am. **88**, 199–125.
Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). ''Detection in noise by spectro-temporal pattern analysis,'' J. Acoust. Soc. Am. **87**, 269–283.
Henning, G. B. (**1974**). ''Detectability of interaural delay in high-frequency complex waveforms,'' J. Acoust. Soc. Am. **55**, 84–90.
Hill, N. I., and Darwin, C. J. (**1996**). ''Lateralization of a perturbed harmonic: Effects of onset asynchrony and mistuning,'' J. Acoust. Soc. Am. **100**, 2352–2364.
Hukin, R. W., and Darwin, C. J. (**1995**). ''Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification,'' Percept. Psychophys. **57**, 191–196.
Isabelle, S. K., and Colburn, H. S. (**1995**). ''Binaural detection with reproducible narrow-band noise maskers,'' J. Acoust. Soc. Am. **97**, 3278(A).
Jeffress, L. A. (**1948**). ''A place theory for sound localization,'' J. Comp. Physiol. Psychol. **61**, 468–486.
Klumpp, R. G., and Eady, H. R. (**1956**). ''Some measurements of interaural time difference thresholds,'' J. Acoust. Soc. Am. **28**, 859–860.
Kubovy, M., and Daniel, J. E. (**1983**). ''Pitch segregation by interaural phase momentary amplitude disparity and by monaural phase,'' J. Audio Eng. Soc. **31**, 630–634.
Kuhn, G. (**1977**). ''Model for the interaural time differences in the azimuthal plane,'' J. Acoust. Soc. Am. **62**, 157–167.
Kuhn, G. (**1987**). ''Physical acoustics and measurements pertaining to directional hearing,'' in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York), pp. 3–25.
Long, J. S. (**1997**). ''Nominal outcomes: Multinomial logit and related models,'' in *Regression Models for Categorical and Limited Dependent Variables* (Sage, Thousand Oaks, CA) pp. 148–178.
Lorenzi, C., Gatehouse, S., and Lever, C. (**1999**). ''Sound localization in noise in normal-hearing listeners,'' J. Acoust. Soc. Am. **105**, 1810–1820.
Macpherson, E. A., and Middlebrooks, J. C. (**2002**). ''Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited,'' J. Acoust. Soc. Am. **111**, 2219–2236.
McFadden, D., and Pasanen, E. G. (**1976**). ''Lateralization at high frequencies based on interaural time differences,'' J. Acoust. Soc. Am. **59**, 634–639.
Middlebrooks, J. C., and Green, D. M. (**1991**). ''Sound localization by human listeners,'' Annu. Rev. Psychol. **42**, 135–159.
Moore, B. C. J. (**1997**). *Introduction to the Psychology of Hearing* (Academic, San Diego).
Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1986**). ''Thresholds for hearing mistuned partials in determining the pitch of complex tones,'' J. Acoust. Soc. Am. **77**, 1853–1860.
Nuetzel, J. M., and Hafter, E. R. (**1981**). ''Discrimination of interaural delays in complex waveforms: Spectral effects,'' J. Acoust. Soc. Am. **69**, 1112–1118.
Raatgever, J., and Bilsen, F. A. (**1986**). ''A central spectrum theory of binaural processing Evidence from dichotic pitch,'' J. Acoust. Soc. Am. **80**, 429–441.
Rasch, R. A. (**1978**). ''The perception of simultaneous notes such as in polyphonic music,'' Acustica **40**, 21–33.
Rayleigh, Lord (**1907**). ''On the perception of sound direction,'' Philos. Mag. **13**, 214–232.
Roberts, B., and Moore, B. C. J. (**1991**). ''The influence of extraneous sounds on the perceptual estimation of first formant frequency in vowels under conditions of asynchrony,'' J. Acoust. Soc. Am. **89**, 2922–2932.
Shaw, E. (**1974**). ''Transformation of sound pressure level from the free field to the eardrum in the horizontal plane,'' J. Acoust. Soc. Am. **56**, 1848–1861.
Siegel, R. A., and Colburn, H. S. (**1989**). ''Binaural processing of noisy stimuli: Internal/external noise ratios for diotic and dichotic stimuli,'' J. Acoust. Soc. Am. **86**, 2122–2128.
SPSS (**1999**). *SPPS Regression Models 10.0* (SPSS, Chicago, IL).
Trahiotis, C., and Bernstein, L. R. (**1986**). ''Lateralization of bands of noise and sinusoidally amplitude modulated tones: Effects of spectral locus and bandwidth,'' J. Acoust. Soc. Am. **79**, 1950–1957.
Wightman, F. L., and Kistler, D. J. (**1992**). ''The dominant role of low-frequency interaural time differences in sound localization,'' J. Acoust. Soc. Am. **91**, 1648–1661.
Woods, W. S., and Colburn, H. S. (**1992**). ''Test of a model of auditory object formation using intensity and interaural time difference discrimination,'' J. Acoust. Soc. Am. **91**, 2894–2902.
Yost, W. A., and Sheft, S. (**1993**). ''Auditory Perception,'' in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 193–236.
Zurek, P. M. (**1993**). ''Binaural advantages and directional effects in speech intelligibility,'' in *Acoustic Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston), 255–277.
Zwisloski, J., and Feldman, R. S. (**1956**). ''Just noticeable differences in dichotic phase,'' J. Acoust. Soc. Am. **28**, 860–864.

Application of loudness models to sound processing for cochlear implants

Hugh J. McDermott[a)]
Department of Otolaryngology and The Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation, The University of Melbourne, 384-388 Albert Street, East Melbourne 3002, Australia

Colette M. McKay
Department of Otolaryngology, The University of Melbourne, 384-388 Albert Street, East Melbourne 3002, Australia

Louise M. Richardson
The Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation and The Bionic Ear Institute, 384-388 Albert Street, East Melbourne 3002, Australia

Katherine R. Henshall
Department of Otolaryngology, The University of Melbourne, 384-388 Albert Street, East Melbourne 3002, Australia

(Received 31 December 2002; revised 31 July 2003; accepted 4 August 2003)

A new paradigm for processing sound signals for multiple-electrode cochlear implants is introduced, and results are presented from an initial psychophysical evaluation of its effect on the perceived loudness of complex sounds. A real-time processing scheme based on this paradigm, called SpeL, has been developed primarily to improve control of loudness for implant users. SpeL differs from previous schemes in several ways. Most importantly, it incorporates a published numerical model which predicts the loudness perceived by implant users for complex patterns of pulsatile electric stimulation as a function of the pulses' physical parameters. This model is controlled by the output of a corresponding model that estimates the loudness perceived by normally hearing listeners for complex sounds. The latter model produces an estimate of the specific loudness arising from an acoustic signal. In SpeL, the specific loudness function, which describes the contribution to total loudness of each of a number of frequency bands (or cochlear positions), is converted to a pattern of electric stimulation on an appropriate set of electrodes. By application of the loudness model for electric stimulation, this pattern is designed to produce a specific loudness function for the implant user which approximates that produced by the normal-hearing model for the same input signal. The results of loudness magnitude estimation experiments with five users of the SpeL scheme confirmed that the psychophysical functions relating overall loudness perceived to input sound level for five complex acoustic signals were, on average, very similar to those for normal hearing. [DOI: 10.1121/1.1612488]

PACS numbers: 43.66.Ts, 43.66.Ba, 43.66.Cb [MRL]

I. INTRODUCTION

Most current users of multiple-electrode cochlear implants report generally satisfactory understanding of speech and many other sounds, at least in favorable listening conditions. However, the dynamic range of acoustic input signals that can be processed effectively by typical sound processors is much narrower than the range of sound levels that is audible and comfortable for listeners with normal hearing. Furthermore, the relation between perceived loudness and input sound level for implant users is not necessarily similar to that for normally hearing listeners, as will be explained below. This paper describes a new sound processing scheme that has been designed to normalize loudness perception for cochlear implant users over a relatively wide input dynamic range. The processor is compatible with the Nucleus CI24 implant system, manufactured by Cochlear Limited. The processing scheme runs in real time, and has been implemented in a small, body-worn, fully digital device. Results of loudness estimation experiments, presented later, confirmed that the scheme is capable of normalizing the perceived loudness of complex sounds varying in level and bandwidth.

The design of a typical sound processor for an existing multiple-electrode implant is outlined in the block diagram of Fig. 1. Input sound signals are converted by the microphone into an electric waveform that is amplified, usually with an automatic gain control (AGC). The AGC restricts the range of levels that is presented to the following stages of the processor, and thereby may reduce the amount of distortion or noise associated with the signal processing. In addition, a suitable fast-acting AGC may improve the perception of speech, especially at low input levels (McDermott *et al.*, 2002). The frequency content of the input waveform is then analyzed using either a Fast Fourier Transform (FFT), or a bank of bandpass filters having partially overlapping frequency responses. The resulting short-term estimate of the

a)Electronic mail: hughm@unimelb.edu.au

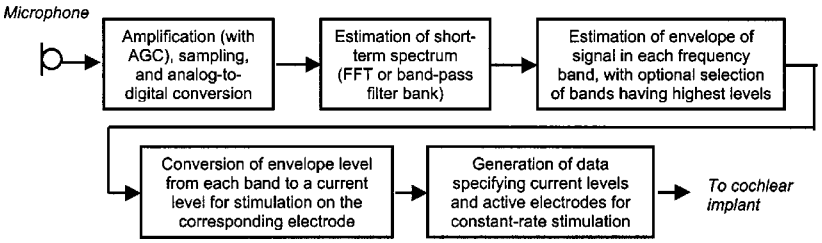


FIG. 1. Block diagram illustrating the function of typical existing sound-processing schemes for cochlear implants.

input power spectrum is subdivided into a number of frequency bands equaling the number of electrodes available for stimulation in the cochlear implant. The frequency bands are assigned to the electrodes such that relatively low input frequencies activate apical electrodes, whereas higher frequencies activate more basal electrodes. The level in each band is converted to an appropriate amplitude for electric stimulation by means of a nonlinear function, as described further below. Typically, the neural stimulation delivered by the implanted electrodes comprises a series of discrete, brief current pulses which are presented in a cyclical sequence.

The conversion of acoustic levels to electric pulse amplitudes is applied independently for each electrode, and therefore for each input frequency band. Conversion functions of various shapes are used in different implant sound processors, and the shape may be altered during programming of the processor for a particular implant user (Zeng and Galvin, 1999). Generally, the conversion function is intended to produce electric stimulation that is just audible for an input sound level that is just above the processor's (or microphone's) internal noise level. This level of stimulation is often referred to as the ''T-level,'' because it is close to the threshold of electric hearing for the implant user. During fitting of the sound processor, a T-level is measured for each active electrode. Usually, a loud but comfortable level (referred to as the ''C-level'') is also measured for each electrode. An increase in the sound level within the processor's input dynamic range causes a corresponding increase in the stimulation amplitude between the T- and C-levels. The conversion function is always compressive, because the input dynamic range in each frequency band is greater than the electric dynamic range (i.e., the ratio between the C-level and the T-level) on the corresponding electrode.

It is well known that both the T-levels and the C-levels vary as a function of the rate of pulsatile electric stimulation (McKay *et al.*, 2001; Skinner *et al.*, 2000). Therefore, these levels are usually measured on each electrode using a steady pulse train with a rate equal to the rate that will be applied by the sound processor to each electrode during normal operation. However, it is also known that the loudness perceived when multiple electrodes are activated at a given rate and level is greater than that of a single electrode activated at the same rate and level (Tong *et al.*, 1983; McKay *et al.*, 1995, 2001). This summation of loudness across multiple active electrodes is not explicitly taken into account in the design or programming of existing sound processors. Thus, it would be expected that the growth of perceived loudness with input sound level may differ from that experienced by listeners with normal hearing, especially for complex acoustic signals. For example, at a relatively high input level, a wideband noise would result in the activation of a larger number of electrodes than a pure tone, and therefore would be perceived as louder even for the same sound-pressure level. Although such loudness summation also occurs in normal acoustic hearing, the detailed characteristics of loudness summation have been shown to be different with electric stimulation. In particular, psychophysical evidence demonstrates that the amount of loudness summation when multiple electrodes are activated (or when the pulse rate on a single active electrode is increased) is much greater for stimulation levels near the T-level than for higher levels (McKay *et al.*, 2001). It is a primary objective of the improved sound-processor design described below to compensate for these effects of loudness summation, and thereby to enable cochlear implant users to perceive the loudness of complex sounds in a way more similar to that of normally hearing listeners.

II. SIGNAL PROCESSING

The main underlying principle of the new processing scheme is to estimate the loudness of incoming sounds using a computational model, and to generate a related pattern of electric stimulation that produces a percept of appropriate loudness for the cochlear implant user. At the same time, the stimulation must be distributed across the available electrodes using an allocation of acoustic frequency to cochlear place that is similar to that used successfully in existing sound processors (outlined briefly above). This is achieved by utilizing an estimate of *specific loudness* that is obtained from the model of normal loudness perception. As explained further below, the specific loudness function describes the distribution of loudness contributions across frequency (or, equivalently, across cochlear positions). The use of this function as the basis for creating the electric stimulation patterns in the new processing scheme is reflected in the scheme's name: *SpeL*.

A functional block diagram of the SpeL scheme appears in Fig. 2. In the present implementation, incoming sound signals transduced by the microphone are amplified with fixed gain, sampled at 16 kHz, and converted into the digital domain. Blocks of samples are windowed using 256 coefficients derived from the product of a Hamming and a sinc function (Harris, 1978). The windowed samples are folded (by adding samples from the second half of the window to corresponding samples from the first half) to produce 128-sample blocks that are updated every 2 ms (Crochiere and Rabiner, 1983). Subsequently, each such block is processed by a 128-point FFT to estimate the short-term spectrum. The output of the FFT comprises bins with center frequencies at

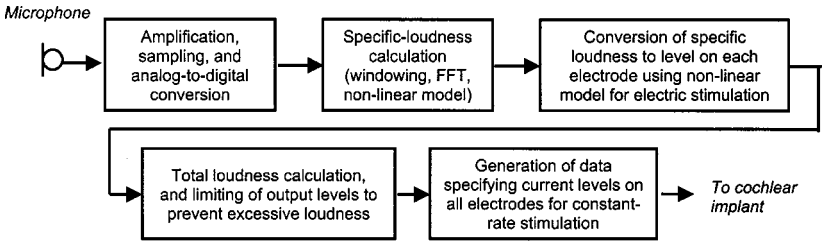


FIG. 2. Block diagram illustrating the function of the SpeL sound-processing scheme.

multiples of 125 Hz. The use of the 256-point window ensures that there is minimal overlap in the frequency domain between adjacent FFT bins (Moore, 1990).

Power spectrum estimates are derived from the FFT outputs by summing the squared real and imaginary components for each bin. These power values are processed by a loudness model derived from the work of Moore and Glasberg (Moore and Glasberg, 1996, 1997). Briefly, the model comprises three main computational stages. In the first stage, the power spectrum is converted to the equivalent rectangular bandwidth (ERB) frequency scale (Glasberg and Moore, 1990). On this scale, each ERB represents an approximately constant spatial extent along the basilar membrane of the cochlea (Moore and Glasberg, 1997). To perform the conversion, the power values from contributing FFT bins are weighted by a set of constant coefficients and then summed into each ERB band. Additionally, a set of correction factors is applied so that the frequency response of the microphone used with SpeL, and that of the normal outer and middle ear, are taken into account. In effect, this correction applies filtering with fixed parameters to the input signal to enable a normal cochlear excitation pattern to be estimated in the second stage of the model. Following Moore and Glasberg (1997), the excitation in each ERB is calculated in the second stage by summing the power passed by a corresponding set of simulated auditory filters. Each such auditory filter has an asymmetric frequency response, with a level-dependent shape (Moore and Glasberg, 1987). Generally, each filter's frequency response has a bandpass characteristic, but with a steeper slope on the high-frequency side of the center frequency than on the low-frequency side. The low-frequency slope becomes shallower with increasing signal level, and the frequency responses among adjacent filters overlap considerably, especially for high signal levels. Finally, in the third stage of the model, the excitation pattern is transformed to a specific loudness function by means of a nonlinear calculation. The result of this transformation is an estimate of the loudness contribution arising from the excitation within each ERB. An estimate of the total loudness of the input signal is also obtained by summing the specific loudness values across ERBs.

In SpeL, the specific loudness values are assigned to appropriate electrodes by allocating one or more ERBs contiguously to each of the available electrodes. For example, if 16 electrodes are available, two ERBs may be assigned to each electrode, starting at ERB number 4 and ending at ERB number 35. These ERBs have center frequencies of approximately 150 Hz and 10 kHz, respectively, and therefore encompass the range of frequencies in which most information about speech and many other types of sound is present. When the SpeL scheme is programed to fit a particular implant user, alternative allocations of ERBs to electrodes may be selected when appropriate. For example, with the CI24 implant (manufactured by Cochlear Limited), it would be possible to utilize up to 22 intracochlear electrodes.

To determine the amplitude of the electric stimulation required on each electrode, the specific loudness estimates are first summed across the ERBs that were allocated to that electrode. These estimates are then converted to current levels using a function derived from a recently published model describing the loudness of pulsatile electric stimulation in terms of the physical parameters of the stimuli (McKay *et al.*, 2003). In applying this general model to the practical implementation of the SpeL scheme, two simplifying assumptions were adopted. First, it is assumed that the amount of loudness summation for multiple-electrode stimulation is approximately independent of the spatial positions of the active electrodes. This assumption is supported by the psychophysical evidence (McKay *et al.*, 2001, 2003). Second, the SpeL processor is designed to activate all of the available electrodes continuously at a constant pulse rate, thus obviating the need to apply an explicit integration process to model temporal aspects of loudness summation (McKay and McDermott, 1998). For example, if 16 electrodes are activated, and the pulse rate on each electrode is 500 Hz (because, as mentioned above, the analysis period is 2 ms), then all the physical parameters of the stimuli except the current levels are fixed, and therefore do not require real-time determination by the computational loudness model. (Note that, with the monopolar stimulation mode and settings of the pulse duration typically required by users of the CI24 implant, a pulse rate of 500 Hz per electrode is easily attained, even when all 22 available intracochlear electrodes are activated.) With these assumptions, a single function can be used to convert the specific loudness values associated with each ERB into current levels on the electrodes stimulating the corresponding cochlear positions.

This function requires a small set of parameters to be adjusted to suit the psychophysical characteristics of each electrode in each implant user (McKay *et al.*, 2003). These parameters include an estimate of the current level corresponding to the threshold of hearing for each available electrode when stimulation is applied at the total rate (e.g., 8 kHz for the above example, in which 16 electrodes are each activated at 500 Hz). This level is the stimulation level produced when an acoustic input signal is present at a level at (or below) the internal noise level of the processor. Thus, it is similar to the conventional T-level described earlier, except that the total pulse rate (rather than the rate per electrode) is used when it is measured. Determining the minimum current

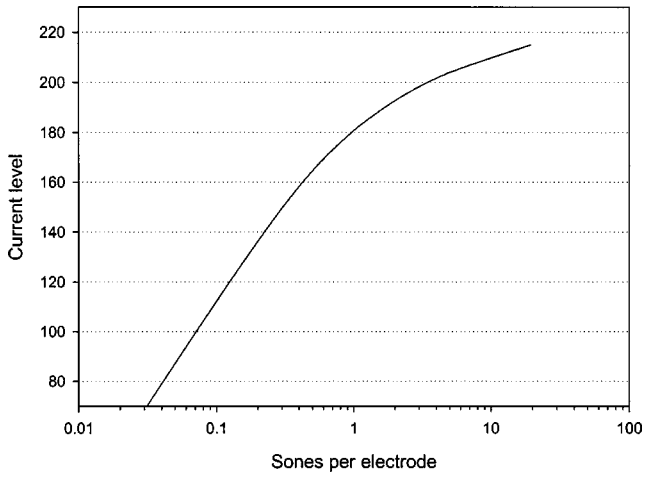


FIG. 3. Example of a typical relationship between the current level on one electrode in the SpeL sound-processing scheme and the required amount of loudness to be contributed by that electrode. The units of the ordinate are current levels (which are closely proportional to the logarithm of current), whereas the units of the abscissa are sones (from the loudness model for normal hearing).

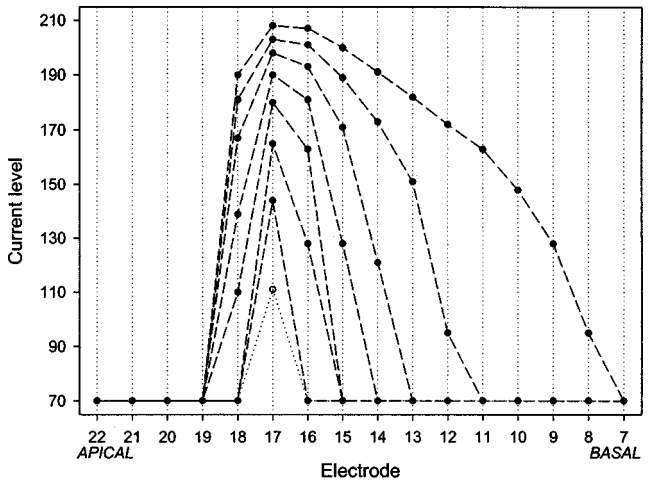


FIG. 4. Example of the outputs from the SpeL sound-processing scheme when the input signal is a pure tone at 1000 Hz. Each curve shows the current level (ordinate) on each of 16 electrodes (abscissa) resulting from a particular level of the input tone. The uppermost curve (dashed line, filled symbols) is for a tone presented at 85 dB SPL, and the lower dashed curves show the effect of reducing the tone level in 10-dB steps. The lowest curve (dotted line, open symbols) shows the output for an input level near the minimum that can be detected by the sound processor, which is approximately 20 dB SPL at 1 kHz. To construct this figure, the relationship between the current level on each electrode and the loudness contribution for that electrode was that shown in Fig. 3.

level in this way accounts for the loudness summation that occurs as a consequence of all the available electrodes being activated continuously regardless of the level of the input signal. Likewise, a higher level similar to the conventional C-level is determined which corresponds to a loud but comfortable hearing sensation produced by the activation of multiple electrodes. However, this level differs from the conventional C-level in that it does not delimit the maximum current output on each electrode, but instead provides a practical means for balancing the loudness across electrodes at a relatively high level.

In fitting SpeL to each implant user, two additional parameters of the above conversion function may also need to be adjusted. These parameters are required because the acoustic loudness model for normal hearing produces outputs in units of sones, whereas it is not practical to obtain loudness estimates from individual implant users that can also be calibrated directly in sones. Therefore, a simple linear function, defined by two parameters, is employed in SpeL to transform the outputs of the acoustic loudness model to suit the loudness preferences of each implant user. In effect, this function compresses or expands (on a ratio scale) the loudness values in sones from the normal-hearing model before applying them to the loudness model for electric stimulation.

A typical function relating the current level of the pulses on one electrode to the corresponding specific loudness estimates obtained from the real-time loudness model for the range of acoustic frequencies assigned to that electrode is illustrated in Fig. 3. The parameter values used to construct the plotted function are representative of the values found appropriate for the fitting of SpeL to the five implant users who participated in the experiments described later.

To ensure that no input sound can result in an uncomfortably loud sensation for each SpeL user, a maximum loudness limit is applied. As shown in Fig. 2, the limiter operates by comparing the estimated total loudness (obtained by summing all the specific loudness values) with a maximum tolerable loudness level obtained when fitting SpeL to the implant user. If the total loudness exceeds this maximum, all of the specific loudness values are reduced by multiplying them by a single factor. The value of this factor, which is always less than or equal to 1, is reduced instantaneously when necessary by the real-time signal processor, but is constrained to increase relatively slowly. The release time constant is approximately 125 ms, which is comparable to that of typical compression limiters employed in conventional acoustic hearing aids. Note that this limiting process is independent of the function described above which generates the current levels for stimulation on each electrode based on specific loudness values before the total loudness is compared with the listener's discomfort level (see Fig. 2).

The above process determines the current levels to be delivered by the active electrodes within each stimulation period (i.e., every 2 ms). The available electrodes are activated in a continuously repeating basal-to-apical cycle. Digital data specifying the parameters of each pulse to be conducted by the electrodes are transmitted to the implant in real time by the processor.

The current levels of the pulses delivered to the electrodes are plotted in Fig. 4 for an illustrative condition in which the SpeL scheme processed pure-tone input signals at 1 kHz. For this condition, the function described above (and shown in Fig. 3) was applied to convert the specific loudness of the input signal into output current levels for each electrode. The 16 most-apical electrodes available in the CI24 implant were used (numbered from 7 to 22 in a basal-to-apical direction), and the stimulation pulse rate was 500 Hz per electrode. The allocation of acoustic frequencies to electrodes was that mentioned previously, in which two ERBs were assigned to each electrode, starting at ERB number 4

TABLE I. Relevant information about each of the subjects who participated in the experiments.

Subject	Age (yrs)	Sex	Duration of profound deafness (yrs)	Etiology	Implant experience (yrs)	Number of electrodes used
S1	70	M	1	Head trauma	24[a]	19
S2	79	M	4	Unknown (progressive)	4	20
S3	55	M	2	Chronic infection	2	20
S4	79	M	1	Noise exposure	2	20
S5	63	M	2	Chronic infection	2	18

[a]S1 first received an implant in 1978, and was reimplanted with the CI24 device in 1998.

and ending at ERB number 35. Thus, the 1-kHz input tone produced the largest output current on electrode 17. For the lowest input levels (up to about 25 dB SPL), only electrode 17 was activated at suprathreshold levels. As the input level was increased, the level on electrode 17 increased, and there was also a spread of activation to surrounding electrodes. This spread was highly asymmetrical, with greater activity on electrodes basal to electrode 17 than on more-apical electrodes. At the maximum input signal level of 85 dB SPL (which is the highest level for which the sinusoidal signal waveform did not get clipped in the sound processor), nine of the ten electrodes that were basal to electrode 17 were activated, whereas in the apical direction, only one additional electrode was activated. This asymmetrical and nonlinear spread of activation across cochlear position with increasing signal level is related to the corresponding effects described for acoustic hearing (e.g., the effect often described as ''upward spread of masking'').

In order to ascertain whether the SpeL scheme is effective at normalizing loudness perception for implant users, an experiment was conducted in which numerical magnitude estimates were obtained when complex acoustic stimuli having several bandwidths were presented via the SpeL processor at a number of input levels. This study is described briefly next.

III. METHOD

A. Subjects

Five experienced, adult users of the Cochlear CI24 implant participated in the experiment. Relevant information about them is provided in Table I. Some of the subjects had participated in earlier psychophysical studies in which the loudness model for electrical stimulation outlined above was developed and verified. In particular, three of the participants in the study of McKay *et al.* (2001), specifically S8, S10, and S11, also participated in the current study, where they are designated S2, S3, and S1, respectively, and four of the participants in the study of McKay *et al.* (2003), specifically S1, S2, S5, and S6, are here designated S3, S5, S2, and S1.

For each subject, the stimulation mode selected was monopolar (using both of the available extracochlear electrodes), the pulse width was set to the minimum possible (approximately 26 μs/phase), and the nonstimulating interval between phases was also set to the minimum possible (approximately 8.4 μs). Subjects were not paid for participation in the research project, although expenses such as travel costs were reimbursed.

B. Stimuli

The main aim of the experiment was to determine whether the SpeL scheme enabled appropriate loudness perception as a function of level for signals at various frequencies, and for complex signals having different bandwidths. Therefore, three stimulus types were selected. These were: (1) 1/3-oct-wide noise bands with center frequencies of 250, 1000, and 4000 Hz; (2) a 1-oct-wide noise band with a center frequency of 1000 Hz; and (3) a 3-oct-wide noise band with a center frequency of 1000 Hz. The noise bands were created by filtering a pink noise using sixth-order Butterworth bandpass filters. Each of these five stimulus types was presented over a range of levels encompassing the acoustic input dynamic range of the SpeL processor with 12 levels in steps of 5 dB. The duration of each of the 60 stimuli was 500 ms, and linear rise/fall ramps of 30 ms were applied.

The SpeL scheme was implemented in a wearable, fully programmable, digital signal processor that had been developed previously in this laboratory for research. In the experiment, the above stimuli were delivered to the processor via a direct audio input. By obviating the use of a microphone, any complications arising from presentation of signals from a loudspeaker (such as interference from background noise, or nonideal characteristics of the acoustic environment) were avoided.

C. Procedure

Six repetitions of the 60 stimuli were presented in a random order via each subject's SpeL processor. The subjects were asked to provide a numerical loudness estimate immediately after the presentation of each stimulus. At the commencement of each test session, the 1/3-oct noise band centered on 1 kHz was presented at a level of 50 dB SPL, and it was suggested to each subject that the loudness of this example stimulus should be assigned a value of about 50. Subjects were instructed to attempt to assign numerical estimates for each stimulus on a ratio scale (i.e., a sound twice as loud as the example stimulus should be assigned a value of 100,

etc.). It was emphasized that the range of the numerical estimates should be unconstrained. These instructions were intended to minimize the likelihood that the results of the experiment would be confounded by the idiosyncratic assignment of numbers by each subject to the loudness perceived. If any stimulus was not detected by a subject after its presentation, that fact was recorded as a loudness estimate of zero. Thresholds of hearing were also obtained for narrowband noise signals with center frequencies of 250, 1000, and 4000 Hz presented in the sound field. For these measurements only, the microphone (rather than the processor's direct audio input) was used to ensure that realistic aided hearing thresholds were obtained. The test signals were delivered from a loudspeaker located about 1.5 m from the subject's listening position in a medium-sized sound-attenuating booth. The threshold data reported below were obtained by averaging three estimates at each frequency, measured using a standard audiometric technique (Carhart and Jerger, 1959).

IV. RESULTS

Averaged across the subjects, the free-field hearing thresholds measured with the SpeL processor at 250, 1000, and 4000 Hz were 30, 28, and 28 dB SPL, respectively. These are approximately 30 dB higher (worse) than the thresholds for normal hearing (ISO, 1996). Although these normal-hearing thresholds apply to measurements made under binaural listening conditions, whereas the implant users' thresholds were obtained with stimulation in only one ear (i.e., the implanted ear), the higher values for the implant listeners are mostly a consequence of inherent electrical noise in the microphone and signal-processing circuitry.

The results of the loudness magnitude estimations, averaged across subjects, are plotted in Fig. 5. Because some of the stimuli presented at the lowest levels were not detected reliably by all of the subjects, the data plotted include all signal presentations for which at least three subjects provided at least three nonzero estimates. The data shown are geometric means. To obtain an indication of the amount of variance in the subjects' responses, the common logarithm of each response was calculated, and the standard deviations of these values were determined for all responses provided by each subject. On this log scale, the standard deviations ranged from 0.076 to 0.43 among the five subjects; averaged across all subjects, the standard deviation was 0.25.

Each curve in Fig. 5 shows the relation between the subjects' averaged estimates (ordinate) and the loudness predictions in sones for normal hearing (abscissa) obtained for each of the five test stimuli. The predicted loudness values were calculated using a non-real-time version of the computational model devised at the University of Cambridge, England. That program, which was developed by Moore and Glasberg (1996, 1997), was used to predict a loudness estimate in sones given the level of each noise signal specified in terms of dB SPL measured in 1/3-oct bands. Values were assigned to the program's input parameters to represent average normal acoustic hearing. Note that, although the computational model's outputs have units of sones, the subjects' average loudness estimates, plotted in Fig. 5, are dimensionless. However, the slopes of the lines, and the amount of overlap among the lines, indicate the effectiveness of the SpeL sound-processing scheme in providing these subjects with normal perception of the loudness of the test stimuli. The five lines representing the subjects' estimates for each of the five stimulus types would be completely coincident if the SpeL scheme compensated exactly for the differences in loudness resulting from the different bandwidths and center frequencies of the noise bands used in the experiments (and there was no random variance in the subjects' estimates). Furthermore, the slopes of each of the lines would be 1.0 if the growth in loudness with level for each of the stimuli was the same as that predicted by the normal-hearing model. A reference line with a slope of 1.0 is included in Fig. 5 to aid visual comparison of the slopes of the lines.

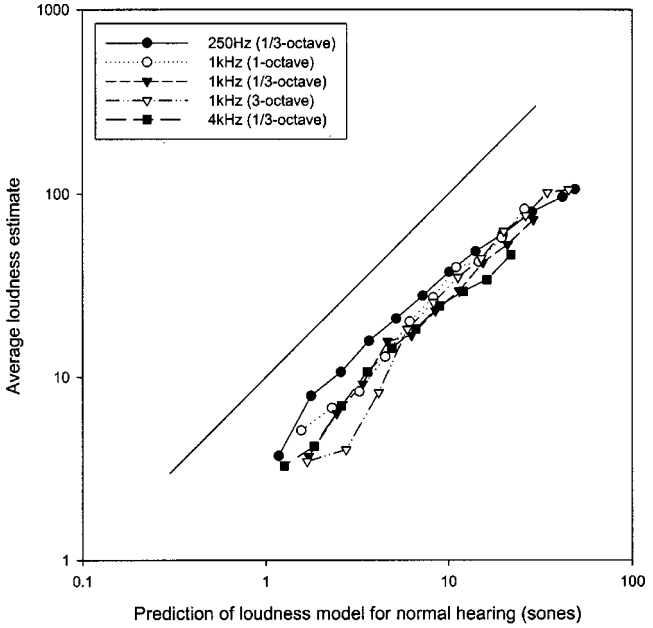


FIG. 5. Results of the loudness estimation experiments. The ordinate shows the loudness estimates, averaged across the five subjects who participated in the experiments, and the abscissa shows the predicted total loudness in sones from a published model for normal acoustic hearing. Each curve and symbol type shows data obtained for each of the five signal types, which are listed in the legend. Each point plotted corresponds to a particular presentation level of the signal. The diagonal line indicates the expected slope of all the curves if the loudness data from the implant users followed the predictions of the loudness model exactly.

V. DISCUSSION

Inspection of the data plotted in Fig. 5 confirms that the loudness functions are generally similar for each stimulus type, with the possible exception of the results for the 1/3-oct noise band centered on 250 Hz. Linear regression was applied to determine objectively whether the slope of each line was close to 1.0. The correlation coefficients (r^2) for each of the lines were all greater than 0.97, confirming that the data are well described by straight lines (on log–log axes). The slopes of the lines for the 1/3-oct noise bands centered on 1 and 4 kHz, and the 1-oct noise band (centered on 1 kHz) were 1.01, 0.94, and 1.02 respectively. The slope for the 3-oct noise band was 1.14, but this was affected by the fact that responses from two of the subjects (S2 and S4) were unreliable when signals were presented at the two lowest

levels. That is, only three of the five subjects contributed loudness estimates to the two lowest data points plotted. When the parameters of the relevant regression line were recalculated with these two data points omitted, the slope was 1.03. However, the slope of the line for the 1/3-oct noise band centered on 250 Hz was only 0.85, and examination of Fig. 5 shows that this is related principally to the fact that the judged loudness was greater than predicted by the model when those signals were presented at low levels. The loudness estimates for the 250-Hz stimulus when presented at higher levels were close to those for the other stimuli.

On average, the experimental data obtained from the five subjects who participated in the study suggest that the SpeL sound-processing scheme is generally effective at restoring near-normal perception of relative loudness to users of cochlear implants, at least for four of the five types of stimuli used. This demonstrates that the SpeL scheme can compensate for the largest effects of loudness summation for electrical stimulation reported in recent psychophysical studies, as mentioned in the Introduction (McKay *et al.*, 2001, 2003). As shown in Fig. 5, the subjects' average loudness estimates for the 1/3-oct noise band centered on 250 Hz tended to be higher than the predictions of the model when the signals were presented at relatively low levels. This is almost certainly a consequence of one of the simplifications adopted to facilitate real-time operation of the SpeL sound-processing scheme. As mentioned earlier, the short-term spectrum of input signals is estimated in the present implementation of SpeL by means of a Fast Fourier transform (FFT). The FFT has 128 points, and the sampling rate applied to the input signal is 16 kHz. Therefore, the FFT spectral estimates are spaced at multiples of 125 Hz. For the real-time loudness model implemented in SpeL, the bank of auditory filters is simulated by combining weighted spectral estimates from the FFT. The 125-Hz spacing of these spectral estimates results in a relatively coarse simulation of the behavior of the auditory filters at frequencies so low that only a very few FFT outputs contribute to each filter. This limitation affects the accuracy of the output of the simulated auditory filters for an input signal such as a narrow-band noise centered on 250 Hz. Furthermore, the accuracy is worse at low signal levels, which correspond to relatively narrow auditory filters, than at higher levels, which correspond to wider auditory filters (Moore and Glasberg, 1987, 1997). Thus, the subjects' overestimates shown in Fig. 5 probably resulted from overestimates of the loudness of these signals by the SpeL sound processor. However, the effect is not large, being equivalent to an overestimate of the input signal level of less than 10 dB for the 250-Hz narrow-band noise at levels near the sound processor's minimum detectable input level. In a future version of SpeL, a modification could be implemented to the processing of information derived from lower-frequency components of input signals to reduce the size of this effect.

One assumption that facilitated the practical implementation of the present version of SpeL was that both the loudness model for normal acoustic hearing and the corresponding model for electric stimulation would provide satisfactory outputs when the input signals are time varying. Initially these models were devised using steady signals. As described above, the simplified implementations of the loudness models in SpeL were also evaluated experimentally with steady signals. Additional studies are required to investigate whether users of SpeL experience appropriate loudness when listening to time-varying signals, such as speech and most other common sounds. However, previous psychoacoustic experiments, and research with sound signals processed in real time using acoustic loudness models, suggest that the overall loudness of amplitude-modulated signals may be related reasonably consistently to their average levels (Moore *et al.*, 1999b; Launer and Moore, 2003). Thus, it seems plausible that the temporal averaging inherent in the SpeL signal-processing functions described earlier, in combination with the temporal integration of loudness that is assumed to occur in perceptual processes more central than the cochlea, would result in appropriate experience of overall loudness for most types of time-varying signals, as well as steady sounds, for SpeL users.

VI. CONCLUSIONS

The SpeL sound-processing scheme is a practical realization of a new paradigm for encoding information about acoustic signals for users of cochlear implants. Unlike previous sound-processing schemes, SpeL compensates explicitly for the largest loudness summation effects associated with nonsimultaneous, pulsatile electric stimulation patterns that have been identified in recently published psychophysical experiments. Based on subjective estimates by implant users of the loudness perceived when acoustic noise bands varying in level, center frequency, and bandwidth were presented, the SpeL scheme can provide near-normal growth of loudness with acoustic level across an input dynamic range of greater than 50 dB (at least on average for the subjects who participated in the reported experiments). In particular, the slopes of all but one of the loudness functions obtained in the experiments were very close to those expected for normal hearing.

Importantly, the fundamental principles underlying the implementation of the SpeL scheme are applicable more generally. For example, it might not always be optimal to control the loudness perceived by implant users such that it closely follows the loudness perceived by normally hearing listeners for the same input signals. The loudness model for electric stimulation implemented in SpeL enables alternative loudness relations to be obtained easily. For instance, to provide maximum speech understanding for implant users, it might be preferable to compress the range of overall loudness perceived relative to normal, or to reduce differences across frequency in the long-term average of the specific loudness function, rather than simply to normalize loudness perception as described above. Maximum intelligibility of speech might be attained if the specific loudness function was modified to increase the loudness contributions arising from those acoustic frequencies that are likely to contain most information about speech at the expense of other frequencies where less speech information is present, or where that information is assumed to be unavailable (e.g., as a consequence of masking). Similar techniques have been applied in various fitting prescriptions devised for acoustic hearing aids (e.g.,

Byrne and Dillon, 1986; Moore and Glasberg, 1998; Moore *et al.*, 1999a; Dillon, 2001), and, more recently, in real-time amplitude compression processing for hearing aids (Launer and Moore, 2003). For users of cochlear implants, the loudness models implemented in SpeL would enable such modifications to be applied without necessarily affecting the overall perceived loudness of the signals. However, it is possible that modifying the specific loudness function in such ways might reduce the perceived quality or naturalness of some sounds. This issue will be addressed in further experiments.

An alternative application of the general loudness-modeling paradigm would be to improve the performance of existing sound-processing schemes for cochlear implants. As noted in the Introduction to this article, sound processors commonly used with the implant devices available presently from commercial manufacturers do not attempt to compensate in real time for the loudness summation that has been shown to occur when varying numbers of electrodes are activated by discrete current pulses delivered in rapid succession to several cochlear locations. Instead, fixed functions are used to convert the acoustic level in each partially nonoverlapping analysis frequency band to an appropriate electric level for activation of the corresponding electrode. The parameters of those functions are selected when the sound processor is fitted to each implant user by seeking a compromise between providing adequate audibility of low-intensity sounds and avoiding loudness discomfort for sounds having high overall intensities. It may be possible to circumvent any perceptual disadvantages resulting from this compromise by application of real-time loudness models, such as those described above, to existing processing schemes. Further research will help determine whether improvements in loudness perception similar to those reported in this article can be provided to implant users by modifying existing processing schemes, or whether the SpeL scheme provides additional benefits related to the direct use of an estimated specific loudness function as the basis for calculating the electric stimulation parameters.

ACKNOWLEDGMENTS

This work was supported by The Garnett Passe and Rodney Williams Memorial Foundation, The Cooperative Research Center for Cochlear Implant and Hearing Aid Innovation, and The Bionic Ear Institute. The authors would particularly like to thank Justin Zakis, Mark Harrison, Andrew Vandali, and Rodney Millard. Brian Moore, Brian Glasberg, and WaiKong Lai provided some of the software used in the data analysis. We are also most grateful to the cochlear implant users who volunteered to participate in the experiments. Finally, we acknowledge the constructive comments made by Brian Moore and an anonymous reviewer on a previous version of the manuscript.

Byrne, D., and Dillon, H. (**1986**). ''The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid,'' Ear Hear. **7**, 257–265.

Carhart, R., and Jerger, J. F. (**1959**). ''Preferred method for clinical determination of pure-tone thresholds,'' J. Speech Hear Disord. **24**, 330–345.

Crochiere, R. E., and Rabiner, L. R. (**1983**). *Multirate Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).

Dillon, H. (**2001**). *Hearing Aids* (Boomerang, Sydney).

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

Harris, F. J. (**1978**). ''On the use of windows for harmonic analysis with the discrete Fourier transform,'' Proc. IEEE **66**, 51–83.

International Standards Organization (ISO) (**1996**). *Acoustics—Reference zero for the calibration of audiometric equipment. Part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions*, Standard 389-7 (ISO, Geneva).

Launer, S., and Moore, B. C. J. (**2003**). ''Use of a loudness model for hearing aid fitting. V. On-line gain control in a digital hearing aid,'' Int. J. Audiol. **42**, 262–273.

McDermott, H. J., Henshall, K. R., and McKay, C. M. (**2002**). ''Benefits of syllabic input compression for users of cochlear implants,'' J. Am. Acad. Audiol. **13**, 14–24.

McKay, C. M., and McDermott, H. J. (**1998**). ''Loudness perception with pulsatile electrical stimulation: The effect of interpulse intervals,'' J. Acoust. Soc. Am. **104**, 1061–1074.

McKay, C. M., Henshall, K. R., Farrell, R. J., and McDermott, H. J. (**2003**). ''A practical method of predicting the loudness of complex electrical stimuli,'' J. Acoust. Soc. Am. **113**, 2054–2063.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1995**). ''Loudness summation for two channels of stimulation in cochlear implants: Effects of spatial and temporal separation,'' Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 230–233.

McKay, C. M., Remine, M. D., and McDermott, H. J. (**2001**). ''Loudness summation for pulsatile electrical stimulation of the cochlea: Effects of rate, electrode separation, level, and mode of stimulation,'' J. Acoust. Soc. Am. **110**, 1514–1524.

Moore, B. C. J., and Glasberg, B. R. (**1987**). ''Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns,'' Hear. Res. **28**, 209–225.

Moore, B. C. J., and Glasberg, B. R. (**1996**). ''A revision of Zwicker's loudness model,'' Acustica **82**, 335–345.

Moore, B. C. J., and Glasberg, B. R. (**1997**). ''A model of loudness perception applied to cochlear hearing loss,'' Aud. Neurosci. **3**, 289–311.

Moore, B. C. J., and Glasberg, B. R. (**1998**). ''Use of a loudness model for hearing aid fitting. I. Linear hearing aids,'' Br. J. Audiol. **32**, 317–335.

Moore, B. C. J., Glasberg, B. R., and Stone, M. A. (**1999a**). ''Use of a loudness model for hearing aid fitting. III. A general method for deriving initial fittings for hearing aids with multi-channel compression,'' Br. J. Audiol. **33**, 241–258.

Moore, B. C. J., Vickers, D. A., Baer, T., and Launer, S. (**1999b**). ''Factors affecting the loudness of modulated sounds,'' J. Acoust. Soc. Am. **105**, 2757–2772.

Moore, F. R. (**1990**). *Elements of Computer Music* (Prentice-Hall, Englewood Cliffs, NJ).

Skinner, M. W., Holden, L. K., Holden, T. A., and Demorest, M. E. (**2000**). ''Effect of stimulation rate on cochlear implant recipients' thresholds and maximum acceptable loudness levels,'' J. Am. Acad. Audiol. **11**, 203–213.

Tong, Y. C., Blamey, P. J., Dowell, R. C., and Clark, G. M. (**1983**). ''Psychophysical studies evaluating the feasibility of a speech processing strategy for a multiple-channel cochlear implant,'' J. Acoust. Soc. Am. **74**, 73–80.

Zeng, F. G., and Galvin, J. J. III (**1999**). ''Amplitude mapping and phoneme recognition in cochlear implant listeners,'' Ear Hear. **20**, 60–74.

Nonlinear dynamics of phonations in excised larynx experiments

Jack J. Jiang,[a] Yu Zhang, and Charles N. Ford
Department of Surgery, Division of Otolaryngology Head and Neck Surgery, University of Wisconsin Medical School, Madison, Wisconsin 53792-7375

(Received 16 August 2002; accepted for publication 19 July 2003)

Nonlinear dynamic methods including correlation dimension and Lyapunov exponents are applied to quantitatively analyze phonations in excised larynx experiments. Irregular phonations are typically characterized by aperiodic waveforms and broadband spectra. Finite correlation dimensions and positive Lyapunov exponents of irregular phonations demonstrate the existence of chaos in excised larynx phonations. Furthermore, the correlation dimension, maximal Lyapunov exponent, jitter, shimmer, and peak prominence ratio are used to statistically distinguish irregular phonations from normal phonations. The correlation dimension and maximal Lyapunov exponent indicate a significant difference between irregular and normal phonations; however, jitter, shimmer, and peak prominence ratio do not reveal such a significant difference and thus are unsuitable to differentiate between irregular phonations and normal phonations. These findings might potentially assist investigators in understanding rough phonations and developing clinically valuable methodologies for the diagnosis of voice disorders. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1610462]

PACS numbers: 43.70.Aj, 43.70.Bk, 05.45.Tp, 05.45.Pq [AL]

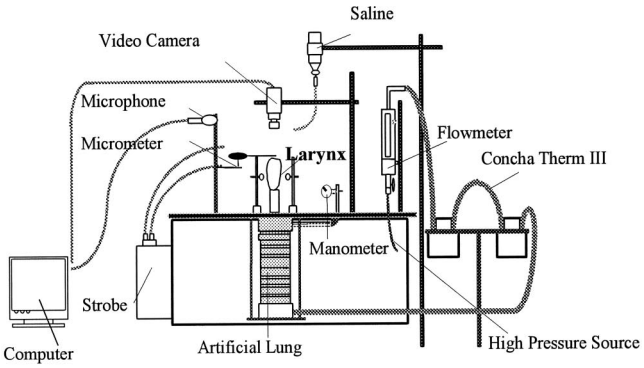
I. INTRODUCTION

Excised larynx experiments facilitate the direct observation and measurement of vocal-fold vibrations, and have proven to be advantageous in the study of voice physiologies.[1–7] Unlike computer models, an excised larynx with its natural morphology represents a physiological model. In excised larynx experiments, the parameters controlling phonations can be systematically monitored and independently controlled. Extreme phonation conditions that are difficult to achieve in human subjects, such as extremely high subglottal pressure, can be applied. Because of this, excised larynx experiments have been applied as a valuable methodology to study irregular phonations. Irregularly curved shapes of the glottis were observed at high subglottal pressures by van der Berg and Tan.[1] Irregular vibratory patterns also were found by Isshiki *et al.*[2] when they studied the effects of asymmetric tension on vocal-fold vibrations with excised larynges as well as in an asymmetric computer model. To determine the maximal threshold pressure of normal phonations, Jiang and Titze[4] defined and measured the phonation instability pressure (PIP). When subglottal pressure was above the PIP, vocal-fold vibrations became irregular and phonations were rough.

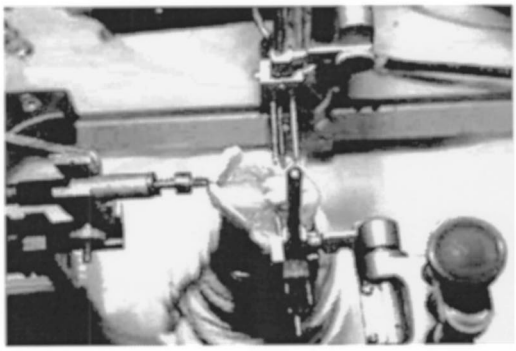
In recent years, nonlinear dynamic methods have been successfully applied in voice analyses and computer models to describe irregular phonation phenomena.[8–18] Nonlinear dynamic analyses of pathological voices from patients with laryngeal pathologies have suggested that voice irregularities are associated with low-dimensional vocal dynamics.[9,11–15] Computer modeling of the vocal folds has also revealed chaotic vibratory patterns.[16–18] Nonlinear dynamics of irregular phonations in excised larynx experiments have not yet been well addressed, although recent studies have qualitatively shown some nonlinear characteristics of irregular phonations. Berry *et al.*[5] applied a bifurcation analysis to study phonation onset and voice instabilities. Various phonatory regimes were observed to overlap when subglottal pressure and vocal-fold asymmetries were changed. Svec *et al.*[19] studied chest-falsetto register jumps in human subjects and excised human larynges; they found that a small change in vocal-fold tension might cause an abrupt change in register and pitch. However, a quantitative study of chaos in excised larynx phonations has not yet been performed, nor has it been determined whether or not the observed voice irregularity can be regarded as a high-dimensional process, described appropriately as a stochastic process, or as low-dimensional chaos. Such quantitative studies in excised larynx experiments are important since established quantitative results obtained with excised larynges might be helpful in determining the dynamic mechanism of irregular voice production, examining computer modeling and clinical voice analysis methods, and designing corrective treatments necessary to improve phonations.

For the purpose of this paper, we apply nonlinear dynamic methods to quantitatively analyze phonations in excised larynx experiments. The Lyapunov exponent and correlation dimension are employed to describe the dynamics of excised larynx phonations. The capability of the correlation dimension and Lyapunov exponent to differentiate between normal and irregular phonations is examined in excised larynx experiments. Some traditional voice analysis parameters, including the jitter, shimmer, and peak prominence ratio (PPR) are also used. We expect that nonlinear dynamic methods are capable of describing the dynamics of excised larynx phonations; however, the traditional parameters may not effectively differentiate between normal and irregular excised larynx phonations.

[a]Electronic mail: Jiang@surgery.wisc.edu



(a)



(b)

FIG. 1. Experimental description of excised larynx experiments. (a) The experimental setup of an excised larynx. (b) The photograph displays an excised canine larynx mounted atop an artificial lung.

II. EXPERIMENTAL MATERIALS AND METHODS

The experimental setup is illustrated in Fig. 1(a). Briefly, 9 larynges were harvested 15 min postmortem from large (25–30 kg), healthy laboratory dogs after their demise from other causes. Excision of the larynx was performed according to the procedure described by Jiang and Titze.[4] The excised larynges were placed in a 0.67% saline solution and stored in a refrigerator set at 40°F (≈4.4 °C). The larynges were used in an experimental trial 12–36 h after excision. Before the experiment, the larynges were trimmed and the resting length of the vocal folds and the position of the arytenoid cartilages were established. The larynx was then mounted with a section of trachea on top of a pipe, and the trachea was tightly clamped to the pipe with a hose clamp made of a 1 mm thick steel strip [see Fig. 1(b)].[4,6] A three-pronged device was used to stabilize the arytenoid cartilages bilaterally, allowing for micrometer control of adduction and abduction. A second micrometer system was attached by stitching a rod to the anterior tip of the thyroid lamina. Turning this micrometer system controlled the elongation of vocal folds. Vocal-fold elongation and adduction were held constant within each larynx for all experimental trials.

We ensured symmetry of the laryngeal vibration system using the following operations. First, the excised larynx was carefully examined to make sure both vocal folds had a similar thickness and length. When mounting the larynges on the setup, we carefully positioned arytenoid cartilages at the same level. Finally, the micrometer systems were applied to symmetrically control the elongation and adducation of the larynges.

An Ingersoll-Rand (Type 30) conventional air compressor was used to generate the airflow. A valve controlled the amount of airflow. The flow was measured as it passed through a Gilmont rotameter-type flow meter (J197). The input air was conditioned to 35 °C–38 °C and 95%–100% relative humidity by two ConchaTherm III heater-humidifiers (Respiratory Care, Inc.) placed in series. A Cole-Parmer Digi-Sense Thermo-couple Thermometer (Model 8529-00) and a humidity probe (Check-It Electronics 424) were used to monitor the humidity 2 cm above the level of the vocal folds. We held vocal-fold elongation and adduction constant and manually increased subglottal pressure in the experiments. The subglottal pressure in the artificial lung was measured with an open-ended water manometer (Dwyer No. 1211). When a certain pressure level was reached, we kept the pressure at a constant level for at least 5 s in order to obtain sustained phonations. PIP was the maximal threshold pressure of normal phonations.[4] When subglottal pressure was below the PIP, we determined normal phonations as nearly regular and periodic patterns. The 20 normal phonations from 9 excised larynges were recorded for analyses. However, when subglottal pressure was higher than the PIP, vocal-fold vibrations became irregular and phonations were rough. We recorded 20 irregular phonations that are clearly aperiodic.

The acoustic signal was recorded using a microphone positioned 15 cm from the glottis. The axis of the microphone was at 45° from the long axis of the larynx in order to avoid airflow impinging directly onto the microphone. The signal picked up by the microphone was A/D converted with the sampling frequency 20 kHz and processed on a computer. The stationarity and length of signals were easy to ensure since the parameters controlling laryngeal phonations, such as subglottal pressure, could be sustained for a long time. We selected the middle segments of the phonation samples with lengths of 2 s for analyses. Some traditional methods of voice analysis, including the spectrogram, jitter, shimmer, and PPR, were applied. The percent jitter is a cycle-to-cycle frequency perturbation measure and the percent shimmer is a cycle-to-cycle amplitude perturbation measure. Jitter and shimmer measure the short-term irregularities of the pitch period and the peak-to-peak amplitude of voices, respectively. PPR is defined as the energy of the peak of interest divided by the total spectral energy of the signal.[20] The software Cspeech 4.0[21] was used to extract the measures of percent jitter and percent shimmer.

Nonlinear dynamic methods, including phase space reconstruction, correlation dimension D_2, and Lyapunov exponent λ_i were applied. Excised larynges do not have supraglottal resonance, which cause excised larynx phonations to have narrow frequency ranges. In the experiments, we found that low-frequency components including fundamental frequency had significantly higher energies than high-frequency components (>2.5 kHz). Using the oscilloscope Tektronix TDS-460, the observed amplitudes at the high-frequency components (>2.5 kHz) were about 60 dB less than the fundamental frequency. There was no high-energy source for

high-frequency components (>2.5 kHz). We thus downsampled the microphone signals with the sampling rate $f_s = 5$ kHz when calculating D_2 and λ_i. The sampling rate of 5 kHz also had been previously used in calculating the fractal dimension[12] and the largest Lyapunov exponent[13] of pathological voice signals. However, we would not recommend using a much lower sampling rate since an excessive undersampling might produce obvious aliasing.

III. NONLINEAR DYNAMIC ANALYSIS FROM THE TIME SERIES

When it is difficult to obtain models in exact detail, reconstructing the dynamics from a measured time series becomes important. For a measured time series, nonlinear dynamic methods, such as the phase space reconstruction, correlation dimension, and Lyapunov exponent, can be applied to study its nonlinear dynamics.

A. Phase space reconstruction

The time delay technique can be used to reconstruct the phase spaces of systems.[22] A time series with length N is measured and recorded as $x(t_1), x(t_2), x(t_3), \ldots$, where $x(t_i) \in \mathbf{R}$, $t_i = t_0 + i\tau$ $(i = 1,2,...,N)$ at the discrete time interval τ. Subsequently, the time delay vector creates the reconstructed phase space as

$$\mathbf{X}(t) = \{x(t), x(t-\tau), \ldots, x(t-(m-1)\tau)\}, \tag{1}$$

where m is the embedding dimension and τ is the time delay. Takens[23] has demonstrated that when $m > 2D+1$ (D is the Hausdorff dimension), the embedding is a diffeomorphism from the original phase space to the reconstructed delay space. That is, the reconstructed phase space preserves the topological properties and geometrical invariance of the original attractor. Therefore, the system dynamics can be studied in the reconstructed phase space.

Takens's embedding theorem does not require the constraint of choosing the time delay τ. However, because of the finite length and finite precision of a time series, choosing the proper time delay is important when reconstructing a phase space. If τ is too small, the time delay vector will be strongly correlated so that the reconstructed attractor is stretched around a diagonal line. On the other hand, if τ is too long, the reconstructed attractor will produce a significant self-intersection. A proper choice of delay time will yield an optimal phase space reconstruction and good estimations of the correlation dimension and Lyapunov exponents. In this study, the mutual information method was used to estimate the proper delay time.[24]

B. Correlation dimension

The correlation dimension describes the complexity of dynamic behaviors. A more complex system with a higher correlation dimension may require more state variables to describe dynamic behaviors, while a simpler system with a lower correlation dimension may need fewer degrees of freedom to describe the dynamic law. Using the correlation dimension method, chaos has been found to distinguish different dynamic characteristics from white noise: the estimate of D_2 of white noise does not converge with the increase of embedding dimension m; however, the estimate of D_2 of a chaotic system converges to a finite value with the increase of m. The correlation dimension can be calculated from a time series with length N,[25]

$$D_2 = \lim_{r \to 0} \lim_{N \to \infty} \frac{\log C(N,r)}{\log r}, \tag{2}$$

where r is the radius around $\mathbf{X}_i$, and the correlation integral $C(N,r)$ is

$$C(N,r) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} \theta(r - \|\mathbf{X}_i - \mathbf{X}_j\|), \tag{3}$$

where the Heaviside function $\theta(x)$ satisfies

$$\theta(x) = \begin{cases} 1, & x > 0, \\ 0, & x \leqslant 0. \end{cases}$$

Furthermore, for finite datasets, the correlated points in a reconstructed phase space would produce an "anomalous shoulder" [26] in the correlation integral curve, which might result in inaccurate estimates of D_2. To avoid this problem, an improved formula is applied,[27]

$$C(W,N,r) = \frac{2}{(N+1-W)(N-W)} \times \sum_{n=W}^{N-1} \sum_{i=0}^{N-1-n} \theta(r - \|\mathbf{X}_i - \mathbf{X}_{i+n}\|), \tag{4}$$

where the W correlated points of the time series are discarded in calculations. When $W = 1$, Eq. (4) is reduced to the standard Grassberger–Procaccia formula in Eq. (3). In our study, we set W to be the proper time delay τ, where τ is obtained using the mutual information method.[24] D_2 is estimated by calculating the local slope of the curve $\log C(N,r)$ vs $\log r$. When the length N is significantly large, in a certain scaling region, the estimate of D_2 converges with the increase of the embedding dimension m. We calculated the correlation dimensions of the Lorenz attractor and the Henon map as $D_2 = 2.03 \pm 0.02$ and $D_2 = 1.24 \pm 0.01$, respectively, which are consistent with the literature values[25] and thus verify our numerical procedures.

C. Lyapunov exponents

Lyapunov exponents are the most effective dynamic descriptors of chaos. They describe the average divergent or convergent rates of neighboring orbits in phase space. Generally, a system with one positive Lyapunov exponent is chaotic, while a system with no positive exponents is regular. An n-dimensional dynamic system has the Lyapunov exponent spectrum $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ with the order $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$. The Kaplan–Yorke dimension is determined by

$$D_L = k + \frac{\sum_{i=1}^{k} \lambda_i}{|\lambda_{k+1}|}, \tag{5}$$

where k satisfies $\sum_{i=1}^{k} \lambda_i \geqslant 0$ and $\sum_{i=1}^{k+1} \lambda_i < 0$.[28] Regarding the relationship of the Kaplan–Yorke dimension to the correlation dimension, generally, this inequality holds: $D_2 \leqslant D_L$.[25]

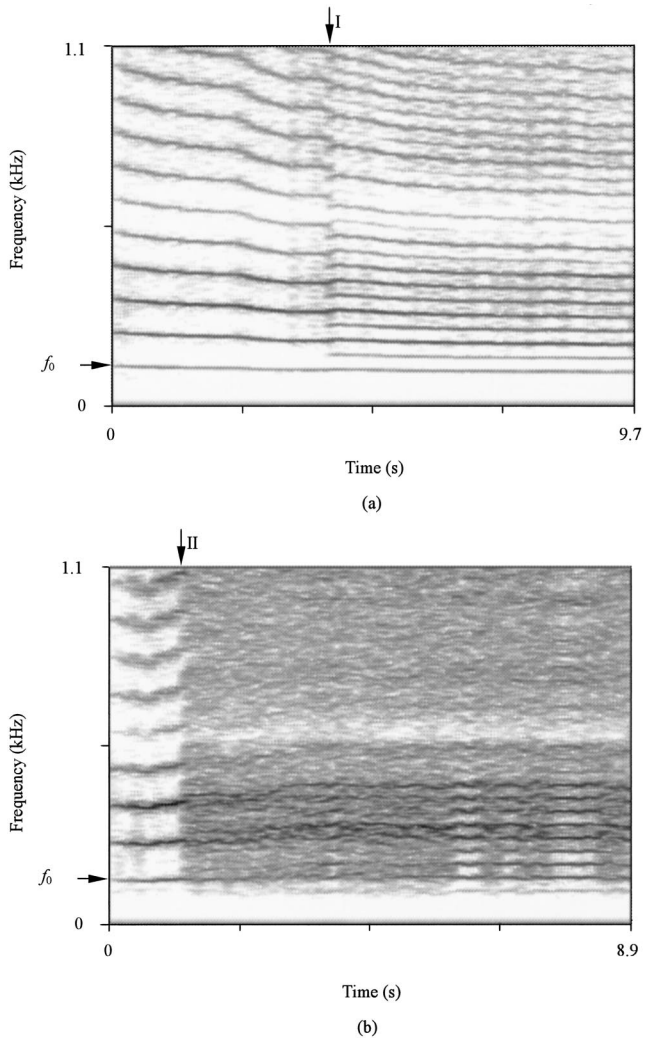


FIG. 2. Narrow-band spectrograms. (a) The period-doubling bifurcation, where arrow I indicates the period-doubling bifurcation. (b) The transition from regular phonation to irregular phonation, where arrow II indicates the transition. (c) The routine from period-doubling bifurcation to chaos, where arrows I and II show the evolution from periodic motion to period-doubling bifurcation to chaos.

For a time series, we used the method by Holzfuss and Lauterborn[29] to calculate the Lyapunov exponent spectrum. For detailed descriptions of this method, we refer to the original work by Holzfuss and Lauterborn.[29] We calculated Lyapunov exponents as the mean rate of divergence for each sampling period; and the rate per second can be obtained by multiplying these values of Lyapunov exponents by the sampling rate $f_s = 1/T_s$. To examine the numerical procedure, we calculated the Lyapunov exponent spectrum based on the Lorenz equation: $\dot{x} = 16(y-x)$, $\dot{y} = x(45.92-z)-y$, $\dot{z} = xy - 4z$. For the Lorenz chaotic attractor, we estimated that the Lyapunov exponents from a time series converged to $\lambda_1 \approx 1.5$, $\lambda_2 \approx 0$, and $\lambda_3 \approx -22.5$. These values of Lyapunov exponents were equal to the results obtained by solving the differential equation by Wolf's method.[30] It should be noted that in some cases, this numerical procedure might not correctly estimate the most negative exponent. This limitation does not produce a problem in this study, since we are not interested in the most negative Lyapunov exponent.

IV. RESULTS

A. Experimental observation of bifurcation and chaos

We studied phonation patterns of excised larynges by increasing subglottal pressure. The self-oscillation of vocal folds occurs when the subglottal pressure exceeds a threshold value.[1,4] In excised larynx experiments, subglottal pressure $P_s = 8$ cm H_2O drove regular vibrations of the larynges. There was a range of subglottal pressures for normal phonations in which phonations sustained a nearly regular and periodic pattern. However, when P_s exceeded this normal range, an irregular phonation was reached and sound became rough. The minimal value of P_s needed to produce phonation instability had been defined as the phonation instability pressure (PIP).[4] For an excised larynx in Fig. 2(a), we measured the PIP at about 16 cm H_2O. When subglottal pressure was increased above the PIP, an irregular phonation, associated with chaos, was produced and then recorded for analysis. PIP was found to be variant for different larynges and was usually high for symmetric larynges, which had been reported by Jiang and Titze in their original work.[4] When P_s increased from 8 to 12 cm H_2O [Fig. 2(a)] and from 8 to 16 cm H_2O [Fig. 2(b)], the vibratory patterns of the excised larynx typically showed a period-doubling bifurcation and a quick transition from periodic motion to chaos, respectively. f_0 is the fundamental frequency. Arrow I in Fig. 2(a) indicated the period-doubling bifurcation, and arrow II in Fig. 2(b) indicated the transition from regular phonation to irregular phonation in the narrow-band spectrograms. The quick transitions from periodic motions to chaos were observed when

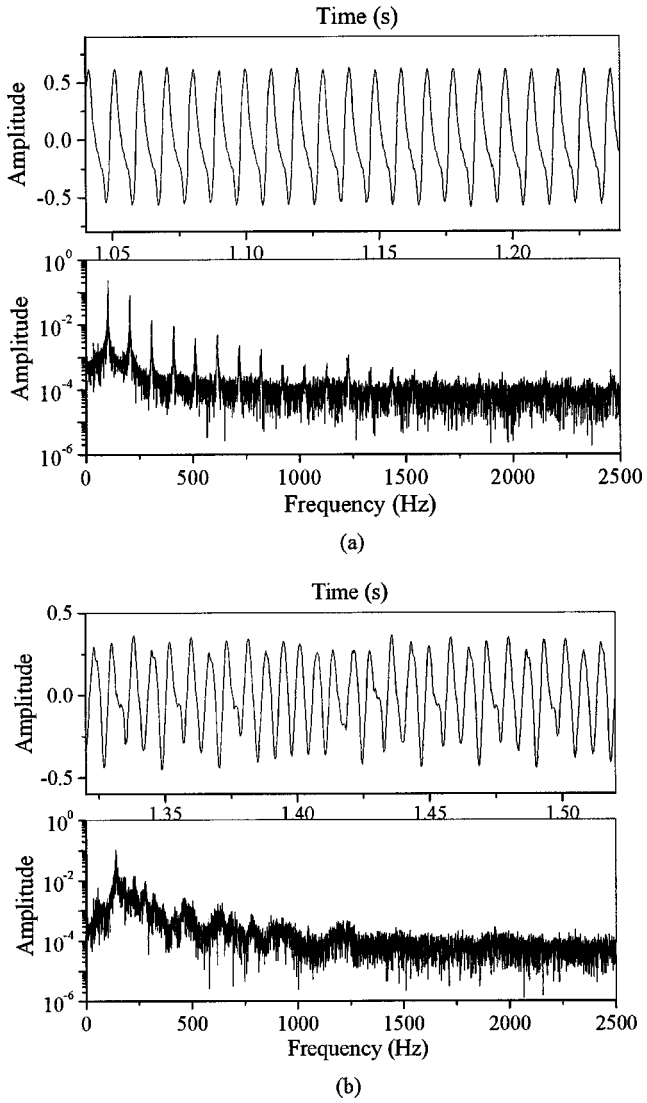


FIG. 3. (a) The time series (the upper curve) and frequency spectrum (the lower curve) of the normal phonation. (b) The time series (the upper curve) and spectrum (the lower curve) of the irregular phonation.

subglottal pressures were incrementally increased. In addition, the routine from period-doubling bifurcation to chaos could also be observed. Arrows I and II in Fig. 2(c) indicate the evolution from periodic motion to period-doubling bifurcation to chaos. Random fluctuations from aerodynamic and biomechanical factors may play an important role in affecting bifurcation characteristics.[31] The bifurcations and chaos in Fig. 2 show that it is important to apply nonlinear dynamic methods to analyze excised larynx phonations. Thus, for 20 normal phonations and 20 irregular phonations from 9 excised larynges, we will employ nonlinear dynamic methods to quantitatively study their nonlinear characteristics.

B. Nonlinear dynamic analysis of phonation data

In this section, we present the typical results of applying a nonlinear dynamic analysis to a normal phonation and an irregular phonation, where the microphone signals are normalized within the interval $[-1,1]$. The normal phonation shows the periodic time series [the upper curve in Fig. 3(a)] and discrete frequency spectrum [the lower curve in Fig. 3(a)]. The irregular phonation shows the aperiodic time series [the upper curve in Fig. 3(b)] and the noiselike broadband spectrum [the lower curve in Fig. 3(b)]. After reconstructing the phase space, we estimate the correlation dimensions D_2 and Lyapunov exponents λ_i, as shown in Fig. 4. The length of all phonation signals is $N=10\,000$ (2 s). By using the mutual information method, the proper time delays of the normal and irregular phonations are estimated at $13T_s$ and $8T_s$, respectively. Figure 4(a) shows the correlation integral of the normal phonation, where the curves from the top correspond to the embedding dimension $m=1,2,...,10$, respectively. We calculate the correlation dimension and its standard deviation by using a curve fit to the curve of $\log_2 C(r)$ vs $\log_2 r$ in the scaling region ($2^{-3}<r<2^{0}$). In Fig. 4(b), the Lyapunov exponents of the normal phonation asymptotically approach constants with the evolution of time, where the embedding dimension $m=4$, and the curves from the top correspond to λ_1, λ_2, λ_3, and λ_4, respectively. The dependences of D_2 and λ_i on the embedding dimension m can be shown in Figs. 4(c) and (d), respectively. With the increase of the embedding dimension m, the estimates of D_2 and the first two Lyapunov exponents (λ_1,λ_2) of the normal phonation converge to $D_2=1.05\pm0.02$, $\lambda_1=0.001\pm0.002$ bits/T_s, and $\lambda_2=-0.02\pm0.001$ bits/T_s; however, the irregular phonation gives $D_2=2.68\pm0.03$, $\lambda_1=0.039\pm0.002$ bits/T_s, and $\lambda_2=0.002\pm0.001$ bits/T_s. The irregular phonation is found to have a higher dimension than the normal phonation, demonstrating its more complex dynamics. From the Lyapunov exponent spectrum in Fig. 4(d), the Kaplan–Yorke dimensions of the normal and irregular phonations are $D_L=1$ and $D_L=3.23$, respectively, which satisfies the inequality $D_2\leqslant D_L$ by Grassberger and Procaccia.[25] As found, the periodic orbit, discrete frequency spectrum, correlation dimension $D_2\approx1.05$, and maximum Lyapunov exponent $\lambda_1\approx0$ of the normal phonation show its regular characteristics. However, the aperiodic waveform, broadband spectrum, correlation dimension $D_2\approx2.68$, and positive Lyapunov exponent $\lambda_1\approx0.039$ of the irregular phonation indicate chaos.

C. Statistical analysis

Figure 3 is a typical example of the difference between a normal phonation and an irregular phonation revealed by nonlinear dynamic methods. However, the variances of different phonation samples require statistical analysis on nonlinear dynamic parameters to determine whether normal phonations are significantly different from irregular phonations. Correlation dimensions and maximal Lyapunov exponents of 20 normal phonations and 20 irregular phonations from 9 larynges are calculated, as shown in Fig. 5. We also calculate the jitter, shimmer, and PPR to assess their capabilities to discriminate between irregular and normal phonations. *t*-tests are performed on these five parameters. Figure 5 shows a comparison of irregular phonations (IP) and normal phonations (NP) on these five parameters, where the curves from the left correspond to jitter, shimmer, PPR, correlation dimension D_2, and maximal Lyapunov exponent λ_1, respectively. The statistical results of jitter ($t=2.35$, $p=0.031$), shimmer ($t=1.58$, $p>0.1$), and PPR ($t=2.25$, $p=0.036$)

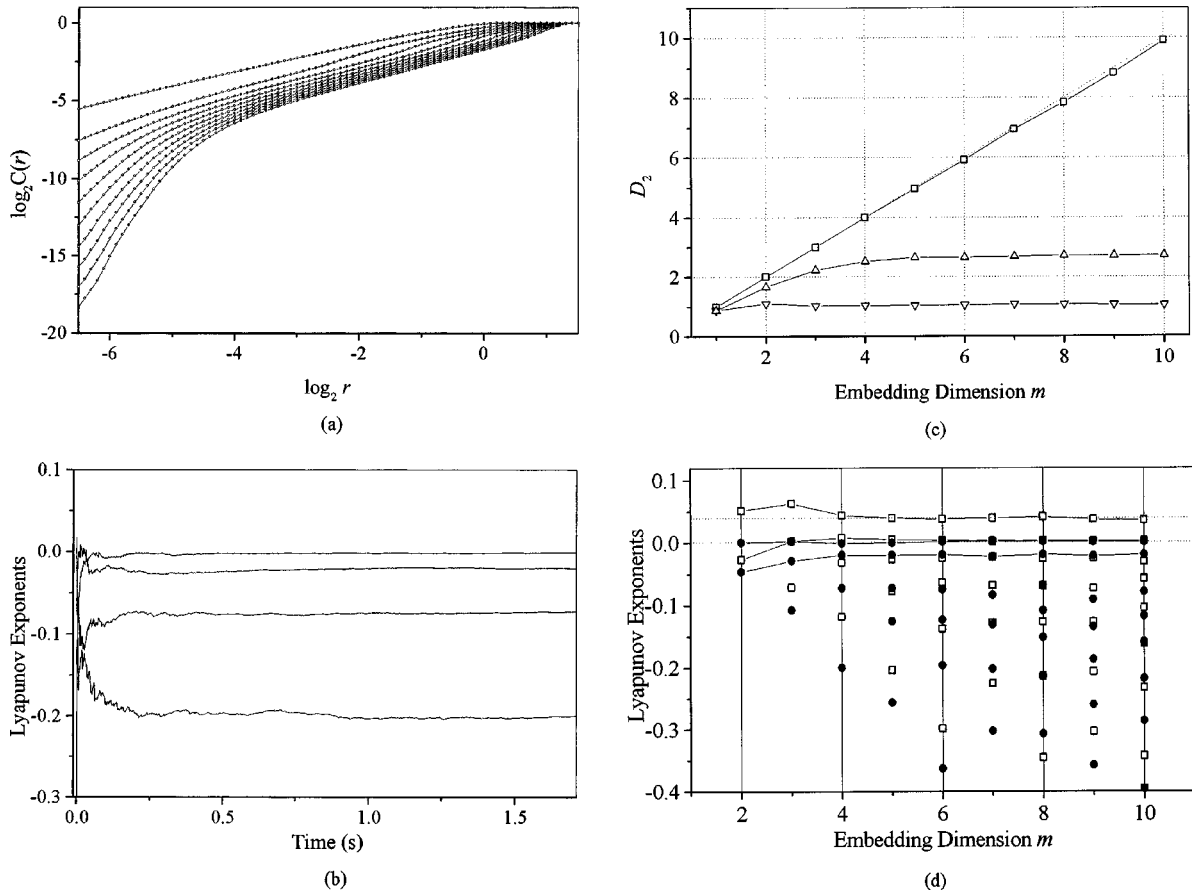


FIG. 4. Nonlinear dynamic characteristics of phonation data. (a) $\log_2 C(r)$ vs $\log_2 r$ of the normal phonation, in which the curves from the top correspond to the embedding dimension $m=1,2,...,10$, respectively. (b) The time evolution of the Lyapunov exponents of the normal phonation, where $m=4$ and the curves from the top correspond to λ_1, λ_2, λ_3, and λ_4, respectively. (c) D_2 vs m of white noise (□), irregular phonation (△), and normal phonation (▽). (d) λ_i vs m, where the first two Lyapunov exponents (λ_1,λ_2) of normal phonation (●) and irregular phonation (□) converge to (0, −0.02) and (0.04, 0), respectively, with the increase of m.

are not significant at the 0.01 level; however, D_2 ($t=5.48$, $p<0.0001$) and λ_1 ($t=8.67$, $p<0.0001$) reveal significant differences between irregular and normal phonations. Jitter and shimmer are determined by estimating the fundamental frequency and peak amplitude of each phonatory cycle. For nearly periodic normal phonations, the variances of fundamental frequency and peak amplitude are small enough that jitter and shimmer could present reasonable analyses. However, for irregular phonations, fundamental frequencies and peak amplitudes of consecutive phonatory cycles appear in an irregular fashion. The ill-defined fundamental frequencies and peak amplitudes prevent the stable estimations of jitter and shimmer. The large variance and poor reliability make jitter and shimmer unsuitable for irregular excised phonations with perceptibly aperiodic or chaotic features. In Fig. 5, jitter, shimmer, and PPR reveal no

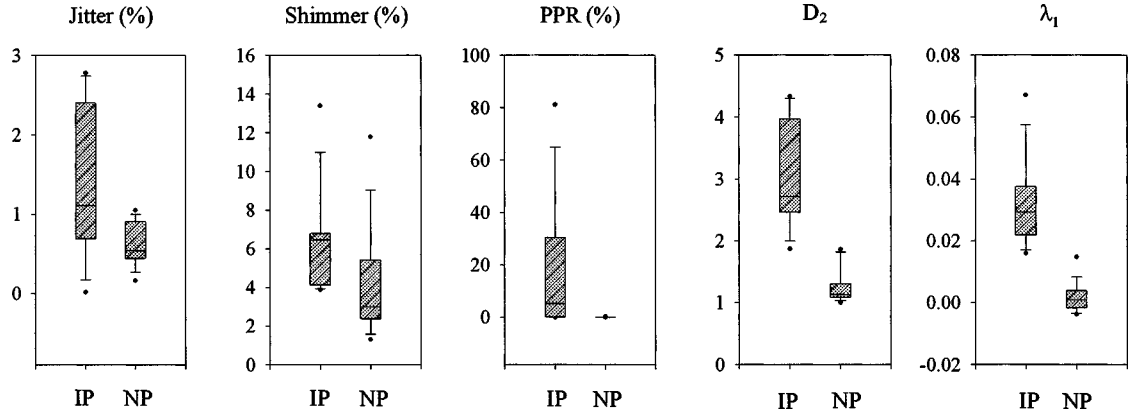


FIG. 5. Comparisons of irregular phonations (IP) and normal phonations (NP) using jitter, shimmer, PPR, D_2, and λ_1, where the curves from the left correspond to jitter, shimmer, PPR, D_2, and λ_1, respectively. At the 0.01 level, no significant difference between irregular and normal phonations has been revealed with jitter, shimmer, and PPR; however, a significant difference between irregular and normal phonations has been found with D_2 and λ_1.

significant differences between irregular and normal phonations, showing that they are not capable of differentiating irregular phonations from normal phonations. However, nonlinear dynamic parameters are not restricted by the limitations of these traditional methods. Theoretically, correlation dimension and Lyapunov exponents represent invariant measures of a chaotic attractor. For irregular excised larynx phonations, the correlation dimensions and maximal Lyapunov exponents are stable. Unlike traditional analysis methods, correlation dimensions and maximal Lyapunov exponents reveal a significant difference between irregular phonations and normal phonations. Furthermore, finite correlation dimensions reveal that all these excised larynx phonations are low dimensional, and positive maximal Lyapunov exponents demonstrate the existence of chaos in excised larynx phonations.

V. DISCUSSION

Voice instabilities and irregularities have previously been found in excised larynx studies.[1–5] However, a quantitative study of irregular laryngeal phonations in nonlinear dynamics has not yet been performed. The effects of nonstationarity and finite signal length on calculating the correlation dimension of pathological voices have been studied by Behrman and Baken.[11] In our excised larynx study, the stationarity and length of the phonation signal are easy to ensure since we can conveniently and steadily control laryngeal phonation conditions, such as subglottal pressure, for a long time. This feature of excised larynx experiments makes it easier to calculate the correlation dimension and Lyapunov exponent than in clinical voice studies. As found in Fig. 3 and Fig. 4, the correlation dimensions reveal that excised larynx phonations are low dimensional; however, the estimate of correlation dimension of white noise has no saturation with the increase of the embedding dimension *m*. Thus, finite state variables may be needed to describe irregular vocal-fold vibrations, which supports the finding of the finite element model by Berry *et al.*;[16] that is, vocal-fold vibrations are largely dominated by the first two to three vibratory modes. This result also suggests that a vocal-fold model with finite state variables, such as the two-mass model, may serve as a useful tool to capture vibratory characteristics of vocal folds. Furthermore, the positive Lyapunov exponents reveal the evidence of chaos in laryngeal phonations. Chaos with positive Lyapunov exponents has been found in speech signals[8,13,32] and computer models,[16–18] while chaotic phonations have been shown in our excised larynx study. We produced irregular phonations by increasing the subglottal pressure. The desynchronization of vibratory modes has also been found to play an important role in producing voice irregularities.[2,5,17] The existence of chaos is associated with nonlinear factors, including the nonlinear relationship of glottal pressure and airflow, nonlinear stress–strain characteristics of vocal-fold tissues, and nonlinear vocal-fold collision.

Early studies of human voice pathology have applied jitter and shimmer to evaluate the voice quality of vocal disorders and document clinical treatment effects.[33] Generally, jitter and shimmer are appropriate for nearly periodic Type I signals.[34] A clinical application of jitter and shimmer has found that these perturbation measures may also be appropriate for Type II signals with subharmonics.[35] Currently, whether jitter and shimmer are suitable for irregular phonations has not been clinically or experimentally determined. Recent studies have suggested that these perturbation methods may be unsatisfactory for the diagnosis of moderate to severe voice disorders in which large variances and poor reliability are observed;[34–36] however, nonlinear dynamic analyses may represent potentially valuable methods.[8–15] In this study, we investigated this issue by using excised larynx experiments. We calculated the jitter, shimmer, PPR, correlation dimension, and maximal Lyapunov exponent and assessed their capability to discriminate between normal and irregular phonations (Fig. 5). We have found that for normal phonations, the jitter, shimmer, correlation dimension, and maximal Lyapunov exponent are capable of analysis; however, for irregular or chaotic phonations, jitter and shimmer are not suitable for analysis, which is in agreement with the theoretical prediction by Titze.[34] No significant difference between irregular and normal phonations has been revealed with jitter and shimmer; however, correlation dimensions and maximal Lyapunov exponents effectively describe voice irregularities and indicate a significant difference between irregular and normal phonations.

Experiments with excised larynges could also be used to examine our previous theoretical study with a two-mass vocal-fold model. Detailed discussions of this symmetric model are presented by Jiang *et al.*[18] In the symmetric computer model, chaotic above-range phonations may occur when subglottal pressure is much higher than the normal pressure range.[18] In excised larynx experiments, PIP has been found to be variant and usually high for symmetric larynges. When subglottal pressure is experimentally increased above the PIP, irregular laryngeal phonations with positive Lyapunov exponents are produced. Thus, under a high driving pressure, irregular phonations observed in excised larynx experiments might be related to the chaotic above-range phonation found in the computer model. In addition, the quick transition from periodic motion to chaos as well as the routine from period-doubling bifurcation to chaos, as revealed in the theoretical model,[18,31] was also observed in the excised larynx experiments (Fig. 2). The experimental observations in the excised larynx showed some qualitative similarities with the computer simulations with the two-mass model of vocal folds.

VI. CONCLUSION

In this paper, we have studied nonlinear dynamic characteristics of excised larynx phonations. The finite correlation dimensions and positive Lyapunov exponents showed low-dimensional chaotic characteristics of irregular laryngeal phonations. The correlation dimension and maximal Lyapunov exponent were significantly more effective at differentiating between irregular phonations and normal phonations than some traditional voice analysis methods, including the jitter, shimmer, and peak prominence ratio. These results showed that nonlinear dynamic methods were applicable to the description of the nonlinear dynamics of excised larynx

phonations. Irregular sound patterns have been widely observed in animal vocalizations[37–41] and disordered human voices;[9,11–15] thus, our study on excised larynx phonations might be potentially helpful for studying these irregular vocalizations and clinically evaluating human voice disorders.

ACKNOWLEDGMENTS

We thank X. J. Yu and M. M. Gintoft for their assistance and Dennis Connaughton for his editorial comments. This study was supported by NIH Grant No. (1-RO1DC05522-01) from the National Institute of Deafness and other Communication Disorders.

[1] J. van den Berg and T. S. Tan, ''Results of experiments with human larynxes,'' Pract. Otorhinolaryngol. (Basel) **21**, 425–450 (1959).
[2] N. Isshiki, M. Tanabe, K. Ishizaka, and D. Broad, ''Clinical significance of asymmetrical vocal cord tension,'' Ann. Otol. Rhinol. Laryngol. **86**, 58–66 (1977).
[3] D. S. Cooper, ''Research in laryngeal physiology with excised larynges,'' in *Otolaryngolgy-Head and Neck Surgery*, edited by C. W. Cummings (C. V. Mosby, St. Louis, 1986), Vol. 3, pp. 1766–1776.
[4] J. J. Jiang and I. R. Titze, ''A methodological study of hemilaryngeal phonation,'' Laryngoscope **103**, 872–882 (1993).
[5] D. A. Berry, H. Herzel, I. R. Titze, and B. H. Story, ''Bifurcations in excised larynx experiments,'' J. Voice **10**, 129–138 (1996).
[6] J. Jiang, K. Verdolini, J. Ng, B. Aquino, and D. Hanson, ''Effects of dehydration on phonation in excised canine larynges,'' Ann. Otol. Rhinol. Laryngol. **109**, 568 (2000).
[7] J. J. Jiang, A. G. Shah, M. M. Hess, K. Verdolini, F. M. Banzali, and D. G. Hanson, ''Vocal fold impact stress analysis,'' J. Voice **15**, 4–14 (2001).
[8] I. R. Titze, R. Baken, and H. Herzel, ''Evidence of chaos in vocal fold vibration,'' in *Vocal Fold Physiology: New Frontier in Basic Science*, edited by I. R. Titze (Singular, San Diego, 1993), pp. 143–188.
[9] H. Herzel, D. Berry, I. R. Titze, and M. Saleh, ''Analysis of vocal disorders with methods from nonlinear dynamics,'' J. Speech Hear. Res. **37**, 1008–1019 (1994).
[10] M. Mazo, D. Erickson, and T. Harvey, ''Emotion and expression: Temporal data on voice quality in Russian lament,'' in *Vocal Fold Physiology—Voice Quality Control*, edited by O. Fujimura and M. Hirano (Singular, San Diego, 1995), pp. 173–184.
[11] A. Behrman and R. J. Baken, ''Correlation dimension of electroglottographic data from healthy and pathologic subjects,'' J. Acoust. Soc. Am. **102**, 2371–2379 (1997).
[12] I. Hertrich, W. Lutzenberger, S. Spieker, and H. Ackermann, ''Fractal dimension of sustained vowel productions in neurological dysphonias: An acoustic and electroglottographic analysis,'' J. Acoust. Soc. Am. **102**, 652–654 (1997).
[13] A. Giovanni, M. Ouaknine, and J.-M. Triglia, ''Determination of largest Lyapunov exponents of vocal signal: application to unilateral laryngeal paralysis,'' J. Voice **13**, 341 (1998).
[14] A. Behrman, ''Global and local dimensions of vocal dynamics,'' J. Acoust. Soc. Am. **105**, 432–443 (1999).
[15] J. J. Jiang and Y. Zhang, ''Nonlinear dynamic analysis of speech from pathologic subjects,'' Electron. Lett. **38**, 294–295 (2002).
[16] D. A. Berry, H. Herzel, I. R. Titze, and K. Krischer, ''Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions,'' J. Acoust. Soc. Am. **95**, 3595–3604 (1994).
[17] I. Steinecke and H. Herzel, ''Bifurcations in an asymmetric vocal-fold model,'' J. Acoust. Soc. Am. **97**, 1874–1884 (1995).
[18] J. J. Jiang, Y. Zhang, and J. Stern, ''Modeling of chaotic vibrations in symmetric vocal folds,'' J. Acoust. Soc. Am. **110(4)**, 2120–2128 (2001).
[19] J. G. Svec, H. K. Schutte, and D. G. Miller, ''On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges,'' J. Acoust. Soc. Am. **106**, 1523–1531 (1999).
[20] J. J. Jiang, E. Lin, J. Wu, C. Gener, and D. G. Hanson, ''Effects of simulated source of tremor on acoustic and airflow voice measures,'' J. Voice **14**, 47–57 (2000).
[21] P. Milenkovic and C. Read, *Cspeech Version 4 User's Manual*, Madison, WI, 1992.
[22] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, ''Geometry from a time series,'' Phys. Rev. Lett. **45**, 712 (1980).
[23] F. Takens, ''Detecting strange attractors in turbulence,'' in *Lecture Notes in Mathematics*, edited by D. A. Rand and B.-S. Young (Springer-Verlag, Berlin, 1981), Vol. 898, pp 366–381.
[24] A. M. Fraser and H. L. Swinney, ''Independent coordinates for strange attractors from mutual information,'' Phys. Rev. A **33**, 1134–1140 (1986).
[25] P. Grassberger and I. Procaccia, ''Measuring the strangeness of strange attractors,'' Physica D **9**, 189–208 (1983).
[26] J. Theiler, ''Spurious dimension from correlation algorithms applied to limited time series data,'' Phys. Rev. A **34**, 2427–2432 (1986).
[27] J. Theiler, ''Estimating fractal dimension,'' J. Opt. Soc. Am. A **7**, 1055–1073 (1990).
[28] J. L. Kaplan and J. A. Yorke, ''Chaotic behaviour of multidimensional difference equations,'' in *Functional Differential Equations and Approximation of Fixed Point*, Lecture Notes in Mathematics 730, edited by H.-O. Peitgen and H.-O. Walther (Springer, Verlag, Berlin, 1979), p. 204.
[29] J. Holzfuss and W. Lauterborn, ''Liapunov exponents from a time series of acoustic chaos,'' Phys. Rev. A **39**, 2146–2152 (1989).
[30] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, ''Determining Lyapunov exponents from a time series,'' Physica D **16**, 285–317 (1985).
[31] J. J. Jiang and Y. Zhang, ''Chaotic vibration induced by turbulent noise in a two-mass model of vocal folds,'' J. Acoust. Soc. Am. **112**, 2127–2133 (2002).
[32] S. S. Narayanan and A. A. Alwen, ''A nonlinear dynamical systems analysis of fricative consonants,'' J. Acoust. Soc. Am. **97**, 2511 (1995).
[33] P. Lieberman, ''Some acoustic measures of the fundamental periodicity of normal and pathologic larynges,'' J. Acoust. Soc. Am. **35**, 344–353 (1963).
[34] I. R. Titze, ''Summary statement: Workshop on acoustic voice analysis,'' National Center for Voice and Speech, Denver, Colorado, 1995.
[35] M. P. Karnell, A. Chang, A. Smith, and H. Hoffman, ''Impact of signal type of validity of voice perturbation measures,'' NCVS Status and Progress Report, 1997, Vol. 11, pp. 91–94.
[36] A. Behrman, C. J. Agresti, E. Blumstein, and N. Lee, ''Microphone and electroglottographic data from dysphonic patients: Type 1, 2, and 3 signals,'' J. Voice **12**, 247 (1998).
[37] M. S. Fee, B. Shraiman, B. Pesaran, and P. P. Mitra, ''The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird,'' Nature (London) **395**, 67–71 (1998).
[38] P. Mergell, W. T. Fitch, and H. Herzel, ''Modeling the role of nonhuman vocal membranes in phonation,'' J. Acoust. Soc. Am. **105**, 2020–2028 (1999).
[39] I. Wilden, H. Herzel, G. Peters, and G. Tembrock, ''Subharmonics, biphonation, and deterministic chaos in mammal vocalization,'' Bioacoustics **9**, 171–196 (1998).
[40] T. Riede, H. Herzel, D. Mehwald, W. Seidner, E. Trumler, G. Tembrock, and G. Bohme, ''Nonlinear phenomena in the natural howling of a dog-wolf mix,'' J. Acoust. Soc. Am. **108**, 1435–1442 (2000).
[41] I. Tokuda, T. Riede, J. Neubauer, M. J. Qwren, and H. Herzel, ''Nonlinear analysis of irregular animal vocalizations,'' J. Acoust. Soc. Am. **111**, 2908–2919 (2002).

A quasi-glottogram signal

Greg Kochanski[a)]
Oxford University Phonetics Laboratory, Oxford OX1 2JF, United Kingdom

Chilin Shih
Departments of Linguistics and East Asian Languages and Culture, University of Illinois, Urbana, Illinois 61801

(Received 23 August 2000; accepted for publication 2 June 2003)

A novel, noninvasive experiment is proposed that reliably shows the strength of glottal oscillations. The quasi-glottogram (QGG) signal is generated from a microphone array that is trained to approximate the electroglottogram signal. The QGG may be useful to improve estimates of whether speech is voiced, to quantify partial voicing, and to reduce the phoneme effect when measuring the amplitude of speech signals. The technique is well adapted to the generation of text-to-speech systems, as it allows an estimate of the glottal flow during undisturbed, natural speech. For prosody studies, it can be used to provide an estimate of amplitude which is relatively unaffected by changes in phonemes, and is at least as reliable as standard estimators of amplitude. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1608964]

PACS numbers: 43.70.Jt, 43.70.Fq, 43.70.Aj [DOS]

I. INTRODUCTION

The source-filter model of speech production explains the acoustic speech signal as a convolution of the time-varying glottal airflow (due to the vibrations of the vocal folds) with the impulse response of the acoustic filter formed by the vocal tract. This model shows why a measurement of the glottal airflow is desirable: it would allow the source and the filter to be studied independently. In particular, such a measurement would allow a determination of voicing, and allow the amplitude and harmonic content of the glottal oscillations to be determined. The source-filter model of speech is embedded in speech coders, automatic speech recognition systems, and speech synthesizers; better understanding of the source and filter separately can lead to better algorithms. For example, speech synthesizers need reliable, precise indications of glottal oscillation because voiced speech is often processed differently from unvoiced speech. Human listeners will easily detect errors in the database of a synthesizer where the speech has a voicing indication that is wrong for more than 30 ms, a sensitivity that requires improvement in current techniques if one wants to create synthesizers automatically without tedious manual checking.

Further, a relatively unexplored area of speech science is the study of prosody. We use the term ''prosody'' to include all the acoustic properties of speech that are not part of the lexical information. Prosody is typically implemented by properties such as the pitch, the amplitude, and the spectral tilt of the sound. Prosody is used to mark boundaries, to emphasize words or phrases, and to help control dialogs, among other functions. However, quantitative measurements of prosody are nontrivial, because, other than pitch, all the candidate acoustic features are strongly influenced by the phoneme: for example, an emphatic /m/ may be quieter than emphasized /a/. Thus, to be able to study and understand prosody, there is a need for measurements of prosody that are relatively independent of the particular phonemes, so that one can compare prosody of different words. Again, measurements of the glottal flow would be useful, because, on one hand, the glottal flow is much less dependent on the choice of phoneme than the far-field acoustic signal is, and, on the other hand, because it can be related to physiological parameters like subglottal pressure and muscle tensions.

Invasive estimates of the glottal flow are possible, using an acoustically matched tube,[1,2] but the tube interferes with lip and jaw movements. This interference rules out applications that require simultaneous recording of natural, undisturbed speech. One such application is the recording of a database for a text-to-speech-system (TTS). Most commercial TTS systems operate by piecing together segments of recorded speech. Natural, high-quality recordings are a basic requirement, but one also needs to automatically and reliably estimate acoustic properties of the speech signals. Other estimates of glottal parameters, such as intubation to measure subglottal pressure,[3] direct photography of the vocal folds,[4,5] or photoglottography[6,7] to measure the glottal open area are also invasive and incompatible with many applications. Plethysmography[8] can measure subglottal pressure, but is cumbersome, and the acoustic properties of the box in which the subject sits need to be carefully considered in order to get clean speech. Finally, electroglottography (EGG)[9–16] is noninvasive, but measures vocal fold contact rather than glottal flow; the folds can be contacting while the glottis is partly open or noncontacting when the glottis is nearly closed. The EGG signal is only weakly correlated with the width of the glottal opening or the glottal flow, and thus is not directly related to acoustic measurements.

Historically, the glottal flow has been estimated by ''inverse filtering.'' Inverse filtering is based on first estimating the vocal tract transfer function from a microphone or anemometer signal, inverting it, and using the inverted filter to remove the effect of the vocal tract.[17] If the estimated trans-

a)Electronic mail: gpk@alum.mit.edu

 0001-4966/2003/114(4)/2206/11/$19.00

fer function is accurate, the result will be close to the actual glottal flow. However, the problem is intrinsically indeterminate, as one is trying to estimate both a time series of glottal flow and a time series of vocal tract parameters from a single input time series. In practice, assumptions are made that the microphone signal is quasi-stationary, that the spectrum of the glottal flow is simple (e.g., a periodic pulse train or a pulse train fed into a preemphasis filter), that the vocal tract can be modeled as an all-pole spectrum, and sometimes that the formant frequencies change smoothly with time. While the resulting algorithms work reasonably well, none of the approximations are perfect, and the resulting estimate of the glottal flow is only approximate, with ill-determined errors.

The goal of this study is to develop a new, noninvasive technique that allows an estimate of the volume flow of air through the glottis, *U*, based on an array of microphones and an EGG signal. We partially realize that goal by computing an estimate for *U* which is convolved with an unknown (but time-invariant) transfer function. In the remainder of the paper, this is what we mean when we speak of the QGG as an ''estimator for *U*.''

This paper will first justify why such an algorithm is possible. Second, it will lay out the details of our algorithm. Third, it will describe tests of the algorithm. Since we do not have invasive flow measurements available for a direct comparison, we bring in two lines of indirect evidence to show that the quasiglottogram (QGG) signal is closer to *U* than the standard signals used for voicing estimation. The first test is qualitative: we compare the QGG's behavior to other signals in ''difficult'' regions of speech, and show that the QGG behaves well under conditions where one or another of the standard signals misbehaves. The second test is quantitative, though indirect: To prepare, we introduce a simple ''toy'' model of speech, and show that in that model, the signals that lead to the least variable amplitude estimates are the signals that are closest to *U*. We then show that the QGG signal allows a very steady estimation of amplitude, generally less variable than the result of other standard linear estimators. This provides evidence that the QGG signal is close to *U*.

A good estimator for glottal flow should be noninvasive. It should distinguish between voiced sounds and sounds generated in other constrictions of the vocal tract. It should also be linear, and should be related to the actual glottal airflow through a time-invariant transfer function. Linearity helps simplify connecting the estimate of *U* to physiologically important parameters like the subglottal pressure and the glottal open area, and it allows straightforward quantification of partial voicing. The output of the algorithm should be related to *U* via a time-invariant transfer function, so that we can meaningfully compare signals at one time to signals at another. Particularly, one would like to compare glottal flows between different phones.

We approach these goals by building a signal from linear combinations of several filtered microphone signals. If several microphones are placed near the head and neck, they will capture different signals, containing different information. We choose the linear combinations and filters to make the best possible match to the EGG signal, which provides an instantaneous measurement of whether the vocal folds are contacting or not, and thus gives some indication whether the glottis is open or closed. We also use data from a microphone near the base of the throat to pick up a signal from the subglottal cavities. The subglottal cavities have an acoustic transfer function from the glottis through the throat wall to an outside microphone that is relatively independent of time, as (unlike the vocal tract) the trachea is not surrounded by muscles used to articulate speech. The dimensions of the trachea and bronchi are also largely unchanged during breathing, and even motions of the larynx are expected to lead to only modest changes in the acoustic resonances of the trachea.[18,19]

The algorithm requires at least two microphones, one to pick up the throat signal, and one for sound from the vocal tract. The throat signal contains a mixture of sound that propagated from the glottis down through the trachea then through the skin of the throat, mixed with sound that propagated from the glottis up through the vocal tract, then back down through free air to the throat microphone. The other microphones are used primarily to cancel out the component via the vocal tract. This cancellation results in a signal that can be thought of as the glottal waveform, filtered by its propagation through the neck. We show that this signal is less variable than commercial inverse-filtered signals, in the sense that the signal shows less phoneme-dependent variation. This stability is what one expects of the glottal waveform; it should be only weakly influenced by the vocal tract configuration.[20]

It is necessary to cancel out the sound from the vocal tract because we want an estimator of the glottal flow, and we do not want our measurement to be disturbed by the dramatic changes in the transfer function of the vocal tract that occur during normal speech. This is a different approach from an inverse-filter estimator, which attempts to dynamically estimate and invert the vocal tract transfer function: a nontrivial, multiplicative operation on one signal. We look for a signal that has a time-invariant relationship to the glottal flow via a linear operation on several signals.

II. EXPERIMENTAL METHOD

In the experiments described below, we mounted four Brüel and Kjær type 4165 omnidirectional microphones on the face guard of a hockey helmet. The microphones were mounted near the nose (4 cm lateral from the end of the nose), the side of the mouth (2 cm lateral of the corner of the mouth), near the forehead (centered, 11 cm above nose level), and near the base of the throat (on centerline, 2 cm from skin, 2 cm above the top of the sternum). The microphones were checked to be slightly outside breath streams, and were protected by 4 mm of windscreen foam. In other experiments, we found that only the placement of the throat microphone was critical: it should be placed as close as possible to where the trachea can be palpated (the *fossa jugularis*), so long as the microphone does not collide with the subject during normal head motions. The placement of the other microphones is not critical, and we have obtained similar signals from a six-microphone array (including cheek and

back-of-the-head microphones) and a three-microphone array (using a gradient microphone for the throat signal).

The EGG signal was obtained from a Portable Laryngograph.[21] The electrodes were placed to maximize the signal strength for long vowels, in modal speech, over the normal range of f_0. Data was digitized at 12 kHz per channel with an antialiasing filter. The microphone and EGG signals are high-pass filtered at 40 Hz to reduce room noise and the large-amplitude, slow components of the EGG that correspond to motions of the larynx.

All inverse filtered signals were produced by ESPS/waves.[22] They are derived from microphone **M** (near the mouth). ESPS/waves computes the linear predictor all-poles filter that will whiten each frame of **M**, then inverts the filter and applies it to **M**.

In the examples that follow, we used Mandarin speech acquired from a female subject, a native Mandarin speaker fluent in English. Because Mandarin is a tone language, Mandarin speakers can conveniently control f_0 movements. Therefore we expect that vocal fold vibration will be more repeatable, reducing unpredictable fluctuations in *U*.

The data presented in Sec. IV were taken from a random (e.g., selected for other purposes) set of English vowels and words, read at normal volume in modal speech, along with some unplanned spontaneous phrases. The data for Sec. V is a database of 979 utterances in the form "*Ta shuo X san tian*" (meaning "He says *X* for three days"). In the database, the frame syllables all have high level tone (tone 1) and *X* is a randomly selected 979 of the ~1300 allowable Mandarin syllables (here we include the tone as part of the syllable). The subject pushed a key for each prompt, and Chinese characters were presented on a screen. Speech began 0.72 ± 0.1 s after the prompt. Overall, f_0 had a mean of 235 Hz and a standard deviation of 26 Hz, with most of the variation on *X*.

Calibration utterances to train the algorithm (i.e., fit the filter coefficients) for both sections were a broad mixture of sounds, totaling 197.3 and 232.5 s for the two sessions. We used 25 calibration utterances from the beginning of the day's recording (randomly selected from ~40 recorded utterances) and another 25 at the end, ~6 h later. Half were single-syllable English words (e.g., "nap," "sap," "bee," "gold," "hold," "moo,"...). The words provided a variety of vowels and consonants in natural contexts. One quarter of the utterances were repeated, unvoiced fricatives ("ch," "f," "s," "sh," "th") with no vowel context, and repeated, aspirated "p" and "h" (again, no vowel context). These voiceless sounds are useful to force cancellation of the mouth signal and to check the operation of the algorithm, because their *U* is a relatively steady flow, compared to voiced speech.

The remainder of the calibration utterances were long vowels, nasals, and voiced fricatives ("a," "u," "v," "z," "m") where f_0 was swept either up or down. The speaker was asked to sweep at least half of her pitch range in each utterance, catching either the lowest or highest comfortable pitch. These f_0 sweep utterances provide voiced speech with the speaker's full pitch range. It is useful for the calibration utterances to cover all the f_0 range of the speech used in the experiments. The algorithm will not produce good estimates for speech that has f_0 outside the range of the training data in *x*. Ours did, having mean f_0 of 248 Hz and standard deviation 97 Hz. The sweep rates were 150 ± 70 Hz/s (average slope over the central 1 s of the utterance).

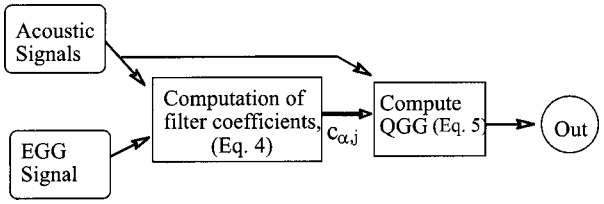


FIG. 1. Diagram of the data and signal flow in the computation of the QGG.

III. THE ALGORITHM

The overall structure of the algorithm is shown in Fig. 1. To explain the design of the algorithm, we can consider a simplified form of the microphone array as shown in Fig. 2: one microphone, **M**, near the mouth, and another, **T**, at the base of the throat. During voiced sounds, the signals are excited by the flow, *U*, through the glottis. The signal at **T** is made of two main components, one traveling through the vocal tract (transfer function *V*) to **M**, the other traveling directly through the neck[23] via a transfer function *N*. Note that *N* includes the resonances of the trachea, in addition to the transmission through the throat, which acts as a low-pass filter. Sound coming out of the mouth and then propagating down through the air to **T** has a transfer function $A \cdot V$.

In this paper, we treat these transfer functions as matrix multiplications in any complete basis, not yet specializing to a time series, frequency representation, or some intermediate basis, following Ref. 24. The throat microphone signal is then $T=(N+A\cdot V)\cdot U$. We take $M=V\cdot U$, neglecting the

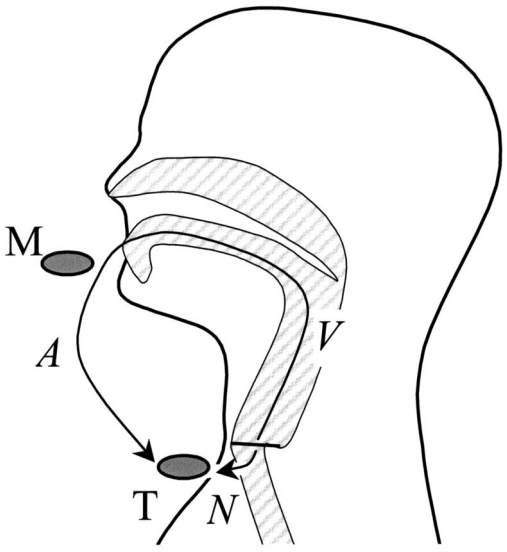


FIG. 2. Schematic of microphone placement and signal paths. **M** is a microphone near the mouth; **T** is a microphone near the throat. *V* refers to the transfer functions from the larynx along the vocal tract to **M**, and $A \cdot V$ is the transfer function from the larynx, through the vocal tract, to **T**. *N* is the transfer function from the larynx through the throat wall to **T**.

small amount of sound that comes out through the neck and propagates upwards through the air to **M**.

We expect that N and A should be nearly independent of the phoneme being produced. On the other hand, V varies dramatically and systematically as a function of the phoneme. We can express this variation by writing $V=\overline{v}+\alpha\cdot\tilde{v}$, where V depends on the phoneme only through α, and $\langle\alpha\rangle=0$ (the angle brackets denote an average over the corpus of speech). In other words, we decompose V into a constant transfer function, $\overline{v}$, and a variable part. Because human speech is the result of several independently controlled articulators, α is normally a vector. Then, $T=(N+A\cdot\overline{v})\cdot U+\alpha\cdot A\cdot\tilde{v}\cdot U$.

Hypothetically, if we had a reference signal, Q, which was derived from U via some time-invariant transfer function, q, we could find a linear filter that would reproduce Q from the microphone signals. (Note that we assume off-line processing, so our filters need not be causal.) The general form for that linear filter would be $C=c_M\cdot M+c_T\cdot T$. To find the filter matrices (coefficients) c_M and c_T, we would minimize the difference between C and Q: $c_M, c_T=\arg\min E(c_M, c_T)$, where

$$E(c_M, c_T)=(C-Q)^T\cdot(C-Q)=\|C-Q\|^2 \tag{1}$$

is the sum of squared errors between our target signal Q and our reconstruction.

If $\alpha=0$, Eq. (1) is degenerate, and its solutions are all filter coefficients on a line which we will call **L**, which goes through $c_{M1}=0$, $c_{T1}=q\cdot(N+A\cdot\overline{v})^{-1}$ and $c_{M2}=q\cdot\overline{v}^{-1}$, $c_{T2}=0$. The first of these two solutions corresponds to using a single filter on the throat microphone to match Q, while ignoring M. The second solution is the reverse: using a single filter on the mouth microphone while ignoring N. Points on **L** correspond to linear combinations of these two filters, and all points on **L** are equally good solutions and match $C=Q$ exactly, so long as $\alpha=0$ (assuming that the necessary inverses exist). This derivation can be extended to allow white additive noise in the microphones, in which case the necessary inverses always exist. The results are qualitatively similar, although the derivation and results become substantially more complex.

If the speaker starts talking, instead of just vocalizing with a stable vocal tract, the transfer functions will vary from phoneme to phoneme, and we will not be able to match Q perfectly at every moment, so E will be positive. Not every solution gets the same increment of error, though. Because V changes from phone to phone, solutions that depend predominantly on V will fit worse and have larger errors than solutions that depend predominantly on N. This difference breaks the degeneracy and typically picks out a single best solution. In general, the best solution will use signals from all microphones, and it will provide a better approximation to Q than could be obtained from any linear filter operating on any single microphone in the array.

If we assume that V varies slowly enough, we can write down the change in error due to the difference between V at a given moment and the average of V (i.e., $\overline{v}$):

$$\Delta E=\langle\|(c_T\cdot A+c_M)\cdot\alpha\cdot\tilde{v}\cdot U\|^2\rangle, \tag{2}$$

where the average (written as angle brackets) is taken over all phonemes in the corpus. The change in error is always non-negative and is normally nonzero everywhere except on a hyperplane we will call **P**, defined by $c_M=-c_T\cdot A$. Not coincidentally, this relationship between the filter coefficients is exactly what is needed to cancel out the part of the signal that came from the mouth, leaving only the part that came through the throat wall. The intersection of **L** and **P** then specifies the best estimator for Q (the one that is least sensitive to changes in the transfer function). These results generalize to arrays of more than two microphones. They also can be generalized to the case where all the transfer functions vary with time, though Eqs. (1) and (2) will change in detail.

So, given a reference signal, we can start with an array of microphones and find the linear combination of filtered microphone signals that best matches the reference signal. The resulting signal is less variable than any signal derived by a linear operator from a single microphone. Loosely speaking, the best estimator is obtained by canceling out the highly variable signal from the mouth microphone and using the part of the signal that did not propagate through the vocal tract.

In the real world, one does not have a perfect reference signal, Q. The best we can obtain noninvasively is the EGG signal. The EGG is related to U in a nonlinear and variable manner, because the larynx moves up and down relative to the electrodes used to measure the EGG. Repeating the above analysis shows that the variability of EGG measurements are not important, so long as changes in the EGG signal are not correlated with phonemes. This is true, by and large, as the larynx moves in response to pitch changes and inhalations, neither of which are correlated with most phonemes in most languages. Glottal stops and pharyngeal sounds are an exception, but they are not particularly common, typically comprising just a few phonemes in a language.

Nonlinearities in the relationship between U and the EGG signal are difficult to analyze analytically. We have investigated their effect empirically in Secs. IV and V.

A. Algorithm introduction

We first build the data matrix, X, where each row contains the signal from one of the microphones, and each column corresponds to one moment in time. X is an n by m matrix when there are n microphones, each digitized to produce m samples of audio. The EGG signal is a one by m matrix.

We then select a set of taps (i.e., taps on a delay line) for each microphone. Physically, the closure of the glottis is the cause of the acoustic signals: when it closes, a sound wave propagates up the vocal tract and down the trachea, reaching the microphones roughly a millisecond later. If one tries to observe a glottal closure at time t, one will need to use microphone data from later times, when the sound waves from the closure reach the microphones. Thus, we use the taps to build a finite impulse response (FIR) filter which will predict the present EGG signal from future microphone signals.

To select the span of the taps, we need to consider the purposes of the filter we are building. It needs to cancel out

the mouth signal that is picked up by the throat microphone, and it needs to match the impulse response of the remaining part to the EGG signal.

The taps span the range of delays beginning with the earliest propagation from the glottis to the microphone in question, ending when the impulse response of the vocal tract goes below 1% of its peak value.

Vocal tract formant bandwidths can be as small as 40 Hz[25,26] when the glottis is closed, though in real speech a bandwidth of 80 Hz is more realistic.[27] Such a bandwidth implies that the vocal tract resonances will take about $3 \cdot 1/(2\pi B) \approx 6$ ms to decay. Bandwidths for the tracheal resonances are wider, 200–400 Hz,[28] and so are not the limiting factor for the window width in the time domain. Matching the acoustic to the EGG signal requires a window length of about $1/f_0$, which can be slightly longer. We choose the range of taps to cover the longest of these times (13 ms in these experiments), and we use the same number of taps for every microphone.

The resulting QGG signal does have some imperfections. It has a nonzero (though relatively small) response to fricatives and plosives. Also, the QGG algorithm does not estimate *U* directly, but estimates *U* times a transfer function, where the transfer function depends on the subject and the experimental configuration. Finally, the algorithm does not adapt to changes in microphone positions relative to the mouth, nose, and throat, so stably mounted microphones are presumably required for good results.

Note that we do not claim any absolute calibration for the QGG signal. Because the QGG filter coefficients involve the EGG signal, the QGG amplitude will differ from person to person and session to session, depending on neck structure and the placement of the EGG electrodes. However, the QGG signal depends on the EGG only through its average properties during the calibration/training session. So, when one is actually using the QGG (as opposed to calibrating it), the EGG is entirely unused and may be disconnected. Therefore, factors that affect the EGG signal during use (such as the momentary position of the larynx with respect to electrodes) will have no effect on the QGG.

There are similarities between the QGG signal and the low-pass filtered mouth signal. Both are generated by linear, time-independent filters. Indeed, in practice, the QGG signal also has less high-frequency power than microphone **M** (though more than the output of a high-order low-pass filter). The steep QGG spectrum has its origins in three places: the EGG signal that it is matching has relatively less high-frequency power than speech, and high-frequency components radiated from the mouth tend to have phases that are strongly dependent on the phone, consequently there is little cross-correlation between the high-frequency EGG components and the high-frequency components from **M** and thus they are deemphasized. Finally, microphone **T** picks up some signal with a stable phase relationship to the EGG, but high frequencies are not effectively coupled out through the throat wall.

However, we expect that the QGG should give a more consistent amplitude measurement than the low-pass filter. Since the low-pass filter signal essentially measures the amplitude of f_0 in microphone **M**, it will be strongly boosted when the first formant is close to f_0. The QGG signal should not be affected by this problem, as it is measuring a combination of signals chosen to match the EGG's relatively stable amplitude: roughly speaking, it is measuring the subglottal cavities through the throat wall, and the formant structure of the vocal tract has little effect down below. The QGG ought also to be more consistent than a low-pass filter if the speaker has a pitch range that swings through more than an octave. In that case, the low-pass filter will sometimes pass the second harmonic, and sometimes not, leading to more amplitude fluctuation.

B. Implementation

We build a set of linear equations corresponding to the FIR filter that best predicts the EGG signal, using straightforward least-squares linear prediction techniques.[29] The predicted EGG signal at each time is a linear combination of $n \cdot q$ values (q taps on each of n microphones). To start, we define the covariances between shifted signals from the ith and jth microphones:

$$\phi_{\alpha,i,j} = \frac{1}{m} \sum_t x_{i,t-\alpha} \cdot x_{j,t}, \tag{3}$$

where α is the time shift between the ith and jth microphones (we neglect end effects, for simplicity) and t indexes the time. These covariances are estimated from the data and contain noise covariances. Analogously, we will write $\phi_{*,\alpha,j}$ for the covariances between the EGG signal and the shifted microphone signals. The filter coefficients that minimize the mean squared error are then the solution to

$$\sum_i \phi_{\alpha,i,j} \cdot c_{\alpha,i} = \phi_{*,\alpha,j}, \tag{4}$$

which is a set of $n \cdot q$ linear equations. We prepare to solve the equations by stacking the $c_{\alpha,i}$ to make a single vector C, stacking $\phi_{*,\alpha,j}$ to make a single vector P (of size $n \cdot q$ by 1), and placing the elements $\phi_{\alpha,i,j}$ into the corresponding places $H_{\alpha n+i,j}$ to make a $n \cdot q$ by n matrix. We solve the resulting matrix equation, $HC = P$, with a singular value decomposition algorithm to allow for degeneracies and near-degeneracies. C can then be unpacked to yield the $c_{\alpha,i}$, which are the filter coefficients that give the best prediction of the EGG signal from the set of microphone signals.

We can now calculate the QGG signal,

$$p_t = \sum_{\alpha,i} x_{i,t-\alpha} \cdot c_{i,\alpha}, \tag{5}$$

from the $c_{\alpha,i}$ and microphone signals. The prediction is not at all precise because the EGG signal is a strongly nonlinear function of the glottal opening: it contains little information beyond the simple fact of whether the vocal folds are touching or not. It would be surprising indeed if one could build a linear filter that would exactly match the EGG signal.

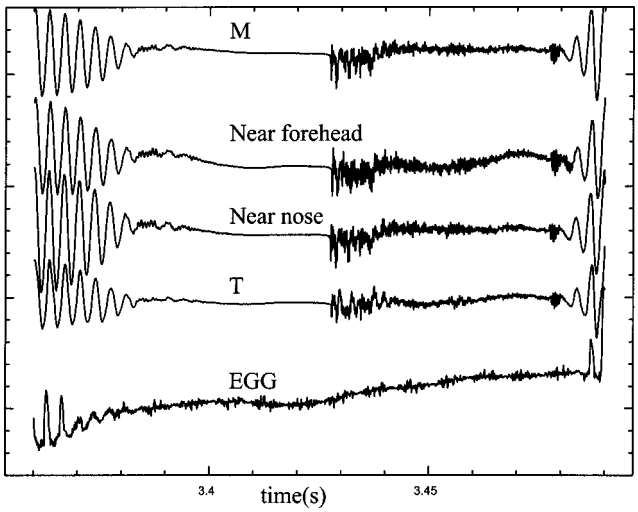


FIG. 3. Raw signals, used as input to the QGG algorithm. From top to bottom, the signals are from microphones near the mouth, near the forehead, near the nose, and near the throat, and (bottom) the electroglottogram. This section of speech corresponds to the ''t'' in the midst of ''...what is... .'' The throat microphone signal, **T**, has a noticeably smaller ratio of high-frequency power near 3.45 s to power in the voiced regions (edges) than the other microphones. That is because the high frequencies get to **T** only from the mouth, whereas the low frequencies also come through the throat wall.

C. Examples

Figure 3 shows the raw data for all four microphones and the EGG signal for a small section of speech. The display is centered about the burst of a ''t'' and shows voiced regions (at the edges), a burst, and then a fricativelike region after the burst. Note that the voiced regions are quite similar in all four microphones (though the throat microphone, **T**, is phase shifted by about 30°). On the other hand, the frictive-like region (3.44–3.46 s) has a substantially different structure at microphone **T**, and also has an amplitude that is relatively small, about 3 dB lower, relative to the voiced regions than the other three microphones.

Figure 4 shows the output of the QGG and several comparison signals in the same region around the ''t'' burst. Figure 5 shows the QGG and comparison signals on a longer stretch of speech.

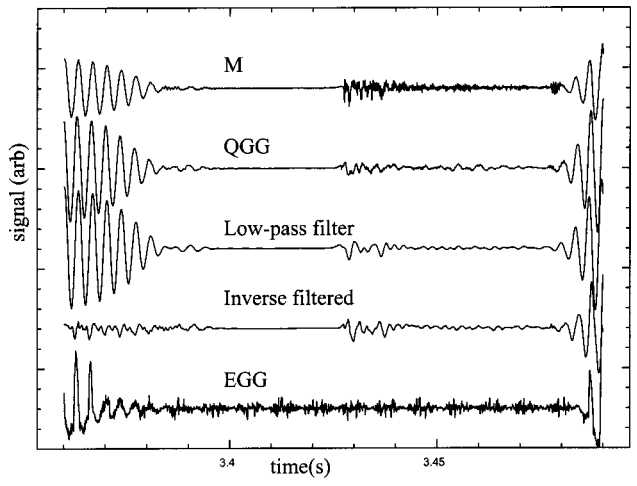


FIG. 4. Output of the QGG algorithm, compared to other standard signals. From top to bottom, the signals are from the microphone near the mouth (**M**), the QGG, a low-pass filter set just above peak f_0, a standard inverse-filtering algorithm, and the EGG signal. This section of speech corresponds to the ''t'' in the midst of ''...what is... ,'' and matches Fig. 3.

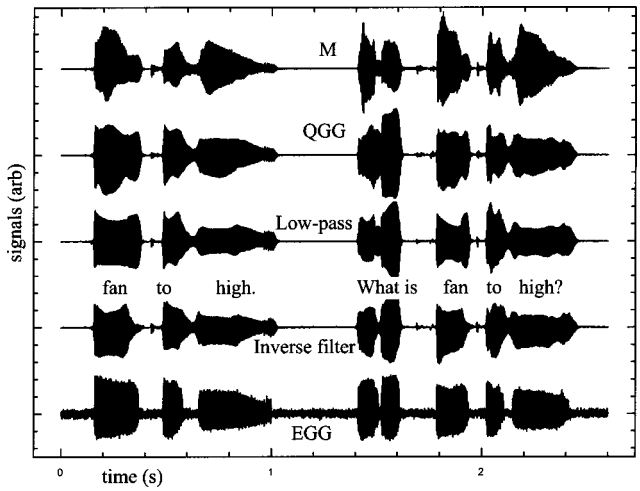


FIG. 5. Output of the QGG algorithm, compared to other standard signals. From top to bottom, the signals are from the microphone near the mouth (**M**), the QGG, a low-pass filter set just above peak f_0, a standard inverse-filtering algorithm, and the EGG signal. The text is marked on the figure, approximately aligned with the audio. (The pause at 1.2 s has been shortened for a better visual display.)

IV. QGG AS FODDER FOR VOICING ESTIMATION

An engineering evaluation of the QGG signal as input for a voicing estimation algorithm is beyond the scope of the paper. Instead, we will show that (at least under some conditions) the QGG signal can provide a more reasonable indication of the presence of voicing than either the EGG or the inverse-filtered mouth signal. This is the first, qualitative, test of the QGG.

The advantages of the QGG signal follow from its construction: because it is a linear function of the pressure near the glottis, it is well behaved during startup and shutdown of the glottal oscillator. So, unlike the EGG, it may be able to quantify partial voicing and mark onsets of voicing precisely. Because the QGG is constructed from a time-invariant filter operating on acoustic signals, it may be more robust than algorithms based on an inverse filter (we do not discuss manual adjustment of inverse filter coefficients here, as such techniques are impractical for large speech corpora). Any time the spectral estimation step of an inverse filter fails to produce a good result, or any time the speech signal is not well represented by an all-poles transfer function, one expects the inverse-filtered signal will not reflect the glottal state. The QGG does not suffer from those problems.

Figures 6–8 show examples of speech signals where the glottal oscillation is starting or stopping. The figures show that the QGG signal can sometimes provide a much better explanation of the acoustic signal than does the EGG signal. Limitations of EGG signals have previously been described elsewhere.[30,31]

In Fig. 6, the envelope of the QGG signal tracks acoustic power (the mouth signal), while the EGG signal shows an unnaturally sharp onset/ending. If the glottal oscillation stopped with the EGG signal, one would have to assume a

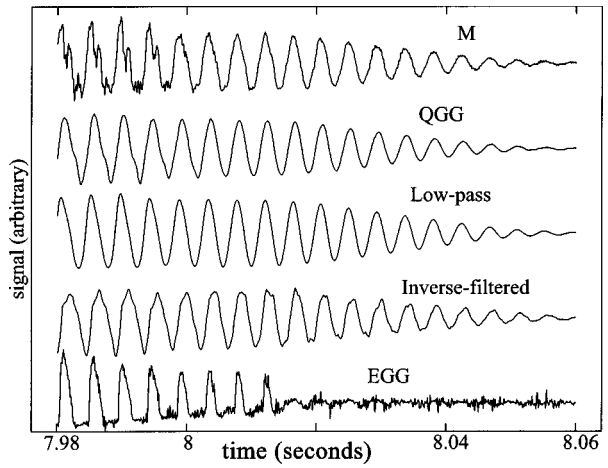


FIG. 6. Comparisons of estimators of the glottal waveform. The signal is the off-glide [aɹ] in "high." **M** is the raw signal from the microphone near the mouth, while "low-pass" and "inverse-filtered" are processed versions of **M**. QGG is derived from a microphone array, and EGG is the electroglottogram. The EGG stops more than 30 ms before the end of the glottal oscillation.

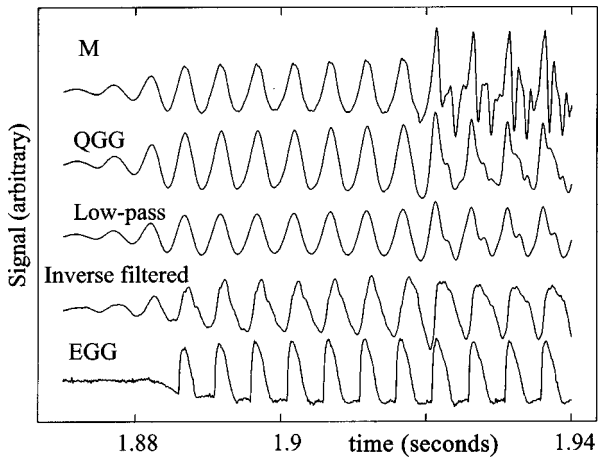


FIG. 8. Beginning of the word "mosey," at normal amplitude. Note that the EGG signal starts late. **M** is the raw signal from the microphone near the mouth, while "low-pass" and "inverse-filtered" are processed versions of **M**. QGG is derived from a microphone array, and EGG is the electroglottogram.

bandwidth for the first formant of only 5 Hz for the acoustic signal to persist as long as it does, which is incompatible with the known bandwidths.[25–27] The vocal tract simply is not a good enough resonator for the sound to persist 30 ms after the end of glottal oscillation. Therefore, glottal oscillation must be continuing at a lower level (so the vocal folds do not touch) without showing up on the EGG. Titze[32] and Stevens[33] have discussed this kind of small oscillation.

Other observations, such as the events around $t \approx 8.015$ s in, again, Fig. 6 can also be explained similarly. Imagine comparing two similar vowels, one uttered with amplitude just small enough so that the glottal folds do not collide, and the other uttered with slightly more amplitude so that the folds do collide on each cycle. If we consider a decomposition of the signal into a stack of harmonics at f_0, $2f_0, 3f_0, 4f_0, \ldots$, the lowest harmonic will primarily measure the total air flow per cycle, and will change only gently and continuously when the vocal folds begin to collide.

However, the higher harmonics do not behave smoothly. Below the collision threshold, the oscillation is close to a simple harmonic oscillator, and there is little power in the harmonics. Above threshold, there is a strong nonlinearity when the vocal folds collide; large amounts of power suddenly start to be generated in the harmonics. This is the situation that is described by the two-mass model,[34,35] which typically gives a spectrum where the amplitude of the nth harmonic scales as $a_n \propto n^{-1}$, or a 6 dB/oct slope. Such a change can be seen in the spectra shown in Fig. 7.

The effect is not confined to the low-amplitude tails of voiced sounds. For example, in Fig. 8, an acoustic signal begins two periods before the first EGG activity. Again, one must assume that the vocal folds are oscillating but not yet completely closing.

Figure 9 shows a section of a low-amplitude, sustained "o" as an extreme example where the EGG fails to explain the acoustic signal. Several times, the amplitude of the EGG signal jumps up dramatically for a few milliseconds. Little effect is seen in the acoustic signal, other than an increase in the power of the higher harmonics, nor is much change seen in the QGG signal. Hypothetically, the vocal folds could be oscillating within a whisker of touching, and some perturbation intermittently reduces the spacing just enough to make

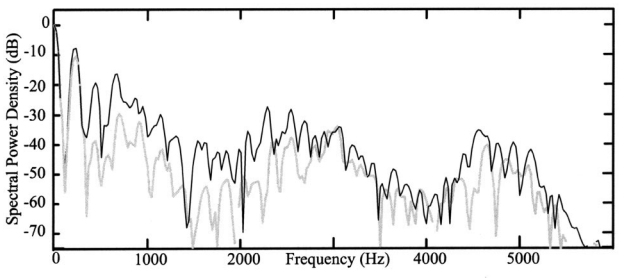


FIG. 7. DFT spectra (16-ms windows) on either side of 8.015 s in Fig. 6. The thick, black curve is before 8.015 s (the window ends at 8.015 s), when the glottal folds are colliding, and the wide, gray curve is after (window begins at 8.015 s). The fundamental (225 Hz) is essentially unchanged in amplitude, but the power in the harmonics drops by about 10 dB.

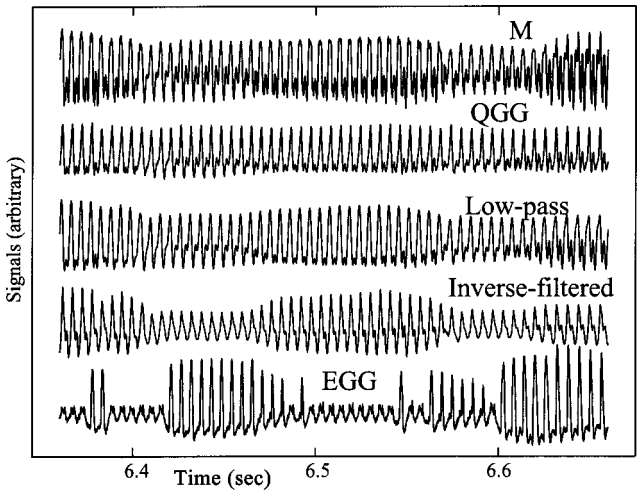


FIG. 9. The middle of sustained low-amplitude "o" phonation showing major EGG changes (bottom) without large changes in the speech signal (top). The various signals are labeled.

them collide. The noteworthy observation here is that the QGG signal is a better predictor than the EGG of whether the acoustic signal shows voicing. This should not be too surprising, as the QGG is constructed from a set of acoustic signals.

These problems we have displayed are not hard to find, occurring at these levels in 6 of 304 voiced syllables inspected. Because the problems seem to be associated most with glottal oscillations where the vocal folds do not contact, we expect that EGG problems should be much more common in languages that make more extensive use of murmurs (i.e., a ''breathy'' voice quality), most notably Hindi. The inverse filtered signal also tends to behave badly for low-amplitude voicing or other conditions where the power in the higher harmonics is very low. Among the displayed signals, the QGG provides signals that display a strong contrast between voiced and unvoiced regions, and have most of their power in the fundamental to reduce the likelihood of octave errors in any pitch tracker to which the QGG signal is fed.

As input to a voicing estimator, many of the advantages of the QGG signal are shared with the low-pass filtered mouth signal. However, we would expect that the suppress bursts (e.g., p, t, k) and be a closer representation of the state of the larynx during closures.

V. QGG AS A MEASURE OF AMPLITUDE OR EMPHASIS

The quasi-glottogram signal is valuable for more than correcting voicing errors. It also provides an estimate of the amplitude of the oscillatory flow through the glottis. We expect that this amplitude will be a better predictor of prosodic emphasis and a better measure of the speech effort being expended by the speaker than is the total acoustic power of the mouth signal, the inverse filtered mouth signal, or the EGG.

Amplitude has been known to be a significant component of prosody since the 1950s,[36–40] and into more recent literature.[41–45] However, all these studies have been severely limited by the large intrinsic variability of speech amplitude measurements. The experimental designs (e.g., ANOVA analysis on p. 190 of Ref. 41) invariably compare the prosodic effect in question to the unpredictable variations. Reducing this variability can be seen to be just as good as having a larger effect to measure. This is one value of the QGG: it allows a cleaner, low-variance amplitude measurement, and should lead to more conclusive experiments.

A. Model of amplitude variance

To justify our intuition that the QGG signal will allow better amplitude measurements, consider a toy model of the speech apparatus: a glottal source that drives the vocal tract, which we model as a time-varying filter. Loosely speaking, the variability of the amplitude outside the mouth comes from two sources: intrinsic variability in U and changes in the coupling through the vocal tract transfer function, V. Since the two variances add, the variability of the mouth power will be greater than the variability of glottal power. Consequently, we expect that the best linear estimators of the glottal source should have the lowest variability. We can use this as a figure of merit to compare algorithms: less variable estimators are better and probably closer to the glottal signal.

We will work in a short-time Fourier transform basis to conveniently describe speech like signals. Signals are then indexed with two parameters: a time index, t, which locates the transform's window, and a frequency index, ω, for the low-resolution spectrum in the window. In this toy model, the glottal source changes its amplitude but not its spectral shape: $U(t,\omega)=f(t)\cdot g(\omega)$, and we assume that the amplitude, $f(t)$, changes slowly compared to structure in the glottal spectrum, $g(\omega)$. We can safely assume $\Sigma_\omega|g(\omega)|^2=1$ without restricting the model, as the overall amplitude can go into $f(t)$.

Next, we can write a time- and frequency-dependent transfer function for the vocal tract: $h(\omega,\phi(t))$, where $\phi(t)$ is the vocal tract configuration (roughly, the phone) at time t. The pressure outside the mouth is then $s(t,\omega)\approx f(t)\cdot g(\omega)\cdot h(\omega,\phi(t))$, and we can sum over frequency to get the rms amplitude of the mouth signal: $a^2(t)=\Sigma_\omega|s(t,\omega)|^2=f^2(t)\cdot z^2(\phi(t))$, where $z^2(\phi)=\Sigma_\omega|g(\omega)\cdot h(\omega,\phi)|^2$ shows how efficiently power is coupled from the glottal source out through the mouth for a particular phone ϕ.

We can now take the log of the power to write

$$\log(a(t))=\log(f(t))+\log(z(t)). \tag{6}$$

As long as variations of the vocal tract are uncorrelated with changes in the larynx, the variances of the two right-hand terms add, and we can conclude that $\mathrm{var}(\log(a))=\mathrm{var}(\log(f))+\mathrm{var}(\log(z))$. Since $\mathrm{var}(\log(z))>0$, $\mathrm{var}(\log(a))>\mathrm{var}(\log(f))$. In other words, the amplitude outside the mouth is always (in this toy model) more variable than the amplitude of U. The same conclusion follows if you consider h to include the vocal tract plus an arbitrary linear operator: thus any filtered version of the mouth signal will still be more variable than U.

The limits to the assumption that f and z are uncorrelated come from two sources: First, the speaker's intentional muscle motions can cause correlations between the glottis and the vocal tract (e.g., a hypothetical language might specify that high vowels are always spoken in a pressed voice). Second, some vocal tract configurations with tight constrictions can change the glottal waveform. However, neither circumstance seems common.

As a concrete example of this, consider a vowel where f_0 matches the first formant frequency, F_1. Acoustic power will then be efficiently coupled from the glottis out the mouth, because a peak of $g(\omega)$ matches with a peak of $h(\omega,\phi(t))$, z will be large, and the amplitude at the mouth will therefore be large. On the other hand, if $3/2f_0=F_1$ with the same amplitude of U, the fundamental frequency will be below the first resonance of the vocal tract, z will be small, and the amplitude at the mouth will be small. Comparing the two cases, we see that substantial variance in amplitude can be generated as the sound wave propagates through different configurations of the vocal tract.

This toy model contains several loose assumptions and should not be taken too far, but it does give important clues for finding good algorithms, since the mathematics remains valid if the transfer function, h, includes the behavior of the

TABLE I. Variability of speech amplitude, after processing by four algorithms. Region *X* is the variable syllable, and shows larger variability because the amplitude is a function of the phoneme in *X*.

Fractional standard deviation of amplitude	Region 1 *ta*	Region 2 *shuo*	*X* *X*	*X′* (tone 1 only)	Region 3 *san*	Region 4 *tian*
Raw	0.22	0.16	0.48	0.32	0.31	0.29
Low-pass filtered	0.29	0.23	0.42	0.34	0.32	0.35
Inverse filtered	0.14	0.14	0.42	0.24	0.22	0.24
QGG	0.09	0.09	0.37	0.16	0.19	0.20

microphones and a linear signal processing algorithm. For example, if h is time-invariant, var(log(z)) will be zero, and the variability of the amplitude estimate, var(log(a)), will be as small as possible. Conversely, if s is always close to g, the transfer function must be near unity, so that var(log(z)) must be small and thus var(log(a)) will be as small as possible. So, we expect that algorithms that are good glottal estimators will give stable amplitude estimates and vice versa. Note that for further arguments, we do not require any of the details of the toy model, merely this conclusion, which is independently testable, and likely to be truly independent of the model.

Purely pragmatic considerations will also lead us to the same figure of merit. If one is studying prosody, then any variation of amplitude that is a function of the phoneme should be considered noise: it prevents one from comparing prosodies of different words. Good comparisons are only possible if the amplitude measure is independent of phoneme. So, we would like to improve the signal-to-noise ratio of prosody experiments by reducing the noise, which again means finding an amplitude estimator that is less variable.

B. Comparison of QGG and acoustic power variance

We conducted a second, quantitative test. This test directly establishes the usefulness of the QGG signal as a measure of amplitude prosody. We also show that the QGG signal allows a very steady estimation of amplitude, less variable than the result of other standard linear estimators. Following the logic in Sec. V A, this test indirectly establishes that the QGG signal is a reasonable estimator of a filtered version of *U*.

We used the database of 979 utterances in the form "*Ta shuo X san tian*" described. We calculated the QGG for all the utterances, along with an inverse-filtered **M**, low-pass filtered **M**, and the raw **M**. The boundaries of the variable syllable (*X*) were hand-segmented, and an algorithm (ESPS/WAVES *get_f*$_0$) was run to find the two voiced regions on the left side of the segmented area (assumed to correspond to the vowels in "*ta*" and "*shuo*"), the voiced region inside *X*, and the two voiced regions to the right of *X* (assumed to correspond to the vowels in "*san*" and "*tian*"). Four of the utterances were voiced through between "*ta*" and "*shuo*" and were dropped because we could not assume that the voiced region matched the vowels. We then calculated the mean power near the center of the five voiced regions, using a cosine window.

Table I shows the standard deviation of log(power) for each combination of region and signal. In every case, the QGG signal is more reproducible than the others, yielding (on average) a standard deviation 46% smaller than the corresponding low-pass filtered speech, 38% smaller than unprocessed speech, and 20% less than the inverse filtered signal. These improvements in SNR are conservative limits, as the speech contains some intrinsic variability that cannot be removed by signal processing.

Comparisons of cells in Table I have several implications. First, one expects the variation of the frame (regions 1–4) to be small, where we always have the same syllable in the same position. In contrast, the variation in the *X* region should be bigger, because the syllable identity changes. This is reflected in Table I, where the *X* region shows the largest variation under all conditions.

Why is the QGG variability not smaller in the *X* region, if it indeed removes the effect of changes in the vocal tract? It is not smaller because, unfortunately, the QGG is not an estimator for *U*, but rather (see Sec. III) a filtered version of *U*, $q \cdot U$, where the filter, q, is time invariant but can be dependent on frequency. The QGG estimate can therefore vary with the f_0 of the speech.

We tested this by choosing a subset of the data that all have similar f_0 and recalculating the standard deviation of region *X*. We chose all 260 syllables that have Mandarin tone 1. This is a high, level tone, which is the same tone that occurs in regions 1–4. In this subset of the data, the f_0 trajectory is relatively flat in each syllable, as well as across the whole utterance. The results are displayed in column *X′* of Table I. The standard deviation of the amplitude of all the signals is lower, but the QGG drops most dramatically.

One can also see both the pitch dependence of the QGG and its relative insensitivity to the vocal tract configuration in Fig. 10. This is a scatter plot of the mean f_0 of all syllables versus the measured amplitude of the QGG signal. We used the same window (as defined above) to calculate the mean f_0 as was used for the amplitude measurement. Syllables with tone 1 are seen as a tight cluster in the upper right corner. That cluster spans the full range of phonemic variation, covering all vowels in combination with a variety of consonants and glides. The other syllables in Mandarin (shown as dots) have either low pitch or they are rising from or falling to low pitch, thus they have an average pitch below that of tone 1. There is a clear trend of increasing amplitude measurement with increasing pitch, perhaps as the result of q. One could empirically correct for this trend, if one knew how much of the effect is the result of q, and how much is the result of the speaker's glottal flow changing as a function of pitch.

While the QGG algorithm dramatically reduced the variation of regions 1 and 2 (to around 10% in Table I), it was less successful in regions 3 and 4. This suggests that

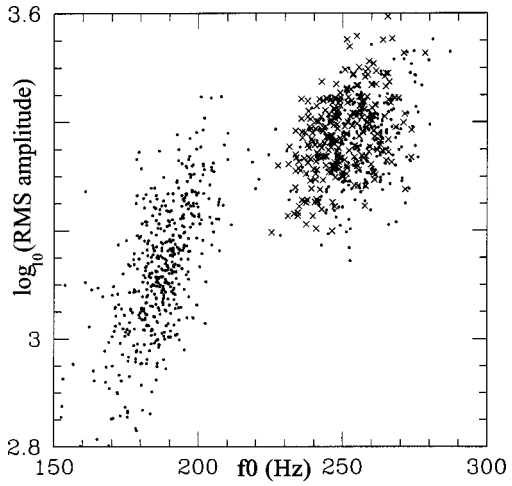


FIG. 10. Amplitude measurement (from QGG signal) versus mean pitch of a syllable. Syllables with tone 1 (used in column X' of Tables I and II) are shown as "×." Other tones are shown as dots. In this plot, the pitch dependencies (from one tone to another) are larger than the dependence on the syllable, within the tone 1 syllables.

there is more inherent variability in regions 3 and 4, which may well be some carry-over from the pitch and phoneme change in region X. This larger inherent variation can also be seen for all the signals, although not as clearly because the other algorithms do not yield as reproducible an amplitude signal as the QGG.

It is possible that much of the variability in regions 1 and 2 is external to the speaker. The major source of variability outside the speaker is expected to be the motion of the throat microphone relative to the throat if the speaker tilts the head forward or back, because the throat microphone was fixed to the helmet. We do not have accurate estimates of this effect, as head motions were not measured. However, we estimate (after the fact) that 2-mm changes in throat to throat-mic distance are quite plausible, and these could yield amplitude changes of 5%–10%. A secondary effect is that the design of our microphone array allows the microphones to move about 1 mm relative to the skull. This movement is expected to lead to ~3% changes in values of the transfer functions (changing some more than others), which would lead to comparable changes in measured amplitude.

C. QGG for amplitude prosody

As a further test with the same database (Table II), we attempt to eliminate any changes in amplitude by predicting the amplitude of each region in terms of the amplitudes of the other regions. Essentially, this normalizes the measured amplitudes to the rest of the utterance, and would eliminate the effect of a uniform change in amplitude from utterance to utterance. The goal here is to reduce the variability deriving from the experimental subject (e.g., from changes in the volume of inspired air) and focus more tightly on variations that result from the signal processing.

We fit a least-squares linear predictor to the logs of the amplitudes, e.g.,

$$\log(\hat{A}_X) = b_0 + b_1 \cdot \log(A_1) + b_2 \cdot \log(A_2) + \cdots, \qquad (7)$$

TABLE II. Unpredictable variability of speech amplitude, after processing by four algorithms. Here, we use the subset of the data where X has tone 1 (a high level tone), so that the pitch matches regions 1–4. The amplitude in each region was predicted in terms of the other four regions; the table shows the remainders.

Fractional standard deviation of amplitude, after linear prediction	Regions 1 and 2 *ta shuo*	X' (tone 1 only)	Regions 3 and 4 *san tian*
Raw	0.16	0.32	0.11
Low-pass filtered	0.09	0.16	0.06
Inverse filtered	0.14	0.23	0.10
QGG	0.08	0.14	0.09

and then measure the rms size of the residual, $\log(A_X) - \log(\hat{A}_X)$. Here, A_R is the amplitude of the Rth region (a rms average of the signal inside a cosine window), and $\hat{A}$ is the predicted amplitude, based on the other regions. The predictor is a five-parameter linear fit, and we fit separate predictors for regions 1, 2, X', 3, and 4. Again, the QGG leads to a nicely small variance: it has predictable amplitude. Surprisingly enough, the low-pass filtered signal is comparably predictable, even though its performance before prediction (Table I) is quite poor.

The most important terms are generally those that predict A_3 in terms of A_4 and vice versa. After prediction, the frame regions on both sides of X have similar variabilities, approximately 8% for the QGG signal. This remainder seems intrinsic to the speaker. The excess variation in regions 3 and 4 is gone, even though the variation in regions 1 and 2 is practically unchanged. This suggests that amplitude variations in X, which are driven by phoneme and f_0 changes, carry forward into the following syllables.

Because region X' contains a diverse set of different syllables while the frame regions (regions 1–4) always have the same syllable in the same position, we expect more variability in the amplitude of X'. Table II shows this. However, changing syllables only disturbs QGG amplitude measurements by 11% (beyond the intrinsic 8%), so long as the pitch is reasonably stable.

VI. SUMMARY

We have shown that the QGG algorithm can produce a useful, noninvasive estimate of the glottal flow (more precisely, the glottal flow filtered by an unknown linear filter). When used to analyze speech, it can be well behaved under conditions where EGG and inverse-filtered signals would lead to errors in voicing estimation. It also yields substantially more stable amplitude measurements than other techniques. Amplitude measurements of repeated words using the QGG signal can have a fractional standard deviation as small as 9% (0.4 dB), and the fractional standard deviation across different syllables in the same location in the sentence is just 37% (1.4 dB), with much of that standard deviation being attributable to intonation differences.

The QGG algorithm should find applications in studies of the amplitude part of prosody. We also see applications in

text-to-speech systems, where there is a need for reliable automatic processing of speech data, and possibly in medical screening or diagnostics of voice disorders.

ACKNOWLEDGMENTS

We acknowledge Robert Kubli and Gary Elko for equipment and advice, Mohan Sondhi and Joseph Hall for discussions, and Ronald J. Baken and three reviewers for substantial comments. This work was begun at Bell Laboratories, and we gratefully acknowledge Lucent Technologies' financial support. G.P.K. would like to thank the Rutgers University Department of Chemistry and Biological Chemistry.

[1] M. M. Sondhi, ''Measurement of the glottal waveform,'' J. Acoust. Soc. Am. **57**, 228–232 (1975).
[2] R. J. Baken and R. F. Orlikoff, *Clinical Measurement of Speech and Voice* (Singular, San Diego, CA, 2000), pp. 431–432 and references therein.
[3] R. J. Baken and R. F. Orlikoff, *Clinical Measurement of Speech and Voice* (Singular, San Diego, CA, 2000), pp. 311–322 and references therein.
[4] D. W. Farnsworth, ''High-speed Motion Pictures of the Human Vocal Cords,'' Bell Lab. Rec. **18**, 203–208 (1940).
[5] R. J. Baken and R. F. Orlikoff, *Clinical Measurement of Speech and Voice* (Singular, San Diego, CA, 2000), pp. 394–406 and references therein.
[6] B. Sonneson, ''On the anatomy and vibratory pattern of human vocal cords with special reference to a photo-electrical method for studying the vibratory movements,'' Acta Oto-Laryngol., Suppl. **156**, 1–81 (1960).
[7] B. R. Gerratt, D. G. Hanson, G. S. Berke, and K. Precoda, ''Phottoglottography: A clinical synopsis,'' J. Voice **5**(2), 98–105 (1989).
[8] R. J. Baken and R. F. Orlikoff, *Clinical Measurement of Speech and Voice* (Singular, San Diego, CA, 2000), pp. 322–324 and references therein.
[9] R. J. Baken and R. F. Orlikoff, *Clinical Measurement of Speech and Voice* (Singular, San Diego, CA, 2000), pp. 413–427 and references therein.
[10] P. Fabre, ''Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation: glottography de haute fréquence. Premiers résultats (A procedure for the electrical recording of the glottal state during phonation: first results from high frequency glottography),'' Bull. Acad. Natl. Med. **141**, 66–69 (1957).
[11] R. F. Orlikoff, ''Assessment of the Dynamics of Vocal Fold Contact from the Electroglottogram: Data from Normal Male Subjects,'' J. Speech Hear. Res. **34**, 1066–1072 (1991).
[12] B. Sonesson, ''On the anatomy and vibratory pattern of the human vocal folds,'' Acta Oto-Laryngol., Suppl. **156**, 1–80 (1960).
[13] R. Timcke, H. von Leden, and P. Moore, ''Laryngeal vibrations: Measurements of the glottic wave: Part 1. The normal vibratory cycle,'' Arch. Otolaryngol. **68**, 1–19 (1958).
[14] T. Baer, A. Löfqvist, and N. S. McGarr, ''Laryngeal vibrations: A comparison between high speed filming and glottographic techniques,'' J. Acoust. Soc. Am. **73**, 1304–1308 (1983).
[15] T. Baer, I. R. Tize, and H. Yoshioka, ''Multiple simultaneous measures of vocal fold activity,'' in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. M. Bless and J. M. Abbs (College Hill, San Diego, CA, 1983), pp. 229–237.
[16] H. R. Gilbert, C. R. Potter, and R. Hoodin, ''Laryngograph and a measure of vocal fold contact area,'' J. Speech Hear. Res. **27**, 556–565 (1984).
[17] R. J. Baken and R. F. Orlikoff, *Clinical Measurement of Speech and Voice* (Singular, San Diego, CA, 2000), pp. 433–436, and references therein.
[18] K. Ishizaka, M. Matsudaira, and T. Kaneko, ''Input acoustic-impedance measurement of the subglottal system,'' J. Acoust. Soc. Am. **60**, 190–196 (1976).
[19] Z. Milutinović, M. Mijić, and S. Djurica, ''Activity of the subglottic voice ('chest resonator'): an echo-tomographic and acoustic study,'' Eur. Arch. Otorhinolaryngol. **254**, 292–297 (1997).
[20] K. Stevens, *Acoustic Phonetics* (MIT, Cambridge, 1998), pp. 66–68, 76, 94.
[21] Laryngograph, Ltd., 1 Foundry Mews, Tolmers Square, London NW1 2PE, U.K.
[22] Entropic Research Laboratory, ESPS/Waves release 5.31. Code attributed to Brian Sublett, John Shore, David Talkin, and Derek Lin. It does a 14th order LPC analysis with a Hanning window, with a preemphasis coefficient of 0.97. A descendant of the code is available at http://www.speech.kth.se/wavesurfer (2003).
[23] K. Stevens, *Clinical Measurement of Speech and Voice* (Singular, San Diego, CA, 2000), pp. 200–202.
[24] S. Weider, *The Foundations of Quantum Theory* (Academic, New York, 1973), pp. 44–59.
[25] G. Fant, ''Formant Bandwidth Data,'' *Speech Transmission Laboratory Quarterly Progress and Status Report 2–3* (Royal Institute of Technology, Stockholm, 1962), pp. 1–3.
[26] O. Fujimura and J. Lindqvist, ''Sweep-tone measurements of vocal tract characteristics,'' J. Acoust. Soc. Am. **49**, 541–558 (1971).
[27] K. Stevens, *Clinical Measurement of Speech and Voice* (Singular, San Diego, CA, 2000), pp. 152–167, 264.
[28] K. Stevens, *Clinical Measurement of Speech and Voice* (Singular, San Diego, CA, 2000), pp. 196–198 and references therein.
[29] W. H. Press, S. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge U.P., Cambridge, 1992), pp. 564–565.
[30] R. H. Colton and E. G. Conture, ''Problems and pitfalls of electroglottography,'' J. Voice **4**, 10–24 (1990).
[31] D. G. Hanson, B. R. Gerratt, and G. S. Berke, ''Frequency, intensity, and target matching effects on photoglottic measures of open quotient and speed quotient,'' J. Speech Hear. Res. **33**, 45–50 (1990).
[32] I. R. Titze, ''The physics of small-amplitude oscillation of the vocal folds,'' J. Acoust. Soc. Am. **83**, 1536–1552 (1988).
[33] K. Stevens, *Clinical Measurement of Speech and Voice* (Singular, San Diego, CA, 2000), p. 91 (1988).
[34] K. Ishizaka and M. Matsudaira, ''What makes the vocal cords vibrate?'' in *Proceedings of Sixth International Congress of Acoustics*, Tokyo (1968), pp. B1–B3.
[35] K. Ishizaka and J. L. Flanagan, ''Synthesis of voiced sounds from a two-mass model,'' Bell Syst. Tech. J. **51**, 1233–1268 (1972).
[36] D. B. Fry, ''Duration and intensity as physical correlates of linguistic stress,'' J. Acoust. Soc. Am. **30**, 765–769 (1955).
[37] D. B. Fry, ''Experiments in the perception of stress,'' Lang. Speech **1**, 126–152 (1958).
[38] D. L. Bolinger, ''A theory of pitch accent in English,'' Word **14**, 109–149 (1958).
[39] P. Lieberman, ''Some acoustic correlates of word stress in American-English,'' J. Acoust. Soc. Am. **32**, 451–454 (1960).
[40] K. Hadding-Koch, *Acoustico-Phonetic Studies in the Intonation of Southern Swedish* (Gleerup, Lund, Sweden, 1961).
[41] K. E. Pollock, D. M. Brammer, and C. F. Hageman, ''An acoustic analysis of young children's productions of word stress,'' J. Phonetics **21**, 183–203 (1990).
[42] M. Kehoe, C. Stoel-Gammon, and E. H. Buder, ''Acoustic Correlates of Stress in Young Children's Speech,'' J. Speech Hear. Res. **38**, 338–350 (1995).
[43] J. A. Sereno and A. Jongman, ''Acoustic Correlates of Grammatical Class,'' Lang. Speech **38**(1), 57–76 (1995).
[44] M. Beckman, *Stress and Non-stress Accent* (Foris, Dordrecht, 1986).
[45] J. Sundberg, N. Elliot, P. Gramming, and L. Nord, ''Short-term variation of subglottal pressure for expressive purposes in singing and stage speech—a preliminary investigation,'' J. Voice **7**(3), 227–234 (1993).

Objective measures of breathy voice quality obtained using an auditory model[a)]

Rahul Shrivastav[b)] and Christine M. Sapienza
Department of Communication Sciences and Disorders, University of Florida, Gainesville, Florida 32611

(Received 13 February 2003; revised 25 June 2003; accepted 14 July 2003)

While several acoustic measures have been proposed to quantify listener ratings of breathy voice quality, most have failed to give a consistent and high correlation with perceptual ratings of breathiness. One reason for these limitations is that most acoustic measures do not address the nonlinear processes that occur in the peripheral auditory system during the auditory perceptual process. It was hypothesized that modeling such nonlinear events during signal processing may provide objective parameters that better correspond to perceptual ratings of breathy voice quality. Ten listeners rated 27 voice stimuli using a five-point rating scale. Acoustic measures were determined from these stimuli and were selected based on their history of having a moderate to strong correlation to perceptual ratings of breathiness. The stimuli were also analyzed using an auditory model proposed by Moore, Glasberg, and Baer [J. Audio Eng. Soc. **45**(4), 224–239 (1997)], and new measures were calculated from the output of this model. These measures included the partial loudness of the signal and the loudness of the aspiration noise. Measures obtained from the output of the auditory model were found to account for a high amount of variance in the perceptual ratings of breathiness.
[DOI: 10.1121/1.1605414]

PACS numbers: 43.71.Bp, 43.71.Gv, 43.70.Gr [AL]

I. INTRODUCTION

Two approaches, subjective and objective, are commonly used to measure and describe breathy voice quality. Subjective methods usually take the form of listeners' ratings of voice quality, often made on an equal-appearing interval or visual analog scale. However, subjective ratings of voice quality using these traditional methods often have poor inter-judge reliability, particularly when used in a clinical situation (Kreiman *et al.*, 1993). Objective measures, on the other hand, make specific measurements from the vocal acoustic signal or from other physiological signals associated with voice production. These measures either relate to vocal fold physiology underlying the production of breathy voices (Hanson, 1997; Hanson and Chuang, 1999; Stevens and Hanson, 1995), or are based upon acoustic findings from voices that were rated as being breathy by a group of listeners (Eskenazi *et al.*, 1990; Feijoo and Hernandez, 1990; Hillenbrand *et al.*, 1994; Wolfe *et al.*, 1997). Therefore, changes in these measures are believed to reflect changes in voice quality. An ideal objective measure of voice quality should be specific to the vocal quality it is supposed to measure and be sensitive to changes in listener's perception of that quality. Unfortunately, although several acoustic measures have been proposed to describe and quantify breathy voice quality, few have shown a consistent and high correlation with perceptual ratings of breathiness.

Kreiman and Gerratt (2000a) note three possible reasons that may lead to a poor correspondence between acoustic measures and perceptual ratings of voice quality. These include (*i*) incorrect or inadequate methods used to calculate the acoustic measures, (*ii*) the use of incorrect methods to determine the relationship between the acoustic and perceptual measures, and (*iii*) an erroneous hypothesized relationship between the acoustic measures and their proposed perceptual consequences. Despite significant improvements in the methods used to calculate acoustic measures (Frohlich *et al.*, 2000; Hillenbrand *et al.*, 1994; Michaelis *et al.*, 1998; Milenkovic, 1987; Qi and Hillman, 1997), and greater insights on the nature of perceptual ratings of voice quality (Gerratt and Kreiman, 2001; Kreiman and Gerratt, 2000b; Kreiman *et al.*, 1993), a clear understanding of how specific acoustic changes affect the perception of voice quality is lacking. For example, the perception of breathy voice quality has been suggested to result from changes in perturbation, spectral slope as well as signal-to-noise ratio (Hillenbrand and Houde, 1996; Martin *et al.*, 1995; Shrivastav, 2001; Wolfe and Martin, 1997). However, the relative importance of these changes in the perception of breathiness is not clear (see Kreiman and Gerratt, 2000a, for a more detailed discussion).

Experiments to determine how changes in the vocal acoustic signal affect the perception of voice quality often assume a direct and linear relationship between the acoustic and perceptual measures of voice quality. However, the relationship between a physical stimulus and its perceptual attribute are often nonlinear. For example, the relationship between sound intensity and loudness or between frequency and pitch has been shown to be nonlinear. Because these characteristics are inherent to the auditory perceptual process, the acoustic-perceptual relationship for voice quality may also be nonlinear. Such nonlinear relationships between various acoustic characteristics and their psychoacoustic percepts have been successfully accounted for by auditory mod-

[a)]This paper was presented at the 3rd Biennial International Conference on Voice Physiology and Biomechanics, Denver, CO, 13–16 September 2002.
[b)]Electronic mail: rahul@csd.ufl.edu

els that simulate the underlying physiological processes in the peripheral auditory system (Moore *et al.*, 1997; Patterson *et al.*, 1995; Zwicker and Scharf, 1965). Such models have been used to explain a variety of psychoacoustic phenomena such as monaural and binaural thresholds, equal loudness contours, intensity discrimination, and vowel formant discrimination (Florentine, 1983; Florentine and Buus, 1981; Kewley-Port, 1991; Kewley-Port and Zheng, 1998; Moore *et al.*, 1997). Therefore, the application of such auditory processing models to the study of voice quality may help us understand how the vocal acoustic signal elicits a specific perceptual response and aid in the development of perceptually relevant measures of breathy voice quality.

Shrivastav (2001) found that modeling such nonlinear processes appears to be useful for developing objective measures of voice quality. In this experiment, the auditory model proposed by Moore and his colleagues (Moore *et al.*, 1997) was used as a signal processing front-end for analyzing vowels produced by 13 speakers with dysphonia. Several measures obtained from the output of the auditory model were compared to the perceptual ratings of breathiness obtained using multidimensional scaling techniques. It was found that the perceptual ratings of breathiness were best predicted by one auditory measure, the partial loudness of the periodic signal. This measure, which depends upon both the aspiration noise and spectral slope, was also found to account for significantly greater variance in the perceptual data than several other acoustic measures obtained without the auditory front-end. The use of an auditory front-end was found to help obtain objective measures of voice that correspond closely with listeners' perception of breathiness.

The present study was undertaken with two primary objectives. The first goal was to replicate the experiment done by Shrivastav (2001) with a larger set of stimuli to determine the consistency of previous findings which supported the use of an auditory model as a signal processing front-end when calculating objective measures of breathiness. This was considered essential especially because previous research to determine objective measures of voice quality has been plagued with inconsistent results. Second, this experiment obtained perceptual ratings of breathiness using an equal-appearing interval scale. This scale simulates a clinical voice quality rating task more closely than the pair-wise comparisons used by Shrivastav (2001).

II. METHODS

A. Listeners

Ten female volunteers served as listeners in this experiment. All listeners were enrolled in a graduate program in speech-language pathology and had taken at least one class on voice disorders. Therefore, all listeners had approximately the same degree of experience in listening to and rating voice quality. The average age of these listeners was 27 years and ranged from 23 to 48 years. All listeners were native speakers of American English and were screened to ensure normal hearing. The listeners were paid for their participation in the experiment.

B. Stimuli

The stimuli used in this experiment were chosen from the Kay Elemetrics Disordered Voice Database (Kay Elemetrics, Inc.). The stimuli were chosen by selecting voices from female patients who were diagnosed as having one of the following conditions that are known to present breathy voice quality: (a) vocal fold paralysis or paresis (recurrent laryngeal nerve or superior laryngeal nerve) and (b) vocal fold bowing associated with presbyphonia. The stimuli were first screened by an independent judge who was unaware of the purpose of the study. The screening was done to discard certain stimuli from the test set. These included voices that demonstrated any evidence of bifurcations or chaos (Titze, 1994) or those which were perceived to have a nasal quality. Voices with bifurcation or chaos were discarded as they were likely to provide invalid acoustic measures. The presence of these characteristics was ascertained manually by studying a narrow-band spectrogram for these stimuli. These criteria led to the selection of 27 vowel stimuli (vowel /a/) from the Kay Elemetrics Disordered Voice database. The mean age of the selected speakers was 59 years and ranged from 19 to 84 years. All stimuli were digitized with 16-bit sampling at the rate of 25 kHz. A stable, 500-ms segment was obtained from each of the voice stimuli. Using Cool Edit (Syntrillium Software), these segments were equated to have equal root mean square energy and were shaped to have a rise-time and decay-time of 25 ms. All acoustic, auditory, and perceptual measures were obtained from these edited stimuli.

C. Acoustic analyses

Acoustic analysis of the stimuli selected for the experiment was done with two objectives. First, acoustic analyses was done to obtain measures that have previously been reported to correlate with perceptual ratings of breathy voice quality (Hanson, 1997; Hillenbrand *et al.*, 1994; Klatt and Klatt, 1990; Kreiman *et al.*, 1990; Stevens and Hanson, 1995). These measures included the difference between the amplitude of the first and second harmonics (H1*−H2*), the difference between the amplitude of the first harmonic and the first formant (H1*−A1), the difference between the amplitudes of the first harmonic and the third formant frequency (H1*−A3*), an estimate of the aspiration noise, the cepstral peak prominence (Hillenbrand *et al.*, 1994), the mean fundamental frequency (F0), short-term perturbation in fundamental frequency and intensity (jitter and shimmer, respectively) and the signal-to-noise ratio (SNR). To allow comparison across different speakers, the spectral measures (H1*−H2*, H1*−A1 and H1*−A3*) were corrected using formulas described by Hanson (1997). The second objective of the acoustic analysis was to estimate the spectrum of the harmonic energy (henceforth referred to as "*signal*") and the spectrum of the aspiration noise (henceforth referred to as "*noise*") in each stimulus. It was necessary to separate the two for use as input to the auditory model described in the next section.

The voice stimuli were analyzed using CSpeechSP (Milenkovic, 1997). The spectral measures H1, H2, A1, and A3 were obtained from the average FFT spectra of stimuli,

calculated using successive 20-ms Hamming windows with 10-ms overlap. These were then used to determine H1*−H2*, H1*−A1 and H1*−A3*. The time waveform of the aspiration noise in these stimuli was estimated by using an algorithm described by (Milenkovic, 1987, 1995) and implemented in the software CSpeechSP. This algorithm decouples aspiration noise from the effects of jitter and shimmer by first estimating the fundamental period using short time windows and parabolic interpolation between data points. The short-time window is used to calculate the cross-correlation function between successive pitch periods, and parabolic interpolation refines the pitch-period estimate between cross-correlation sample points. A running pitch estimate is used to adapt the window to the length of the pitch period. The amplitude modulation of each glottal cycle is also calculated by minimizing the mean square error between successive glottal cycles. Finally, the aspiration noise is estimated by calculating the difference between the actual and the predicted signal with known period and amplitude modulation. The ''NewJit'' algorithm (Milenkovic, 1995) estimates the aperiodicity noise by using both the pitch period and amplitude change estimates to compute a weighted difference between the signal and the signal a pitch period later. Low-pass filters are used to compute samples of a filtered signal and samples of that same signal sampled a fractional number of sample intervals later. The spectrum of this aspiration noise was also obtained by calculating the average FFT using 20-ms Hamming windows with 10-ms overlap between successive windows. Finally, the spectrum of the signal alone was estimated by subtracting the spectrum of the aspiration noise from the overall FFT spectrum of the voice stimuli.

D. Auditory model

The auditory model proposed by Moore *et al.* (1997) was implemented in MATLAB® 6 (Mathworks, Inc.). The model has different stages that represent the transfer function of the outer and middle ear, the excitation pattern elicited on the basilar membrane within the cochlea, and finally the transduction of this excitation pattern into neural activity in the fibers of the auditory nerve. These stages have been briefly described below. Readers are referred to Moore and Glasberg (1983) and Moore *et al.* (1990, 1997) for more details about the auditory model.

The first stage of this model accounts for the effects of the outer and the middle ears on the input spectrum. Because these structures act as simple resonators, they are modeled as linear filters with specific center frequency and bandwidths. Correction factors to account for the transfer function of the transducer used to deliver the stimuli is also included in the calculation of these filter characteristics. The next stage calculates the excitation pattern elicited on the basilar membrane by the sound input. The cochlea is modeled as a bank of asymmetrical band-pass filters with bandwidths that increase with the center frequency of the filters. These auditory filters are represented in the equal rectangular bandwidth (ERB) scale which was obtained through masking experiments using notched noise (Moore *et al.*, 1990; Patterson, 1976; Patterson and Nimmo-Smith, 1980). The ERB filters are numerically labeled with lower numbers representing the low frequency (at the apical end of the cochlea) and higher numbers representing the higher frequencies (at the basal end of the cochlea). Finally, the total neural excitation in response to the excitation pattern is calculated to reflect the total loudness elicited by the input sound. To estimate the total loudness, the excitation pattern is first transformed to *specific loudness* (N'). This represents the loudness elicited per ERB and is obtained by applying a nonlinear compressive function to the excitation pattern. The total loudness of a sound (measured in Sones) is obtained by calculating the total area under the specific loudness curve. This auditory model can also be used to estimate the loudness of a sound when it occurs in the presence of another competing sound (masker). Loudness elicited by a sound in the presence of a masker is called *partial loudness* and is calculated by subtracting the specific loudness elicited by the signal from that elicited by the masker.

The frequency range between 50 Hz and 10 kHz was divided into 33 ERB filters. To obtain an accurate excitation pattern, successive filters were overlapped by 0.1 ERB units and a total of 330 filters were generated. The excitation patterns of the voice stimuli were obtained by summing the output of all the ERB filters in response to the input sound spectrum. Excitation patterns were obtained separately for the harmonic energy (signal), aspiration noise (noise) and the total energy for each voice stimuli. These excitation patterns were then used to calculate two measures from the vocal acoustic signal—the partial loudness of the harmonic energy (N'_{SIGNAL}) and the loudness of aspiration noise (N'_{NOISE}). While N'_{NOISE} does not take masking into account, calculation of N'_{SIGNAL} assumes that the loudness of the signal (harmonic energy) is affected by a masker (aspiration noise). These measures were obtained because past research has suggested that perception of breathiness may be related to the intensity of the harmonic energy in the voice stimuli and/or the intensity of the aspiration noise (Childers and Lee, 1991; Klatt and Klatt, 1990). Aspiration noise was assumed to act as a masker when calculating the N'_{SIGNAL} because harmonic energy and aspiration noise coexist in breathy voice stimuli.

E. Perceptual ratings

Listeners were seated in a sound treated room and the stimuli were presented binaurally using TDH-39 headphones. Listeners were asked to rate each voice stimulus using a five-point rating scale, where a value of 0 indicated minimal breathiness and a value of 4 indicated maximum breathiness. Listeners were strictly instructed to avoid making judgments based on pitch or loudness, and only to rate breathiness. A formal definition of breathiness was not provided. All ratings were made using a computer monitor and keyboard.

Ten randomized lists of voice stimuli were prepared. Each of the 27 voice stimulus was presented once in each list. Therefore, every listener rated each voice stimulus ten times. Listeners controlled the rate of presentation of stimuli and successive stimuli were presented 2 s after the rating for

TABLE I. Descriptive statistics for the acoustic and auditory measures for the stimuli used in this experiment.

	F0 (Hz)	Jit. (%)	Shmr %	SNR(dB)	CPP	H1*−H2*	H1*−A1	H1*−A3*	N'_{SIGNAL}	N'_{NOISE}
Mean	207.3	0.95	5.81	16.00	13.65	7.36	6.84	31.50	10.88	10.26
s.d.	23.64	0.65	2.47	5.16	2.64	6.26	9.49	6.80	3.81	3.35

the previous stimulus was made. To minimize fatigue, a short break (approximately 2 min) was provided after each list and a longer break (approximately 5 min) was provided mid-way through the perceptual task. The rating task took approximately 1 h for each listener. The ten ratings obtained for each voice stimulus from a listener were averaged to obtain a single rating. These ratings were then averaged to obtain a group mean rating for each voice stimulus, and were used as the dependent variable in a stepwise regression described below.

Multiple ratings were obtained for each voice stimulus from each listener to estimate a ''true'' measure of voice quality. Theories of psychometric measurement suggest that listeners' responses to a stimulus in a psychoacoustic experiment result from stimulus characteristics as well as other factors, such as psychological variables (attention, memory, etc.) and factors related to the rating task (such as order effects) (Guilford, 1954; Watson, 1973). These factors introduce ''noise'' in the perceptual data. Because this noise is normally distributed, it may be minimized by averaging several ratings from each listener. The use of average judgments is commonly used in most psychophysical experiments that seek to investigate the relation between a physical stimulus and its perceptual consequence. Listeners in this experiment took between 6 and 10 min to rate each list of 27 stimuli and did not report any fatigue during the listening task.

F. Statistical analysis

To determine the interjudge reliability, Pearson's correlation coefficients were obtained between the average ratings from each listener. In addition, the interclass correlation (ICC) was also calculated. To determine which objective measures best correspond to perceptual data, two stepwise regression functions were determined. In the first regression, the acoustic measures were used to predict the perceptual ratings of breathiness. In the second regression, the same breathiness ratings were predicted using the two auditory measures. For both of these stepwise regression functions the convergence criteria were set to a p value of 0.05 for entering variables and 0.01 for removing variables from the regression equation. The stepwise regression always started by entering one variable at a time, based upon the entry criteria. Finally, to allow comparison of the auditory spectra of stimuli with different degrees of perceived breathiness, the voice stimuli were divided into four categories. These categories were made based upon the average perceptual rating for each stimulus. All voices with a mean breathiness rating between 0 and 1 were grouped as ''normal,'' between 1 and 2 were grouped as ''mild,'' between 2 and 3 were grouped as ''moderate,'' and those with a mean breathiness score between 3 and 4 were grouped as ''severe.''

III. RESULTS

The acoustic and auditory measures described previously were calculated and are described in Table I. The correlations between each of these measures and the perceptual ratings of breathiness are shown in Table II. Pearson's correlation between the average ratings of each listener and the group mean rating are shown in Table III. The average correlation between ratings made by all listeners was 0.880, and the average correlation between ratings made by each listener and the group mean ratings was 0.944. The interclass correlation for these ratings was 0.863.

The stepwise regression predicting the average breathiness rating from the acoustic measures retained three variables. The single best predictor of breathiness ratings was CPP, which accounted for 75.1% of variance in the perceptual data (adjusted R-Square$=0.751$, $F_{(1,25)}=79.402$, $p<0.001$). The addition of two other variables, SNR and shimmer, improved the results significantly, increasing the adjusted R-square to 0.781 (CPP+SNR) and 0.809 (CPP+SNR+Shimmer), respectively. No other acoustic measure added significantly to the variance in perceptual data accounted for by the regression equation. The results of this stepwise regression are summarized in Table IV. The scatterplot showing the relationship between CPP and average breathiness ratings is shown in Fig. 1.

The second stepwise regression predicting the average breathiness ratings from auditory measures retained both auditory measures. N'_{SIGNAL} accounted for 78.6% of the variance in the perceptual data (Adjusted R-Square$=0.786$, $F_{(1,25)}=96.219$, $p<0.001$). Adding N'_{NOISE} to the regression equation accounted for an additional 6.6% of the variance (Adjusted R-Square$=0.852$, $F_{(1,25)}=76.024$, $p<0.001$), thereby accounting for a total of 85.2% of the variance in the perceptual data. These results are shown in Table V. The scatterplot between average breathiness ratings and N'_{SIGNAL} is shown in Fig. 2 and that between breathiness ratings and N'_{NOISE} is shown in Fig. 3.

Finally, to determine how changes in N'_{SIGNAL} and

TABLE II. Correlation between the various acoustic and auditory measures and average breathiness ratings.

	Breathiness
F0 (Hz)	−0.501
Jitter (%)	0.863
Shimmer (%)	0.511
SNR (dB)	−0.829
H1*−H2* (dB)	0.553
H1*−A1 (dB)	0.575
H1*−A3* (dB)	0.417
CPP	−0.872
N'_{SIGNAL} (Sones)	−0.891
N'_{NOISE} (Sones)	0.849

TABLE III. Correlation matrix for perceptual ratings of breathiness obtained from each listener and the group mean rating.

Listener no.	Listener 1	2	3	4	5	6	7	8	9	10	Group mean
1	1.000										
2	0.757	1.000									
3	0.900	0.910	1.000								
4	0.879	0.920	0.923	1.000							
5	0.905	0.896	0.935	0.920	1.000						
6	0.840	0.944	0.942	0.931	0.910	1.000					
7	0.959	0.784	0.914	0.889	0.902	0.874	1.000				
8	0.913	0.800	0.871	0.843	0.859	0.858	0.895	1.000			
9	0.947	0.850	0.931	0.917	0.931	0.899	0.935	0.887	1.000		
10	0.902	0.683	0.810	0.799	0.815	0.795	0.925	0.864	0.830	1.000	
Group mean	0.955	0.901	0.967	0.955	0.961	0.950	0.964	0.927	0.967	0.896	1.000

N'_{NOISE} affect the perception of breathiness, the ''average N'_{SIGNAL}'' and ''average N'_{NOISE}'' were calculated for the stimuli in each of the four groups. The averaged values of signal partial loudness and noise specific loudness show a clear trend with increasing breathiness. Voices perceived to be more breathy show a decrease in signal partial loudness particularly in frequencies above approximately 3000 Hz. This change in partial loudness is probably strongly influenced by an increase in the loudness of the aspiration noise. The averaged N'_{SIGNAL} and N'_{NOISE} are shown in Figs. 4 and 5, respectively.

IV. DISCUSSION

In this experiment an auditory model proposed by Moore *et al.* (1997) was used as a signal-processing front-end prior to calculating measures to characterize the breathiness for 27 vowels selected from a commercially available database. Perceptual ratings of breathiness were obtained from ten listeners using carefully designed experimental tasks that aimed at minimizing the noise in the perceptual data. The average perceptual ratings for these stimuli indicate a wide range in severity of breathiness. These ratings were then correlated with several objective measures, with and without the use of the auditory processing front-end, to determine which measure(s) could best predict the perceptual ratings made by a group of listeners.

The acoustic measure that accounted for the greatest variance in perceptual ratings of breathiness was CPP, accounting for 75% of the variance in the perceptual data. In combination with SNR and Shimmer, the acoustic measures accounted for 80.9% of variance in the listeners' ratings of breathiness. In contrast, the auditory measure N'_{SIGNAL}, accounted for 79.5% of the variance of the same perceptual data. When combined with N'_{NOISE}, the auditory measures accounted for 85.2% of the variance of breathiness ratings. While these results show that the auditory measures were marginally better at predicting breathiness ratings, these differences may not be significant. However, these findings lend strong support to those of a previous experiment (Shrivastav, 2001) and verify that the results are maintained with a larger set of stimuli and using a scale that resembles a clinical rating task more closely.

TABLE IV. Stepwise regression predicting breathiness ratings from acoustic measures.

	R	Adjusted *R*-square	F (degrees of freedom)	Significance (*p*)
CPP	0.872	0.751	79.402 (1,25)	<0.001
CPP+SNR	0.893	0.781	47.366 (2,24)	<0.001
CPP+SNR +Shimmer	0.912	0.809	37.805 (3,23)	<0.001

The use of an auditory model as a signal processing front-end for objective measurement of voice quality can be justified only if the obtained results improve upon those obtained through the use of conventional techniques for acoustic analyses. A good objective measure of voice quality needs to have three properties—high sensitivity to changes in perceptual ratings, high specificity to changes in perceptual ratings, and high consistency across different stimuli and groups of listeners. Unfortunately, many of the conventionally used acoustic measures for measuring voice quality fall short on these criteria. First, these measures do not show good sensitivity to changes in perceptual ratings of voice quality. This often results in poor correlations between

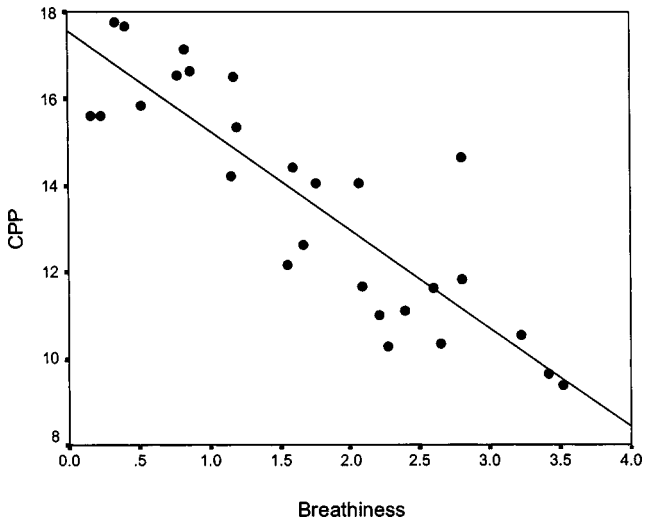


FIG. 1. Scatter plot between breathiness and CPP.

TABLE V. Stepwise regression predicting breathiness ratings from auditory measures.

	R	Adjusted *R*-square	F (degrees of freedom)	Significance (*p*)
N'_{SIGNAL}	0.891	0.786	96.219 (1,25)	<0.001
N'_{SIGNAL} + N'_{NOISE}	0.929	0.852	76.024 (2,24)	<0.001

acoustic and perceptual measures of voice quality. Second, these measures have not shown good specificity to changes in voice quality. This often leads to the same acoustic measure being correlated to two or more different voice qualities (for example, jitter and shimmer have been shown to have high correlations with breathy as well as rough voices). And finally, the correlation between the acoustic and perceptual data has been found to be highly inconsistent across different experiments.

Amongst the various acoustic measures used to describe breathiness, CPP probably best fulfills the three criteria described above. The findings of this experiment support those of past research that suggests CPP to be a good indicator of perceived breathiness and demonstrates high sensitivity, specificity and consistency across different experiments (Heman-Ackah *et al.*, 2002; Hillenbrand *et al.*, 1994; Hillenbrand and Houde, 1996). The auditory measures described in this experiment, N'_{SIGNAL} and N'_{NOISE}, account for slightly greater variance in perceptual ratings than CPP. A close examination of the scatterplots shown in Figs. 1–3 show that N'_{SIGNAL} is better than CPP in predicting perceptual data for the mid-range of the scale used in this experiment. However, for voices with minimal breathiness, N'_{SIGNAL} tends to underestimate the perceptual ratings of breathiness and for stimuli rated most breathy the errors occur on both the positive and negative sides of the predicted line.

The underestimation of perceptual data at the low end of the scale may have resulted from the particular formulas used for the implementation of the auditory model in this experiment. The auditory model described by Moore *et al.* (1997) has specific parameters to reflect the steep rise in loudness at levels just above the absolute threshold of hearing. However, several of the empirically derived constants used in these formulae are based upon data obtained from masking pure tones using notched noise and may not be accurate for broadband signals used in this experiment. Appropriate modifications of these values may result in a lower value of the N'_{SIGNAL} for certain stimuli, and a better fit for stimuli with very low levels of aspiration noise. The prediction errors for voices with the greatest noise (and rated most breathy) may have resulted from two possible factors. First, the algorithm used for estimating the aspiration noise may not be accurate for the most severely disordered voices. In this case, the poor prediction of breathiness for these stimuli may reflect incorrect estimation of aspiration noise. Second, the perceptual cues that listeners use to make judgments of voice quality may be somewhat different for the severely disordered voices. For example, for mild and moderately breathy voices, listeners may respond to changes in N'_{SIGNAL}, but for the severely disordered voices they may respond to changes in N'_{NOISE}. The fact that N'_{NOISE} is a good predictor

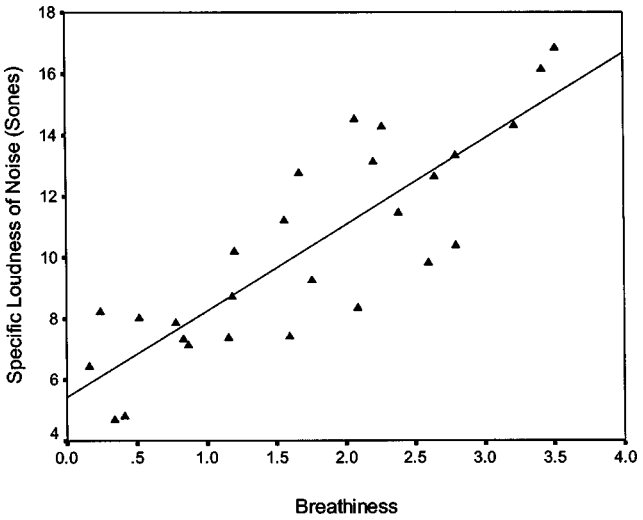


FIG. 3. Scatter plot between breathiness and N'_{NOISE}.

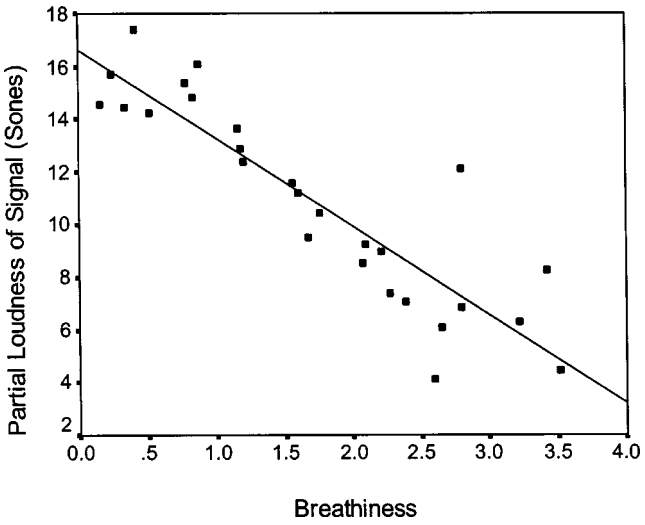


FIG. 2. Scatter plot between breathiness and N'_{SIGNAL}.

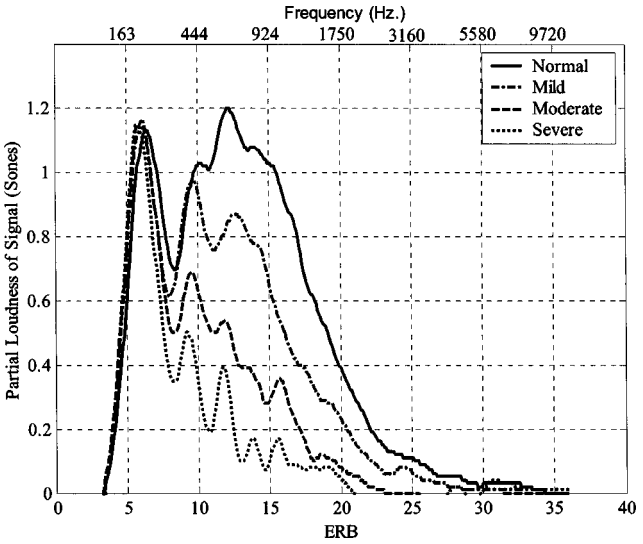


FIG. 4. Partial loudness of signal for voices in the four groups.

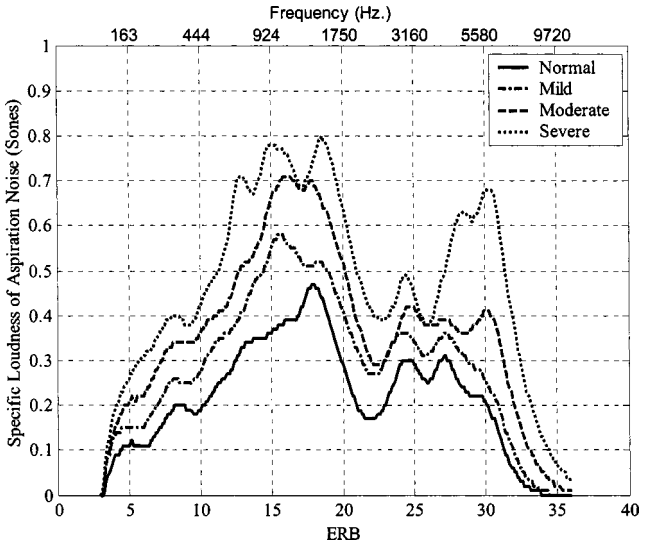


FIG. 5. Specific loudness of aspiration noise for voices in the four groups.

of breathiness ratings for the stimuli rated as most breathy (Fig. 3) supports the latter hypothesis, but there is too little data for this to be conclusive and further research in this direction is necessary. However, these findings demonstrate the sensitivity and consistency of N'_{SIGNAL} and N'_{NOISE} as objective measures of breathy voice quality.

''Partial loudness'' results from an interaction between the aspiration noise and the complex quasi-periodic wave generated by vocal fold vibration. Therefore, this measure is influenced by changes in the spectral slope of the glottal wave as well as the intensity and frequency of the aspiration noise. In this regard, this measure is similar to measures of SNR in the vocal acoustic signal. However, unlike SNR, this measure accounts for the nonlinear relationship between intensity and loudness as well as auditory masking that occurs when two competing sounds are present simultaneously. Although previous research has found both aspiration noise and spectral slope to show a high correlation with perceived breathiness (Childers and Lee, 1991; Huffman, 1987; Klatt and Klatt, 1990), the results of these studies have not been consistent (Kreiman and Gerratt, 2000a). By accounting for the nonlinear relationship between some acoustic characteristics and their perceptual consequences, the use of an auditory front-end for signal processing may provide objective measures that can quantify the degree of perceived breathiness more consistently.

The partial and specific loudness patterns obtained by averaging the auditory spectra of stimuli in each of the four groups show a specific pattern of change with increasing breathiness. Voices perceived to have minimal breathiness are characterized by a gradually falling partial loudness pattern that elicits at least some excitation in frequencies above 24 ERB (which corresponds to approximately 3000 Hz). Voices that are perceived to be more breathy show a steeper slope of the partial loudness curve, with the most severely breathy voices showing almost no partial loudness in frequencies above 24 ERB. These changes likely arise from two factors: (a) voices that are perceived more breathy show greater aspiration noise (Fig. 5), leading to greater masking of the harmonics, and (b) voices that are perceived more breathy have greater spectral slope, which further reduces the intensity of the harmonics in the higher frequencies. Because aspiration noise is typically centered around 3000 Hz (Klatt and Klatt, 1990), and because of the positive slope of the glottal source spectrum, the changes in the partial loudness patterns are most prominent in frequencies above 3000 Hz. The relative increase in the first harmonic region associated with increasing breathiness can also be seen in the partial loudness patterns for the four groups of voice stimuli (Fig. 4).

The findings of this experiment are limited by a few factors. The stimuli used in the present experiment were selected from a larger database of disordered voices using criteria to obtain a set of voices that were expected to show a relatively wide range of breathiness. However, to avoid potential complications in acoustic analyses, all stimuli for which the fundamental frequency could not be reliably determined were discarded. Therefore, certain voice stimuli, potentially those with most severe breathiness, were not included in the analyses. A second limitation of the present experiment relates to the use of stimuli that were only 500 ms in duration. A short but stable portion of the voice acoustic waveform was deliberately chosen to avoid smearing the FFT spectrum. Using a larger window or averaging the spectrum across a larger time frame may have resulted in an incorrect representation of the acoustic spectrum. The auditory model implemented in this experiment does not permit time-domain analyses and, therefore, temporal changes in the acoustic signal cannot be adequately described. However, temporal cues may play an important role in the perception of voice quality and need to be investigated further.

An additional reason for discarding the most severely disordered voices is a limitation imposed by the signal conditioning required prior to the use of the auditory front-end. Estimation of aspiration noise in voice is a critical step in the calculation of partial loudness and the success of this algorithm was also dependent upon the correct estimation of the fundamental frequency of an acoustic waveform. When the voice acoustic waveform is temporally unstable or when it has a poor SNR, the estimation of the fundamental frequency (and consequently the estimation of aspiration noise) is affected significantly. Improvements in signal processing techniques to estimate the aspiration noise in severely disordered voices may further aid the use of an auditory front-end for measurement of voice quality.

These limitations described above restrict the extent to which the findings of this experiment can be generalized. Specifically, these results are likely applicable to voices that have a relatively stable acoustic waveform over time and where the fundamental frequency of the waveform can be reliably ascertained. Despite these limitations, however, the present experiment has shown promising results. The use of an auditory-processing model as a signal processing front-end can help us to understand how listeners make judgments of voice quality and to generate objective measures that correspond well to listener judgment of breathiness. The limitations of the present study also guide further research. For example, future research needs to focus upon methods to

analyze and characterize voices that lack a clear periodic structure and investigations to modify existing auditory-processing models for the study of voice quality.

V. CONCLUSIONS

Several acoustic measures have been proposed to quantify the perception of breathy voice quality. However, these measures often fail to account for a large amount of variance in perceptual ratings and show inconsistent results across different experiments. One reason for these inconsistencies is that most acoustic measures fail to account for the nonlinearity in the transduction processes that occur in the peripheral auditory system. Because these processes are an inherent part of the auditory-perceptual process, modeling such nonlinear processes during signal processing can help develop objective measures that better correspond with perceptual data. For relatively periodic vocal signals, the perceived breathiness was found to correlate best with the partial loudness of the periodic glottal signal. This measure was determined when the aspiration noise in voices is modeled as an auditory masker. Measures obtained from the output of the auditory model were found to account for a large amount of variance in perceptual ratings of breathiness.

ACKNOWLEDGMENTS

We would like to thank Karen Wheeler and Archie Harmon for help with data collection and analyses. We also wish to thank Dr. Anders Löfvquist and an anonymous reviewer for their suggestions on an earlier version of this manuscript.

Childers, D. G., and Lee, C. K. (**1991**). ''Vocal quality factors: analysis, synthesis, and perception,'' J. Acoust. Soc. Am. **90**(5), 2394–2410.

Eskenazi, L., Childers, D. G., and Hicks, D. M. (**1990**). ''Acoustic correlates of vocal quality,'' J. Speech Hear. Res. **33**(2), 298–306.

Feijoo, S., and Hernandez, C. (**1990**). ''Short-term stability measures for the evaluation of vocal quality,'' J. Speech Hear. Res. **33**(2), 324–334.

Florentine, M. (**1983**). ''Intensity discrimination as a function of level and frequency and its relation to high-frequency hearing,'' J. Acoust. Soc. Am. **74**(5), 1375–1379.

Florentine, M., and Buus, S. (**1981**). ''An excitation pattern model for intensity discrimination,'' J. Acoust. Soc. Am. **70**, 1646–1654.

Frohlich, M., Michaelis, D., Strube, H. W., and Kruse, E. (**2000**). ''Acoustic voice analysis by means of the hoarseness diagram,'' J. Speech Lang. Hear. Res. **43**(3), 706–720.

Gerratt, B., and Kreiman, J. (**2001**). ''Measuring voice quality with speech synthesis,'' J. Acoust. Soc. Am. **110**(5), 2560–2566.

Guilford, J. P. (**1954**). *Psychometric Methods*, 2nd ed. (McGraw–Hill, New York).

Hanson, H. M. (**1997**). ''Glottal characteristics of female speakers: acoustic correlates,'' J. Acoust. Soc. Am. **101**(1), 466–481.

Hanson, H. M., and Chuang, E. S. (**1999**). ''Glottal characteristics of male speakers: acoustic correlates and comparison with female data,'' J. Acoust. Soc. Am. **106**(2), 1064–1077.

Heman-Ackah, Y. D., Michael, D. D., and Goding, Jr., G. S. (**2002**). ''The relationship between cepstral peak prominence and selected parameters of dysphonia,'' J. Voice **16**(1), 20–27.

Hillenbrand, J., and Houde, R. A. (**1996**). ''Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech,'' J. Speech Hear. Res. **39**(2), 311–321.

Hillenbrand, J., Cleveland, R. A., and Erickson, R. L. (**1994**). ''Acoustic correlates of breathy vocal quality,'' J. Speech Hear. Res. **37**(4), 769–778.

Huffman, M. K. (**1987**). ''Measures of phonation type in Hmong,'' J. Acoust. Soc. Am. **81**(2), 495–504.

Kewley-Port, D. (**1989**). ''Detection thresholds for isolated vowels,'' J. Acoust. Soc. Am. Suppl. 1 **85**, S51.

Kewley-Port, D., and Zheng, Y. (**1998**). ''Auditory models of formant frequency discrimination for isolated vowels,'' J. Acoust. Soc. Am. **103**(3), 1654–1666.

Klatt, D. H., and Klatt, L. C. (**1990**). ''Analysis, synthesis, and perception of voice quality variations among female and male talkers,'' J. Acoust. Soc. Am. **87**(2), 820–857.

Kreiman, J., and Gerratt, B. (**2000a**). ''Measuring voice quality,'' in *Voice Quality Measurement*, 1st ed., edited by R. D. Kent and M. J. Ball (Singular, San Diego), pp. 73–101.

Kreiman, J., and Gerratt, B. R. (**2000b**). ''Sources of listener disagreement in voice quality assessment,'' J. Acoust. Soc. Am. **108**(4), 1867–1876.

Kreiman, J., Gerratt, B. R., and Precoda, K. (**1990**). ''Listener experience and perception of voice quality,'' J. Speech Hear. Res. **33**(1), 103–115.

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., and Berke, G. S. (**1993**). ''Perceptual evaluation of voice quality: review, tutorial, and a framework for future research,'' J. Speech Hear. Res. **36**(1), 21–40.

Martin, D., Fitch, J., and Wolfe, V. (**1995**). ''Pathologic voice type and the acoustic prediction of severity,'' J. Speech Hear. Res. **38**(4), 765–771.

Michaelis, D., Frohlich, M., and Strube, H. W. (**1998**). ''Selection and combination of acoustic features for the description of pathologic voices,'' J. Acoust. Soc. Am. **103**(3), 1628–1639.

Milenkovic, P. (**1987**). ''Least mean square measures of voice perturbation,'' J. Speech Hear. Res. **30**(4), 529–538.

Milenkovic, P. (**1995**). ''Rotation based measures of voice aperiodicity,'' in *Workshop of Acoustic Voice Analysis: Proceedings*, edited by D. Wong (National Center for Voice and Speech, Iowa City), pp. 1–10.

Milenkovic, P. (**1997**). CSpeech (Version 4.0), University of Wisconsin—Madison.

Moore, B. C., and Glasberg, B. R. (**1983**). ''Suggested formulae for calculating auditory-filter bandwidths and excitation patterns,'' J. Acoust. Soc. Am. **74**(3), 750–753.

Moore, B. C., Peters, R. W., and Glasberg, B. R. (**1990**). ''Auditory filter shapes at low center frequencies,'' J. Acoust. Soc. Am. **88**(1), 132–140.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). ''A model for the prediction of thresholds, loudness and partial loudness,'' J. Audio Eng. Soc. **45**(4), 224–239.

Patterson, R. D. (**1976**). ''Auditory filter shapes derived with noise stimuli,'' J. Acoust. Soc. Am. **59**(3), 640–654.

Patterson, R. D., and Nimmo-Smith, I. (**1980**). ''Off-frequency listening and auditory-filter asymmetry,'' J. Acoust. Soc. Am. **67**(1), 229–245.

Patterson, R. D., Allerhand, M. H., and Giguere, C. (**1995**). ''Time-domain modeling of peripheral auditory processing: A modular architecture and software platform,'' J. Acoust. Soc. Am. **98**(4), 1890–1894.

Qi, Y., and Hillman, R. E. (**1997**). ''Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals,'' J. Acoust. Soc. Am. **102**(1), 537–543.

Shrivastav, R. (**2001**). ''Perceptual Structure of Breathy Voice Quality and Auditory Modeling of its Acoustic Cues,'' unpublished Doctoral Dissertation, Indiana University, Bloomington.

Stevens, K. N., and Hanson, H. M. (**1995**). ''Classification of glottal vibrations from acoustic measurements,'' in *Vocal Fold Physiology: Voice Quality Control*, edited by O. Fujimura and M. Hirano (Singular, San Diego), pp. 147–170.

Titze, I. (**1994**). *Workshop on Acoustic Voice Analysis*, National Center for Voice and Speech.

Watson, C. S. (**1973**). ''Psychophysics,'' in *Handbook of General Psychology*, edited by B. B. Wolman (Prentice-Hall, Englewood Cliffs, NJ).

Wolfe, V., and Martin, D. (**1997**). ''Acoustic correlates of dysphonia: type and severity,'' J. Commun. Disord. **30**(5), 403–415; quiz 415–406.

Wolfe, V., Fitch, J., and Martin, D. (**1997**). ''Acoustic measures of dysphonic severity across and within voice types,'' Folia Phoniatr. Logop. **49**(6), 292–299.

Zwicker, E., and Scharf, B. (**1965**). ''A model of loudness summation,'' Psychol. Rev. **72**, 3–26.

Effects of contrast between onsets of speech and other complex spectra

Jeffry A. Coady[a)]
Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706-1611

Keith R. Kluender[b)]
Department of Psychology and Department of Physiology, University of Wisconsin, Madison, Wisconsin 53706-1611

William S. Rhode[c)]
Department of Physiology, University of Wisconsin, Madison, Wisconsin 53706

(Received 25 June 2003; accepted for publication 10 July 2003)

Previous studies using speech and nonspeech analogs have shown that auditory mechanisms which serve to enhance spectral contrast contribute to perception of coarticulated speech for which spectral properties assimilate over time. In order to better understand the nature of contrastive auditory processes, a series of CV syllables varying acoustically in $F2$-onset frequency and perceptually from /ba/ to /da/ was identified following a variety of spectra including three-peak renditions of [e] and [o], one-peak simulations of only $F2$, and spectral complements of these spectra for which peaks are replaced with troughs. Results for three-versus one-peak (or trough) precursor spectra were practically indistinguishable, suggesting that effects were spectrally local and not dependent upon perception of precursors as speech. Effects of complementary (trough) spectra had complementary effects on perception of following stops; however, effects for spectral complements were particularly dependent upon the interval between precursor and CV onsets. Results from these studies cannot be explained by simple masking or adaptation of suppression. Instead, they provide evidence for the existence of processes that selectively enhance contrast between onset spectra of neighboring sounds, and these processes are relevant for perception of connected speech. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1608955]

PACS numbers: 43.71.Es, 43.71.Pc, 43.66.Dc [CWT]

I. INTRODUCTION

All sensorineural systems are particularly responsive to change; perception of stimulus energy is always relative to properties of temporally or spatially proximal energy. The broad descriptor for many instances of responsivity to change is ''contrast.'' Demonstrations of perceptual contrast are ubiquitous and have a long history of being recognized as central to understanding perception (e.g., Locke, 1706/1974). For vision, contrast has been demonstrated for perception of lightness (Koffka, 1935; Wallach, 1948), line orientation (Gibson, 1933, 1937; Gibson and Radner, 1937), size, position, and curvature (Kohler and Wallach, 1944), spatial frequency (Blakemore and Sutton, 1969), depth (Ames, 1935; Bergman and Gibson, 1959; Kohler and Emery, 1947), and color (Cathcart and Dawson, 1928–1929). Contrast has been found for tempo of behavior (Cathcart and Dawson, 1927–1928) and lifting of weights (Guilford and Park, 1931; Sherif, Taub, and Hovland, 1958). For hearing, contrast effects for pitch (Cathcart and Dawson, 1928–1929; Christman, 1954) and spatial location of sounds (Flügel, 1920–1921) have been demonstrated.

Not surprisingly, contrastive processes have been introduced as putative explanations for a variety of perceptual effects found with speech sounds. For example, Diehl and colleagues (Diehl, Kluender, and Walsh, 1990) have posited durational contrast as an explanation for effects of vowel length preceding medial stops and following initial stops and glides. For the former case, listeners are more likely to report hearing short-closure [+voice] stops [b] following longer vowels, and to report long-closure [−voice] stops [p] following shorter vowels (Denes, 1955; Port and Dalby, 1982; Raphael, 1972). For stops with short formant transitions such as [b] versus longer-transition glides such as [w], longer following vowels encourage perception of the shorter [b], while [w] is more likely to be perceived preceding shorter vowels (Miller and Liberman, 1979). In each of these cases, analogous patterns of results were found for nonspeech stimuli that maintained durational characteristics equivalent to those for speech syllables tested (Diehl and Walsh, 1989; Kluender and Diehl, 1988; Pisoni, Carrell, and Gans, 1983).

Perceptual contrast between adjacent spectral composition has been suggested to explain some aspects of perception of coarticulated speech (e.g., Holt and Kluender, 2000). Coarticulation, the spatial and temporal overlap of adjacent articulatory activities, is reflected in the acoustic signal by severe context dependence. Acoustic information specifying one speech sound varies substantially depending on surrounding sounds. This context always follows the same pattern; adjacent sounds assimilate toward spectral characteristics of one another. Owing to mass and inertia of articulators (as well as planning), articulatory movements are compro-

[a)]Electronic mail: jacoady@facstaff.wisc.edu
[b)]Electronic mail: krkluend@facstaff.wisc.edu
[c)]Electronic mail: rhode@physiology.wisc.edu

mises between where articulators have been and where they are going. Because the acoustic signal directly reflects these articulatory facts, the spectrum assimilates in the same fashion that speech articulation assimilates in coarticulation.

Lindblom (1963) and Öhman (1965) provided some of the best early descriptions of how context systematically influences speech production. Lindblom reported that the frequency of the second formant ($F2$) was higher in the productions of [dɪd] and [dʊd] than for the vowels [ɪ] and [ʊ] in isolation, and that $F2$ for each vowel was lower in the productions of [bɪb] and [bʊb]. In both contexts, $F2$s approached the $F2$ of flanking consonants, which are higher for [d] than for [b]. In a subsequent study, Lindblom and Studdert-Kennedy (1967) demonstrated how perception of coarticulated vowels is complementary to these facts of articulation. For CVC syllables with $F2$ of the vowel varying acoustically from higher to lower and perceptually from [ɪ] to [ʊ], listeners reported hearing [ɪ] (higher $F2$) more often in the [wVw] (low $F2$) context, and [ʊ] (lower $F2$) more often in the [yVy] (high $F2$) context. Lindblom and Studdert-Kennedy (1967) wrote: "...mechanisms of perceptual analysis whose operations contribute to enhancing contrast in the above-mentioned sense are precisely the type of mechanisms that seem well suited to their purpose given the fact that the slurred and sluggish manner in which human speech sound stimuli are often generated tends to reduce rather than sharpen contrast." (p. 842).

Findings from more recent studies more clearly demonstrate the role of general auditory processes for providing this contrast. Using CVCs and hybrid speech–nonspeech analogs, Holt, Lotto, and Kluender (2000) demonstrated that replacing syllable-initial and syllable-final consonantal information with single frequency-modulated sine waves (following $F2$ trajectories) or with constant-frequency sine waves (set at onset and offset frequencies of $F2$) resulted in the same pattern of perceived vowel quality as was found for full-spectrum CVCs.

For VCCVs, Mann and Repp (1981) entertained the hypothesis that contrast might account in part for their finding that listeners were more likely to report hearing higher-frequency [s] following [r] (low $F3$) and lower-frequency [ʃ] following [l] (high $F3$) in syllables varying from [arʃa] to [arsa] and from [alʃa] to [alsa]. Mann (1980) had earlier shown that preceding [r] and [l] had differential effects of perception of following stops [d] and [g], such that listeners are more likely to report hearing higher-$F3$ [d] following low-$F3$ [r] and hearing lower-$F3$ [g] following high-$F3$ [l]. Lotto and Kluender (1998) later demonstrated that the same pattern of identification for a series of stops varying from [d] to [g] was obtained when preceding [ar] and [al] were replaced either by single frequency-modulated sine waves following the trajectory of $F3$ for [r] and [l], or by single constant-frequency sine waves set to the offset frequency of $F3$ for [r] and [l].

Data from the foregoing studies are consistent with the universal principle that sensorineural systems respond most vigorously to stimulus change. Results are consistent with an account by which spectral contrast between adjacent speech or nonspeech spectra predicts changes in perception as a function of neighboring spectral composition. Alternatively, based upon results from a McGurk-type experiment investigating bimodal perception of disyllables, Fowler, Brown, and Mann (2000) argued that their audio-visual demonstration of contrast effects rendered the auditory spectral contrast account unviable, or at least incomplete. However, the auditory process of spectral contrast is not purported to be an exclusive explanation of speech perception, auditory or audio-visual, and other processes are expected to be involved. More recent evidence (Holt, Stephens, and Lotto, unpublished; Stephens and Holt, 2002) serves to question the original Fowler *et al.* (2000) results and conclusions. Holt and colleagues could only replicate Fowler *et al.*'s original findings with the original stimulus materials, but not with similar materials. Upon examining the original video materials, Holt and colleagues found a critical confound in the video coinciding with test syllables. Different visual properties of the test CVs alone were sufficient to produce Fowler *et al.*'s original results, even without contextual precursors.

There are a large number of experimental precedents for spectral contrast. Summerfield *et al.* (1984) established the existence of an "after effect" in vowel perception. When a uniform harmonic spectrum was preceded by a spectrum that was complementary to a particular vowel with troughs replacing peaks and vice versa, listeners reported hearing a vowel during presentation of the uniform spectrum. The vowel percept (for the uniform spectrum) was appropriate for a spectrum with peaks at frequencies where there were troughs in the preceding spectrum. Summerfield *et al.* (1984) noted that perceiving vowel sounds in uniform spectra (following appropriate complementary spectral patterns) has a well-known precedent in psychoacoustics. This oft-reported finding is that, if just one member of a set of harmonics of equal amplitude is omitted from a harmonic series and is reintroduced, then it stands out perceptually against the background of pre-existing harmonics (Cardozo, 1967; Green, McKey, and Licklider, 1959; Houtgast, 1972; Schouten, 1940; Viemeister, 1980).

Summerfield and colleagues (Summerfield *et al.*, 1984; Summerfield, Sidwell, and Nelson, 1987) suggested that either simple adaptation or adaptation of suppression could account for their findings of enhanced perception of energy in spectral regions where previously there had been relatively little energy. By one simple adaptation account, neurons adapt (become less sensitive), and the enhanced prominence of the added harmonic(s) is due to the fact that neurons tuned to its frequency were not adapted prior to its onset. In contrast, some researchers (e.g., Houtgast, 1974; Moore and Glasberg, 1983) have suggested that adaptation serves mostly to enhance onsets selectively. Suppression, then, is hypothesized to be a process through which differences in level of adjacent spectral regions in complex spectra (e.g., formants in speech signals) are preserved and/or enhanced. Viemeister and Bacon (1982) showed that not only was an "enhanced" target tone more detectable; the tone also served as a more effective masker of a following tone. They suggested that suppression must be included in an adaptation scenario to place it in closer accord to this finding.

Different frequency components of a signal serve to sup-

press one another, and Viemeister and Bacon (1982) suggest that what is relevant is not so much adaptation to absolute amplitude of specific frequency components, but rather individual spectral channels are less able to suppress neighboring channels. This explanation is consistent with studies of two-tone suppression which has been cast as an instance of lateral inhibition in hearing (Houtgast, 1972). Investigators have argued that suppression helps to provide sharp tuning (e.g., Festen and Plomp, 1981; Wightman, McGee, and Kramer, 1977). With respect to speech perception, Houtgast (1974) has argued that this process serves to sharpen the neural projection of a vowel spectrum in a fashion that effectively provides formant extraction. One way to conceptualize these processes is that they serve to provide simultaneous spectral contrast, enhancing prominences versus spectral regions of lesser energy.

There exist several neurophysiological observations that bear upon enhancement effects. In particular, a number of neurophysiological studies of auditory-nerve (AN) recordings (e.g., Smith, 1979; Smith, Brachman, and Frisina, 1985; Smith and Zwislocki, 1971) provide evidence for peripheral adaptation. Delgutte and colleagues (Delgutte, 1980, 1986, 1996; Delgutte *et al.*, 1996; Delgutte and Kiang, 1984) have established the case for a much broader role of peripheral adaptation for perception of speech. They note that peaks in AN discharge rate correspond to spectro-temporal regions that are rich in phonetic information, and that adaptation increases the resolution with which onsets are represented. This role of adaptation for encoding onset information is consistent with earlier observations noted above. Delgutte and colleagues (1996) note neurophysiological evidence that ''adaptation enhances spectral contrast between successive speech segments'' (p. 3.) This enhancement arises because a fiber adapted by stimulus components close to its CF is relatively less responsive to subsequent energy at that frequency, while stimulus components not present immediately prior are encoded by fibers that are unadapted—essentially the same process offered by psychoacousticians but now grounded in physiology. Delgutte also notes that adaptation takes place on many time scales. In general, adaptation extends over longer intervals with increasing level in the auditory system. Some of the effects described above (particularly those that are temporally extended) may be less likely to have very peripheral origin.

Inspired by Summerfield and his colleagues' vowel after-effect studies (1984, 1987), the present series of experiments investigates the extent to which the same or similar processes may be responsible for the findings with very simple sine-wave flanking stimuli (e.g., Holt *et al.*, 2000; Lotto and Kluender, 1998) and those earlier studies which used rich spectra that were complementary to those for vowel sounds. Although there is an extended history of using sine waves and FM glides as nonspeech proxies for formants, such sounds have only limited resemblance to speech formants. While it is true that spectrograms illustrate formants as bands of energy and formant transitions as bands of energy traversing frequency, such pictorial descriptions can be misleading. For example, if fundamental frequency ($f0$) is constant, individual harmonics of the fundamental do not change frequency; only relative amplitudes of harmonics change. Individual frequency components of the speech spectrum change frequency no more than $f0$ changes. In an early report concerning effects of simple nonspeech context (FM and constant-frequency sine waves), Lotto and Kluender (1998) used the term ''frequency contrast'' to describe their effects. This may be a misnomer because the frequency of spectral components did not change in their speech stimuli. Describing effects as frequency contrast implied effects due to frequency differences rather than as differences in relative amplitude or spectral envelope.

The present experiments bridge this gap between simple sine waves as weak renditions of speech and the empirical precedents of Summerfield and his colleagues by investigating whether ''after-image'' effects translate into the same sorts of changes in phonetic perception brought about by simpler nonspeech precursors. They examine the perception of VCV sequences for which characteristics of the initial vowel have dramatic effects on the spectral characteristics of the following consonant in production (Öhman, 1965). And, like the findings reported above for perception of vowels in CVCs, perception of spectral information for the consonant in VCVs is complementary to the facts of production (Holt, 1999).

Experiments reported below examine the effects of vowels and spectral complements of vowels on perception of a subsequent consonant–vowel series. Instead of simple sine-wave nonspeech analogs of vowel precursors, spectral complement precursors are complex harmonic spectra that here are hypothesized to affect consonant perception in a manner complementary to vowel precursors. First, for typical speech VCVs, spectral prominences (formants) in vowel precursors should influence perception of following stops by enhancing the perceptual prominence of spectral energy away from prominences (at contrasting frequencies.) For example, greater energy in the higher-frequency region of $F2$ in [e] should encourage perception of a lower-frequency $F2$ stop [b] following the vowel, and greater energy in the lower-frequency region of $F2$ in [o] should encourage perception of the higher-frequency $F2$ stop [d]. Following the same assumptions, spectral complements of [e] and [o] should provide complementary patterns of perception. The absence of spectral energy in the spectral complement precursors should enhance perception of those frequencies not represented in the precursors.

II. EXPERIMENT 1

A. Method

1. Subjects

Eighteen native-English speaking undergraduates at the University of Wisconsin-Madison participated in return for course credit in Introductory Psychology. All subjects reported normal hearing.

2. Stimuli

Half of the stimuli consisted of VCV disyllables with digitally created renditions of vowels [e] or [o] followed by each member of a six-step series of CV syllables varying

perceptually from [ba] to [da]. The other stimuli consisted of the same six-step CV series, except these were preceded by spectral complements [~e] and [~o]. To create precursors, a 500-ms harmonic spectrum with a fundamental frequency of 120 Hz and approximately 6-dB/octave roll-off was created using a MATLAB© implementation of the Klatt (1980) speech synthesizer (Kiefte, Kluender, and Rhode, 2002). The source output was extracted before going through formant filters. Spectra were filtered using FIR filtering in MATLAB©. A 2000-order filter with a Blackman window was used to ensure maximal stopband attenuation. To create the vowel-like precursors, the harmonic spectrum was passed through a passband filter, leaving harmonics corresponding to formants. For the [e] precursor, $F1$ was represented by harmonics at 360 and 480 Hz, $F2$ by harmonics at 1800, 1920, and 2040 Hz, and $F3$ by harmonics at 2520 and 2640 Hz. The [o] precursor was the same as [e] except harmonics for $F2$ were at 720, 840, and 960 Hz. To create the spectral complement precursors [~e] and [~o], the harmonic spectrum was passed through stopband filters at the same values, leaving no energy in regions occupied by harmonics in the vowel-like precursors. Owing to the length of the impulse response for such steep filters, the initial portions of the original 500-ms waveform are not filtered as accurately as the latter portions. Consequently, the final 100 ms of each precursor was excised at zero crossings at the beginning and end of pitch pulses.

Duration of precursors was varied because it may influence the magnitude of effects upon perception of following CV. In particular, effects of some processes of adaptation may be expected to be larger with increasing duration of precursor. There was a second unexpected benefit to using variable-duration precursors. Pilot testing revealed that some listeners may actively attend away from nonspeech precursors to the extent that precursors seem superfluous to the their task of identifying the following CV. In some related earlier work (e.g., Holt, 1999; Holt *et al.*, 2000) listeners were asked to explicitly label precursors in some fashion, so listeners did not have the opportunity to neglect precursors. The use of precursors of variable durations made it more difficult for listeners to neglect precursors because the CV to be labeled could occur at different intervals following the onset of the precursor. While study of effects of attention may be of interest as a separate question, in this case, variable precursor durations can simply be taken as more representative of normal continuous processing of connected speech for which information can continually be gleaned from the signal. Precursors were therefore 100-, 200-, and 300 ms in duration. The 200- and 300-ms precursors were created by digitally iterating the 100-ms samples. Each of the 12 precursors ([e], [o], [~e], and [~o] at 3 durations) had 5-ms linear ramps at onset and offset.

A six-step [ba–da] CV series varying only in $F2$-onset frequency was synthesized at a 10-kHz sampling rate with 12-bit resolution on the cascade branch of the Klatt speech synthesizer (1980). Syllables were 250 ms in total duration with a constant $f0$ of 120 Hz. Nominal $F2$ onset frequency varied in 120-Hz steps from 1080 to 1680 Hz. Over the first 40 ms, $F2$ ramped linearly from onset frequency to 1200 Hz, and then held constant. Frequency of $F1$ rose linearly from 400 to 700 Hz over 40 ms and was constant for the remainder of the syllables. Frequencies of $F3$, $F4$, and $F5$ were set at constant values of 2580, 3500, and 4500 Hz, respectively. This series was used as the test series in all of the present experiments. Precursors and CV syllables were rms matched in amplitude, and each of the 12 precursors ([e], [o], [~e], and [~o] at 3 durations) was concatenated with each of the six CV syllables separated by 10-ms intervening silence. Figure 1 provides schematic spectrograms for each precursor with the CV series.

3. Procedure

Stimulus presentation and response collection were under the control of an 80486-25 microcomputer. Following D/A conversion (Ariel DSP-16), stimuli were low-pass filtered (4.8 kHz cutoff frequency, Frequency Devices, #677), amplified (Stewart HDA4), and presented to subjects via headphones (Beyer DT-100) at a level of 75 dB SPL.

Listeners participated in a 2AFC identification task. One to three subjects were tested concurrently in individual sound-attenuated chambers during a single experimental session. In a completely mixed design ([eCa], [oCa], [~eCa], and [~oCa]), participants identified CVs as [ba] or [da] by pressing either of two buttons on a handheld electronic response box with buttons labeled ''BA'' and ''DA.'' Over two blocks of 360 presentations, participants responded to each stimulus ten times. The entire session lasted approximately 40 min. After the session, participants were informally asked to identify the precursor sounds. Vowel precursors were easily identified, while spectral complement precursors were more difficult to label. Those listeners who attempted to identify them reported hearing a buzz.

B. Results and discussion

To ensure that listeners performed competently, data from listeners who respond with at least 90-percent consistency on unambiguous endpoint stimuli were analyzed. Only two listeners failed to meet this criterion. Figure 2 displays mean consonant identification for the vowel and vowel complement precursors in experiment 1. Data for vowel precursors at each duration were analyzed via paired t-tests.[1] Consistent with previous results (e.g., Holt, 1999), there was a significant effect of vowel context on consonant identification for all precursor durations: 100 ms, $t(15)=3.48$, $p<0.01$; 200 ms, $t(15)=2.68$, $p<0.05$; 300 ms, $t(15)=4.92$, $p<0.001$. Results of t-tests performed on probit data revealed the same pattern of results.[2] Listeners were much more likely to report hearing [da] in the [oCa] context than in the [eCa] context.

Data for spectral complement precursors were analyzed in the same fashion. The 100-ms spectral complement precursors [~e] and [~o] had a significant effect on consonant identification, $t(15)=3.42$, $p<0.001$. However, 200- and 300-ms spectral complement precursors did not, $t(15)=-0.07$, $p=0.94$, and $t(15)=-0.38$, $p=0.66$, respectively.

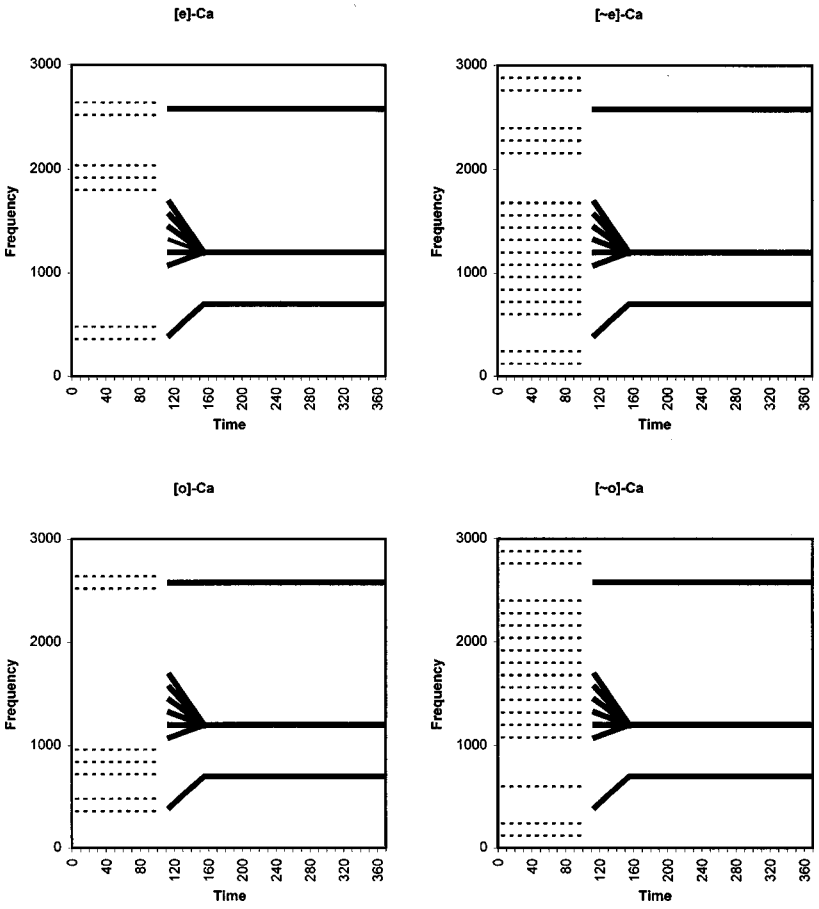


FIG. 1. Schematic renditions of stimuli for experiment 1. All harmonics of precursors are illustrated, and are represented by thin dotted lines. Formant center frequencies of the [ba]–[da] series are represented by thick solid lines (all harmonics present in test stimuli). The left column includes [e] (top) and [o] precursors preceding the [ba–da] series, and the right column includes spectral complement [~e] and [~o] precursors.

Only for the briefest precursors, listeners were more likely to report hearing [da] after spectral complement [~e] than after [~o].

While complementary effects for spectrally complementary precursors were predicted, the presence of this effect for only the briefest precursor durations was not fully expected. The lack of effect for longer precursors is not expected if one assumes either adaptation or adaptation of suppression (e.g., Viemeister and Bacon, 1982) as an explanation for this effect. However, adaptation of suppression was but one of several instances of adaptation in the auditory system. In experiment 2, the importance of precursor duration will be investigated *vis à vis* other instances of adaptation and their putative effects on encoding of stimulus onsets.

There are other reasons one might expect that nonspeech precursors should provide less effect on perception of following stops. Listeners have a tremendous amount of experience hearing coarticulated VCVs. Repp (1982) argues that ''listeners make continuous use of their tacit knowledge of speech patterns'' (p. 81) in perceiving coarticulated speech. Coarticulation yields multiple covariances in the signal that are orderly in as much as they reflect dependable regularities in physical constraints upon articulators. Perceptual learning would be expected to seize upon such dependable relationships between stimulus qualities, and one would expect results from experienced listeners to reflect perceptual experi-

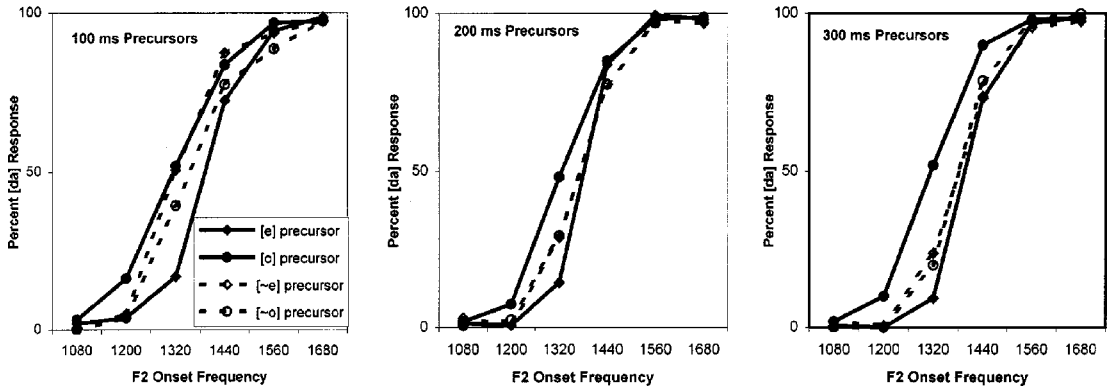


FIG. 2. Data from identification (/b/,/d/) task in experiment 1 for each precursor duration. Percent /da/ responses as a function of second formant onset frequency are plotted by precursor context. Solid lines denote the functions for [e] and [o] precursors. Data for spectral complements [~e] and [~o] are shown with dashed lines.

ence with speech in addition to less domain specific auditory processes (Holt and Kluender, 2000; Lindblom and Studdert-Kennedy, 1967).

C. Additional conditions

A series of follow-up experiments was conducted to examine whether spectral properties of the precursors might explain why longer spectral complement precursors failed to elicit the predicted effects. For the sake of brevity, and because the results across conditions are very consistent, the follow-up experiments are summarized below.

1. Single peak or trough

First, the spectral contrast effects were replicated with precursors with just a single spectral prominence or trough. A spectral contrast account suggests that effects in the previous experiment result, in whole or in part, from the presence or absence of energy in the frequency range of $F2$. While precursors in experiment 1 differ only in the frequency of the second spectral peak or trough, they are nevertheless acoustically rich stimuli. In the case of renditions of [e] and [o], precursors were clearly perceived as /e/ and /o/. And, listeners came to the laboratory with ample experience about acoustic properties of consonants following these two vowel sounds. Single-band (pass and stop) precursors that share much less acoustic resemblance to speech sounds were used in the interest of attenuating effects that are due to experience with coarticulated speech sounds. In these conditions, results matched those for the three-formant vowel and spectral complement precursors. Effects were significant for single-passband precursors at all durations, while effects for single-stopband precursors were only significant at the shortest duration.

2. Equalized cochlear distance

Next, cochlear area was considered as a potentially confounding variable. Because the cochlea is organized in a roughly logarithmic fashion above 1 kHz, a three-harmonic spectral prominence or trough centered at a lower frequency (typical of $F2$ for [o]) will map onto a greater cochlear area than one centered at a higher frequency (typical of $F2$ for [e]). Therefore, bandwidths were adjusted using equal rectangular bandwidth (ERB) estimates of auditory filter width (Moore and Glasberg, 1983). The ERB scale also corresponds well to estimates of human cochlear distance (Greenwood, 1990). In this instance, three harmonics at a lower frequency (typical of $F2$ for [o]) map onto an equivalent cochlear area as five harmonics at higher frequencies (typical of $F2$ for [e]). Therefore, three-formant vowel and spectral complement precursors, along with single spectral prominence and trough precursors were created with and without ERB-rate normalization. Because of the large number of precursors, 300-ms precursors were omitted from this condition. For the vowel and single-spectral prominence precursors, effects were significant for 100- and 200-ms precursors, replicating experiment 1. For the spectral complement precursors, results generally matched those for experiment 1. When comparing ERB-rate normalized three-formant and single-formant spectral complement precursors, spectral contrast effects were significant for shorter, 100-ms precursors, but not for longer 200-ms ones. However, when comparing three-formant spectral complement precursors with and without ERB-rate normalization, spectral contrast effects were significant for both 100- and 200-ms precursors. Because the same non-ERB-normalized stimuli were used in experiment 1 and this follow-up, it is not clear what accounts for this difference. It may be due to subject variance or to some unknown effect of removing 300-ms precursors from the task. Data from additional experiments, however, indicate that the likely explanation is that the temporal extent of these effects begins to wane at 200 ms, but is not yet extinguished.

3. Random phase harmonics

Finally, we consider that listeners may somehow ''fill in'' spectral troughs over the longer time course of 300-ms, and sometimes 200-ms, precursors. For natural harmonic spectra, harmonics are in phase. Duifhuis (1970) and Wightman (1973) observed that, when listeners are presented with harmonic complexes from which specific harmonics have been removed, missing harmonics could potentially be derived from the temporal waveform. Horst, Javel, and Farley (1990) found that, when a center component of a harmonic complex was left out, this missing component was restored in the pattern of discharges in AN at higher stimulus levels. In recent studies of basilar-membrane mechanics, Rhode and Recio (2001) presented equal-interval seven-tone complexes. When the center component was deleted and only six tones were presented, the center component was restored in the basilar-membrane response as a result of distortion-product generation in the nonlinear cochlea.

In all experiments reported here thus far, precursors were filtered versions of a harmonic spectrum created via a MATLAB© implementation of the Klatt80 synthesizer (Kiefte *et al.*, 2002; Klatt, 1980), and all harmonics had the same phase. It is possible that listeners in the experiments above perceptually restored missing harmonics for longer precursor stimuli, but not for the shortest precursors. Consequently, spectral troughs may have been only a temporary characteristic of these stimuli as represented in the auditory spectrum. Randomizing phase of harmonics makes it impossible to recover missing harmonics. Spectral complement precursors synthesized by Summerfield and colleagues (1984, 1987) had harmonics in random phase. Their effects maintained even though precursors were 1 s length. To control for this possibility, new precursors were created such that their harmonics were in random phase. Results for vowel precursors match previous results, with significant spectral contrast effects for precursors of all durations. For spectral complement precursors, spectral contrast effects were significant for the shorter 100- and 200-ms precursors, but not for the longer 300-ms precursors.

Thus far, the results for three- (vowel) and one-passband precursors are consistent with a general description of spectral contrast in speech perception. Spectral energy at higher frequencies makes subsequent spectral energy at lower frequencies more effective, and vice versa. This is true at each of the three precursor durations tested. These effects are amenable to explanation via processes of adaptation or adap-

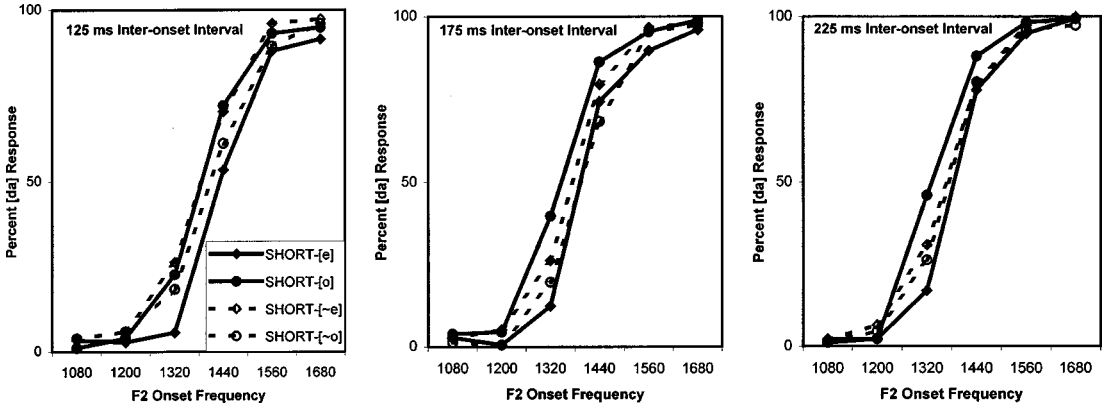


FIG. 3. Data from identification (/b/,/d/) task in experiment 2 for very short (25-ms) precursors and three interonset intervals. Percent /da/ responses as a function of second formant onset frequency are plotted by precursor type. Solid lines indicate data for three-peak SHORT-[e] and SHORT-[o] precursors, and data from spectral complement SHORT-[~e] and SHORT-[~o] precursors are shown with dashed lines.

tation of suppression in a relatively straightforward way. Results for three- and single-stopband (complementary) precursors appear more difficult to interpret. Effects of troughs in the spectrum are consistent across conditions when duration is 100 ms. These effects are less reliable for 200-ms precursors, and are functionally absent for 300-ms precursors.

One hypothesis for the effect of spectral troughs at short but not long precursor durations is that effects are due to some change in the processing of onset properties of stimuli. While, up to this point, longer precursors may be thought of as providing greater adaptation, what has been neglected is the fact that longer precursors result in onsets of precursors being displaced further in time from onsets of CVs. It is widely appreciated that onset properties of speech sounds are very important for listeners (e.g., Blumstein and Stevens, 1981). As noted above, Delgutte (1996) argues that adaptation in AN increases the spectral resolution with which onsets are represented. From psychoacoustics, others (e.g., Houtgast, 1974; Moore and Glasberg, 1983) have suggested that adaptation serves mostly to enhance onsets selectively. Experiment 2 is designed to test the importance of contrast between successive (precursor and CV) onsets. To the extent that precursor onset properties give rise to contrast effects with the following CV, precursor duration *per se* does not matter. Experiment 2, therefore, uses only very short (25 ms) precursors presented at variable intervals preceding CVs.

III. EXPERIMENT 2

A. Method

1. Subjects

Two groups of native-English speaking undergraduates at the University of Wisconsin–Madison participated in return for course credit in Introductory Psychology. All reported normal hearing. The first group of 22 listeners heard only the CV series with SHORT-[e] and SHORT-[o] precursors. A second group of 23 listeners heard only the CV series with spectral complement SHORT-[~e] and SHORT-[~o] precursors.

2. Stimuli

Precursors were created from the same three-passband ERB-normalized vowels [e] and [o] and their spectral complements [~e] and [~o] used previously. Based on the 120-Hz fundamental frequency, each pitch pulse is 8.33 ms in duration. Precursors that were 25 ms in duration were created from three pitch pulses. Three interonset intervals of 125, 175, and 225 were created by inserting 100, 150, or 200 ms of silence between precursor offset and CV onset.

3. Procedure

Stimulus presentation and response collection were identical to experiment 1. Each group of listeners heard only one set of stimuli, either with spectral peak or spectral trough precursors. Each stimulus was presented 8 times in random order for a total of 288 presentations (2 precursors×3 intervals×6 CV syllables×4 presentations×2 blocks). Experimental apparatus and procedure were identical to those for experiment 1. The experimental session lasted approximately 20 min.

B. Results and discussion

Figure 3 displays mean consonant identification curves for experiment 2. Data for spectral peak and spectral trough precursors for each interval were examined in paired t-tests. Results for SHORT-[e] and SHORT-[o] precursors correspond well with those from all experiments reported above with significant effects for all onset–onset durations: 125 ms, $t(21)=4.03$, $p<0.001$; 175 ms, $t(21)=5.57$, $p<0.0001$; and 225 ms, $t(21)=4.63$, $p<0.0001$. As in all previous conditions, listeners heard [d] more often in the [o] context than in the [e] context, even with these very short precursors. In spectral complement conditions, four listeners failed to reach 90-percent consistency on unambiguous endpoint stimuli, and results from the remaining 18 listeners were analyzed. Onset properties of these 25-ms SHORT-[~e] and SHORT-[~o] precursors significantly alter perception of subsequent CVs as they did in previous experiments when precursor onsets were separated by 125 ms, $t(18)=2.80$, $p<0.05$, and by 175 ms, $t(18)=3.40$, $p<0.01$. Consistent with the inconsistent effects for 200-ms precursors in the previous experi-

ments, the difference between precursors was not significant when 225 ms intervened between onsets, $t(18)=1.75$, $p=0.10$.

Even very short 25-ms precursors affect perception of subsequent CVs in a manner consistent with contrast between onset spectra. Across all conditions, effects of vowel precursors have been constant across variations in precursor duration, bandwidth in the region of $F2$, presence or absence of lower and higher spectral energy (in the region of $F1$ and $F3$), phase of harmonics, and duration of intervening silence. In all cases, no matter the manipulation, listeners are more likely to report hearing higher-frequency [da] following lower-frequency energy corresponding to $F2$ in [o].

Effects for spectral complement precursors, however, appear more circumscribed. As is the case for one- and three-peak passband stimuli, patterns of results for one- and three-trough stopband precursors are strikingly similar. However, unlike effects for spectral peaks, the strength of the effects of spectral complements appears to depend critically upon the duration of the interval between onset of precursor and onset of CV. Aside from this difference, the patterns of findings correspond well to prediction on the basis of spectral contrast between precursor and CV onsets. Listeners are more likely to hear higher-frequency [da] following [~e] for which there is an absence of energy in the higher-frequency region corresponding to $F2$ in [e]. This effect of spectral troughs wanes as the time between the onset of the precursor and the onset of the test syllable increases. Interestingly, it does not seem to be influenced by the presence or absence of intervening acoustic energy. Identification of subsequent consonants was influenced by spectral complement precursors when precursors were 100 ms in duration followed by 10 ms of silence, and when they were 25 ms in duration followed by 100-ms silence.

For a number of reasons, the fact that onset properties have such a great effect for both vowel and spectral complement stimuli in experiment 2 and earlier experiments is not entirely surprising. Perhaps most obviously, it is the perception of CV onsets that is being affected by precursors. The [ba–da] series is defined by variation in onset frequency of $F2$. Transitions were 40 ms in duration, and all members of the series become increasing similar over the first 40 ms until all six stimuli share common spectral properties. More generally, it is well known that onsets of sounds are physiologically significant, not only in AN neurons, but also in neurons in cochlear nucleus (e.g., Rhode, 1991; Winter and Palmer, 1995) and at successive stages of the auditory system through to auditory cortex (Heil, 1997a, 1997b; Phillips and Hall, 1990).

The spectral complement precursors used in the current series of experiments provide a strong test of spectral contrast accounts for some phenomena of speech perception. They are acoustically rich signals, created from harmonic spectra modeled after the human voice source. They do not mimic speech, but rather, are complementary to it. However, listeners seldom hear such sounds outside of the speech laboratory. Nevertheless, one may expect that processes underlying effects demonstrated above should be well suited to the perception of running speech. In fluent speech, onsets are temporally displaced, and acoustic energy typically exists between onsets. In order to test the extent which the present findings generalize to more complex speech signals that are more representative of connected speech, experiment 3 was conducted to test the extent to which the onset spectra of a preceding syllable ([ya] and [wa]) affect perception of the initial stop in following syllables ([ba–da]). Glide-vowel syllables [ya] and [wa] were chosen because onset spectra could be crafted to mimic spectral properties of precursors used above.

IV. EXPERIMENT 3

A. Method

1. Subjects

Forty-one native-English speaking undergraduates at the University of Wisconsin-Madison participated in return for course credit in Introductory Psychology. All reported normal hearing.

2. Stimuli

Precursor syllables were synthesized at a 10-kHz sampling rate with 12-bit resolution on the cascade branch of the Klatt speech synthesizer (1980). Syllables were 100 ms in duration with a constant fundamental frequency of 120 Hz. Across 50 ms, $F1$ increased from 400 to 700 Hz before remaining at 700 Hz until the end of the syllable. Onset of $F2$ for [ya] was 1920 Hz, and that for [wa] was 840 Hz. These values match those for the center frequencies of the vowels [e] and [o], respectively, used in the previous experiments. $F2$ changed linearly from onset frequency to 1200 Hz over 50 ms and then held constant. Values for $F3$, $F4$, and $F5$ were held constant at 2580, 3500, and 4500 Hz, respectively, throughout the syllables. Owing to the relatively brief (50-ms) vowel portion of the [ya] and [wa] syllables, these were clearly heard as glides, not stops. Perception of these syllables as glides is consistent with earlier findings by Miller and Liberman (1979) demonstrating perception of stops and semivowels at different durations of transitions and following vowels. Precursor glide-vowel syllables and following stop-glide syllables were rms matched in amplitude before being concatenated with 40-ms intervening silence. Thus, the interval between syllable onsets is 140 ms. Figure 4 provides formant tracks for both precursors with the CV series.

3. Procedure

Stimulus presentation and response collection were identical to experiment 1. Each stimulus was presented 10 times in random order for a total of 120 presentations (2 precursors [ya], [wa])×6 CV syllables×5 presentations×2 blocks). Experimental apparatus and procedure were identical to those for experiment 1. The experimental session lasted approximately 15 min.

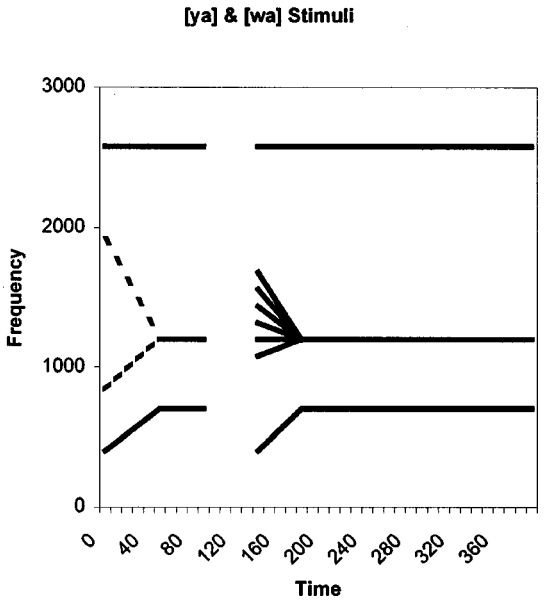


FIG. 4. Schematic depictions of center frequencies of the first three formants of stimuli for experiment 3. Onsets of glide-vowel syllable precursors (left) have identical center frequencies as [e] and [o] precursors used in previous experiments.

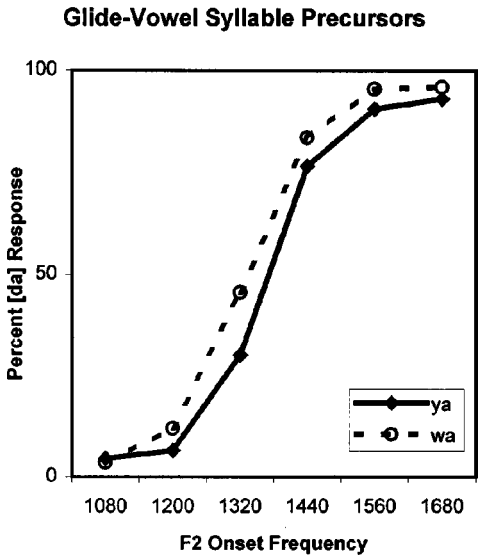


FIG. 5. Data from identification (/b/,/d/) task following [ya] (filled diamonds) and [wa] (open circles) syllables in experiment 3.

B. Results and discussion

Figure 5 displays the mean consonant identification curves for experiment 3. Data from the 20 listeners who met the endpoint criterion were submitted to a one-sample t-test.[3] Listeners reported hearing [da] more often after [wa] than after [ya], $t(19)=2.99$, $p<0.01$. Spectral characteristics of the onsets of the precursor stimuli provided sufficient spectral contrast to affect the perception of subsequent speech, even following considerable intervening acoustic information.

V. DISCUSSION

Significance of spectral contrast for perception of coarticulated speech previously had been demonstrated for a variety of phonetic environments (e.g., CVVC: Lotto and Kluender, 1998; CVC: Holt, Lotto, and Kluender, 2000; VCV: Holt, 1999). In all previous work, very simple sine-wave flanking stimuli were used. Here, investigation of spectral contrast was extended to flanking stimuli with richer spectral composition. Single bands of spectral energy, approximating a single speech formant ($F2$), proved adequate to shift perception of a following stop to virtually the same extent as was found for replicas of vowels [e] and [o].

The most novel aspect of the present effort concerned the use of complementary precursor spectra for which spectral peaks were replaced with spectral troughs. Inasmuch as they were harmonic complexes, these complementary spectra were comparable to speech in complexity. Yet, they rendered no speech percept. As would be predicted by a spectral contrast account, complementary spectral precursors affected perception of following stops in a manner complementary to that for speech stimuli upon which they were modeled. These spectral complements were similarly effective in modifying perception of the following stop with either one or three stopbands, as was the case for one- and three-passband stimuli. From this, it can be inferred that the principal source of effects on perception of following stops rests in characteristics of the preceding spectrum in the region of $F2$—the principal acoustic determinant of perception of labial versus alveolar stops in the test series.

Results for complementary spectra differed in two important ways. First, contrast between spectral properties of the onsets of precursors versus CV syllables was most important. Second, due to the importance of onsets, the temporal distance between successive onsets proved critical. Effects of contrast between onsets appears to have a temporal extent that exceeds 100 ms and is less than 300 ms. Across experiments, it can be inferred that effects of precursor onsets are diminishing and become less reliable at around 200-ms interonset interval. The importance of interonset interval, versus precursor duration *per se*, appears less consistent with adaptation of suppression (Viemeister and Bacon, 1982) and more consistent with the perspective that adaptation serves mostly to enhance characteristics of onsets (e.g., Delgutte, 1996; Houtgast, 1974; Moore and Glasberg, 1983). It is important to note, however, that adaptation and adaptation of suppression are not mutually exclusive models. All that can be said is that adaptation provides the best explanation of the present data for spectral complements, and adaptation of suppression likely remains the best explanation for the many instances of auditory enhancement. This probably includes Summerfield and colleagues' demonstrations of vowel aftereffects and some of the effects of flanking speech (bandpass) and nonspeech (sine-wave) energy on perception of adjacent speech sounds.

Results from experiment 2 demonstrate that contrast between onsets is perceptually significant for both spectral peaks and spectral troughs, and results from experiment 3 provide evidence that these effects are sufficiently robust to play a role in connected speech for which there exists ample energy intervening between onsets.

Contrast effects are ubiquitous in perception, occurring

in every modality. This universality finds its cause in the fact that sensorineural systems respond predominantly to change. Within a modality, one may expect to find effects of contrast for every dimension within that modality. This certainly appears to be the case for vision, for which contrast plays a role in perception of lightness, line orientation, size, position, curvature, spatial frequency, depth, and color. One may similarly expect processes of contrast to be present for every auditory dimension including contrast between spectral properties of onsets.

ACKNOWLEDGMENTS

This work was supported by the National Institute of Deafness and Communicative Disorders, Grant No. R01 DC 04072. The authors thank Chris Darwin, Christopher Turner, and an anonymous reviewer for insightful suggestions on an earlier version of this manuscript.

[1]Data for number of [da] responses were entered into a 2 ([e], [o])×2 (passband, stopband)×3 (duration)×6 (stimulus steps) within-subjects ANOVA. Consistent with the large effect apparent in Fig. 2, statistical analyses reveal a significant interaction between passband/stopband and duration.

[2]There is no reason to believe that effects of spectral contrast are restricted to regions near the identification crossover (boundary). Consequently, total response is the most appropriate measure of effect size (see also Samuel, 1986). This is especially true for the series of CV stimuli used here because all six members of the series are within three steps from identification crossover points. Nevertheless, crossover points were estimated using probit analysis (Finney, 1971) for every experiment reported here, and in only four cases did probit values result in statistical patterns different from those for mean response data. In all but one case (200-ms spectral complements in the second follow-up experiment, $p=0.106$, two-tail), probit differences were significant in a one-tail comparison but fell short of significance for two-tail.

[3]It is not clear why there was such a high attrition rate in this study. Syllables [ya] and [wa] are the most realistic stimuli used across all experiments reported here. Apparently listeners found this task very difficult, perhaps because of the similarity between precursor and test syllables. Some listeners fell just short of the 90-percent endpoint criterion. Others responded based on the precursor syllable identities. Still others seemed to respond randomly for unknown reasons. However, one can be confident that the remaining listeners heard the syllables as intended, because those who did not were removed based on the endpoint criterion. This study was replicated, with 13 of 28 listeners failing to meet the endpoint criterion. The results of the remaining 15 listeners replicate the original finding, $t(14)=2.58$, $p<0.05$.

Ames, A. (**1935**). ''Aneiseikonia—A factor in the functioning of vision,'' Am. J. Ophthalmol. **28**, 248–262.

Bergman, R., and Gibson, J. J. (**1959**). ''The negative aftereffect of a surface slanted in the third dimension,'' Am. J. Psychol. **72**, 364–374.

Blakemore, C., and Sutton, P. (**1969**). ''Size adaptation: A new aftereffect,'' Science **166**, 245–247.

Blumstein, S. E., and Stevens, K. N. (**1981**). ''Phonetic features and acoustic invariance in speech,'' Cognition **10**, 25–32.

Cardozo, B. L. (**1967**). ''Ohm's Law and masking,'' in *IPO Annual Progress Report, Institute for Perception Research* (Eindhoven, The Netherlands), **2**, pp. 59–64.

Cathcart, E. P., and Dawson, S. (**1928–1929**). ''Persistence (2),'' Br. J. Psychol. **19**, 343–356.

Christman, R. J. (**1954**). ''Shifts in pitch as a function of prolonged stimulation with pure tones,'' Am. J. Psychol. **67**, 484–491.

Delgutte, B. (**1980**). ''Representation of speech-like sounds in the discharge patterns of auditory nerve fibers,'' J. Acoust. Soc. Am. **68**, 843–857.

Delgutte, B. (**1986**). ''Analysis of French stop consonants with a model of the peripheral auditory system,'' in *Invariance and Variability of Speech Processes*, edited by J. S. Perkell and D. H. Klatt (Erlbaum, Hillsdale, NJ), pp. 131–177.

Delgutte, B. (**1996**). ''Auditory neural processing of speech,'' in *The Handbook of Phonetic Sciences*, edited by W. J. Hardcastle and J. Laver (Blackwell, Oxford), pp. 507–538.

Delgutte, B., Hammond, B. M., Kalluri, S., Litvak, L. M., & Cariani, P. A. (**1996**). ''Neural encoding of temporal envelope and temporal interactions in speech,'' in *Auditory Basis of Speech Perception*, edited by W. Ainsworth and S. Greenberg (European Speech Communication Association), pp. 1–9.

Delgutte, B., and Kiang, N. Y. S. (**1984**). ''Speech coding in the auditory nerve IV. Sounds with consonant-like dynamic characteristics,'' J. Acoust. Soc. Am. **75**, 897–907.

Denes, P. (**1955**). ''Effect of duration on perception of voicing,'' J. Acoust. Soc. Am. **27**, 761–764.

Diehl, R. L., Kluender, K. R., and Walsh, M. A. (**1990**). ''Some auditory bases of speech perception and production,'' in *Advances in Speech, Hearing, and Language Processing*, edited by W. A. Ainsworth (JAI, London), pp. 243–267.

Diehl, R. L., and Walsh, M. A. (**1989**). ''An auditory basis for the stimulus-length effect in the perception of stops and glides,'' J. Acoust. Soc. Am. **85**, 2154–2164.

Duifhuis, H. (**1970**). ''Audibility of high harmonics in a periodic pulse,'' J. Acoust. Soc. Am. **48**, 888–893.

Festen, J. M., and Plomp, R. (**1981**). ''Relations between auditory functions in normal hearing,'' J. Acoust. Soc. Am. **70**, 356–369.

Finney, D. J. (**1971**). *Probit Analysis* (Cambridge University Press, New York).

Flügel, J. C. (**1920–1921**). ''On local fatigue in the auditory system,'' Br. J. Psychol. **11**, 105–134.

Fowler, C. A., Brown, J. M., and Mann, V. A. (**2000**). ''Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans,'' J. Exp. Psychol. Hum. Percept. Perform. **26**, 877–888.

Gibson, J. J. (**1933**). ''Adaptation, after-effect and contrast in the perception of curved lines,'' J. Exp. Psychol. **16**, 1–31.

Gibson, J. J. (**1937**). ''Adaptation with negative after-effect,'' Psychol. Rev. **44**, 222–244.

Gibson, J. J., and Radner, M. (**1937**). ''Adaptation, after-effect, and contrast in the perception of tilted lines. I. Quantitative studies,'' J. Exp. Psychol. **20**, 453–467.

Green, D. M., McKey, M. J., and Licklider, J. C. R. (**1959**). ''Detection of a pulsed sinusoid in noise as a function of frequency,'' J. Acoust. Soc. Am. **31**, 1146–1152.

Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Guilford, J. P., and Park, D. G. (**1931**). ''The effect of interpolated weights upon comparative judgments,'' Am. J. Psychol. **43**, 589–599.

Heil, P. (**1997a**). ''Auditory cortical onset responses revisited. I. First-spike timing,'' J. Neurophysiol. **77**, 2616–2641.

Heil, P. (**1997b**). ''Auditory cortical onset responses revisited. II. Response strength,'' J. Neurophysiol. **77**, 2642–2660.

Holt, L. L. (**1999**). ''Auditory constraints on speech perception: An examination of spectral contrast,'' unpublished Ph.D. dissertation, University of Wisconsin–Madison.

Holt, L. L., and Kluender, K. R. (**2000**). ''General auditory processes contribute to perceptual accommodation of coarticulation,'' Phonetica **57**, 170–180.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2000**). ''Neighboring spectral content influences vowel identification,'' J. Acoust. Soc. Am. **108**, 710–722.

Holt, L. L., Stephens, J. D., and Lotto, A. J. (unpublished).

Horst, J. W., Javel, E., and Farley, G. R. (**1990**). ''Coding of fine spectral structure in the auditory nerve. II. Level-dependent nonlinear responses,'' J. Acoust. Soc. Am. **88**, 2656–2681.

Houtgast, T. (**1972**). ''Psychophysical evidence for lateral inhibition in hearing,'' J. Acoust. Soc. Am. **51**, 1885–1894.

Houtgast, T. (**1974**). ''Auditory analysis of vowel-like sounds,'' Acustica **31**, 320–324.

Kiefte, M. J., Kluender, K. R., and Rhode, W. S. (**2002**). ''Synthetic speech stimuli spectrally normalized for nonhuman cochlear dimensions,'' ARLO **3**, 41–46.

Klatt, D. H. (**1980**). ''Software for a cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**(3), 971–995.

Kluender, K. R., and Diehl, R. L. (**1988**). ''Vowel-length differences before voiced and voiceless consonants: An auditory explanation,'' J. Phonetics **16**, 153–169.

Koffka, K. (**1935**). *Principles of Gestalt Psychology* (Harcourt, Brace & World, New York).
Kohler, W., and Emery, D. A. (**1947**). ''Figural aftereffects in the third dimension of visual space,'' Am. J. Psychol. **60**, 159–201.
Kohler, W., and Wallach, H. (**1944**). ''Figural aftereffects: An investigation of visual processes,'' Proc. Am. Philos. Soc. **88**, 269–357.
Lindblom, B. E. F. (**1963**). ''Spectrographic study of vowel reduction,'' J. Acoust. Soc. Am. **35**, 1773–1781.
Lindblom, B. E. F., and Studdert-Kennedy, M. (**1967**). ''On the role of formant transitions in vowel recognition,'' J. Acoust. Soc. Am. **42**, 830–843.
Locke, J. (**1706/1974**). *An Essay Concerning Human Understanding*, 5th ed., edited by A. D. Woolzley (New American Library, New York). Original work published in 1706.
Lotto, A. J., and Kluender, K. R. (**1998**). ''General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification,'' Percept. Psychophys. **60**, 602–619.
Mann, V. A. (**1980**). ''Influence of preceding liquid in stop-consonant perception,'' Percept. Psychophys. **28**, 407–412.
Mann, V. A., and Repp, B. H. (**1981**). ''Influence of preceding fricative on stop consonant perception,'' J. Acoust. Soc. Am. **69**, 548–558.
Miller, J. L., and Liberman, A. M. (**1979**). ''Some effects of later-occurring information on the perception of stop consonant and semivowel,'' Percept. Psychophys. **25**, 457–465.
Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Suggested formulas for calculating auditory-filter bandwidths and excitation patterns,'' J. Acoust. Soc. Am. **74**, 750–753.
Öhman, S. E. G. (**1965**). ''Coarticulation in VCV utterances: Spectrographic measurements,'' J. Acoust. Soc. Am. **39**(1), 151–168.
Phillips, D. P., and Hall, S. E. (**1990**). ''Response timing constraints on the cortical representation of sound time structure,'' J. Acoust. Soc. Am. **88**, 1403–1411.
Pisoni, D. B., Carrell, T. D., and Gans, S. J. (**1983**). ''Perception of the duration of rapid spectrum changes in speech and nonspeech signals,'' Percept. Psychophys. **34**, 314–322.
Port, R. F., and Dalby, J. (**1982**). ''Consonant/vowel ratio as a cue for voicing in English,'' Percept. Psychophys. **32**, 141–152.
Raphael, L. F. (**1972**). ''Preceding vowel duration as a cue to the perception of the voicing characteristics of word-initial consonants in English,'' J. Acoust. Soc. Am. **51**, 1296–1303.
Repp, B. (**1982**). ''Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception,'' Psychol. Bull. **92**, 81–110.
Rhode, W. S. (**1991**). ''Physiological–morphological properties of the cochlear nucleus,'' in *Neurobiology of Hearing: The Central Auditory System*, edited by R. A. Altschuler (Raven Press, New York).
Rhode, W. S., and Recio, A. (**2001**). ''Basilar membrane response to multicomponent stimuli in chinchilla,'' J. Acoust. Soc. Am. **110**, 981–994.
Samuel, A. G. (**1986**). ''Red herring detectors and speech perception: In defense of selective adaptation,'' Cognit. Psychol. **18**, 452–499.
Schouten, J. F. (**1940**). ''The residue, a new component in subjective analysis,'' Proc. K. Ned. Akad. Wet. **43**, 356–365.
Sherif, M., Taub, D., and Hovland, C. I. (**1958**). ''Assimilation and contrast effects of anchoring stimuli on judgments,'' J. Exp. Psychol. **55**, 150–155.
Smith, R. L. (**1979**). ''Adaptation, saturations, and physiological masking in single auditory-nerve fibers,'' J. Acoust. Soc. Am. **65**, 166–178.
Smith, R. L., Brachman, M. L., and Frisina, R. D. (**1985**). ''Sensitivity of auditory-nerve fibers to changes in intensity: A dichotomy between decrements and increments,'' J. Acoust. Soc. Am. **78**, 1310–1316.
Smith, R. L., and Zwislocki, J. J. (**1971**). ''Responses of some neurons of the cochlear nucleus to tone-intensity increments,'' J. Acoust. Soc. Am. **50**, 1520–1525.
Stephens, J. D., and Holt, L. L. (**2002**). ''Are context effects in speech perception modulated by visual information?'' 43rd Annual Meeting of the Psychonomic Society, Kansas City, MO.
Summerfield, Q., Haggard, M. P., Foster, J., and Gray, S. (**1984**). ''Perceiving vowels from uniform spectra: Phonetic exploration of an auditory aftereffect,'' Percept. Psychophys. **35**, 203–213.
Summerfield, Q., Sidwell, A., and Nelson, T. (**1987**). ''Auditory enhancement of changes in spectral amplitude,'' J. Acoust. Soc. Am. **81**, 700–707.
Viemeister, N. F. (**1980**). ''Adaptation of masking,'' in *Psychophysical, Physiological, and Behavioral Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (University Press, Delft), pp. 190–197.
Viemeister, N. F., and Bacon, S. P. (**1982**). ''Forward masking by enhanced components in harmonic complexes,'' J. Acoust. Soc. Am. **71**, 1502–1507.
Wallach, H. (**1948**). ''Brightness constancy and the nature of achromatic colors,'' J. Exp. Psychol. **38**, 310–324.
Wightman, F. L. (**1973**). ''Pitch and stimulus fine structure,'' J. Acoust. Soc. Am. **54**, 397–406.
Wightman, F., McKee, T., and Kramer, M. (**1977**). ''Factors influencing frequency selectivity in normal and hearing-impaired listeners,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 295–310.
Winter, I. M., and Palmer, A. R. (**1995**). ''Level dependence of cochlear nucleus onset unit responses and facilitation by second tones or broadband noise,'' J. Neurophysiol. **73**, 141–159.

Speech segregation based on sound localization

Nicoleta Roman[a)]
Department of Computer and Information Science, The Ohio State University, Columbus, Ohio 43210

DeLiang Wang[b)]
Department of Computer and Information Science and Center for Cognitive Science, The Ohio State University, Columbus, Ohio 43210

Guy J. Brown[c)]
Department of Computer Science, University of Sheffield, Sheffield S1 4DP, United Kingdom

(Received 25 June 2002; accepted for publication 18 July 2003)

At a cocktail party, one can selectively attend to a single voice and filter out all the other acoustical interferences. How to simulate this perceptual ability remains a great challenge. This paper describes a novel, supervised learning approach to speech segregation, in which a target speech signal is separated from interfering sounds using spatial localization cues: interaural time differences (ITD) and interaural intensity differences (IID). Motivated by the auditory masking effect, the notion of an ''ideal'' time–frequency binary mask is suggested, which selects the target if it is stronger than the interference in a local time–frequency (T–F) unit. It is observed that within a narrow frequency band, modifications to the relative strength of the target source with respect to the interference trigger systematic changes for estimated ITD and IID. For a given spatial configuration, this interaction produces characteristic clustering in the binaural feature space. Consequently, pattern classification is performed in order to estimate ideal binary masks. A systematic evaluation in terms of signal-to-noise ratio as well as automatic speech recognition performance shows that the resulting system produces masks very close to ideal binary ones. A quantitative comparison shows that the model yields significant improvement in performance over an existing approach. Furthermore, under certain conditions the model produces large speech intelligibility improvements with normal listeners. [DOI: 10.1121/1.1610463]

PACS numbers: 43.72.Ew, 43.66.Ba, 43.66.Qp [DOS]

I. INTRODUCTION

The perceptual ability to detect, discriminate, and recognize one utterance in a background of acoustic interference has been studied extensively under both monaural and binaural conditions (Bregman, 1990; Blauert, 1997; Bronkhorst, 2000). The human auditory system is able to segregate a speech signal from an acoustic mixture using various cues, including fundamental frequency ($F0$), onset time and location, in a process that is known as *auditory scene analysis* (ASA) (Bregman, 1990). $F0$ is widely used in computational ASA systems that operate upon monaural input—however, systems that employ only this cue are limited to voiced speech (Brown and Cooke, 1994; Wang and Brown, 1999). On the other hand, localization (binaural) cues have the advantage of being generally independent of the signal content and can be used to track a sequence of voiced and unvoiced components that originates from the same location in space.

It is widely acknowledged that for human audition, interaural time differences (ITD) are the main localization cue used at low frequencies ($<$1.5 kHz), whereas in the high-frequency range both interaural intensity differences (IID) and interaural time differences between the envelopes of the signals (IED) are used (Blauert, 1997). The resolution of the binaural cues has implications for both localization and recognition tasks. Headphone experiments show that listeners can reliably detect 10–15 μs ITDs from the median plane, which correspond to a difference in azimuth of between 1 and 5 deg. On the other hand, the smallest detectable change in IID by the human auditory system is about 0.5 to 1 dB at all frequencies. Resolution deteriorates as the reference ITD gets larger, and the difference limen can be as much as 10 deg when the ITD corresponds to a source located far to the side of the head (Blauert, 1997).

Classical models for processing binaural cues compare the acoustic signals at the two ears, although they explain the binaural interaction through different mechanisms. These include extensions of the Jeffress coincidence model (Jeffress, 1948; Lindemann, 1986; Gaik, 1993), the equalization and cancellation (EC) theory (Durlach, 1972; Breebaart *et al.*, 2001), and auditory-nerve-based models (Colburn, 1977; Stern and Colburn, 1978). The goal of this line of research is to explain experimental data for a number of psychoacoustical phenomena including lateralization, binaural masking levels, and the precedence effect (for a review see Stern and Trahiotis, 1995).

Increased speech intelligibility in binaural listening compared to the monaural case has also prompted research in designing cocktail-party processors based on psychoacoustic principles (Lyon, 1983; Slatky, 1993; Bodden, 1993; Liu *et al.*, 2001; Whittkop and Hohmann, 2003). Most cocktail-party-processor designs utilize the following observation: as

[a)]Electronic mail: niki@cis.ohio-state.edu
[b)]Electronic mail: dwang@cis.ohio-state.edu
[c)]Electronic mail: g.brown@dcs.shef.ac.uk

 0001-4966/2003/114(4)/2236/17/$19.00

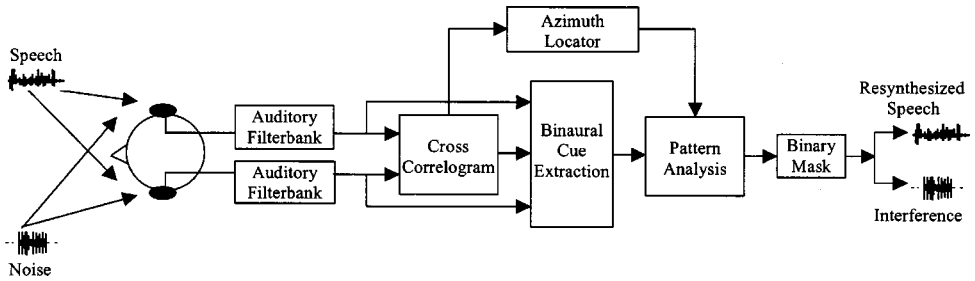


FIG. 1. Schematic diagram of the model. Binaural signals are obtained by convolving input signals with measured head related impulse responses (HRIR) from a KEMAR dummy head. A model of the auditory periphery is employed. Azimuth localization for all the sources is based on a cross-correlation mechanism. ITD and IID are computed independently for different frequency channels. A pattern analysis block produces an estimation of an ideal binary mask, which enables the reconstruction of the target signal and the interfering sound.

the relative strength of the interference with respect to the target increases, certain attributes of the auditory event including location and spatial extent change systematically compared to the case of the target source alone. In particular, building on a previous cross-correlation model for sound localization, Bodden (1993) proposed a model that estimates optimal time-varying Wiener coefficients for all critical bands by comparing desired cross-correlation patterns to observed ones. Bodden's model has shown that psychoacoustically motivated auditory mechanisms can produce substantial enhancement in speech intelligibility (Bodden, 1996).

In this study, we propose a sound segregation model using binaural cues extracted from the responses of a KEMAR dummy head that realistically simulates the filtering process of the head, torso, and external ear (Burkhard and Sachs, 1975). Such a model can be applied to, among other things, enhancing speech recognition in noisy environments and improving binaural hearing aid design. A typical approach for signal reconstruction uses a time–frequency (T–F) mask: T–F units are weighted selectively in order to enhance the target signal. Here, we introduce the notion of an ideal binary mask, which is defined *a priori* as follows. An element in the mask is 1 if the corresponding T–F unit contains target energy that is stronger than interference energy, and 0 otherwise. When target and intrusion are available before mixing, as is the case during our evaluation, ideal binary masks can be readily constructed. We call such a mask *ideal* because its construction requires the knowledge of individual sound sources before mixing. In addition, from a theoretical ASA perspective, an ideal binary mask gives a performance ceiling for all binary masks. Note that an ideal mask remains well-defined for situations when more than one target needs to be segregated. The ideal mask notion is motivated by the human auditory masking phenomenon, in which a stronger signal masks a weaker one in the same critical band (Moore, 1997). Moreover, such masks generate high-quality reconstruction for a variety of signals, and have been recently shown to provide a highly effective front end for robust speech recognition (Cooke *et al.*, 2001). Furthermore, as will be shown later, deviations from ideal binary masks lead to gradual degradation in speech recognition performance. Hence, our model aims to estimate an ideal binary mask using information about the spatial configuration of sound sources.

Statistics for the relationship between the relative strength of sources and the pattern of binaural cues are at the core of our system. We show for mixtures of multiple sound sources that there exists a strong correlation between the relative strength of target and interference and estimated ITD/IID, resulting in a characteristic clustering across frequency bands. Our aim is to maximize the performance of the system independently for different spatial configurations. Consequently, we employ a nonparametric classification method to determine decision regions in the ITD–IID feature space that correspond to an optimal estimate for an ideal mask. An objective evaluation of the system with both SNR (signal-to-noise ratio) and ASR (automatic speech recognition) measures shows that the results of our system are comparable with those obtained using ideal binary masks. In addition, a speech intelligibility evaluation using normal listeners shows a large improvement under certain conditions.

The rest of the paper is organized as follows: the next section contains an overview of the model. Section III describes the peripheral auditory model. Section IV describes the azimuth localization algorithm. Section V is mainly devoted to the ideal binary mask estimation, which constitutes the core of the model. Section VI presents the evaluation results of the system and a quantitative comparison with the Bodden model. In the last section we give further discussions and future directions.

II. MODEL ARCHITECTURE

Our model consists of the following four stages: (1) a model of the auditory periphery; (2) binaural cue extraction and azimuth localization for both target and interference based on a cross-correlation mechanism; (3) estimation of an ideal binary mask; and (4) reconstruction of the target signal. Figure 1 illustrates the model architecture for the case of two sound sources.

The input to our model is a mixture of two or more signals at different, but fixed, locations: target speech and acoustic interference. Measurements of head-related transfer functions (HRTF) are a standard method for realistic binaural synthesis. We utilize here a catalog of HRTF measurements collected by Gardner and Martin (1994) from a KEMAR dummy head under anechoic conditions. The measurements consist of left/right KEMAR responses from a distance of 1.4 m in the horizontal plane, resulting in 128 point impulse

responses at a sampling rate of 44.1 kHz. Binaural signals are obtained by filtering monaural signals with HRTFs corresponding to the direction of incidence. The responses to multiple sources are added at each ear. HRTFs introduce a natural combination of ITD and IID into the signals that is extracted by subsequent stages of our model.

The auditory periphery is simulated using a filterbank that models the cochlear filtering mechanism. In addition, the gains of the filters are adjusted to account for middle-ear transfer, which is direction independent. The output of each filter is processed using a simple model for hair-cell transduction, which performs half-wave rectification and square-root compression. The output of the model gives a firing rate representation of auditory-nerve activity.

Simulated auditory-nerve responses from both ears are evaluated independently for all frequency bands in order to extract interaural differences. The most common method to determine ITD is cross correlation of the corresponding left and right signals within individual frequency bands, which is calculated for time lags equally distributed in the plausible range. Our localization stage uses only ITD information. Consequently, the system cannot tell front from back. We restrict our model to the half-horizontal plane with azimuth in the range $[-90°,90°]$. Due to some diffraction effects, a frequency-dependent nonlinear transformation from the time-lag axis to the azimuth axis is necessary. The set of cross correlations for all frequency bands and at all times results in a 3D structure called the ''cross-correlogram,'' where the coordinates are given by frequency, azimuth, time. A cross-correlogram is further evaluated to extract spatial information. Assuming fixed sources, the source locations are obtained as the positions of the maxima in a pooled cross-correlogram (Shackelton *et al.*, 1992)—obtained by integrating the cross-correlogram across time and frequency. Further stages of our model use this spatial information: the number of sources, their locations, and the target source location.

At the core of our system are decision rules that determine whether the target source is stronger than the interference in individual T–F units. The system is based on observed characteristic clustering of extracted ITD and IID features. The novelty of our approach lies in the introduction of supervised learning for different spatial configurations and across all frequency bands in a joint ITD–IID feature space. For a given frequency channel and a stimulus configuration, conditional probabilities are estimated from samples of ITD, IID, and the corresponding relative strength based on a corpus of training data. Therefore, auditory grouping is implemented based on proximity in the ITD–IID space. The output of this pattern analysis is a time–frequency mask, which is an estimate of an ideal binary mask. The time–frequency resolution for the current implementation is 20-ms time frames with a 10-ms frame shift (see, e.g., Wang and Brown, 1999), and 128 frequency channels that cover the range of 80 Hz to 5 kHz.

The last stage of the model is a reconstruction path, which allows the target signal to be recovered from the acoustic mixture by nullifying the T–F units dominated by interference. The method employed here is the same in principle as that described by Weintraub (1986) (see also Brown and Cooke, 1994). The target signal is reconstructed from the output of the gammatone filterbank. To remove across-channel phase differences, the output of a filter is time reversed, passed through the gammatone filter, and time reversed again. Furthermore, the output for each filter is divided in 20-ms sections with 10-ms overlap that correspond to T–F units in the binary mask, and windowed with a raised cosine. Binary weights estimated in the previous stage are then applied to each section to remove the interference. This method achieves high-quality reconstruction (Weintraub, 1986; Brown and Cooke, 1994; Wang and Brown, 1999).

III. AUDITORY PERIPHERY

It is widely acknowledged that cochlear filtering can be modeled by a bank of bandpass filters. The filterbank employed here consists of 128 fourth-order gammatone filters (Patterson *et al.*, 1988) following an implementation by Cooke (1993). The impulse response of the *i*th filter has the following form:

$$g_i(t) = \begin{cases} t^3 \exp(-2\pi b_i t)\cos(2\pi f_i t + \phi_i), & \text{if } t \geqslant 0, \\ 0, \quad \text{otherwise} \end{cases} \tag{1}$$

where b_i is the decay rate of the impulse response, related to the bandwidth of the filter, f_i is the center frequency of the filter, and ϕ_i is the phase (here, we set ϕ_i to zero).

The equivalent rectangular bandwidth (ERB) scale is a psychoacoustic measure of auditory filter bandwidth. The center frequencies f_i are equally distributed on the ERB scale between 80 Hz and 5 kHz, and specifically for each filter we set the bandwidth according to the following equations (Glasberg and Moore, 1990):

$$\mathrm{ERB}(f_i) = 24.7(4.37 f_i/1000 + 1), \tag{2}$$

$$b_i = 1.019\,\mathrm{ERB}(f_i). \tag{3}$$

Since the HRTF reflects the filtering effects due to pinna and meatus but not the middle ear, we adjust the gains of the gammatone filters in order to simulate the middle-ear transfer function; such data are provided by Moore *et al.* (1997). We include this middle-ear processing for the purpose of physiological plausibility. In the final step of the peripheral model, the output of each gammatone filter is half-wave rectified in order to simulate firing rates of the auditory nerve. Saturation effects are modeled by taking the square root of the rectified signal.

Psychophysical models for sound localization generally employ envelopes of the responses in the high-frequency range. This is supported by discrimination experiments using transposed stimuli, suggesting similar sensitivity to ITD for both low- and high-frequency ranges (Bernstein and Trahiotis, 2001). Therefore, we additionally extract the envelopes using the Hilbert transform for channels with center frequencies above 1.5 kHz. Note that the envelope is not actually used in our current implementation; rather, it is used in Sec. V as part of a comparison of the effectiveness of different interaural cues.

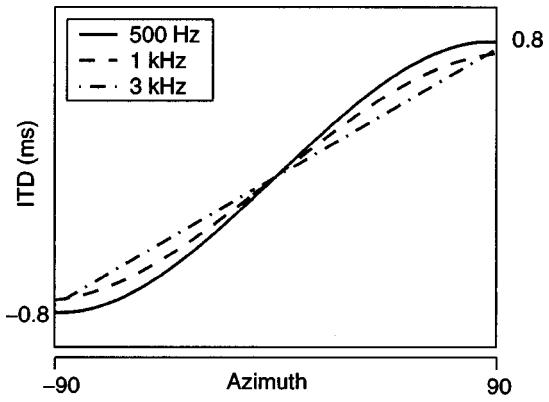


FIG. 2. Functions relating azimuth to ITD for three auditory channels with center frequencies of 500 Hz, 1 kHz, and 3 kHz.

IV. AZIMUTH LOCALIZATION

Current models of azimuth localization almost invariably employ cross correlation, which is functionally equivalent to the coincidence detection mechanism proposed by Jeffress (1948). Cross correlation provides excellent time delay estimation for broadband stimuli, and for narrow band stimuli in the low-frequency range. However, for high-frequency narrow band signals it produces multiple ambiguous peaks. Here, we use the normalized cross correlation computed at lags equally distributed from -1 to 1 ms ($-44 < \tau < 44$) using a rectangular integration window of 20 ms (corresponding to $K = 880$ samples). The cross correlation is computed for all frequency channels and updated every 10 ms, according to the following formula for frequency channel i, time frame j, and lag τ:

$$C(i,j,\tau) = \frac{\sum_{k=0}^{K-1}(l_i(j-k)-\overline{l_i})(r_i(j-k-\tau)-\overline{r_i})}{\sqrt{\sum_{k=0}^{K-1}(l_i(j-k)-\overline{l_i})^2}\sqrt{\sum_{k=0}^{K-1}(r_i(j-k-\tau)-\overline{r_i})^2}}, \quad (4)$$

where l_i, r_i refer to the left and right auditory periphery output of the ith channel, and $\overline{l_i}$, $\overline{r_i}$ refer to their mean values estimated over the integration window.

For each frequency channel, ITD is estimated as the lag corresponding to the position of the maximum in the cross-correlation function. Diffraction effects introduce weak frequency dependences for ITD (MacPherson, 1991). As a result, we derive frequency-dependent nonlinear transformations to map the time-delay axis onto the azimuth axis, resulting in a cross-correlogram $C(i,j,\varphi)$, where φ denotes azimuth. The mappings are obtained based on the cross-correlation output in response to white noise presented systematically at locations in the azimuth range $[-90°,90°]$. Figure 2 shows three ITD-azimuth mappings, for channels with center frequencies of 500 Hz, 1 kHz, and 3 kHz. The functions are monotonic, being sigmoidal at low frequencies where diffraction effects are greater and increasingly linear at high frequencies.

In addition, a "skeleton" $S(i,j,\varphi)$ is formed by replacing the peaks in the cross-correlogram with Gaussians whose widths are narrower than the original peaks. That is, each local peak generates an impulse of the same height and then the obtained impulse train is convolved with a Gaussian. Here, the width is linear with respect to the center frequency of the channel. This technique sharpens the cross-correlogram, an effect similar to a lateral inhibition mechanism (Arbib, 2003).

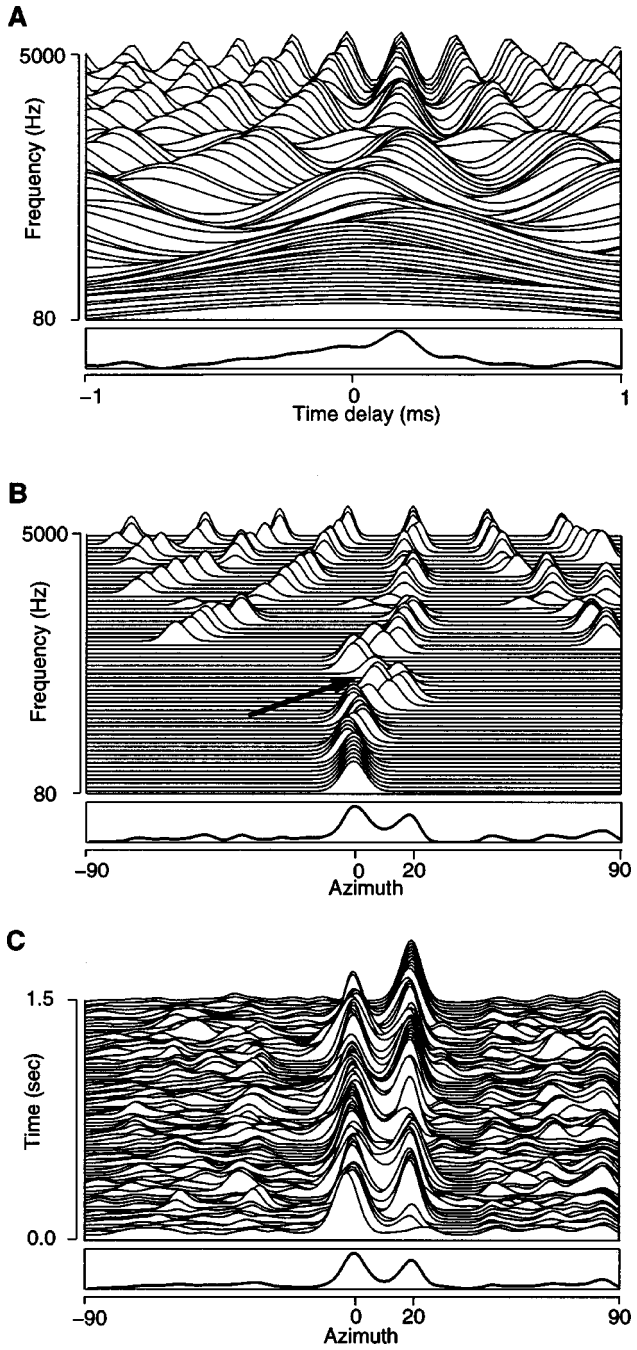


FIG. 3. Azimuth localization for a mixture of male utterance at 0° and female utterance at 20°. The bottom plot in each panel shows a summation across all rows. (A) Cross-correlation functions for 128 frequency channels in the range 80 Hz–5 kHz at time frame 40 (i.e., 400 ms after the start of the stimulus). For clarity, only every other channel is shown, resulting in 64 channels. (B) Skeleton cross-correlogram for the same time frame. The arrow indicates channels that contain roughly equal energy from both target and interference. (C) Pooled cross-correlogram for a stimulus of duration 1.5 s, shown every 20 ms.

The cross-correlation method provides inconsistent results when two acoustic sources are present. Figure 3 shows

the cross-correlation functions [Fig. 3(A)] and the skeleton cross-correlogram [Fig. 3(B)] for a mixture of male speech presented at 0° and female speech presented at 20°. Here, the width of the Gaussians in the skeleton cross-correlogram ranges from 4° at the low-frequency end to 2° at the high-frequency end. For frequency channels where one source is much stronger, activity is observed near the true location of that source. For T–F units where the two sources overlap the peak deviates, generally being closer to the more intense source. Peaks at both locations can occur in high-frequency channels—this ambiguity is due to the periodicity of the cross-correlation function. Hence, if little overlapping occurs for a sufficient number of channels a good estimate of the two source locations can be obtained at every time frame by pooling the cross-correlogram across all frequency channels. At time frame j and azimuth φ, this yields the following pooled cross-correlogram:

$$p(j,\varphi)=\sum_i S(i,j,\varphi). \tag{5}$$

Improved localization results are obtained using the skeleton cross-correlogram proposed here over the standard cross correlation. Summing across frequencies produces sharper peaks on the skeleton cross-correlogram; in the case of Fig. 3, the skeleton cross-correlogram gives a good estimate of source locations, whereas the conventional cross-correlogram does not [compare the bottom plots in Fig. 3(A) and Fig. 3(B)]. In Fig. 3(C) we display the pooled cross-correlogram for a signal of duration 150 frames (i.e., 1.5 s). Peaks in the pooled cross-correlogram indicate the locations of active sources at every frame. Assuming fixed sources, multiple locations can be reliably determined by further summating the pooled cross-correlogram across time as shown in the bottom plot of Fig. 3(C). This represents our method for azimuth localization.

V. IDEAL MASK ESTIMATION

The objective of this stage of the model is to develop an efficient mechanism for estimating an ideal binary mask, which selects the T–F units where the estimated signal energy is greater than the noise energy (i.e., greater than 0-dB SNR). Note that different SNR criteria are possible for defining an ideal binary mask (see Cooke *et al.*, 2001). In the absence of evidence for a better SNR measure, we choose the 0-dB criterion for simplicity. We propose an estimation method based on the following observation regarding the auditory interaction of multiple sources. In a narrow band, the ITD and IID corresponding to the target source exhibit azimuth-dependent characteristic values. As the interference from additional sound sources increases, ITD and IID systematically shift away from these values. Consequently, in a local T–F unit both binaural cues can be potentially used to determine whether the target signal dominates.

In what follows, we analyze this phenomenon for the case of pure tones (see Slatky, 1993, for an extensive study of binaural cues with sinusoidal signals). Although in real-world scenarios the conditions of this simplified model are generally not fulfilled, our experimental results show that a similar trend holds for a variety of natural signals when analyzed in narrow frequency bands. This analysis also serves to motivate the introduction of our proposed algorithm for the general case in subsection B.

A. Pure tones

We consider a simple model of two sources emitting pure tones in a narrow band. In this case, the left-ear and the right-ear responses are given by

$$\begin{aligned} l(t)&=|H_1^l(\omega_1)|A_1\sin(\omega_1 t)\\ &\quad+|H_2^l(\omega_2)|A_2\sin(\omega_2 t+\Delta\varphi),\\ r(t)&=|H_1^r(\omega_1)|A_1\sin(\omega_1 t+\omega_1 d_1)\\ &\quad+|H_2^r(\omega_2)|A_2\sin(\omega_2 t+\omega_2 d_2+\Delta\varphi), \end{aligned} \tag{6}$$

where A_i is the amplitude, ω_i is the frequency, d_i corresponds to the interaural time delay (equivalent to the phase difference between left and right HRTFs at frequency ω_i), and $H_i^r(\omega_i)$ and $H_i^l(\omega_i)$ represent, respectively, the right and left HRTF, for the ith source. $\Delta\varphi$ is the sum of phase differences between the initial signals and those due to the arrival times of the signals at the left ear.

To simplify, we neglect the magnitude of the HRTF response in analyzing ITD, which represents a reasonable assumption only in a narrow band low-frequency range. The cross-correlation function for infinite-duration signals is obtained by

$$c(\tau)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} r(t)l(t+\tau)dt. \tag{7}$$

Observe that in approximating the cross-correlation function in a finite duration, there exists a trade-off between the difference in frequency $|\omega_1-\omega_2|$ and the total integration time. Therefore, we study the cross correlation under the following two conditions.

1. Case 1: $\omega_1=\omega_2=\omega$

In this case, we have

$$\begin{aligned} c(\tau)&=\frac{A_1^2}{2}\cos(\omega(\tau-d_1))+\frac{A_2^2}{2}\cos(\omega(\tau-d_2))\\ &\quad+A_1A_2\cos\left(\omega\left(\tau-\frac{d_1+d_2}{2}\right)\right)\\ &\quad\cdot\cos\left(\Delta\varphi+\omega\frac{d_2-d_1}{2}\right). \end{aligned} \tag{8}$$

Due to the periodicity of $c(\tau)$, we study the cross-correlation function on a 2π interval centered at $\omega(d_1+d_2)/2$. Without loss of generality, assume that the phase differences ωd_1, ωd_2 are in this interval; otherwise, simply shift the phases with multiples of 2π. To fix the discussion let $d_1<d_2$. By observing the deviation of the peak location $\tau_{\max}$ from the middle of the two sources, $(d_1+d_2)/2$, we obtain the stronger source

$$\tau_{\max}>(d_1+d_2)/2\Leftrightarrow A_1<A_2. \tag{9}$$

This result gives a threshold to decide which source is stronger based on ITD. Furthermore, we want to study how

ITD changes with the relative strength $R=A_2/(A_1+A_2) \in [0,1]$. Hence, we derive the solution for $\tau_{\max}$ as follows:

$$\tau_{\max}=\frac{d_1+d_2}{2} + \frac{1}{\omega}\left(\arctan\left[\frac{(A_2^2-A_1^2)\sin\beta}{(A_1^2+A_2^2)\cos\beta+2A_1A_2\cos(\Delta\varphi+\beta)} \right] + k\pi \right), \tag{10}$$

where $\beta=\omega[(d_2-d_1)/2] \in [0,\pi]$ and k is an integer. The relation obtained in (9) uniquely determines $k \in \{0,\pm 1\}$ for the 2π interval considered. More specifically, $\beta \leqslant \pi/2 \Rightarrow k=0$ and $\beta > \pi/2 \Rightarrow k=1$ when $A_1 < A_2$, and $k=-1$ when $A_1 > A_2$. Furthermore, simulations and derivations show that a good approximation for the mean value $\overline{\tau}_{\max}$ when $\Delta\varphi$ varies uniformly in the range $[-\pi,\pi]$ is given by

$$\overline{\tau}_{\max} \approx \begin{cases} d_1, & R<0.5 \\ \dfrac{d_1+d_2}{2}, & R=0.5 \\ d_2, & R>0.5. \end{cases} \tag{11}$$

2. Case 2: $\omega_1 \neq \omega_2$

In this case, due to the orthogonality of sine waves of different frequencies the cross-correlation function becomes

$$c(\tau)=\frac{A_1^2}{2}\cos(\omega_1(\tau-d_1))+\frac{A_2^2}{2}\cos(\omega_2(\tau-d_2)). \tag{12}$$

A closed-form solution for the peak location in this case does not exist. Instead, we analyze the behavior of the peak location for relatively close angles, i.e., $|\omega_1 d_1-\omega_2 d_2| < \pi/2$. In this interval, we apply a second-order Taylor expansion as an approximation for the cosine, resulting in a simple solution: $\tau_{\max}=(A_1^2\omega_1^2 d_1+A_2^2\omega_2^2 d_2)/(A_1^2\omega_1^2+A_2^2\omega_2^2)$. Note that this is a monotonic function with respect to the relative strength R.

For the general case, we observe that as the frequencies ω_1 and ω_2 vary uniformly in a narrow band centered at ω, a good approximation for the mean of $\tau_{\max}$ is given by

$$\overline{\tau}_{\max}=\frac{d_1+d_2}{2}+\frac{1}{\omega}\left(\arctan\left[\frac{(A_2^2-A_1^2)}{(A_1^2+A_2^2)}\tan\beta \right] + k\pi \right), \quad k \in \{0,\pm 1\}, \tag{13}$$

which is the solution for the maximum position in (12) when $\omega_1=\omega_2$. This function is monotonically increasing with respect to R when $\beta < \pi/2$ and decreasing when $\beta > \pi/2$. Figure 4 shows the results when $\omega=500$ Hz and β equals $\pi/4$ and $3\pi/4$, respectively.

A systematic change in R also results in a corresponding shift in IID. A similar discussion applies here. That is, the frequency difference between the two tones affects the spread of IID distribution. We do not study the case $\omega_1=\omega_2$, since the results for IID distribution are complex and not amenable to the analysis used here. In addition, IID is most reliable at high frequencies where filter bandwidths are large. Therefore, we consider the case $\omega_1 \neq \omega_2$. IID is approximated as the ratio of signal power at the two ears, resulting in the following expression:

$$\mathrm{IID}=10\log_{10}\frac{A_1^2|H_1^r(\omega_1)|^2+A_2^2|H_2^r(\omega_2)|^2}{A_1^2|H_1^l(\omega_1)|^2+A_2^2|H_2^l(\omega_2)|^2}, \tag{14}$$

where the power of a signal $u(t)$ is $\lim_{T\to\infty} 1/2T \int_{-T}^{T} u^2(t)dt$. Note that IID is monotonic with respect to the relative strength R.

The above analysis suggests that the distribution of the binaural cues in a given filter channel is directly influenced by the filter bandwidth. To test this, we simulate left and right signals using Eq. (6), where the relative strength is fixed, $\Delta\varphi$ is uniformly distributed in the range $[-\pi,\pi]$, and $\omega_{1,2}$ in $[\omega-\Delta\omega,\omega+\Delta\omega]$. Figures 5(A) and (B) show the mean and the variance of ITD as a function of R for the condition of $\omega=500$ Hz, 30° azimuth separation, 20-ms integration time, and four $\Delta\omega$ values in the range of 0 to 200 Hz. In the figure, M_1 is the ITD mean as derived in (11) and it approximates well the case $\Delta\omega=0$. M_2 is the ITD mean derived in (13) for the more general case $\Delta\omega \neq 0$. Similarly, Figs. 5(C) and (D) show results for IID when $\omega=2.5$ kHz and five $\Delta\omega$ values in the range of 0 to 400 Hz. Here, M is the IID mean as derived in (14). It is worth noting that the theoretical derivations of M_2 and M approximate well the simulation results when the bandwidth approaches the auditory filter ERB, which is 80 Hz for a 500-Hz center frequency and 300 Hz for 2.5 kHz. In addition, there is a systematic decrease in variance for both ITD and IID as $\Delta\omega$ approaches the ERB. This behavior generalizes to other frequencies as well.

To conclude, our analysis shows that ITD and IID undergo systematic shifts from the ideal target values as the relative strength R of two sinusoidal sources is changed. A

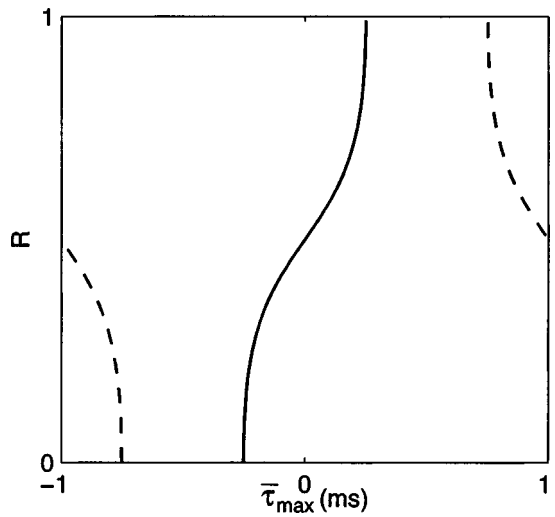


FIG. 4. Theoretical approximation for the mean ITD, $\overline{\tau}_{\max}$, for two pure tones randomly distributed in a narrow band centered at 500 Hz. The *y* axis corresponds to the relative strength *R*. Two cases are shown: $\beta=\pi/4$ (solid line) and $\beta=3\pi/4$ (dashed line).

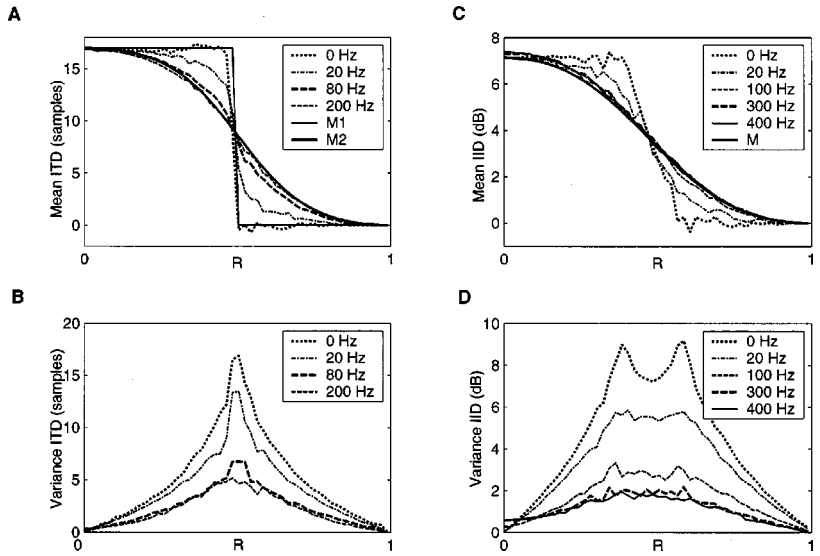


FIG. 5. The influence of filter bandwidth on the mean and variance of ITD and IID with respect to the relative strength R. The data are from simulations of two pure tones uniformly distributed in a narrow band. One tone is at 0° and the another is at 30°. The sampling frequency is 44.1 kHz. (A) Mean ITD as a function of R for 500-Hz center frequency and four bandwidths between 0 and 200 Hz. The auditory filter ERB here is 80 Hz. M_1 and M_2 correspond to the theoretical mean ITD as derived in Eq. (13) and Eq. (14), respectively. (B) ITD variance for the same condition as in (A). (C) Mean IID as a function of R for a 2.5-kHz center frequency and five bandwidths between 0 and 400 Hz. M corresponds to the theoretical mean IID as derived in Eq. (15). The auditory filter ERB is 300 Hz. (D) IID variance for the same condition as in (C).

comparison of the above theoretical derivations with the real data presented in the next subsection shows that the match is very close.

B. Model

The analysis of ITD and IID for pure tones shows relatively smooth changes with the relative strength R in narrow frequency bands. In order to capture this relationship in the context of real signals, statistics are collected for individual spatial configurations during training. Binaural signals are obtained by convolving with KEMAR HRTFs as explained in Sec. II. We employ a training corpus consisting of ten speech signals from the TIMIT database (Garofolo *et al.*, 1993): five male utterances and five female utterances as presented in Table I. The speaker ID in the table uniquely identifies the speaker in the TIMIT database where the first letter indicates the sex of the speaker. In the two-source case, we select S0–S4 to be the target and the rest interference. In the three-source case, we have S0–S3 as target signals and the two interfering sets are S4–S6 and S7–S9.

Estimates for ITD, IID, and R are extracted independently for all frequency channels. Since the cross-correlation function is periodic, resulting in multiple peaks for mid to high frequencies, we consider the following strategy for estimating ITD. We study deviations from the target ITD for individual frequency channels, which is obtained from the ITD-azimuth mappings presented in Sec. IV. Consequently, we compute ITD_i as the peak location of the cross-correlation function in the range $2\pi/\omega_i$ centered at the target ITD, where ω_i indicates the center frequency of the ith channel. IID_i corresponds to the mean power ratio at the two ears, expressed in decibels

$$\mathrm{IID}_i = 20\log_{10}\left(\sum_t r_i^2(t) \Big/ \sum_t l_i^2(t)\right), \tag{15}$$

where l_i and r_i refer to the left and right auditory periphery output of the ith channel, respectively. Note that in computing IID_i, we use 20 instead of 10 in order to compensate for the square-root operation in the peripheral processing stage.

The relative amplitude is a measure of the relative strength between the target source and the acoustic interference, defined using root-mean-square values of the original signals at the "better ear"—the ear with higher SNR (see, e.g., Shinn-Cunningham *et al.*, 2001)

$$R_i = \sqrt{\sum_t s_i^2(t)} \Big/ \left(\sqrt{\sum_t s_i^2(t)} + \sqrt{\sum_t n_i^2(t)}\right), \tag{16}$$

where s_i refers to the response of the ith gammatone filter to the target signal and n_i the response to the acoustic interference (noise).

Figure 6 shows empirical results obtained for a two-source configuration: target source in the median plane and interference at 30°. The scatter plot in Fig. 6(A) shows samples of ITD_i and R_i obtained for the channel with a center frequency of 500 Hz (about 7000 samples in total). In

TABLE I. Speech signals of the training set.

ID	Speaker ID	Utterance
S0	MKLS0	"Primitive tribes have an upbeat attitude"
S1	FCKE0	"Only the best players enjoy popularity"
S2	MCDC0	"Our aim must be to learn as much as we teach"
S3	FEAR0	"Development requires a long-term approach"
S4	FDMS0	"Poets, moreover, dwell on human passions"
S5	FETB0	"Change involves the displacement of form"
S6	FCMM0	"The system works as an impersonal mechanism"
S7	MJWS0	"Most assuredly ideas are invaluable"
S8	MRVG0	"False ideas surfeit another sector of our life"
S9	MJRH0	"But in every period it has been humanism"

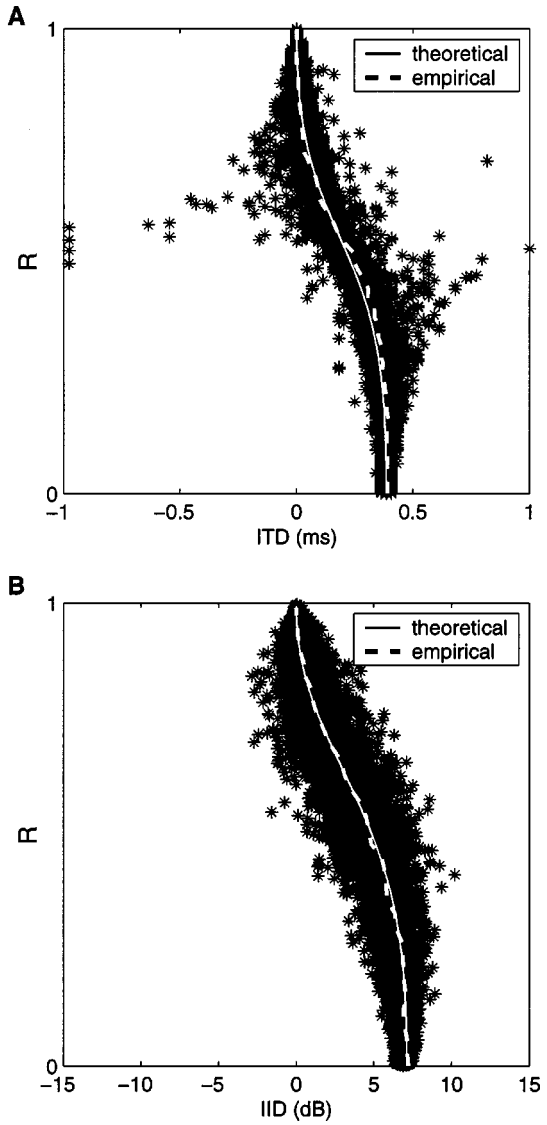


FIG. 6. Relationship between ITD/IID and the relative strength R for a two-source configuration: target in the median plane and interference on the right side at 30°. (A) The scatter plot shows ITD and R estimates from the training corpus for a channel with center frequency of 500 Hz. The solid curve shows the theoretical mean [see Eq. (14)] and the dash curve shows the data mean. (B) Results for IID for a filter channel with center frequency 2.5 kHz. The solid curve shows the theoretical mean [see Eq. (15)] and the dash curve shows the data mean.

addition, we display the empirical mean of the samples and the theoretical one derived in (13). Similarly, Fig. 6(B) shows the results that describe the variation of IID_i with R_i for a channel with a center frequency of 2.5 kHz and compares the empirical mean with the one derived in (14). Note that R_i incorporates the HRTF responses at the better ear. Therefore, the R axis for the theoretical mean is converted accordingly. Figure 6 exhibits a systematic shift of the estimated ITD and IID with respect to R for real signals. Moreover, the theoretical means obtained in the case of pure tones match the empirical ones very well. Similar matches are observed in other frequency channels and other spatial configurations.

The above observation extends to multiple-distracter scenarios. As an example, Fig. 7 displays smoothed histograms that show the relationship between R_i and both ITD_i [Fig. 7(A)] and IID_i [Fig. 7(B)] for a three-source situation. Samples correspond to a frequency channel with a center frequency close to 1.5 kHz for target at 0° (median plane) and two interferences at −30° and 30°. Note that the interfering sources introduce systematic deviations of the binaural cues. Consider a particularly troubling case: the target is silent and two interferences have equal energy in a given T–F unit. This results in binaural cues indicating an auditory event at half of the distance between the two interference locations; for our setup, it is 0°—the target location. However, the data in Fig. 7 suggest a low probability for this case. Figure 7 instead shows a clustering phenomenon, suggesting that in most cases only one source dominates a T–F unit.

By displaying the information in the joint ITD–IID space, we observe a location-based clustering of the binaural cues, which is clearly marked by strong peaks that correspond to distinct active sources as shown in Fig. 7(C). There exists a trade-off between ITD and IID across frequencies, where ITD is most salient at low frequencies and IID at high frequencies. But, a fixed cutoff frequency that separates the effective use of ITD and IID does not exist for different spatial configurations (see Fig. 8 later). This motivates our choice of a joint ITD–IID feature space that optimizes the system performance across different configurations. Differential training seems necessary for different channels, given

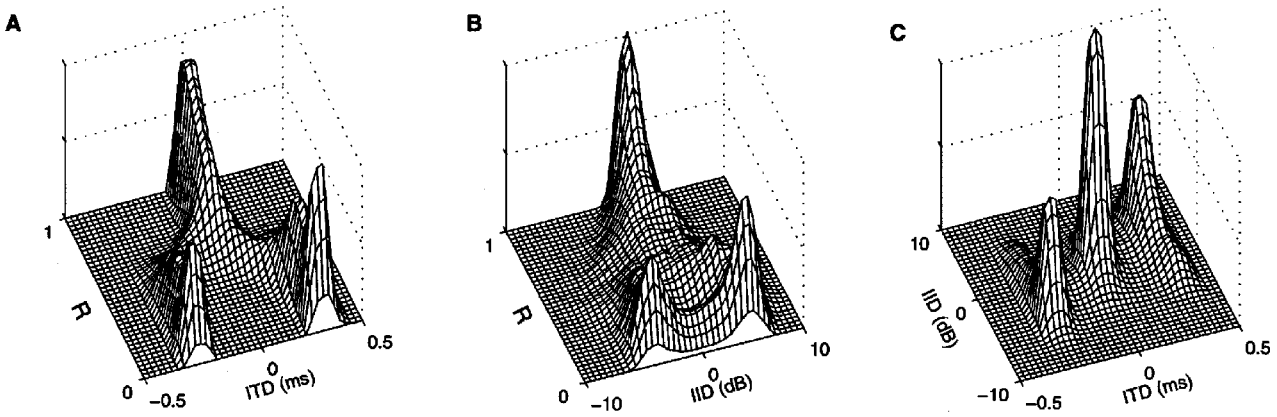


FIG. 7. Relationship between ITD/IID and the relative strength R for a three-source configuration: target source in the median plane and interference at −30° and 30°. Statistics are obtained from the training corpus for a channel with center frequency close to 1.5 kHz. (A) Histogram of ITD and R samples. (B) Histogram of IID and R samples. (C) Clustering in the ITD–IID space.

that there exist variations of ITD and, especially, IID values with different center frequencies.

Since the goal is to estimate an ideal binary mask, we focus on detecting decision regions in the two-dimensional ITD–IID feature space for individual frequency channels. Consequently, standard supervised learning techniques can be applied. For the *i*th channel, we test the following two hypotheses. The first one is H_1: target is dominant or $R_i > 0.5$, and the second one is H_2: interference is dominant or $R_i \leqslant 0.5$. Based on estimates of the bivariate densities $p(x|H_1)$ and $p(x|H_2)$, the classification is done in accordance with the *maximum a posteriori* (MAP) decision rule: $p(H_1)p(x|H_1) > p(H_2)p(x|H_2)$. There exists a plethora of techniques for probability density estimation ranging from parametric techniques (e.g., mixture of Gaussians) to nonparametric ones (e.g., kernel density estimators). We initially tried the EM algorithm for learning Gaussian mixtures (Duda *et al.*, 2001), but this did not prove to be as robust due to the following factors: (i) the true number of mixing components is usually unknown, and (ii) the algorithm tends to be sensitive to parameter initialization. Even for the two-source scenario, the method of computing ITD for mid- to high frequencies can result in two-mode distribution for the H_2 hypothesis. In order to completely characterize the distribution of the data, we use the kernel density estimation method independently for all frequency channels.

Kernel density estimation is well documented in the literature (Silverman, 1986), so we only summarize its essence here. Generally, the multidimensional kernel density estimate for n observations $\mathbf{x}_1,\ldots,\mathbf{x}_n$ of dimensionality d is given by the following formula:

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^{n} \frac{1}{n h_1 \ldots h_d} \prod_{j=1}^{d} K\left(\frac{x_j - x_{ij}}{h_j}\right), \tag{17}$$

where $\mathbf{x} = (x_1,\ldots,x_d)$ is a feature vector, x_{ij} is the jth element of $\mathbf{x}_i$, K is a Gaussian function, and h_j's are parameters called bandwidths that define the amount of smoothing for the empirical distribution. In our case, the ITD–IID feature space has dimensionality $d = 2$. The selection of the smoothing parameters is critical to the success of the estimation process: for too-small values it approximates the data well but generalizes poorly, and for too-large values the structure of the data distribution disappears. One approach for finding optimal values is the least-squares cross-validation method (LSCV) (Silverman, 1986). We employ the LSCV method for high dimensions and the Gaussian kernel given by Sain *et al.* (1994) (p. 808). Optimal smoothing values are chosen as local minima in the range $[n^{-1/6}\sigma_i/4,\ 3n^{-1/6}\sigma_i/2]$, where σ_i represents the variance of the data set in the ith dimension and n is the size of sample data set.

One cue not employed in our model is IED. Auditory models generally use IED in the high-frequency range (see, for example, Bodden, 1993), since the auditory system becomes gradually insensitive to interaural phase differences above 1.5 kHz. In addition, the occurrence of multiple peaks at high frequencies in the cross-correlation function is much reduced for the IED cue. We have compared the individual performance of the three binaural cues: ITD, IID, and IED, for a one-dimensional classification task based on the kernel density estimation method presented above. An error is made whenever the estimated binary mask value for a T–F unit differs from the corresponding ideal value. Figure 8 shows the error rates with respect to frequency channel using the Cooke corpus (see Sec. VI A) as the test set, where we consider two cases: target source in the median plane and the acoustic interference at 5° [Fig. 8(A)] and 30° [Fig. 8(B)]. IED results are given for the frequency range of interest—above 1.5 kHz (i.e., channel number >80). As the source separation increases, error rates for IED and IID improve. On the other hand, ITD loses discriminability for high-frequency channels where the multiple-peak problem results in the same ITD for both target and interference [Fig. 8(B)]. Figure 8 also displays the corresponding error rates for the joint ITD–IID space and the joint IED–IID space, and it shows that the joint ITD–IID space yields the best overall performance across different spatial configurations. As indicated in Fig. 8, we have found no benefit for using IED after incorporating ITD and IID, and hence it is not utilized in our model.

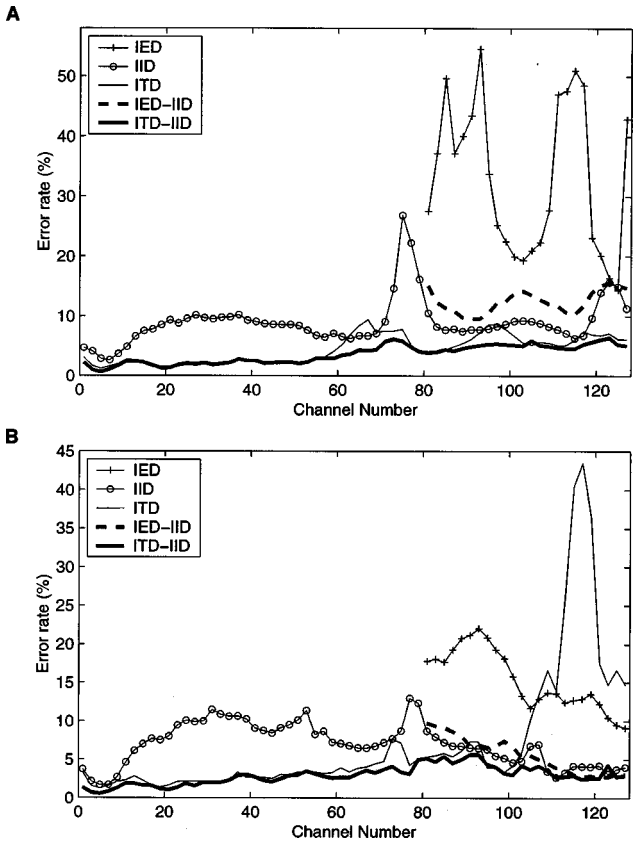


FIG. 8. Discriminability comparison for the three binaural cues, ITD, IID, and IED, the joint IED–IID space, and the joint ITD–IID space. Error rates are displayed as a function of channel number (frequency) for a classification task for two spatial configurations. (A) Target source in the median plane and interference on the right side at 5°. (B) Target source in the median plane and interference on the right side at 30°. IED results are shown for frequencies above 1.5 kHz, i.e., above channel number 80.

VI. EVALUATION AND COMPARISON

A binary mask produced by the model described in the last section approximates very well the corresponding ideal binary mask, which is obtained by comparing the energies of

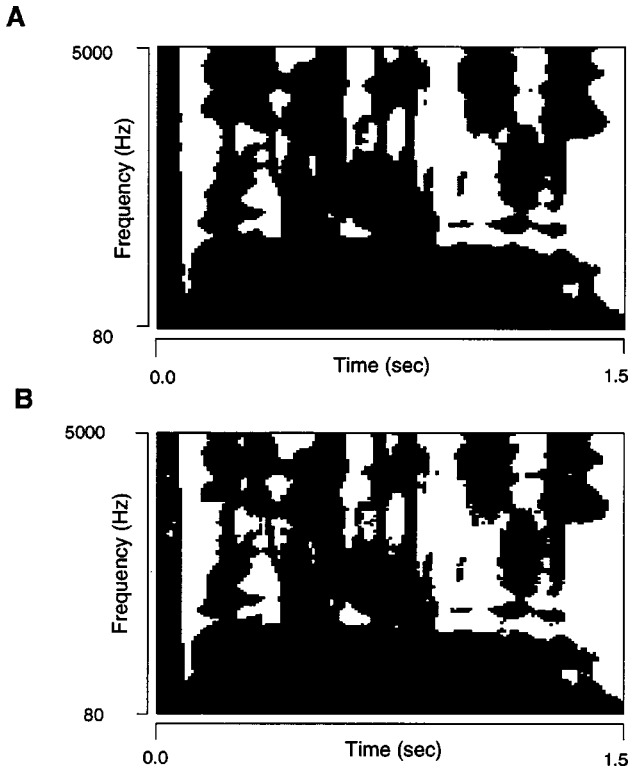


FIG. 9. A comparison between an ideal binary mask (A) and the binary mask resulting from our model (B) for a mixture of male utterance in the median plane (target) and female utterance on the right side at 30° (interference). The black regions indicate those T–F units dominated by target speech.

the original target and interference before mixing. As an example, Fig. 9 shows a comparison between the ideal binary mask and the estimated mask for a mixture of target male speech presented at 0° and interference female speech at 30° at the better ear. In the figure, a blank pixel indicates a T–F unit in which the target dominates. The two masks are very similar, with an SNR difference of only 0.19 dB.

The performance of a segregation system can be assessed in different ways, depending on intended applications. To extensively evaluate our model, we use the following three criteria: (1) an SNR measure using the original target as signal; (2) ASR rates using our model as a front end; and (3) human speech intelligibility tests. Results with each criterion are given below.

A. SNR evaluation

To conduct an SNR evaluation, a segregated signal is reconstructed from a binary mask following the method described in Sec. II. To quantitatively assess system performance, we measure in decibels the SNR using the original target speech before mixing as signal

$$\mathrm{SNR}=10\log_{10}\sum_t s_T^2(t)\Big/\sum_t (s_T(t)-s_E(t))^2, \quad (18)$$

where $s_T(t)$ represents the original target signal reconstructed using an all-one mask and $s_E(t)$ the estimated target reconstructed from the binary mask. With a binary mask, a more conventional SNR measure would use the mask to pass through original target and intrusion in order to obtain signal and noise, as done by Wang and Brown (1999). The problem with such a measure is that loss of target energy is not penalized, and as a result a separate measure of retained target energy needs to be given (Wang and Brown, 1999). Equation (18) provides a single measure, and in the case of an all-one mask yields the original SNR. Though the signal part in (18) is higher than that retained by a binary mask, it is offset by the denominator that is also higher than retained noise energy; the denominator penalizes both retained noise by the binary mask and target distortion. Our measure is more stringent than the conventional SNR measure; indeed, our tests show that (18) gives systematically lower SNR values. To minimize the loss of target energy we take advantage of the higher initial SNR at the better ear. As a result, the reconstructed signal corresponding to the better ear contains more target energy. Therefore, all the following evaluations are performed at the better ear.

The system performance is measured on independent test corpora for different spatial configurations. For the two-source scenario, one test set is the corpus collected by Cooke (1993), chosen because it is commonly used in computational ASA studies (Brown and Cooke, 1994; Wang and Brown, 1999; Wu *et al.*, 2003). The corpus contains ten voiced speech signals and ten noise intrusions, encompassing a variety of common acoustic interferences such as telephone ringing, rock music, and other speech utterances. In addition, we employ a second corpus containing ten normal speech utterances from the TIMIT database (see Table II) as target mixed with the ten intrusions from the Cooke corpus (see Table III). In the case of three sources, we use the Cooke corpus for testing: five speech signals form the target set and the other five form one interference source. The ten intru-

TABLE II. Target signals of the test set.

ID	Speaker ID	Utterance
S0	MWSB0	"Bright sunshine shimmers on the ocean"
S1	MDCD0	"Challenge each general's intelligence"
S2	MDHS0	"The Thinker is a famous sculpture"
S3	MTAA0	"Only lawyers love millionaires"
S4	MRPC1	"Biblical scholars argue history"
S5	FPKT0	"They make us conformists look good"
S6	FJRE0	"Artificial intelligence is for real"
S7	FPAC0	"A good attitude is unbeatable"
S8	FREH0	"Too much curiosity can get you into trouble"
S9	FBCH0	"Clear pronunciation is appreciated"

TABLE III. Noise signals of the test set.

ID	Utterance
N0	1-kHz tone
N1	Random noise
N2	Noise bursts
N3	"Cocktail" party noise
N4	Rock music
N5	Siren
N6	Telephone trill
N7	"Don't ask me to carry an oily rag like..."
N8	"She had your dark suit in greasy wash..."
N9	"Why were we keen to use human..."

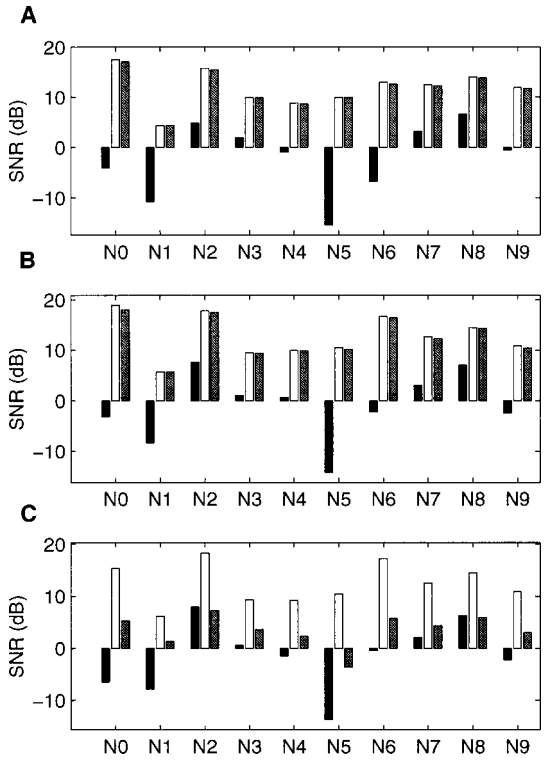


FIG. 10. Systematic results for two-source configuration with 5° azimuth separation. Black bars correspond to the SNR of the initial mixture, white bars indicate the SNR obtained using ideal binary mask, and gray bars show the SNR from our model. Results are obtained for speech mixed with ten types of intrusions (see Table III) for different spatial configurations. (A) Target at 0°, interference at 5°. (B) Target at 40°, interference at 45°. (C) Target at 80°, interference at 85°.

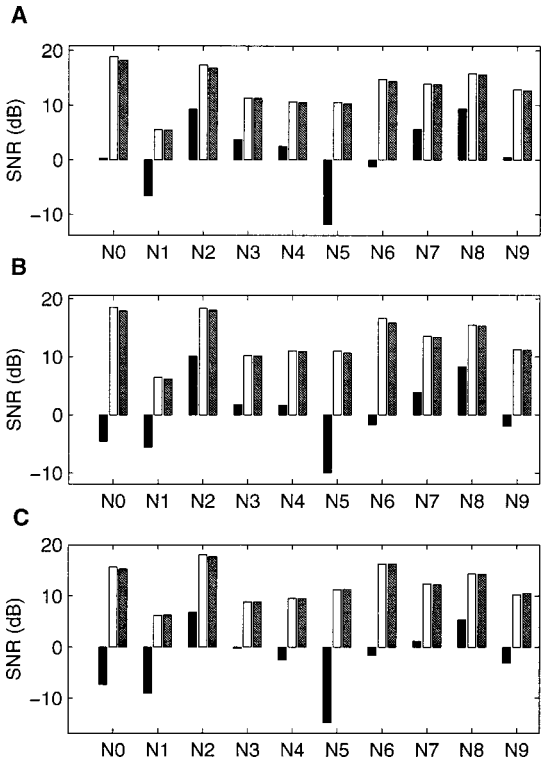


FIG. 11. Systematic results for two-source configuration with 30° azimuth separation. Black bars correspond to SNR of the initial mixture, white bars to the SNR obtained using an ideal binary mask, and gray bars to the SNR from our model. (A) Target at 0°, interference at 30°. (B) Target at 30°, interference at 60°. (C) Target at 60°, interference at 90°.

sions then form the second interference source. Therefore, in this three-source corpus every mixture contains two utterances plus an additional intrusion.

For the two-source case, the model is systematically evaluated at the better ear for various combinations of azimuth angles. We compare the SNR gain obtained by our model against that obtained using an ideal binary mask. For the test corpus of Table II, Fig. 10 shows the results for a spatial separation of 5° and target at azimuth 0°, 40°, and 80°. Results are similar across mixtures in the same noise category; hence, we present the averaged result for each category. Very good results are obtained when the target is close to the median plane for an azimuth separation as small as 5°. Performance degrades when the target source is moved to the side of the head; this is a direct consequence of poorer resolution of the binaural cues at higher azimuth angles. When comparing with the SNR of the initial mixture, there is an average-SNR gain of 13.76 dB for the target in the median plane, and it reduces to 5.04 dB with the target at 80°. When the spatial separation increases, excellent results are obtained across all spatial configurations. Figure 11 shows results for target at 0°, 30°, and 60° and interference at 30° to the right of target. Similar results are obtained for other spatial configurations. Figure 12 shows that the system performs equally well on the Cooke corpus. Figure 12(A) gives the results for a 5° azimuth separation and the average improvement is 13.73 dB. Similarly, Fig. 12(B) gives the results for a 30° separation.

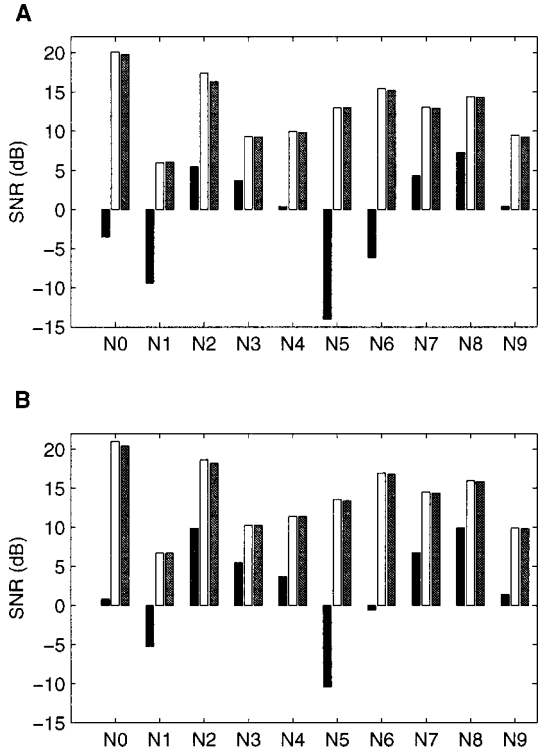


FIG. 12. Systematic results for two-source configuration using the Cooke corpus as the test corpus. Black bars correspond to SNR of the initial mixture, white bars to the SNR obtained using an ideal binary mask, and gray bars to the SNR from our model. (A) Target at 0°, interference at 5°. (B) Target at 0°, interference at 30°.

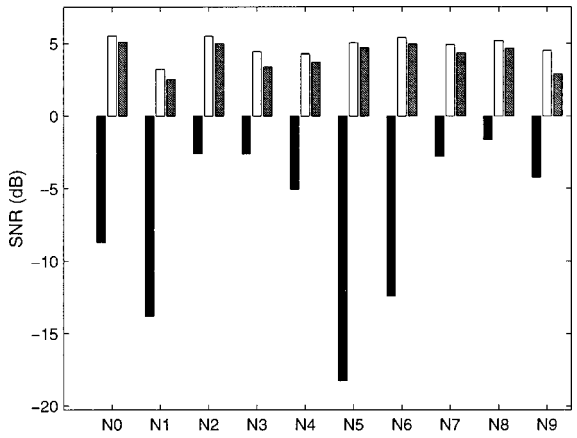


FIG. 13. Evaluation for a three-source configuration. The target is in the median plane and intrusions are at −30° and 30°. Black bars correspond to the SNR of the initial mixture, white bars to the SNR obtained using ideal binary mask, and gray bars to the SNR from our model.

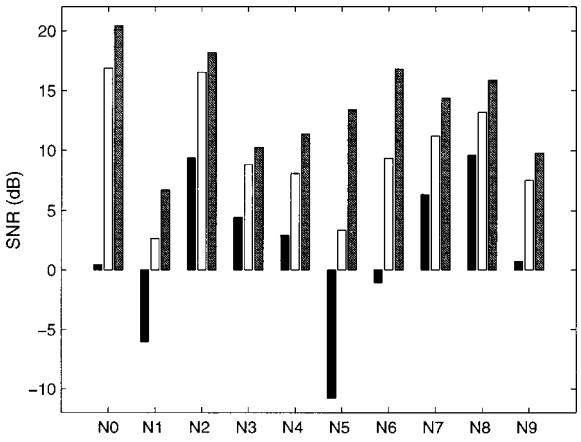


FIG. 14. SNR comparison between the Bodden model (white bars) and our model (gray bars) for a two-source configuration: target in the median plane and interference at 30°. The black bars correspond to the SNR of the original mixture.

Our approach, like other location-based methods using cross correlation, can be extended to cases with more than two sources. With given locations, our model performs target segregation in a similar manner, i.e., estimating an ideal binary mask following the method outlined in Sec. V B. Figure 13 illustrates the performance of the model in a three-source scenario with target located in the median plane and two interfering sources at −30° and 30°. Here, the ten noise intrusions from the Cooke corpus are presented at 30° azimuth and the target is reconstructed based on the right ear mixture. As previously, results are mean values for the ten types of noise intrusion. The performance degrades compared to the corresponding two-source situation, from an average SNR of about 12 to 4.10 dB. Still, the average SNR gain obtained is approximately 11.31 dB.

In order to draw a quantitative comparison with another binaural processing model, we have implemented the Bodden model (Bodden, 1993), which produces good-quality sound separation using source locations. The localization stage of this model uses an extended cross-correlation mechanism based on contralateral inhibition and it adapts to HRTFs. The separation stage of the model is based on estimation of the weights for a Wiener filter. Specifically, for a given T–F unit the weight is given by the ratio between a desired excitation and an actual one. The actual excitation corresponds to the integration of the cross-correlation pattern across the azimuth axis, and the ideal peak shape is used as a window to derive the desired excitation. The Bodden model differs from ours in several aspects. First, his sound localization stage builds on the previous models of Lindemann (1986) and Gaik (1993), which simulate aspects of the precedence effect for reverberant scenarios, whereas our localization stage is simpler and does not address the precedence effect. Second, his model requires only a target azimuth and no training is necessary as spatial configuration changes. Although these aspects add to the flexibility of his model, the estimation of Wiener filter weights appears less robust than our binary estimation of ideal masks. In addition, our configuration- and channel-specific training utilizes more information provided by localization and makes an optimal use of frequency-dependent ITD and IID cues.

Bodden's system uses a 24-channel filterbank intended to simulate critical bands. For a fair comparison, our implementation of the Bodden system uses the same time–frequency resolution employed in our system with a 128-channel gammatone filterbank; we also implemented the Bodden model with 24-channel critical bands and the results are not as good. We find that, when two sources are relatively close, the Bodden model is less robust than ours. Our comparison is based on the Cooke corpus and a spatial configuration of target at 0° and intrusion on the right side at 30°, an azimuth separation in the range where his model performs optimally. As displayed in Fig. 14, our model shows a considerable improvement over the Bodden system, producing 3.5-dB average improvement. The improvement is especially high for a few cases (e.g., N5 and N6) where our estimated masks result in large SNR improvements over the original mixtures.

B. ASR evaluation

As discussed before, an ideal binary mask is defined *a priori*. Similar *a priori* masks have been shown to produce impressive performance when applied to the automatic recognition of noisy speech using a ''missing data'' approach (Cooke *et al.*, 2001). In this approach, a continuous density hidden Markov model recognizer is modified such that only acoustic features indicated as reliable in the mask are used during decoding. Since our ideal binary masks are generated in a similar way to those used by Cooke *et al.*, we would expect them to be an equally effective front end to missing-data ASR.

Our motivation for ASR evaluation is twofold. First, a practical system must estimate such a mask, and as a result deviations from an ideal mask must be considered. Hence, we want to find how tolerant recognition performance is to deviations from an ideal mask. Second, we want to give a quantitative measure of the potential improvement on ASR performance using our speech segregation model as a front end.

We use the missing-data technique (Cooke *et al.*, 2001) for our ASR evaluation. This technique uses a binary time–frequency mask, where 1's indicate reliable T–F units and 0's unreliable or missing ones. Hence, it works seamlessly with the output from our speech segregation system. We have implemented the missing data algorithm with the same 128-channel gammatone filterbank as described in Sec. III. Feature vectors are obtained using the instantaneous Hilbert envelope at the output of each gammatone filter. More specifically, each feature vector is extracted by smoothing the envelope using an 8-ms first-order filter, sampling at a frame rate of 10 ms and finally log compressing. There are different classification methods for missing-data recognition. Here, we use the bounded marginalization method (Cooke *et al.*, 2001). As in the original study, the task domain is recognition of connected digits, and both training and testing are performed using the male speaker dataset in the TIDigits database (Leonard, 1984).

To study the sensitivity of an ideal mask to estimation error, our first test assesses the correctness score and the accuracy score (correctness minus word insertion errors) when a random deviation from an ideal binary mask is introduced. Here, we use for simplicity a monaural condition as in Cooke *et al.* (2001). Deviations are obtained by randomly flipping the same number of bits from 0's and 1's; the number is measured as percentage of the total number of 1's in an ideal mask. The percentages tested are 0%, 5%, 10%, 20%, and 50%. Since the underlying acoustic energy associated with a T–F unit, or a bit, can vary in a large range, we further measure the energy deviation ratio as the ratio of the energy corresponding to flipped bits and the total energy corresponding to the ideal binary mask. The results for a male target speaker mixed with ''car noise'' (Cooke *et al.*, 2001) are given in Fig. 15, where the abscissa indicates the energy deviation ratio. Three SNR levels for the mixture, i.e., −5, 0, and 5 dB, are tested. Figure 15(A) give the correctness score and Fig. 15(B) the accuracy score. Figure 15 shows that both correctness score and accuracy score decrease gradually and systematically as deviation ratio increases. This suggests that ideal binary masks are robust to estimation error. A comparison between Fig. 15(A) and Fig. 15(B) shows that the accuracy score degrades faster than the correctness score. This suggests that word insertions, which result from noise retention or word boundary blurring, are more sensitive to estimation error than recognition of present words.

The second test directly evaluates binary masks estimated by our system for binaural conditions with two sources and three sources. For all tests, the same male target speaker is located at 0°. Both training and testing of the system are performed on acoustic features from the left ear signal (see Fig. 1). Figure 16(A) and Fig. 16(B) show the correctness and accuracy scores for a two-source condition, where the interference is another male speaker at 30°. The performance of our model is compared against the ideal masks systematically for four SNR levels, i.e., 5, 0, −5, and −10 dB. Also shown in the figure is the baseline performance where the recognition is conducted on unprocessed mixtures from the left ear. Similarly, Fig. 16(C) and Fig. 16(D) show the results for the three-source case with an added female speaker at −30°. The results in Fig. 16 show that an ideal binary mask exhibits only slight and gradual degradation in recognition performance with decreasing SNR and increasing number of sources. In the two-source case, the estimated masks perform equally well as the ideal masks. In the three-source case, the estimated masks do not perform as well, and this is to be expected since we know from Sec. VI A that the quality of ideal mask estimation for three sources is not as good as for two sources. Consistent with the observations from Fig. 15, performance degrades more quickly for the accuracy score than for the correctness score. Observe that large improvements over baseline performance are obtained across all conditions (to a lesser degree for the accuracy score in the three-speaker condition). This shows the strong potential of applying our model to robust speech recognition.

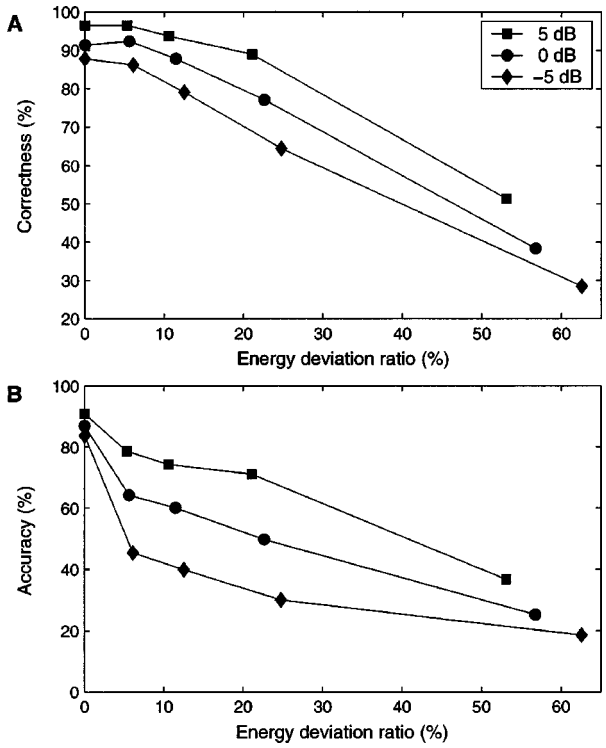


FIG. 15. Degradation of recognition score with deviations from an ideal binary mask evaluated for three SNR values: 5 dB (square), 0 dB (circle), and −5 dB (diamond). (A) Correctness score. (B) Accuracy score.

C. Speech intelligibility evaluation

Finally, we evaluate our model on speech intelligibility with human listeners. Before reporting the results, we should point out that human listeners have a remarkable ability to perform ASA, and their superior ability to recognize speech in the presence of acoustic interference is the very motivation for our model design. Because of this, our tests focus on relatively low SNR conditions; otherwise, scores will be indiscriminately high for both unprocessed mixtures and segregated speech.

We use the Bamford–Kowal–Bench sentence database that contains short semantically predictable sentences (Bench and Bamford, 1979) for intelligibility tests. The score is evaluated as the percentage of keywords correctly identified,

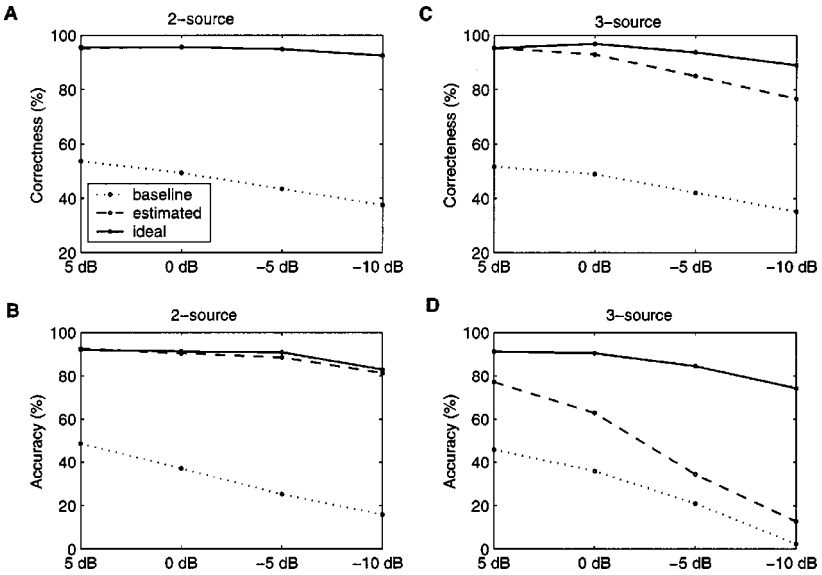


FIG. 16. Recognition performance at different SNR values for original mixture (dotted line), ideal binary mask (solid line), and estimated mask (dashed line). (A) Correctness score for a two-source case. (B) Accuracy score for a two-source case. (C) Correctness score for a three-source case. (D) Accuracy score for a three-source case.

ignoring minor errors such as tense and plurality (Stubbs and Summerfield, 1990). Two different spatial configurations are considered: a two-source configuration at 0° and 5°, and a three-source configuration at −30°, 0°, and 30°. To eliminate potential location-based priming effects (Maljkovic and Nakayama, 1996) we randomly swap the locations for target and interference for different trials. In the unprocessed condition, binaural signals are produced by convolving original signals with the corresponding HRTFs and the convolved signals are presented to a listener dichotically (see Bodden, 1993). In the processed condition, our algorithm is used to reconstruct the target signal at the better ear and results are presented diotically.

Twelve native English speakers with normal hearing, between 24–30 years old, participated in the experiments. The tests were conducted in a sound-insulating booth (IAC model 40a-9) and signals were presented over Sennheiser HD 256 headphones. At the beginning of a test, subjects were familiarized with the voice of a target male speaker and they were free to adjust the sound volume to a comfortable level. The task of a subject during each test run was to report what was comprehended and a human operator marked the result. Each listener participated in a total of 8 conditions. Each condition contained 25 new, randomly chosen sentences, with the first 5 sentences used for practice only and their data discarded.

Figure 17 gives the keyword intelligibility score (median values and interquartile ranges) for the two-source configuration. Three SNR level are tested: 0, −5, and −10 dB, where the SNR is computed at the better ear for each sentence. The interfering source used for this configuration is babble noise. The general finding is that our algorithm improves the intelligibility score for the tested conditions. The improvement becomes larger as the SNR decreases (61% at −10 dB), even though the algorithm introduces more target distortions at lower SNR levels. Our informal observations suggest, as expected, that the intelligibility score improves for unprocessed mixtures when two sources are more widely separated than 5°. Figure 18 shows the results for the three-source configuration, where our model yields a 40% improvement. Here, the SNR is fixed at −10 dB at the better ear. The two interfering sources are one female speaker and a different male speaker. Note that, in this case, azimuth sepa-

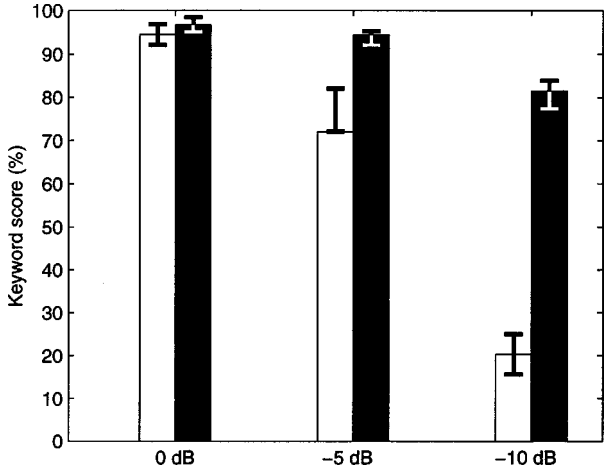


FIG. 17. Keyword intelligibility score (median values and interquartile ranges) before (white bars) and after processing (black bars) for a two-source condition (0° and 5°) at three SNR values: 0, −5, and −10 dB.

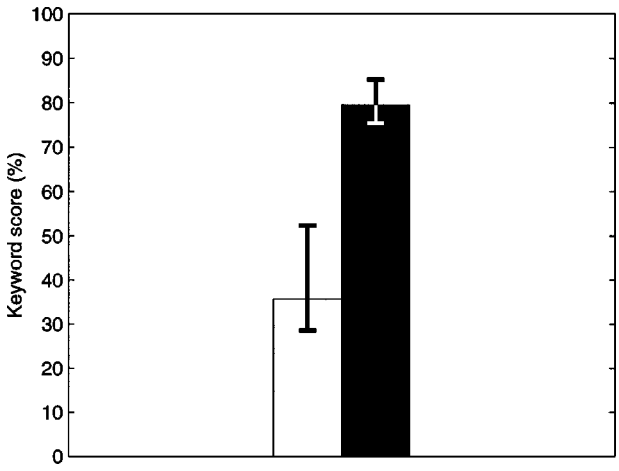


FIG. 18. Keyword intelligibility score (median values and interquartile ranges) before (white bars) and after processing (black bars) for a three-source condition (0°, 30°, and −30°) at −10-dB SNR.

ration is high between the three sources. Though we have not formally tested in the three-source configuration, we would expect that a trend similar to the one in Fig. 17 occurs with respect to SNR levels; that is, the model improvement decreases as SNR increases.

We recognize that comprehensive human subject evaluations of a model would require a separate study (e.g., see Stubbs and Summerfield, 1990), and indeed this is a topic we intend to pursue in the future. Nonetheless, as far as we know, our system is the first binaural model that has been shown to produce a large speech intelligibility improvement for normal listeners (see Kollmeier and Koch, 1994; Shamsoddini and Denbigh, 2001). The configurations and SNR conditions under which improvement occurs will be systematically characterized in future investigation.

VII. DISCUSSION

The human auditory system is capable of adapting to a variety of acoustical situations. A key feature of our model is the introduction of supervised learning for different spatial configurations, and training is conducted independently for different frequency channels. We assume that such training takes place before performing specific segregation tasks, and it would correspond to learning during the development stage. Supervised signals for a spatial configuration of target and intrusion could be supplied in a number of ways, including sound localization, signal estimation from a specific location, and even information extracted from a different modality (e.g., vision). It is worth emphasizing that, unlike a typical supervised learning situation, the training here does not need to capture the specific contents of training signals. As a result the model can be trained equally well using other natural sounds, and estimated distributions generalize in a broad range. In an earlier study (Roman *et al.*, 2002), for example, we employed a different training methodology and a different training corpus, but the system performance was very similar.

While satisfying the demands of an effective computational system, our model is motivated by physiological and psychoacoustical findings regarding the extraction of spatial features (Patterson *et al.*, 1988). The peripheral processing is based on a gammatone filterbank, which has a foundation in physiology and psychoacoustics. Similarly, the cross-correlation mechanism for ITD extraction as well as the across-frequency integration for localization are supported by related physiological findings (Popper and Fay, 1992).

An open question concerns the role of spatial location in perceptual separation of competing sounds. The experiments by Culling and Summerfield (1995), using simulated vowels in which the formants were defined by noise bands, showed that simultaneous grouping across frequencies based on ITD is weak. Later experiments by Darwin and Hukin (1997, 1999) found that ITD plays a weak role in concurrent sound segregation, but a much stronger role in linking acoustic events from a common location over time. The recent experiments of Freyman *et al.* (2001) further showed a sizable improvement in recognizing target speech in the presence of one or two competing speakers based on perceived spatial separation, which suggests a location-based grouping mechanism. Our computational results demonstrate that computed locations can play an effective role in across-frequency grouping. On the other hand, many monaural cues are also important for sound source segregation (see the Introduction), and how to incorporate both monaural and binaural cues in a comprehensive system remains a challenge.

Our approach uses characteristic clustering of the joint ITD–IID space in order to accurately estimate an ideal binary mask. Related models for estimating target masks through clustering have been proposed previously (Tessier and Berthommier, 1997; Lehn, 1997; Glotin *et al.*, 1999, Jourjine *et al.*, 2000). Notably, the experimental results by Jourjine *et al.* (2000) suggest that speech signals in a multiple-speaker condition obey to a large extent disjoint orthogonality in time and frequency. That is, at most one source has a nonzero energy at a specific time and frequency. Such models, however, assume input directly from microphone recordings and head-related filtering is not considered. Simulation of human binaural hearing introduces different constraints as well as clues to the problem. First, both ITD and IID should be utilized, since IID is more reliable for higher frequencies than ITD. Second, frequency-dependent combinations of ITD and IID arise naturally for a fixed spatial configuration. Consequently, channel-dependent training for each frequency band becomes necessary. Our tests with just ITD (as in Glotin *et al.*) or channel-independent classification (as in Jourjine *et al.*) yield considerably inferior performance.

As illustrated in Fig. 13, the proposed model can be used to extract target speech from an acoustic mixture that contains more than one intrusion. Although segregation results are expected to drop as the number of sources increases, this property of our model differs from blind source separation using independent component analysis (Hyväarinen *et al.*, 2001) or spatial filtering using sensor arrays (Krim and Viberg, 1996); such techniques require that the number of sensors increases as the number of acoustic sources increases. A main reason for this difference is that considerations of human audition play a large role in our model design.

Conventional two-microphone adaptive beamformers can develop one deep null which cancels almost perfectly one interference under optimal conditions (Greenberg and Zurek, 2001). The performance, however, degrades when the number of interfering sources increases and is largely affected by the relative SNR of the individual interferences in the reference channel. Weiss (1987) measured the attenuation of individual interferences in acoustical mixtures across different conditions. The experimental results in the anechoic case show attenuation up to 14.5 dB in the two-source case, when both target and interference are active during filter adaptation. For the three-source case, the performance degrades across all interferences by 4 dB, and improvement can be as low as 0 dB. In comparison, our model works for a wide range of spatial configurations with two or more sources; for example, Fig. 13 shows that with three sources our model still obtains an average SNR gain of 11.3 dB. Conditions with high SNRs degrade the performance of adaptive beamforming. Our model, on the other hand, works especially

well for high-SNR scenarios. Subband versions of adaptive beamforming also exist (see, for example, Nordholm *et al.*, 2003). In this case, the signal is analyzed independently in frequency bands, and different directivity patterns are adaptively chosen in each band. This allows cancellation of multiple interferences with nonoverlapping spectra (Cezanne and Pong, 1995). Conventional adaptive beamformers with slow adaptation rate are unable to track fast spectral changes in a multispeaker scenario, resulting in suboptimal performance. Using a frame-by-frame multisource localization scheme, Liu *et al.* (2001) have proposed an equalization and cancellation system that has virtually zero adaptation time. Their two-microphone system exploits the location information in each frame and steers a different null in each frequency band, resulting in 6–7 dB gain in multispeaker scenarios. Our model uses a similar strategy, by employing the localization cue independently in each T–F unit in order to cancel simultaneous interferences. Hence, binaural processing models including ours may have advantages over adaptive beamformers in a range of acoustical situations.

In terms of limitations, our model currently does not address room reverberation or moving sound sources. Observe that supervised training is required for different spatial configurations. This limits the flexibility of our system to cope with, say, diffuse background noise. In addition, the localization of many sources in reverberant conditions with just two sensors is a challenging topic. The situation becomes more complex when source motion is considered. Some tracking mechanism based on measurements of binaural cues across frequency channels, combined with channel selection to discard unreliable T–F units, could be employed to estimate the locations of active sources. For voiced sources, periodicity may provide a measure for the reliability of T–F units (see Wu *et al.*, 2003). Spatial and pitch information have both been utilized to simulate double-vowel recognition, showing added benefits for voiced stimuli (Lehn, 1997; Tessier and Berthommier, 1997). Other auditory mechanisms, such as the precedence effect and forward/backward masking, could also provide important cues to cope with reverberation. Our model also does not address how to define a target in a multisource situation; to address this issue would inevitably require some high-level processes such as attention and task specification. We plan to investigate these and other related issues in future work.

To conclude, we have proposed a model for speech segregation based on spatial location. We have observed systematic deviations of the ITD and IID cues from the reference ones with respect to the relative strength between target and acoustic interference, and configuration-specific clustering in the joint ITD–IID feature space. Consequently, supervised learning of binaural patterns is employed for individual frequency channels and different spatial configurations. Finally, the system estimates a binary mask in order to eliminate acoustic energy in time–frequency units where interference is stronger than target. Our model has been systematically evaluated using both SNR and ASR measures. Evaluation results show that the system estimates ideal binary masks very well and performance degradation is gradual with increasing number and intensity of interferences. In addition, when tested with normal listeners, the model produces large speech intelligibility improvements for two-source and three-source conditions.

ACKNOWLEDGMENTS

The authors wish to thank the three anonymous reviewers for their constructive suggestions/criticisms. This research was supported in part by an AFOSR grant (F49620-01-1-0027) and an NSF grant (IIS-0081058). G.J.B. was supported by EPSRC grant GR/R47400/01. A preliminary version of this work is included in the Proceedings of 2002 ICASSP.

Arbib, M. A., editor (**2003**). *The Handbook of Brain Theory and Neural Networks*, 2nd ed. (MIT Press, Cambridge, MA).

Bench, J., and Bamford, J. (**1979**). *Speech Hearing Tests and the Spoken Language of Hearing-Impaired Children* (Academic, London).

Bernstein, L. R., and Trahiotis, C. (**2001**). ''Transposed stimuli reveal similar underlying sensitivity to interaural timing information at high and low frequencies,'' J. Acoust. Soc. Am. **109**, 2485–2485.

Blauert, J. (**1997**). *Spatial Hearing—The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA).

Bodden, M. (**1993**). ''Modeling human sound-source localization and the cocktail-party-effect,'' Acta Acust. (Beijing) **1**, 43–55.

Bodden, M. (**1996**). ''Auditory demonstrations of a cocktail party processor,'' Acustica **82**, 356–357.

Breebaart, J., van der Par, S., and Kohlrausch, A. (**2001**). ''Binaural processing model based on contralateral inhibition. I. Model structure,'' J. Acoust. Soc. Am. **110**, 1074–1088.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT Press, Cambridge, MA).

Bronkhorst, A. (**2000**). ''The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,'' Acustica **86**, 117–128.

Brown, G. J., and Cooke, M. P. (**1994**). ''Computational auditory scene analysis,'' Comput. Speech Lang. **8**, 297–336.

Burkhard, M. D., and Sachs, R. M. (**1975**). ''Anthropometric manikin for acoustic research,'' J. Acoust. Soc. Am. **58**, 214–222.

Cezanne, J., and Pong, A. N. (**1995**). ''An adaptive subband differential microphone,'' J. Acoust. Soc. Am. **96**, 3262.

Colburn, H. S. (**1977**). ''Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise,'' J. Acoust. Soc. Am. **61**, 525–533.

Cooke, M. P. (**1993**). *Modeling Auditory Processing and Organization* (Cambridge University Press, Cambridge, U.K.).

Cooke, M. P., Green, P., Josifovski, L., and Vizinho, A. (**2001**). ''Robust automatic speech recognition with missing and unreliable acoustic data,'' Speech Commun. **34**, 267–285.

Culling, J. F., and Summerfield, Q. (**1995**). ''Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay,'' J. Acoust. Soc. Am. **98**, 785–797.

Darwin, C. J., and Hukin, R. W. (**1997**). ''Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity,'' J. Acoust. Soc. Am. **102**, 2316–2324.

Darwin, C. J., and Hukin, R. W. (**1999**). ''Auditory objects of attention. The role of interaural time differences,'' J. Exp. Psychol. Hum. Percept. Perform. **25**, 617–629.

Duda, R. O., Peter, E. H., and Stork, D. G. (**2001**). *Pattern Classification*, 2nd ed. (Wiley, New York).

Durlach, N. I. (**1972**). ''Binaural signal detection: Equalization and cancellation theory,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. II.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). ''Spatial release from informational masking in speech recognition,'' J. Acoust. Soc. Am. **109**, 2112–2122.

Gaik, W. (**1993**). ''Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling,'' J. Acoust. Soc. Am. **94**, 98–110.

Gardner, W. G., and Martin, K. D. (**1994**). ''HRTF measurements of a KEMAR dummy-head microphone,'' MIT Media Lab Perceptual Computing Technical Report #280.
Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., and Dahlgren, N. (**1993**). ''Darpa timit acoustic-phonetic continuous speech corpus,'' Technical Report NISTIR 4930, National Institute of Standards and Technology, Gaithersburg, MD.
Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **44**, 99–122.
Glotin, H., Berthommier, F., and Tessier, E. (**1999**). ''A CASA-labeling model using the localisation cue for, robust cocktail-party speech recognition,'' Proc. Eurospeech, Vol. 5, pp. 2351–2354.
Greenberg, J. E., and Zurek, P. M. (**2001**). ''Microphone—array hearing aids,'' in *Microphone Arrays: Signal Processing Techniques and Application*, edited by M. Brandstein and D. Ward (Springer, Berlin), pp. 229–253.
Hyväarinen, A., Karhunen, J., and Oja, E. (**2001**). *Independent Component Analysis* (Wiley, New York).
Jeffress, L. A. (**1948**). ''A place theory of sound localization,'' J. Comp. Physiol. Psychol. **41**, 35–39.
Jourjine, A., Rickard, S., and Yilmaz, O. (**2000**). ''Blind separation of disjoint orthogonal signals: Demixing N sources from 2 mixtures,'' in Proc. ICASSP, Vol. 5, pp. 2985–2988.
Kollmeier, B., and Koch, R. (**1994**). ''Speech enhancement based on physiological and psychoacoustical models of modulation perception and binaural interaction,'' J. Acoust. Soc. Am. **95**, 1593–1602.
Krim, H., and Viberg, M. (**1996**). ''Two decades of array signal processing research: The parametric approach,'' IEEE Signal Process. Mag. **13**, 67–94.
Lehn, K. H. (**1997**). ''Modeling binaural auditory scene analysis by a temporal fuzzy cluster analysis approach,'' Proc. IEEE WASPAA.
Leonard, R. G. (**1984**). ''A database for speaker-independent digit recognition,'' Proc. ICASSP, pp. 111–114.
Lindemann, W. (**1986**). ''Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation for lateralization for stationary signals,'' J. Acoust. Soc. Am. **80**, 1608–1622.
Liu, C., Wheeler, B. C., O'Brien, Jr., W. D., Lansing, C. R., Bilger, R. C., Jones, D. L., and Feng, A. S. (**2001**). ''A two-microphone dual delay-line approach for extraction of a speech sound in the presence of multiple interferers,'' J. Acoust. Soc. Am. **110**, 3218–3230.
Lyon, R. F. (**1983**). ''A computational model of binaural localization and separation,'' Proc. of IEEE ICASSP, pp. 1148–1151.
MacPherson, E. A. (**1991**). ''A computer model of binaural localization for stereo imaging measurement,'' J. Audio Eng. Soc. **39**, 604–622.
Maljkovic, V., and Nakayama, K. (**1996**). ''Priming of pop-out. II. Role of position,'' Percept. Psychophys. **58**, 977–991.
Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, San Diego).
Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). ''A model for prediction of thresholds, loudness, and partial loudness,'' J. Audio Eng. Soc. **45**, 224–240.
Nordholm, S. E., Claesson, I., and Grbic, N. (**2003**). ''Performance limits in subband beamforming,'' IEEE Trans. Speech Audio Process. **11**, 193–203.
Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1988**). ''An efficient auditory filterbank based on the gammatone function,'' Applied Psychology Unit (APU), Report 2341 (Cambridge, UK).
Popper, A. N., and Fay, R. R., editors (**1992**). *The Mammalian Auditory Pathway: Neurophysiology* (Springer, New York).
Roman, N., Wang, D. L., and Brown, G. J. (**2002**). ''Location-based sound segregation,'' Proc. ICASSP, Vol. 1, pp. 1013–1016.
Sain, S. R., Baggerly, K. A., and Scott, D. W. (**1994**). ''Cross-validation of multivariate densities,'' J. Am. Stat. Assoc. **89**, 807–817.
Shackleton, T. M., Meddis, R., and Hewitt, M. J. (**1992**). ''Across frequency integration in a model of lateralization,'' J. Acoust. Soc. Am. **91**, 2276–2279.
Shamsoddini, A., and Denbigh, P. N. (**2001**). ''A sound segregation algorithm for reverberant conditions,'' Speech Commun. **33**, 179–196.
Shinn-Cunningham, B. G., Schickler, J., Kopc-Carono, N., and Litovsky, R. (**2001**). ''Spatial unmasking of nearby speech sources in a simulated anechoic environment,'' J. Acoust. Soc. Am. **110**, 1118–1129.
Silverman, B. W. (**1986**). *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, New York).
Slatky, H. (**1993**). ''Algorithmen zur richtungsselektiven Verarbeitung von Schallsignalen eines binauralen Cocktail-Party-Prozessors,'' Ph.D. thesis, Ruhr-Universität Bochum.
Stern, R. M., and Colburn, H. S. (**1978**). ''Theory of binaural interaction based on auditory nerve data. IV. A model for subjective lateral position,'' J. Acoust. Soc. Am. **64**, 127–140.
Stern, R. M., and Trahiotis, C. (**1995**). ''Models of binaural interaction,'' in *Hearing*, edited by B. C. J. Moore (Academic, New York).
Stubbs, R. J., and Summerfield, Q. (**1990**). ''Algorithms for separating the speech of interfering talkers: Evaluations with voiced sentences, and normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **87**, 359–372.
Tessier, E., and Berthommier, F. (**1997**). ''A model of the cumulative effect of pitch and interaural delay differences for double vowel segregation,'' Proc. ICSP Seoul.
Wang, D. L., and Brown, G. J. (**1999**). ''Separation of speech from interfering sounds based on oscillatory correlation,'' IEEE Trans. Neural Netw. **10**, 684–697.
Weintraub, M. (**1986**). ''A computational model for separating two simultaneous talkers,'' Proc. ICASSP, pp. 81–84.
Weiss, M. (**1987**). ''Use of an adaptive noise canceller as an input preprocessor for a hearing aid,'' J. Rehabil. Res. Dev. **24**, 93–102.
Whittkop, T., and Hohmann, V. (**2003**). ''Strategy-selective noise reduction for binaural digital hearing aids,'' Speech Commun. **39**, 111–138.
Wu, M., Wang, D. L., and Brown, G. J. (**2003**). ''A multipitch tracking algorithm for noisy speech,'' IEEE Trans. Speech Audio Process. **11**, 229–241.

Nonlinear characteristics of single-reed instruments: Quasistatic volume flow and reed opening measurements

Jean-Pierre Dalmont,[a] Joël Gilbert, and Sébastien Ollivier
Laboratoire d'Acoustique de l'Université du Maine (UMR CNRS 6613), 72085 Le Mans cedex 9, France

(Received 20 December 2002; accepted for publication 16 June 2003)

A wind instrument can be described as a closed feedback loop made up of a linear passive element—the resonator—and a lumped nonlinear element—the mouthpiece. A method for measuring the nonlinear characteristics of the mouthpiece—nonlinear flow relationship—in static condition is given. An artificial mouth is used in which the volume flow is deduced from the pressure difference between both sides of a constriction (orifice) which takes place in the resonator. The orifice also plays the role of a nonlinear absorber, thwarting possible reed oscillations. This allows the measurement of the complete characteristics. In addition, the reed opening is measured using an optical device. Results are compared to a model in which the reed is reduced to its stiffness and the flow is governed by the Bernoulli equation. It is shown that the reed stiffness and the ratio of the effective surface of the jet and the reed opening are constant in a large range of openings. Standard range values of embouchure parameters are given. [DOI: 10.1121/1.1603235]

PACS numbers: 43.75.Ef, 43.75.Pq, 43.75.Yy [NHF]

I. INTRODUCTION

Sound production in reed wind instruments is the result of self-sustained oscillations. A mechanical oscillator, the reed, acts as a valve which modulates the air flow entering into the mouthpiece. The destabilization of the mechanical element is the result of a complex aeroelastic coupling between the reed, the air flow into the instrument driven by the mouth pressure of the musician, and the resonant acoustic field in the instrument itself. Following McIntyre *et al.* (1983), wind instruments can be described in terms of a lumped model formed by a closed feedback loop operating as a self-sustained oscillator. In their model the loop is made up of two elements, a lumped nonlinear element—the mouthpiece—and a linear passive element—the resonator, that is the instrument itself. The modeling and the measurement of the resonator have been studied extensively since Bouasse (1929) by many authors (see, for example, references in Nederveen, 1998, or Fletcher and Rossing, 1998). On the contrary, the nonlinear element has only recently been the subject of thorough studies (see, for example, the review given by Hirschberg, 1995) and the knowledge of this element's action is considerably less than that of the resonator. The nonlinear element can be defined by a relationship between the pressure difference across the reed and the volume flow at the inlet of the pipe of the instrument. Neglecting inertia and damping, the volume flow is thus an explicit function of the pressure difference. This function is called the characteristics of the reed. The dynamic aspects related to the reed are certainly not negligible with regard to the spectrum of the instrument (Thomson, 1979). However, when the playing frequency is small compared to the reed resonance frequency, it has been shown that these aspects are not essential from the strict point of view of the auto-oscillation (Wilson and Beavers, 1974; Dalmont *et al.*, 1995). Most authors converge on an elementary model of the characteristics presented in Sec. II A (Wilson and Beavers, 1974; Fletcher, 1979; Fletcher, 1993; Saneyoshi *et al.*, 1987; Kergomard, 1995; Kergomard *et al.*, 2000; Ollivier *et al.*, 2002). It is based on a quasistationary model of the air flow through the mouthpiece, and on a mechanical model of the reed discussed, respectively, in Secs. II B and II C. The aim of the present paper is to obtain experimental data for the nonlinear element in order to check the validity of the elementary model and to find realistic values of the parameters useful for physical modeling synthesis. The data may also give information helpful for the understanding of the physical phenomena involved. Therefore, a method for measuring the characteristics in stationary conditions using an artificial blowing machine has been developed. To perform the volume flow measurements, an orifice (constriction) taking the place of the resonator is used as a pressure-reducing element for a differential pressure flowmeter (Sec. III B). The major advantage of such a device is that the orifice also plays the role of a nonlinear absorber (Ingard and Ising, 1967), which thwarts a possible standing wave in the mouthpiece. Oscillations of the reed are thus in most cases impeded and avoided if the orifice is well chosen. This makes possible the measurement of complete characteristics of real clarinet mouthpieces. The choice and dimensions of the orifice which depends on the reed properties are discussed Sec. III C. The orifice does not induce additional sensors other than those commonly present in an artificial mouth. In addition to the flow measurement, the opening of the reed is measured optically. Then, the measurements of the reed opening and of the pressure difference across the reed give information on the evolution of the reed stiffness, which is important for the understanding of the quasistatic reed mechanics. To obtain relevant results, an accurate calibration of the sensors has to be done. Thus, a specific experimental procedure has been

[a]Electronic mail: jean-pierre.dalmont@univ-lemans.fr

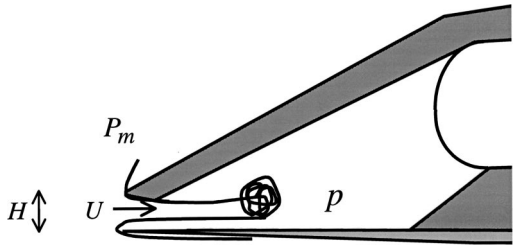


FIG. 1. Schematic view of the mouthpiece with jet formation and turbulent dissipation. H reed opening, U volume flow, P_m mouth pressure, and p pressure in the mouthpiece.

developed which is discussed in Sec. III E. Some typical experimental results are presented in Sec. IV A. Sections IV B and IV C are focused on two aspects of the mechanics of the reed, viscoelasticity and stiffness, respectively. Section IV D is dedicated to the flow aspects. Finally, in Sec. IV E typical values of the parameters of the model are summarized and the accuracy of their determination is discussed.

II. STATE OF THE ART

A. Elementary model

Backus (1963) has presented the first measurements of the characteristics of a single-reed instrument under steady flow conditions. The main result of Backus is a nonlinear expression relating the volume flow U through the reed and two variables: the pressure difference Δp across the reed, and the opening H between the tip of the reed and the mouthpiece. The pressure difference Δp is the difference between the mouth pressure P_m and the pressure in the mouthpiece p (Fig. 1). Backus fitted the experimental data by means of an expression in which U is proportional to $(\Delta p)^{4/3}$ and to $H^{2/3}$. There are good arguments against the adoption of Backus' formula (Hirschberg *et al.*, 1990; Hirschberg, 1995), and most flow models are now based on the assumption of jet formation at the neck of the flow channel with a total turbulent dissipation of the kinetic energy of the jet (Fig. 1). Then, the stationary Bernoulli equation (Wilson and Beavers, 1974; Fletcher, 1979, 1993; Saneyoshi *et al.*, 1987) can be used, which leads to

$$U = S_j \sqrt{\frac{2\Delta p}{\rho}}, \tag{1}$$

where S_j is the cross section of the jet and ρ the density of the air. The steps leading to Eq. (1) are summarized by Hirschberg (1995) and in Sec. II B.

It is furthermore assumed that the section of the jet is equal to the rectangular surface of the front of the reed slit, that is

$$S_j = wH, \tag{2}$$

where w is the width of the reed channel. This hypothesis is discussed in Sec. II B. The position of the reed determines the opening cross-section area and controls the volume flow entering in the mouthpiece. Assuming the mechanical response of the reed to be linear and reduced to its stiffness, the variation of the reed opening $y = H_0 - H$, where H is the reed opening and H_0 the reed opening in the absence of flow,

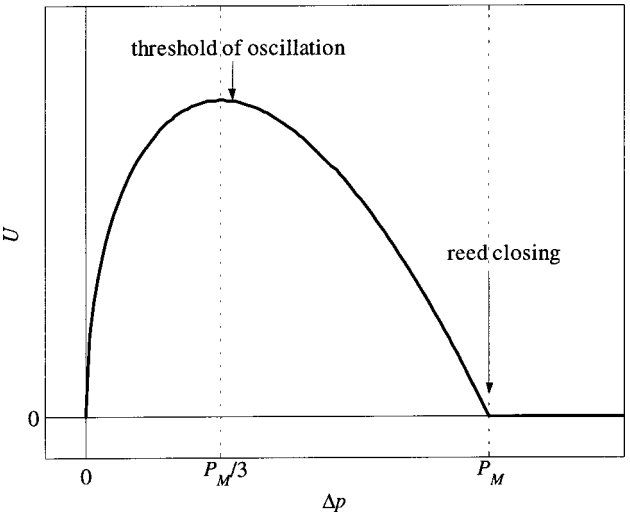


FIG. 2. Theoretical characteristics [Eq. (5)]: volume flow U as a function of the pressure difference ΔP (arbitrary scales). P_M is the value of the pressure difference corresponding to the reed blocked on the mouthpiece.

is proportional to the pressure difference Δp across the reed

$$y = \frac{\Delta p}{k}, \tag{3}$$

where k is the reed stiffness (in Pa/m). This hypothesis is discussed in Sec. II C. Equation (3) is meaningful until H equals zero (reed blocked on the lay). This corresponds to a limit value P_M of the pressure difference Δp given, if k is assumed to be constant, by

$$P_M = kH_0. \tag{4}$$

If the pressure difference is larger than P_M, the reed closes the opening and no flow enters into the mouthpiece. Finally, the volume flow U can be written as a function of the pressure difference ΔP

$$U = w\left(H_o - \frac{\Delta P}{k}\right)\sqrt{\frac{2\Delta p}{\rho}} = wH_o\left(1 - \frac{\Delta p}{P_M}\right)\sqrt{\frac{2\Delta p}{\rho}} \quad \text{if } \Delta p \leqslant P_M$$
$$U = 0 \quad \text{if } \Delta p \geqslant P_M. \tag{5}$$

This nonlinear characteristics is displayed in Fig. 2. Notice that there is a strong localized nonlinearity in the model around the particular value of the pressure difference, $\Delta p = P_M$. The maximum value of the flow $U_{\max} = \frac{2}{3} wH_0 \sqrt{2P_M/3\rho}$ is obtained for $\Delta p = P_M/3$, which is just below the threshold of oscillation (Kergomard *et al.*, 2000)

B. Quasistationary model of air flow

Hirschberg *et al.* (1990) and Van Zon *et al.* (1990) have studied experimentally and theoretically the volume flow control by the motion of the reed in order to explore the limits of the validity of the elementary model presented above, and to provide a better understanding of the results of Backus (1963).

Following Van Zon *et al.* (1990), due to the abrupt transition from the narrow reed channel of height H to the inner part of the mouthpiece of diameter D, flow separation occurs

for sufficiently high Reynolds numbers, that is: $Re = \rho U/(\eta w) > 10$, where η is the shear viscosity of the air. A free jet is formed in the mouthpiece. For large values of $D/H > 10$, which is a typical value for single-reed instruments, the pressure recovery upon deceleration of the flow in the mouthpiece is negligible. Hence, the pressure p in the mouthpiece is assumed uniform and equal to the pressure at the end of the reed channel. Measurements by Van Zon *et al.* (1990) confirm this assumption: the pressure variations within the mouthpiece are less than 3% of the dynamic pressure in the jet.

In the limit of high Reynolds numbers and a short channel ($Re\, H/L > 1000$, where L is the length of the reed channel), the volume flow can be estimated by assuming a uniform flow in the reed channel and by applying Bernoulli's equation. Ignoring the flow separation at the entrance of the channel, the section of the jet is found to be given by Eq. (2). As noted in Hirschberg *et al.* (1990), separation occurs when the edges at the entrance are sharp, which is the case for the clarinet mouthpiece and its reed. A free jet with a section S_j lower than the reed-mouthpiece opening cross-section S is formed within in the channel. For short reed channels ($L/H < 3$) no reattachment of the flow occurs within the channel, and the volume flow U is given by

$$U = \alpha w H \sqrt{\frac{2\Delta p}{\rho}}, \quad \text{where} \quad \alpha = \frac{S_j}{S}. \tag{6}$$

The coefficient α is a ''contraction'' coefficient which is strongly dependent on the geometry of the reed channel inlet. For typical 2D mouthpiece geometry, values in the range $0.5 < \alpha < 0.61$ are expected. Laser Doppler flow measurements and flow visualization experiments by Van Zon (1989) confirm the typical values of α for the geometry considered. For low Reynolds number ($Re < 10$) and long reed channel ($L/H > 10$), the flow is well approximated by a fully developed Poiseuille flow. This corresponds to the case of the reed almost closed, for which the volume flow is thus given by

$$U = \frac{w H^3 \Delta p}{12 \eta L}. \tag{7}$$

Both Poiseuille and Bernoulli limits were also found by Gilbert (1991) and Maurin (1992). The intermediate flows between the two extreme cases mentioned above are discussed in Van Zon *et al.* (1990) and Hirschberg *et al.* (1991). For $L/H > 4$, the jet formed by the separation of the flow from the sharp edge of the reed at the entrance of the reed channel reattaches to the wall after a distance of about $2H$. If the channel is short the friction is negligible; the volume flow U approaches the value given by Eq. (1) and the section S_j of the jet is equal to the reed-mouthpiece opening cross-section S. The quasistationary models described above assume a fixed separation point at the inlet or at the exist of the reed channel and a uniform section of the reed channel. This hypothesis is questionable in the case of the clarinet mouthpiece. The transition between the ''reed channel'' and the mouthpiece can be smooth. In such a case the reed channel height is not uniform, and for $L/H > 4$ the separation point is not easy to determine. As a consequence, the coefficient α is not easy to predict precisely and can take values larger than unity. In other words, in such a case the volume flow U can be larger than predicted with the elementary model [Eqs. (1) and (2)]. Note that, in that case, coefficient α is no more a *vena contracta* coefficient but simply the ratio of the effective jet cross section to the mouthpiece opening cross-section S. In the case of a fully separated jet flow (short reed-channels situation, $L/H < 2$), the channel geometry is not expected to be critical and the result given in Eq. (6) could remain valid. All the theoretical results described before have been successfully compared with experimental results obtained with a two-dimensional mouthpiece geometry. A particularity of clarinet mouthpieces has to be mentioned. The reed channel consists of two parts: (a) the front slit delimited by the edge of the mouthpiece tip and the reed and (b) the lateral slits between the lay and the reed. Then, the effective section S_j can be larger than the opening cross-section S. This is another reason why the ''contraction'' coefficient α as defined by Eq. (6) can be larger than unity.

C. Quasistatic response of the reed

The mechanics of the reed is complex. The material is orthotropic and the dimension irregular. Prediction of the deformation of the reed is difficult because reed is an essentially inhomogeneous material (Heinrich, 1991). Its mechanical properties are variable and also strongly depend on the amount of water in the material (Heinrich, 1991; Obataya and Norimoto, 1999; Marandas *et al.*, 1994). Marandas *et al.* (1994) suggest that a dry reed displays a viscoelastic behavior, whereas a wet reed has a viscoplastic behavior. The reed rolls up (or not) on the lay of the mouthpiece, whose geometry is said to be very critical: from the experience of craftsmen it seems that variations of some hundredths of millimeters on dimensions of the curvature of the lay lead to a change of behavior perceptible by the musician (Hirschberg *et al.*, 1991). In addition, the lips of the musician are pressed on the reed. This means that the mechanical behavior of the reed is also dependent on the coupling with the lip, a mechanical system which is also not easy to characterize.

The reed is usually considered to be a one degree of freedom oscillator, that is the reed tip displacement y is related to the pressure difference by the following equation:

$$\mu \ddot{y} + \mu g \dot{y} + k y = \Delta p, \tag{8}$$

where g is a viscous-damping coefficient (in s^{-1}) and μ a mass per m^2. The dynamic aspects related to the reed are not considered in the present paper and Eq. (8) is then reduced to Eq. (3). The determination of the reed stiffness k is difficult and only orders of magnitude for these parameters can be found in the literature. The difficulty lies in the fact that these parameters are generally found only in an indirect way and under experimental conditions which are not always realistic. Thus, writing

$$\omega_r = \sqrt{\frac{k}{\mu}}, \tag{9}$$

suggests that the stiffness can be deduced from the reed resonance frequency. This is questionable: as Eq. (8) corresponds

to a low-frequency approximation, ω_r might be different from the first reed resonance frequency. In addition, the surface per area μ is itself badly known. Moreover, contrary to what one could suppose, the resonance frequency of the reed itself is difficult to determine. Thus, Facchinetti and Boutillon (2001) showed that the frequency of a squeak depends as much on the resonator as on the reed. The reed stiffness was measured by Nederveen (1998), but not in a playing situation. Another solution for determining this stiffness k consists in measuring the impedance of reed (Boutillon and Gibiat, 1996; Dalmont *et al.*, 1995). Unfortunately, the stiffness is obtained by using an equivalent surface whose value is badly known. A simple method to determine the stiffness consists in determining the closing pressure, P_M, and the opening at rest, H_0. The closing pressure can be estimated with an artificial mouth by seeking the pressure for which the reed starts to oscillate after having been blocked on the lay. This rather inaccurate method (probably $\pm 5\%$ on P_M), based on Eq. (4), is used implicitly by Kergomard (1995). In the present paper, another method allowing a more accurate determination of the beating pressure is given.

In the model presented the reed stiffness k is assumed to be constant. *A priori*, it seems natural to think it is not. The most commonly admitted idea is that the vibrating length of the reed decreases with the opening. This would lead to an increase in the stiffness (Nederveen, 1998). To go further in the analysis, various authors propose to model the reed as a bar (Stewart and Strong, 1980; Sommerfeld and Strong, 1988; Gazengel, 1994; Ducasse, 2001; Van Walstijn, 2002). Their results tend to show (Gazengel, 1994; Ducasse, 2001; Van Walstijn, 2002) that the reed rolls up on the table of the mouthpiece only under certain conditions which are satisfied for only very particular reed geometries. With the dimensions of a real reed and table, Ducasse (2001) showed that the reed deforms without sticking to the table until a given point near the tip touches the table. This result is confirmed by the simulation of Van Walstijn (2002), who shows that the stiffness is nearly constant as long as the end of the reed does not touch the table and takes a larger value afterwards. These studies would justify the ''simplistic'' approach of the model suggested [Eq. (5)] at least until the reed tip touches the lay. It remain to verify whether the transverse bending of the reed which is supported only on the side by the table does not modify this two-dimensional behavior.

III. EXPERIMENTAL DEVICE AND PROCEDURE

A. The artificial mouth

The artificial mouth (Fig. 3) consists of a Plexiglas box with metal reinforcement (Gazengel, 1994). The artificial lip consists of a cylindrical latex balloon of small diameter (10 mm) in which a piece of foam saturated with water is inserted. The lip is fixed on a rigid support which position can be translated vertically by means of a screw. The mouthpiece is inserted in a metal barrel whose horizontal position can be adjusted. Resonators (for example, a clarinet) can be fixed onto the other end of the barrel. The air is supplied by a high-pressure system through a pressure reducing valve.

The pressure P_m in the mouth is measured by a static pressure sensor. A miniature differential pressure sensor mounted in the wall of the mouthpiece measures the pressure difference $\Delta p = P_m - p$ between the mouth cavity and the inside of the mouthpiece. The reed slit opening is measured using a laser beam and a photoelectric diode. This is the method used by Backus (1963) (cf. Fig. 3).

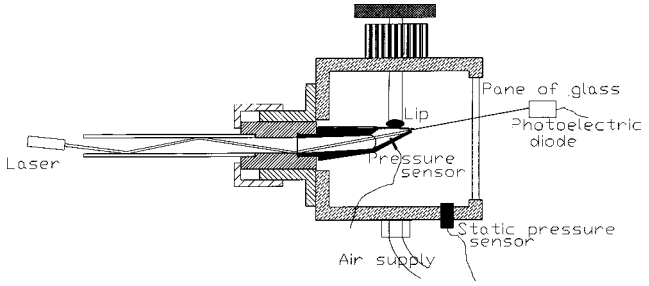


FIG. 3. Experimental device.

B. Flow measurement

To determine the volume flow, the resonator is replaced by a piece of tube with a constriction playing the role of a pressure-reducing element. Using the pressure measurements, the flow through the reed is calculated by using Bernoulli's equation

$$p = \frac{1}{2}\rho \frac{U^2}{S_{\text{dia}}^2}, \tag{10}$$

where the atmospheric pressure is used as a reference, S_{dia} being the section of the opening of the orifice. Compared to another flowmeter placed upstream of the cavity, the orifice, apart from its simplicity of implementation, has several advantages. It makes it possible, if its dimensions are well chosen, to obtain complete characteristics since it avoids oscillations. Indeed, apart from its pressure-reducing effect the orifice plays, for acoustics, the role of a nonlinear resistance damping standing wave inside the barrel. This makes our experiment similar to the one suggested by Benade (1976, p. 437).

C. The choice of the orifice

The diameter of the orifice is the result of a compromise. If this diameter is too large the pressure difference p created by the orifice is too small to be measured. If, on the contrary, it is too narrow the reed closes suddenly above a critical threshold and part of the nonlinear characteristics cannot be explored. This phenomenon is described by Hirschberg (1995): when the pressure increases, the volume flow increases too until the reed position become unstable and closes suddenly (Fig. 4). This occurs when

$$\frac{\partial(P_m)}{\partial U} = \frac{\partial(\Delta p + p)}{\partial U} = 0. \tag{11}$$

Considering that the pressure Δp and p are given, respectively, by Eqs. (5) and (10), it is thus necessary, to avoid phenomenon of sudden closure, that

$$S_{\text{dia}} > wH_0/\sqrt{3}. \tag{12}$$

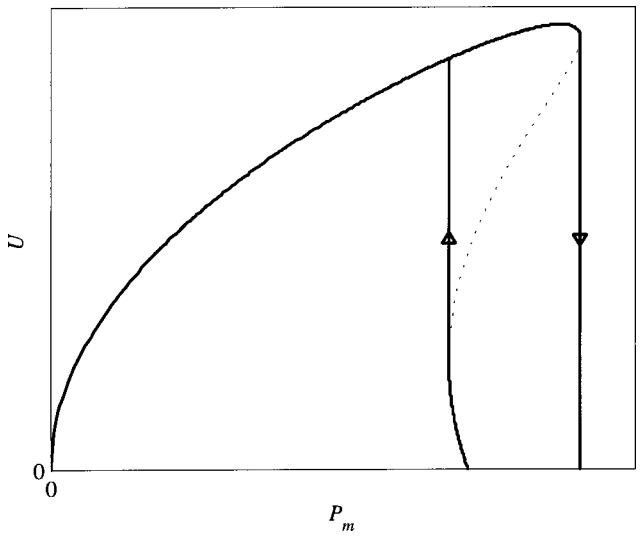


FIG. 4. Volume flow U as a function of the mouth pressure P_m for the clarinet mouthpiece ended with an orifice (arbitrary scales).

In practice, the section of the opening of the orifice can be chosen to be close to the section of the reed slit at rest position. This may avoid the sudden closing of the reed while ensuring a sufficiently large pressure difference. In our experiments orifices of 2, 2.5, 3, 3.5, and 4 mm have been used. The larger orifice for which the sudden closing does not occur is systematically chosen.

The cross-section area of the jet formed by the orifice is assumed here to be equal to the section of the opening of the orifice. This assumption is valid only in the absence of a *vena contracta*. To avoid a *vena contracta* the orifices has been chamfered (Fig. 5). This chamfer suppresses the *vena contracta* and also extends the range in which the coefficient of discharge is constant (OMEGA, 1995). In order to check this, the orifices have been calibrated by means of a volume gas meter. The calibration of the various orifices proves that the *vena contracta* coefficient is constant and equal to unity within 1% uncertainty for all the orifices except for the 3.5-mm diameter orifice, for which it is estimated to be 0.97.

D. Reed opening measurement

Reed opening is measured by means of a laser and a photoelectric diode. This measurement is likely to give information on the behavior of the jet entering in the instrument. In particular, the opening can be compared to the effective section of the jet entering in the mouthpiece. This measurement also makes it possible to relate the opening H and the pressure difference Δp on both sides of the reed. This allows the determination of the evolution of the reed stiffness with the pressure.

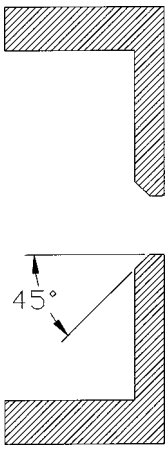


FIG. 5. Orifice geometry.

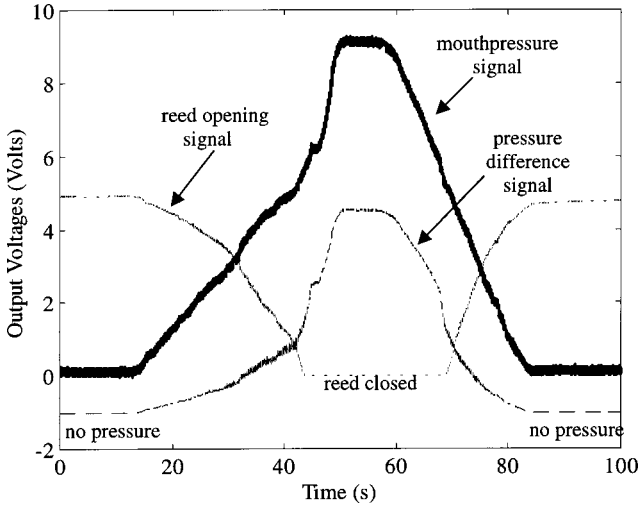


FIG. 6. Experimental signals: ——— static pressure sensor; -·- differential pressure sensor; --- optical sensor.

The sensitivity of the optical system used to measure the opening H is determined by comparison with measurements done with a small ruler placed next to the front of the mouthpiece, by means of visual observation with a camera in macro mode. This device was used for checking the linearity of the optical system and to determine the opening at rest H_0. A difficulty with the optical system is that the light beam can be fully stopped while the reed slit is not completely closed. In other words the optical signal can be zero whether or not the reed is blocked on the lay. To limit this problem, it is checked by visual inspection before each experiment that, when the reed is almost closed, the diode still detects a signal.

E. Experimental procedure

The two pressure signals and the optical signal are collected on a computer via a data acquisition card. The opening at rest having been measured as described above, the experiment starts without blowing pressure ($\Delta p = 0$; see Fig. 6). This state is maintained for a few seconds. This is done for the determination of the zeros of the pressure sensors and the value of the optical signal corresponding to the opening at rest. The pressure in the mouth cavity P_m is then increased gradually until the reed completely closes the opening. This state is also maintained for a few seconds. This is done for the relative calibration of the differential pressure sensor measuring Δp, as in that case $p = 0$ and $\Delta p = P_m$, the sensitivity of the static pressure sensor measuring P_m supposed to be known, the sensitivity of the differential pressure sensor can be deduced. This calibration is necessary because the pressure in the mouthpiece p is obtained by making the difference between the signals from the two pressure sensors

($p = P_m - \Delta p$). The pressure in the mouth is then gradually brought back to zero. A typical duration of such an experiment is 50 to 100 s, with a sampling frequency of 100 Hz (see Fig. 6). For some embouchures, when the reed is almost closed, reed oscillations occur which make the corresponding part of the nonlinear characteristics not exploitable.

IV. EXPERIMENTAL RESULTS

A. A typical experimental result

The nonlinear characteristics were measured for various mouthpieces, reeds, and embouchures. Before each measurement of the characteristics, the embouchure is tested with a cylindrical pipe of 30 cm in order to check that the instrument produces a realistic sound of good musical quality. The phenomena observed are globally reproducible even if one observes a large variation in the numerical values of the various parameters. The goal of this section is to present a typical case in order to stress the most significant results. A more detailed analysis of the experimental results is provided in the following sections.

In Fig. 7(a) a nonlinear characteristic $U = NL(\Delta p)$ is shown for an opening at rest $H_0 = 0.6$ mm. The reed is a Rico Plasticover™ (reed covered with plastic) of strength 3. This type of reed has been chosen because it is less influenced by moisture than a standard reed. The clarinet mouthpiece is a C80 by Selmer. The curve shows a hysteresis, the maximum flow being larger for increasing pressure than for decreasing pressure. This hysteresis can be attributed to the viscoplastic behavior of the reed. This point is discussed in Sec. IV B. The curves are similar to the theoretical curves which are also plotted for reference in Fig. 7(a) (dots) [Eq. (5)]. The parameters of these theoretical curves are chosen so that the maximum of the theoretical curves matches those of the experiments. It appears that the closing is never total and that even when the reed can be considered as closed (beyond 60 hPa) a weak flow remains. This flow decreases when the maximum pressure is maintained a few seconds. The theoretical model is thus valid until the reed is nearly closed. It does not take into account the residual flow when the reed channel is closed.

Figure 7(b) plots the reed opening as a function of the pressure difference. This curve shows that the stiffness of the reed can be considered as roughly independent of the pressure. This is emphasized in Fig. 7(b) by two straight lines of the same slope, showing that the stiffness is approximately the same when the pressure is increasing and when it is decreasing. Only the rest position differs (4% difference). This is confirmed by the fact that between the two theoretical curves of Fig. 7(a) only the values of H_0 are different in the same ratio. This result is explained on the basis of a viscoelastic model in which the return to rest position is delayed (cf. Sec. IV B). It appears from Fig. 7(b) that the zero of the optical signal is reached for a reed aperture, which is still significant. As noted in Sec. III D, this is due to the difficulty of adjustment of the optical setup for which the zero is reached, whereas the reed is still open.

As explained in Sec. II B, the jet cross-section S_j can be deduced from the volume flow using Eq. (1). Figure 7(c) plots the effective cross section as a function of the variation of the reed opening measured by the optical device. The results show systematically that the jet cross section varies linearly with the aperture. In some cases, when the optical adjustment was optimum, a line passing through zero within

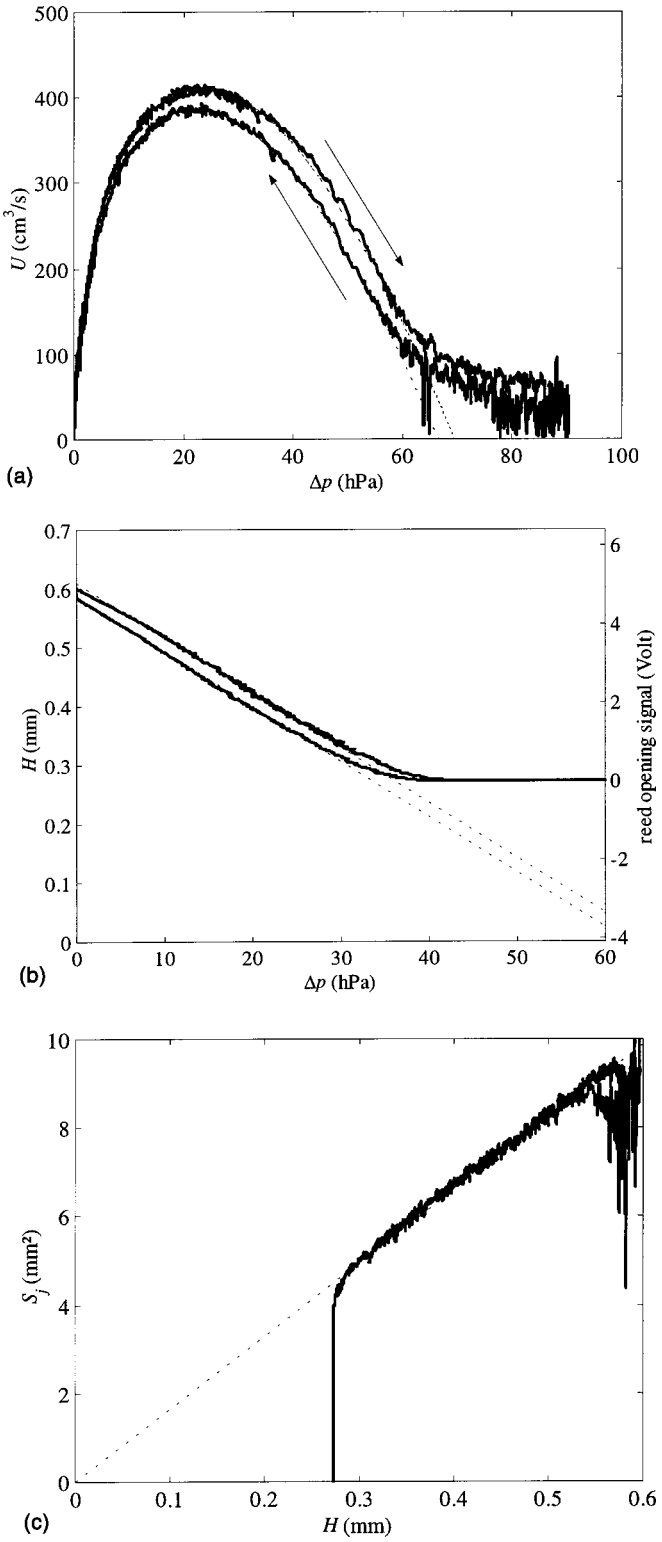


FIG. 7. Typical experimental results (continuous line experiments; dotted lines theory). (a) Nonlinear characteristics, volume flow U versus pressure difference ΔP. (b) reed opening H versus pressure difference ΔP (on the left is the corresponding scale for the reed opening signal); (c) jet cross-section S_j versus reed opening H.

an uncertainty lower than 0.1 mm was obtained. The assumption of a jet cross section proportional to the opening thus seems sensible. This assumption is used for the determination, by extrapolation, of the opening corresponding to the zero value of the optical signal.

B. Viscoelasticity of the reed

As noticed in Sec. IV A, some curves exhibit a hysteresis due to a change of the rest position. This result could seem contradictory with the assumption of a static measurement. Indeed, the duration of the experiment and the speed of the pressure variations are such that the effects of the inertia are negligible during the experiment. The hysteresis can therefore only be explained as the result of a viscoelastic behavior of the reed which only recovers its original rest position after a time delay larger than our experiments (Marandas, 1994). By analyzing the opening as a function of time after the reed has been blocked on the lay and then quickly released, it appears, for a given reed, that this one recovers its rest position in three steps. The reed slit opening reaches almost instantaneously 93% of its maximum value. An exponential decay of the difference between the opening value and the maximum opening with a relaxation time $\tau_1 \approx 8$ s is then observed. At the end of this second phase the reed slit opening reaches 97% of its maximum value. Rest position is finally reached at the end of a last phase, for which the relaxation time is $\tau_2 \approx 900$ s. Taking into account the typical duration of an experiment, it can be asserted that the hysteresis observed in the experimental characteristics is due to a conjugation of these two effects: in particular, the fact that when the reed is closed there remains a flow which tends to decrease if the maximum pressure is maintained a sufficiently long time. These phenomena, which may have some importance for the musicians (Ducasse, 2001; Marandas, 1994), are probably not relevant when considering a physical model of the auto-oscillation process. The actual characteristics might be found somewhere in between the two static characteristics obtained, respectively, upon increasing the pressure and decreasing the pressure. By chance, the hysteresis being rather small, the uncertainty on the relevant parameters of the model is small. Typically it should not exceed few percent of H_0. To limit this effect it is useful, before doing an experiment, to close the reed for a few seconds by applying a large pressure in the mouth volume in order to limit the long-term effect ($\tau_2 \approx 900$ s). This is similar to some musicians' practice, which consists in pressing the reed with the thumb before playing. It is important to notice that the amplitude of the hysteresis can vary considerably with the reed. We observed that, for some reeds, *a priori* not different from the others, the hysteresis did not appear. On the other hand, our experiments have been done with a dry reed covered with plastic. A wet standard reed would probably have emphasized a slightly different behavior (Marandas, 1994).

C. Reed stiffness

In most of our experiments, the reed stiffness can be regarded as a constant until the reed beats. However, with some mouthpieces and reeds a slightly different behavior has been observed. Figure 8 shows a result with a clarinet mouthpiece B40 by Vandoren, a Plasticover reed No. 3 and a reed opening at rest $H_0 = 0.6$ mm. In Fig. 8(a) a plot of the characteristics is compared to the simple basic model [Eq. (5)] in which parameters P_M and wH_0 are chosen so that the maxima of the experimental and theoretical curves coincide. Figure 8(b) plots the reed slit opening as a function of the pressure difference. The reed stiffness can be deduced from the slope of the curve. The following value for the reed stiffness is obtained: $k = 116$ hPa/mm. The reed stiffness can also be deduced from the nonlinear characteristics in which the extrapolated closing pressure is $P_M = 73$ hPa. With $H_0 = 0.6$ mm this leads to $k = P_M / H_0 = 122$ hPa/mm, which is, with regards to the uncertainty, in agreement with the result deduced from reed slit opening measurement. This value is not very different from the result of another experiment plotted in Fig. 7, in which the reed stiffness is found to be $k = 107$ hPa/mm. It is also noticed that the extrapolated line passes through zero for $\Delta p = P_M$, which validates the calibration method of the optical system based on the extrapolation of the function $S_j(H)$ (cf. Sec. IV A). It can be noted that the experimental characteristics and the model [Fig. 8(a)] differ for a pressure slightly lower than $\Delta p = 40$ hPa. This is in agreement with the observation which can be made in Fig. 8(a) that the function $H(\Delta p)$ deviates from a linear behavior above 40 hPa. Beyond this pressure the reed stiffness increases as a result of a reduction of the free-reed length due to contact with the lay. The remarkable result here is that the change in the reed stiffness value appears only for a pressure higher than the threshold of oscillation, $\Delta p \approx P_M/3$. According to various measurements it seems that this phenomenon is general. This result is to be compared with the analysis of various authors (Gazengel, 1994; Ducasse, 2001; Van Walstjin, 2002), according to whom the curling up phenomenon is limited and has a small influence on the reed stiffness, the end of the reed touching the lay without smoothly curling up on the lay. Our measurements do not exclude this scenario but show that it should appear

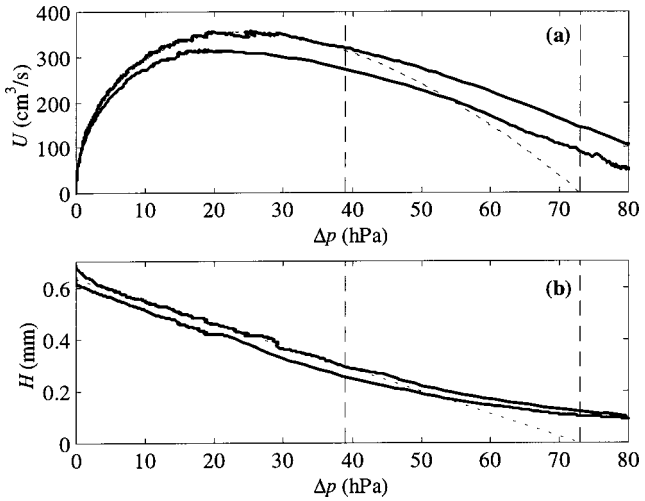


FIG. 8. Experimental results (continuous line experiments; dotted lines theory). (a) nonlinear characteristics, volume flow U versus pressure difference ΔP; (b) reed opening H versus pressure difference ΔP.

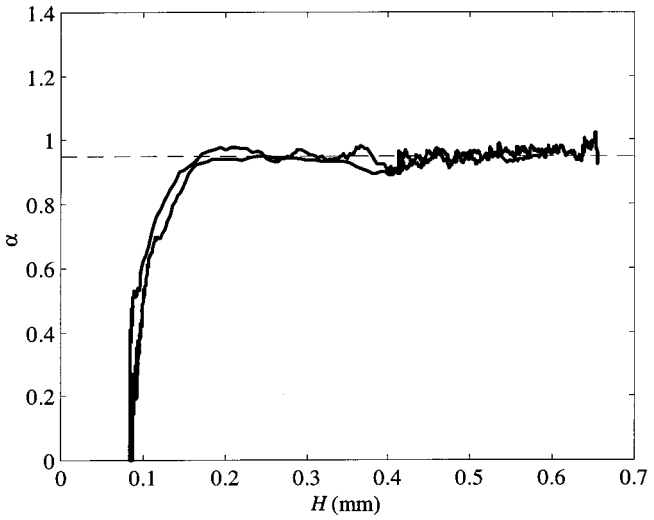


FIG. 9. Contraction coefficient α versus reed opening H (same experiment as in Fig. 8).

only for high pressures. Complementary measurements, allowing to determine directly the points of contact between the reed and the lay, are now in preparation to confirm this result (Ollivier, 2002).

D. Air flow

To evaluate the discrepancy between the model for the volume flow and the measurement, the measured volume flow U can be divided by the theoretical one U_b calculated using Eqs. (1) and (2), where H is measured with the optical device and $w=14$ mm is the external width of the mouthpiece inlet. The dimensionless quantity obtained is the *vena contracta* coefficient α [see Eq. (6)]. It is displayed versus the reed opening in Fig. 9 for the same experiment as in Fig. 8. Parameter α remains more or less constant along a large range of reed openings: $\alpha \approx 0.95$ for $0.2\ \text{mm} < H < 0.65$ mm. For $H < 0.15$ mm the optical measurements are no longer relevant. From the experiment of Fig. 7 the value $\alpha \approx 1.2$ is found. The constant behavior of α has been observed for every embouchure tested, α being in the range 0.85 to 1.30. These results confirm that a volume flow calculated from the Bernoulli law with a constant *vena contracta* coefficient α is a reasonable approximation for sufficiently large reed opening. For a small reed opening the experimental device and particularly the reed opening measurement are, at this stage, not sufficient to draw conclusions.

It is interesting to compare the measured α values of coefficient α with that obtained for simplified geometries of mouthpieces and in less realistic playing conditions. As mentioned in Sec. II B, Van Zon (1989) has obtained values in the range 0.50 to 0.61 for a typical 2D mouthpiece. Furthermore, Maurin (1992) observed larger values, in the range 0.60 to 0.85 for a clarinet mouthpiece mounted in an artificial mouth, the lateral parts of the reed channel being waxed to approach a 2D geometrical situation. The larger value of α could be explained by an effect of the confinement of the flow upstream of the inlet reed channel. This phenomenon would also occur in actual clarinet playing. Valkering (1993) measured volume flows through a reed channel formed by a stiff, flat, metal reed with sharp edges placed on an actual clarinet mouthpiece. It is shown that the flow through lateral sides can increase the flow within 50%, which leads to values for α compatible with our experimental results.

TABLE I. Summary of typical parameters range values for the clarinet.

Quantity	Symbol	Typical values
Reed opening at rest	H_0	0.4–1.0 mm
Beating pressure	P_M	40–100 hPa
Reed stiffness	k	80–130 hPa/mm
Jet effective width	αw	12–18 mm
Maximum flow	$U_{\max}$	200–600 cm^3/s

E. Typical parameters range values

The model formulated in Eq. (5) is based on parameters of which the values have to be defined when dealing with simulations or physical modeling synthesis. Our experiments allow the determination of some of these parameters. In the present section the range of these parameters and the accuracy with which they can be determined using our setup are discussed. Results are summarized in Table I. The reed parameters, that is the opening at rest H_0 and the reed stiffness k, are determined using the optical device (see Secs. III D and IV C). The reed opening is approximately $H_0=1$ mm when the reed is free, that is when the lips do not press the reed. When the lip presses the reed this value decreases but probably not much under $H_0=0.4$ mm, which corresponds to a rather tight embouchure. Because of the small value of the reed opening and of the geometrical irregularities of the reed, this parameter H_0 cannot be determined with great accuracy. The accuracy of this parameter H_0 is estimated to be only 10%. The reed stiffness which is deduced from $k=H_0/\Delta p$ is found to be in the range of 100 hPa/mm. From our measurements it appears that the range in which this parameter varies is rather small. In all the measurements realized, it did not vary more than 50%. This parameter being deduced from H_0, its accuracy is in the same range, that is 10%. The product of the two reed parameters, H_0 and k, is the beating pressure, $P_M=kH_0$. This beating pressure P_M is found to be around 80 mbar for a loose embouchure and around 60 mbar for a tight embouchure. This parameter can be determined with good accuracy from the flow measurement (see Sec. IV A). In some cases an accuracy of 2% can be reached (a little bit more if the hysteresis is considered; see Sec. IV B). From flow measurements the maximum flow can also be deduced, from which the value of the effective surface of the jet at rest, $S_j=\alpha w H_0$, can be deduced with a good accuracy (3%). From this surface the effective width of the reed channel αw can be deduced (Sec. IV D). The uncertainty on this parameter is in the same range as for the reed opening H_0, from which it is deduced. In Table I realistic values of the parameters of the model for the clarinet are summarized. These values are in agreement with those given by other authors (Nederveen, 1998; Stewart and Strong, 1980). We guess that these values are useful for physical modeling synthesis and simulations.

V. CONCLUSION AND PROSPECTS

The experiments presented in the present paper allow a better characterization of the clarinet mouthpiece behavior. It also gives elements useful for a better comprehension of the physical phenomena involved. In the nineties a considerable effort was carried out in parallel at the LAUM (Le Mans, France) and at the TUE (Eindhoven, Netherlands) to obtain a reliable model for the relationship between the jet section and the reed channel height characterized by the coefficient α of Eq. (6) (Hirschberg *et al.*, 1990; Van Zon, 1989; Hirschberg *et al.*, 1994; Hirschberg, 1995). Using simplified 2D geometries, different stationary regimes have been identified experimentally and explained theoretically. We have now proposed an experimental procedure allowing measurement of these characteristics with actual mouthpieces and reeds under conditions close to playing conditions. Our measurements with this new procedure appear to agree qualitatively with the earlier measurements in simplified geometries. The main problem in the interpretation of our results is the uncertainty in the geometry of the reed channel and of the lateral slits between the lay and the reed. Our results confirm that a volume flow calculated from the Bernoulli law with a constant "*vena contracta*" coefficient is a reasonable first approximation for a sufficiently large reed opening. For a small reed opening the effect of friction becomes significant and a correction for viscous effects should be introduced.

In the quasistationary basic model, the behavior of the reed is reduced to a spring with a constant stiffness. Rather surprisingly, experimental results confirm that the latter hypothesis is reasonable, a stiffness value being associated with each embouchure adjustment. Nevertheless, for some of the embouchures, the hypothesis is correct up to a critical pressure threshold above which the equivalent stiffness is increasing when the reed–lay aperture is decreasing. Some authors proposed to consider a variable stiffness $k(H)$ to take into account the curvature of the lay (see, for example, the time-domain simulations of Ducasse, 1990). Our results show that this should be done with care. Indeed, a variation of the stiffness with the opening has been observed for some embouchures, but only when the reed is near closing. Supplementary investigations have to be done with different reeds and mouthpieces in order to check the variability of the results. A comparison with the recent theoretical works of Ducasse (2001) and Van Walstijn (2002) should be done by using a reed and a lip for which characteristics would be measurable and controllable. Incidentally, our measurements confirm the visco-elastic behavior of reeds reported by Marandas *et al.* (1994). It seems, however, that this behavior is not relevant when predicting oscillations because the memory time scales involved are long compared to the oscillation period.

Finally, our study allows us to conclude that the "simplistic" model described in Sec. II is valid at least for the major part of the characteristics. This result confirms the practical interest of theoretical studies based on this model (Wilson and Beavers, 1974; Fletcher, 1979, 1993; Saneyoshi *et al.*, 1987; Kergomard, 1995; Kergomard *et al.*, 2000; Ollivier *et al.*, 2002). Similar studies could be done for other instruments. For instance, an alto saxophone mouthpiece has been tested, too: the same behavior as for a clarinet mouthpiece has been observed both from the point of view of the air flow and the reed mechanics in stationary regime. This suggests that, except for the size, there is no major difference between a clarinet and a saxophone mouthpiece. The case of the double reed of an oboe or a bassoon could also be investigated in the same way. This could allow one to check if the hypothesis by Barjau and Agullo (1989) of a cross section proportional to the opening to the power two is sensible. The conjecture that there is a significant pressure recovery inside the narrow pipes on which double reeds are mounted (Hirschberg, 1995) could also be checked.

ACKNOWLEDGMENTS

The authors are indebted to X. Boutillon, A. Hirschberg, J. Kergomard, and C. J. Nederveen for helpful discussions and to M. Blondet for his participation in experiments.

Backus, J. (**1963**). "Small-vibration theory of the clarinet," J. Acoust. Soc. Am. **35**, 305–313.

Barjau, A., and Agullo, J. (**1989**). "Calculation of the starting transients of a double-reed conical woodwind," Acustica **69**, 204–210.

Benade, A. H. (**1976**). *Fundamentals of Musical Acoustics* (Oxford University Press, Oxford).

Bouasse, H. (**1929**). *Instruments à Vent* (Librairies Delagrave, Paris; reprint Blanchard, Paris, 1985), Vols. I and II.

Boutillon, X., and Gibiat, V. (**1996**). "Evaluation of the acoustical stiffness of saxophone reeds under playing conditions by using the reactive power approach," J. Acoust. Soc. Am. **100**, 1178–1189.

Dalmont, J. P., Gazengel, B., Gilbert, J., and Kergomard, J. (**1995**). "Some aspects of tuning and clean intonation in reed instruments," Appl. Acoust. **46**, 19–60.

Ducasse, E. (**1990**). "Modélisation d'instruments de musique pour la synthèse sonore: Application aux instruments à vent," J. Phys. (Paris), Colloq. **51**, C2-837–C2-840 (text in French).

Ducasse, E. (**2001**). "Modélisation et simulation dans le domaine temporel d'instruments à vent à anche simple en situation de jeu: Méthodes et modèles," thèse de doctorat, Université du Maine, Le Mans, France (text in French).

Facchinetti, M., Boutillon, X., and Constantinescu, A. (**2003**). "Numerical and experimental modal analysis of the reed and pipe of a clarinet," J. Acoust. Soc. Am. **113**, 2874–2883.

Fletcher, N. H. (**1979**). "Air flow and sound generation in musical wind instruments," Annu. Rev. Fluid Mech. **11**, 123–135.

Fletcher, N. H. (**1993**). "Autonomous vibration of simple pressure-controlled valves in gas flows," J. Acoust. Soc. Am. **93**, 2172–2180.

Fletcher, N. H., and Rossing, T. D. (**1998**). *The Physics of Musical Instruments* (Springer, New York).

Gazengel, B. (**1994**). "Caractérisation objective de la qualité de justesse, de timbre et d'émission des instruments à vent à anche simple," thèse de doctorat, Université du Maine, Le Mans, France (text in French).

Gilbert, J. (**1991**). "Etude des instruments de musique à anche simple: Extension de la méthode d'équilibrage harmonique, rôle de l'inharmonicité des résonances, mesure des grandeurs d'entrée," thèse de 3ème cycle, Université du Maine, Le Mans, France (text in French).

Heinrich, J. M. (**1991**). "Recherches sur les propriétés densitométriques du matériau cane de provence et ses similaires étrangers; relation avec laqualité musicale; étude associée d'une mesure de dureté," Technical report, Ministère de la Culture, Direction de la Musique et de la Danse, Paris, France (text in French).

Hirschberg, A., Van de Laar, R. W. A., Marrou-Maurrières, J. P., and Wijnands, A. P. J. (**1990**). "A quasi stationary model of the air flow in the reed channel of single reed woodwind instruments," Acustica **70**, 146–154.

Hirschberg, A., Gilbert, J., Wijnands, A. P. J., and Houtsma, A. J. M. (**1991**). "Non-linear behaviour of single-reed woodwind musical instruments," Ned. Akoestish Genootshap **107**, 31–43.

Hirschberg, A., Gilbert, J., Wijnands, A. P. J., and Valkering, A. M. C. (**1994**). ''Musical aero-acoustics of the clarinet,'' J. Phys. IV, Colloque C5, Supplément au J. Phys. III **4**, C5-559–568.
Hirschberg, A. (**1995**). ''Aero-acoustics of wind instruments,'' in *Mechanics of Musical Instruments*, edited by A. Hirschberg, J. Kergomard, and G. Weinreich, CISM courses and lectures No. 355 (Springer, Wien).
Ingard, U., and Ising, H. (**1967**). ''Acoustic non-linearity of an orifice,'' J. Acoust. Soc. Am. **42**, 6–17.
Kergomard, J. (**1995**). ''Elementary considerations on reed-instruments oscillations,'' in *Mechanics of Musical Instruments*, edited by A. Hirschberg, J. Kergomard, and G. Weinreich, CISM courses and lectures No. 355 (Springer, Wien).
Kergomard, J., Ollivier, S., and Gilbert, J. (**2000**). ''Calculation of the spectrum of self-sustained oscillators using a variable truncation method,'' Acustica **86**, 685–703.
Marandas, E., Gibiat, V., Besnainou, C., and Grand, N. (**1994**). ''Caractérisation mécanique des anches simples d'instruments à vent,'' 3ième Congrés Français d'Acoustique, Toulouse, Suppl. J. Phys. **III**, 4,C5, 633–636.
Maurin, L. (**1992**). ''Confrontation théorie-expérience des grandeurs d'entrée d'un excitateur à anche simple,'' rapport de DEA d'acoustique appliquée, Université du Maine, Le Mans, France (text in French).
McIntyre, M. E., Schumacher, R. T., and Woodhouse, J. (**1983**). ''On the oscillations of musical instruments,'' J. Acoust. Soc. Am. **74**, 1325–1345.
Msallam, R., Dequidt, S., Caussé, R., and Tassart, S. (**2000**). ''Physical model of the trombone including nonlinear effects. Application to the sound synthesis of loud tones,'' Acustica **86**, 725–736.
Nederveen, C. J. (**1998**). *Acoustical Aspects of Woodwind Instruments* (Northern Illinois University Press, De Kalb, IL).
Obataya, E., and Norimoto, M. (**1999**). ''Acoustic properties of a reed (*Arundo donax L.*) used for the vibrating plate of a clarinet,'' J. Acoust. Soc. Am. **106**, 1106–1110.
Ollivier, S. (**2002**). ''Contribution à l'étude des oscillations des instruments à vent à anche simple: Validation d'un modèle élémentaire,'' thèse de doctorat, Université du Maine, Le Mans, France (text in French).
Ollivier, S., Dalmont, J. P., and Kergomard, J. (**2002a**). ''Idealized models of reed woodwinds. I. Analogy with the bowed string,'' Acta Acust. (Beijing) (to be published).
Ollivier, S., Kergomard, J., and Dalmont, J. P. (**2002b**). ''Idealized models of reed woodwinds. II. On the stability of 'two step' oscillations,'' Acta Acust. (Beijing) (to be published).
OMEGA (**1995**). *Complete Flow and Level Measurement Handbook and Encyclopedia®* (OMEGA).
Saneyoshi, J., Teramura, H., and Yoshikawa, S. (**1987**). ''Feedback oscillations in reed woodwind and brasswind instruments,'' Acustica **62**, 194–210.
Sommerfeld, S. D., and Strong, W. J. (**1988**). ''Simulation of a player-clarinet system,'' J. Acoust. Soc. Am. **83**, 1908–1918.
Stewart, S. E., and Strong, W. J. (**1980**). ''Functional model of a simplified clarinet,'' J. Acoust. Soc. Am. **68**, 109–120.
Thompson, S. C. (**1979**). ''The effect of the reed resonance on woodwind tone production,'' J. Acoust. Soc. Am. **66**, 1299–1307.
Valkering, A. M. C. (**1993**). ''Characterization of a clarinet mouthpiece,'' Vakgroep Transportfysica, Eindhoven University of Technology, Eindhoven.
Van Walstijn, M. (**2002**). ''Discrete-Time Modelling of Brass and Reed Woodwind Instruments with Application to Musical Sound Synthesis,'' Ph.D. thesis, University of Edinburgh, Edinburgh, United Kingdom.
Van Zon, J. (**1989**). ''Stromingsgeïnduceerde klepinstabiliteiten,'' Vakgroep Transportfysica, Eindhoven University of Technology (text in Dutch).
Van Zon, J., Hirschberg, A., Gilbert, J., and Wijnands, A. P. J. (**1990**). ''Flow through the reed channel of a single reed music instrument,'' J. Phys. (Paris), Colloq. (Suppl). **51**, C2 821–824.
Wilson, T. A., and Beavers, G. S. (**1974**). ''Operating modes of the clarinet,'' J. Acoust. Soc. Am. **56**, 653–658.

Cutoff frequencies and cross fingerings in baroque, classical, and modern flutes[a)]

Joe Wolfe[b)] and John Smith[c)]
School of Physics, University of New South Wales, Sydney 2052, Australia

(Received 14 June 2002; accepted for publication 4 August 2003)

Baroque, classical, and modern flutes have successively more and larger tone holes. This paper reports measurements of the standing waves in the bores of instruments representing these three classes. It presents the frequency dependence of propagation of standing waves in lattices of open tone holes and compares these measurements with the cutoff frequency: the frequency at which, in an idealized system, the standing waves propagate without loss in such a lattice. It also reports the dependence of the sound field in the bore of the instrument as a function of both frequency and position along the bore for both simple and ''cross fingerings'' (configurations in which one or more tone holes are closed below an open hole). These measurements show how ''cross fingerings'' produce a longer standing wave, a technique used to produce the nondiatonic notes on instruments with a small number of tone holes closed only by the unaided fingers. They also show why the changes from baroque to classical to modern gave the instruments a louder, brighter sound and a greater range. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1612487]

PACS numbers: 43.75.Ef, 43.75.Qr [NHF]

I. INTRODUCTION

The tone holes of woodwind instruments are used to reduce the effective length of their bore. An open tone hole provides a low inertance shunt between the bore and the external radiation field so, at sufficiently low frequency, the acoustic pressure inside the bore near an open tone hole is small. Consequently the bore behaves, at low frequencies, like a simple tube whose end is a little way beyond the first open tone hole. This extra length or end correction is frequency dependent: at higher frequencies, the impedance of the inertive shunt is larger, and so the standing wave in the bore propagates past the first open tone hole with an increasing relative amplitude as the frequency increases. As the relative amplitude of the standing wave propagating beyond the first open tone hole increases, the length of the end correction increasingly depends upon whether tone holes further down the bore are open or closed. This allows what musicians call cross fingering: the closing of tone holes down stream from the first open tone hole so as to change (usually to flatten) the pitch of the note played. In older instruments, cross fingerings are used in all registers. In modern orchestral woodwinds, there is a tone hole for each semitone interval and cross fingerings are principally used in the high registers, although they are also used in the other registers for pitch and timbre adjustments and for contemporary techniques such as multiphonics and microtones.

The tone holes of woodwind instrument became larger, in relation to the bore, from the baroque to classical to modern periods. This change was particularly pronounced in the flute. The larger tone holes had several effects: they made the instruments louder and brighter in timbre and they allowed them to play higher notes. They also eliminated or reduced the use of cross fingering in all but the highest registers. These effects are explained below.

Many aspects of the acoustics of wind instruments are well understood, and reviews are given by Fletcher and Rossing (1998), Nederveen (1998) and others. The effects of individual tone holes have been studied in detail theoretically (Dubos *et al.*, 1999; Keefe, 1982a; Strong *et al.*, 1985; Nederveen *et al.*, 1998) and experimentally (Coltman, 1979; Keefe, 1982b). Benade (1960, 1976) derives approximate theoretical expressions for the cutoff frequency of an array of open tone holes: the frequency above which the standing waves propagate significantly past the first open tone hole (see Sec. III). He also explains how the frequency-dependent propagation past a single open tone hole allows cross fingering to flatten the pitch, and why the effect is greater in the second register than in the first.

Figure 1 is a sketch of two of the standing waves in the spirit of Benade's description. [Figure 1(a) combines features of Figs. 21.1, 21.10 and 22.12 in Benade (1976).] The pressure of the standing wave does not fall to zero at the first open hole, because the inertance of the air in the open hole is not zero. Rather, it penetrates into the lattice, where it is attenuated along the bore. Extrapolating the standing wave past the open hole gives the end effect. Because the impedance of the open tone hole is greater at higher frequency, the standing wave pressure at the open hole is greater, so higher modes propagate further into the lattice and so have a larger end effect. Figure 1(b) is an analogous sketch for a cross fingering, showing the larger end effect due to the lower attenuation under the closed tone holes.

Benade's explanations are in good qualitative agreement with the observed effects on real instruments. However, we know of no experimental studies of the end effects in musical instruments that examine the standing waves as functions of both position and frequency.

The purpose of the current study is to measure these standing waves in examples of flutes from different eras and

[a)]This work has not previously been submitted for publication. A brief account was presented at the International Congress on Acoustics, Rome (2001), ''Tone holes and cross fingerings in wood wind instruments,'' J. Smith and J. Wolfe, Session 8.09, pp. 14–15.

[b)]Electronic mail: j.wolfe@unsw.edu.au
[c)]Electronic mail: jrs@newt.phys.unsw.edu.au

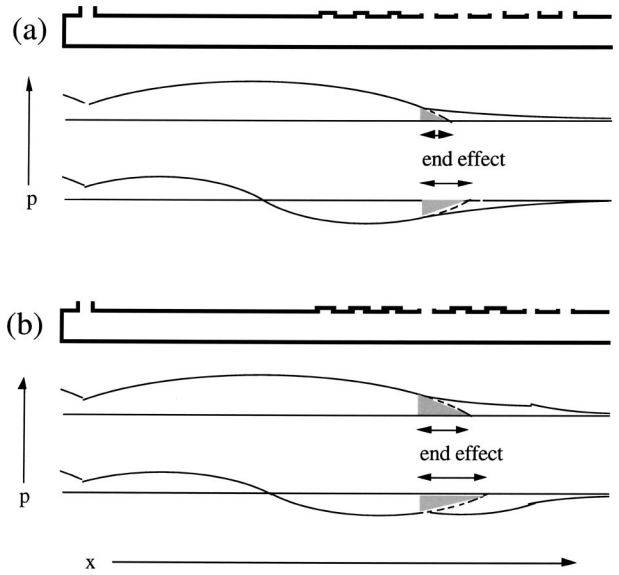


FIG. 1. (a) is a sketch, after Benade (1976), of the acoustic pressure p of the standing waves for the first two harmonics in the bore of a flute for a simple fingering. (b) shows an analogous sketch for a cross fingering. The dotted line indicates the standing wave in a simple tube with a length that gives the same resonant frequency as the flute with this fingering.

in particular to measure their propagation beyond the first open tone hole, and how it varies with frequency. We also measure the effects of cross fingering and show some of the effects produced when the size of tone holes is changed. We chose flutes from the baroque, classical, and modern eras. In the modern flute, the ratio of tone hole diameter to bore diameter approaches the upper practical limit. In the baroque flute, the diameter ratio is smaller by a factor of nearly 2. The classical instrument has intermediate parameters. For the three instruments studied, acoustic impedance spectra measured at the embouchure and sound files and spectra of played notes have been published previously (Wolfe *et al.*, 2001a).

II. MATERIALS AND METHODS

A. The instruments

The modern flute is a production line instrument (Pearl PF-661, open hole, C foot), the same instrument studied previously (Smith *et al.*, 1997; Wolfe *et al.*, 2001b). The classical and baroque flutes were reproductions made by T. McGee of Canberra, Australia. The classical instrument is based on a large-hole Rudall and Rose (R&R #655 from the Bate collection in Oxford) but has been rescaled to play at A $=440$ Hz. The baroque flute is an unscaled replica of an instrument made by J. A. Crone in Leipzig in about 1760. It plays at A$=415$ Hz. The measurements were made at room temperature and humidity (for the different instruments, measured on different days, this varied from 21 °C to 26 °C and 53% to 58% relative humidity, values lower than those in an instrument under playing conditions). The modern Boehm flute is nearly cylindrical, but the head joint tapers to be slightly narrower at the embouchure end. The classical and baroque flutes have cylindrical head joints and approximately conical bodies, narrowing towards the foot. The cone angle is greater in the baroque flute, so its bore is on average smaller than that of the classical instrument, which is in turn narrower than that of the modern flute. Table I gives some of the dimensions of the three instruments.

B. Measurement of impedance and standing waves

The measurements of acoustic impedance were made as described previously (Smith *et al.*, 1997; Wolfe *et al.*, 2001b). The reference for calibration was a semi-infinite cylindrical waveguide, whose impedance was assumed to be real and to have a value given by $\rho c/S$, where ρ is the density of air, c is the speed of sound and S is the cross-sectional area. A compromise is made among frequency range, frequency resolution, and dynamic range. For the experiments measured over the range 0.2 to 3 kHz, the fre-

TABLE I. A table of sizes of holes, positions of their centers, and bore diameters for the three flutes studied. The key naming convention gives numbers to the three long fingers of each hand or the name of the note played when the key is closed. The holes ''tr2'' and ''tr1'' are used for trills and as register holes on the Boehm flute. ''emb'' refers to the embouchure hole, and its stated diameter is that of a circle having the same area.

	Baroque			Classical			Modern		
Hole	Diameter	Bore	Position	Diameter	Bore	Position	Diameter	Bore	Position
cork		18.2	−18.2		19.0	−17.5		15.7	−17.5
emb	8.5	18.2	0	11.1	19.0	0	11.5	17.5	0
tr2							8.0	19.0	201.0
tr1							7.7	19.0	218.3
1	6.6	16.4	229	7.7	17.1	230	7.0	19.0	234.8
B				5.5	16.8	249	13.2	19.0	267.2
A♯							13.2	19.0	286.7
2	6.4	15.8	267	10	16.5	268	13.2	19.0	307.4
3	5.4	15.3	304	7.5	15.8	303	13.2	19.0	330.6
G							13.3	19.0	352.8
F♯							14.2	19.0	377.3
1	5.5	14.6	363	9.0	14.6	361	14.2	19.0	402.9
2	5.6	14.0	398	10.9	14.0	393	14.2	19.0	430.3
E				8.0	13.9	412			
3	4.5	13.3	435	6.4	13.1	429	14.2	19.0	459.0
D	6.2	12.1	497	11.6	12.4	478	15.5	19.0	490.6
C♯				12.3	11.7	514	15.5	19.0	524.2
C				9.7	11.0	548	15.5	19.0	557.8

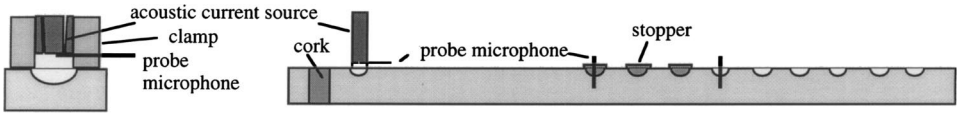


FIG. 2. A schematic of the technique we used to measure the standing waves, with a broad band acoustic current source and probe microphones that can be placed at the embouchure or at closed or open tone holes. The details at the embouchure are enlarged in the sketch at left.

quency spacing was 2.5 Hz. For those over the range 0.2 to 12.5 kHz, the frequency spacing was 10 Hz.

The calibration procedure for measurements of the standing wave in the bore was as follows (Fig. 2). An acoustic current was synthesized from frequency components ranging from 0.2 to 3 kHz in 2.5-Hz increments. This was input to the embouchure hole via a short pipe of diameter 7.8 mm with a length (6 mm) that was chosen so that its impedance approximates the radiation impedance that normally loads the instrument at this point when it is played (Smith *et al.*, 1997; Wolfe and Smith, 2001). The effect of this impedance may be removed using the transfer matrix for a cylindrical waveguide, and this has been done for the high-frequency measurements reported in Fig. 3, which aims to explain the shape of the impedance spectrum rather than to predict playing frequencies. For the other figures, it has been included. Fletcher and Rossing (1998) analyzed the momentum of the region in which the air entering from the jet mixes with the air in the bore and concluded that the embouchure radiation load impedance is effectively in series with the impedance of the bore when it acts on the jet. A probe microphone, outer diameter 1 mm, was placed on the upstream side of this short pipe, at the position of the acoustic current source. The pressure spectrum at this point was measured, and the spectrum of the output acoustic current was adjusted so that the pressure signal measured by this microphone was independent of frequency.

To maintain the same geometry after calibration, the original probe microphone located at the embouchure was replaced with a second microphone of the same type. The original probe microphone was now used to measure the sound pressure in the bore via the tone holes. Because we report only ratios of the pressure in the bore to the pressure at the embouchure, and these have been both measured using the same microphone, any intrinsic frequency dependence of the microphone response will cancel.

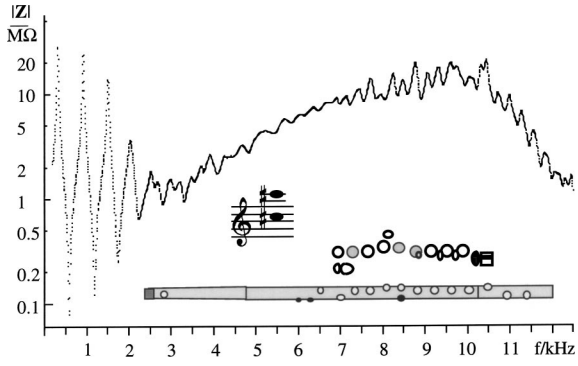


FIG. 3. The input impedance spectrum (in MPa.s.m^{-3} or MΩ) of a modern flute for the fingering used to play C♯5 and C♯6. The upper inset shows a fingering diagram such as a flutist would recognize. The lower inset shows a schematic of the keys. For these notes, all tone holes are open except for two small trill or register keys and one of two similarly placed holes that are used as alternates.

In order to perturb minimally the system under study, we used a probe microphone with a small outer diameter (1 mm). However, the frequency response of a long probe microphone with a small diameter decreases rapidly when its length is increased because of viscous losses. This imposes a short upper limit on the useful length of the probe microphone, and makes it difficult to insert usefully along the axis of the instrument. For this reason, measurements were only taken at tone holes. The tone holes are separated by a few cm, which is rather smaller than the wavelengths studied, so this spacing is adequate for the study. When the tone holes were closed (either by stoppers made to size, or by keys with central holes into which stoppers were inserted), the probe of the microphone was passed through a hole along the axis of the stopper. In all cases, the end of the probe microphone was on the axis of the bore.

C. Measurement of pitch

To measure the pitch change produced by cross fingerings, three experienced flutists were asked to play each of the flutes, with each fingering, in each of two registers. They were told that the fingerings were not standard, and that they played at various pitches between about F4 and G4 in the first register and F5 and G5 in the second. They were asked to blow normally (*mezzo forte*) for a note in the first register and then to blow normally (*mezzo forte*) for a note in the second register. For these measurements, the temperature and humidity of the air were presumably rather higher than ambient, but this was not measured. For each flutist/fingering/register combination, three examples were recorded using a digital tape recorder. The fundamental frequency was determined from a digital Fourier transform, with windows of 2^{16} points, sampling at 44.1 kHz, and averaging over one or more seconds.

III. RESULTS AND DISCUSSION

The physical measurements are made at room temperature and humidity, rather than at the elevated temperatures and humidities of playing conditions, so the speed of sound is expected to be about 1% less than that under playing conditions. The baroque flute is tuned to A=415 Hz, so it plays about one semitone (6%) flatter than the classical and modern instruments. This study, however, is primarily concerned with the relative pitch changes due to cross fingering or between registers, rather than with absolute pitch.

A. Input impedance spectrum of the flute

Figure 3 shows the input impedance spectrum for a modern flute, for a fingering in which all of the tone holes

that are normally used are open. In this plot only, the radiation impedance at the embouchure is not included. This fingering is conventionally used to play C♯5 and C♯6 (554 and 1108 Hz), with the player changing embouchure and jet speed to select the note. It can also play G♯6 (1661 Hz), so it plays the first three notes in a harmonic series, which correspond to the first three minima in $\mathbf{Z}(f)$, approximately equally spaced by 550 Hz. These correspond to resonances of a length of bore about half the length of the flute. This is a little more than the distance to the first open hole, beyond which the standing wave is attenuated strongly. The resonances associated with this harmonic series become weaker with increasing frequency because of increasing viscothermal losses at the walls.

We show in the next section that the calculated cutoff frequency f_c for this instrument is about 2 kHz. At frequencies well above that of the cutoff, the frequency spacing between minima is 270 Hz. These resonances correspond to standing waves along the whole length of the instrument (see especially between 7 and 12 kHz). At such high frequencies, the inertive reactance of the tone holes is so high that the standing wave propagates along the tone hole lattice almost as though the tone holes were closed. This effect is seen more clearly in instruments with lower f_c (discussed later, Figs. 4 and 5).

Above about 3 kHz, the resonances of the bore become rather weak. This can be explained qualitatively by saying that they are effectively in parallel with a Helmholtz resonator formed by the air in the embouchure hole (the mass) and the volume of air that lies between that hole and the sealed end of the flute (the compliance). The expected frequency of this Helmholtz resonator is of the order 3 kHz (the value is approximate because the wavelength is no longer very much larger than the dimensions of the air forming the compliance). The broad maximum between 7 and 11 kHz may correspond to the resonance of the air in the embouchure riser. This is a truncated cone, 5 mm long. A waveguide model for a flanged, truncated cone with this length and the radii of Table I has a broad maximum of about 10 MΩ at around 9.5 kHz.

B. Cutoff frequencies

Benade (1976) applied transmission line theory to the section of bore with open tone holes, and derived the following approximate expression for the ''cutoff'' frequency f_c of this ''open tone-hole lattice:''

$$f_c = 0.110 \frac{b}{a} \cdot \frac{v}{\sqrt{s \cdot t_e}}, \tag{1}$$

where b and a are the radii of the tone hole and the bore respectively, v is the speed of sound, s is half the distance between the centers of adjacent tone holes, and t_e is the effective length of the tone hole, being approximately its geometric length plus 1.5 b. In this model, waves with frequencies below f_c will be attenuated as they propagate along the lattice. However, frequencies above f_c will propagate through the lattice essentially without loss.

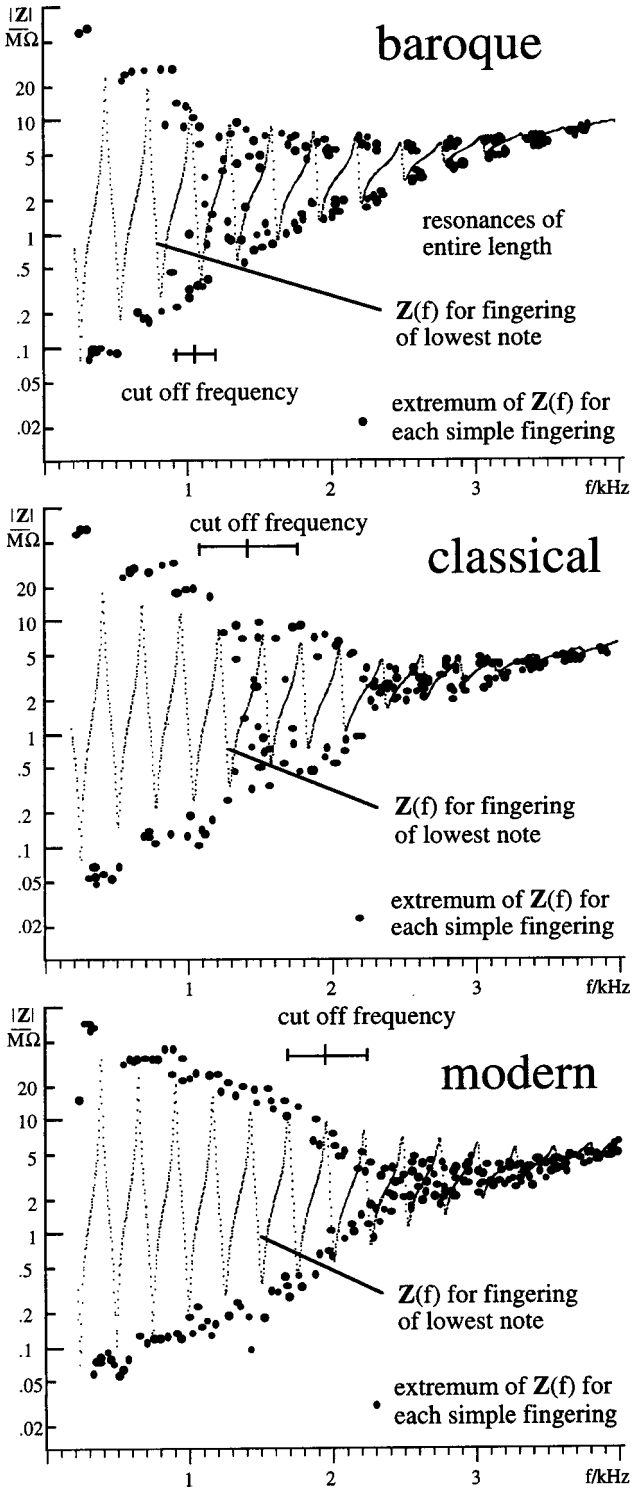


FIG. 4. The effects of the filtering by the open hole lattice, for the three instruments studied. The measured impedance spectrum Z_l of the lowest note is indicated by the faint dots that form an almost continuous line. Superimposed are the extrema of the impedance spectra for all simple fingerings (large dots). Cutoff frequencies calculated from Benade's approximate expression are shown. In practice, the cutoff frequency varies from hole to hole, so means and standard deviations are shown.

The three instruments studied have different size bores and tone holes, so the cutoff frequencies of the lattice of open tone holes are different: lowest for the small-holed baroque flute and highest for the modern instrument. For the baroque and classical flutes, in which adjacent holes usually

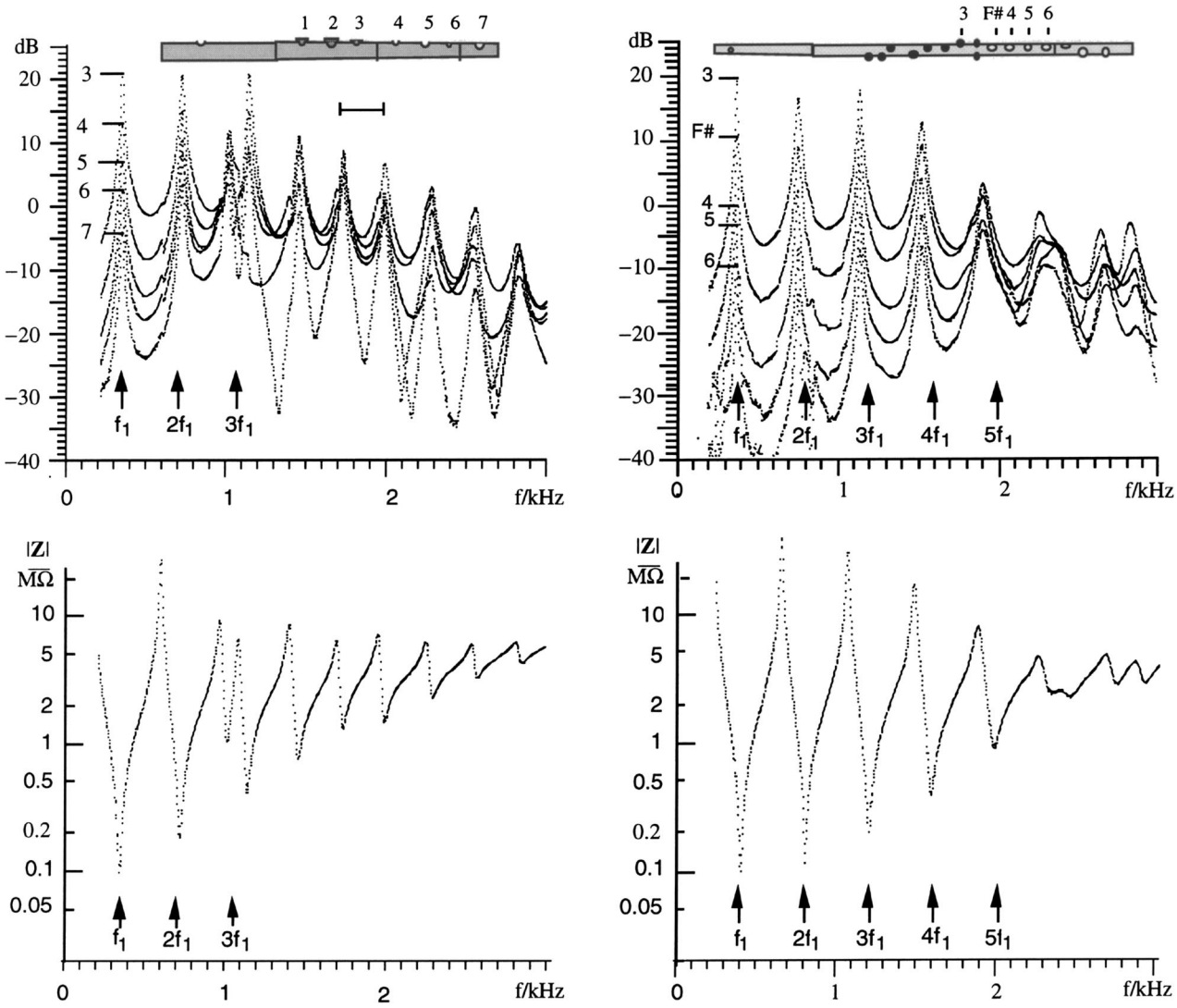


FIG. 5. The upper figures show the sound pressure spectrum in the center of the bore of a baroque flute (left) and a modern instrument (right) for the fingering XXX-OOO for the note G4 or G5. Results are expressed as $p_{\text{bore}}/p_{\text{emb}}$. Different curves were obtained on the bore axis at the positions of different tone holes, as indicated by the numbers on the curves and on the sketch of the instruments. The lower figures show the acoustic impedance spectra measured at the embouchure for these fingerings. The scale bar at the upper right of the figure for the baroque flute indicates the harmonic spacing of resonances of the entire bore. (To allow closer inspection, the individual curves that superpose here are printed separately at http://www.phys.unsw.edu.au/~jw/Crossfingeringfigures.pdf).

have different sizes, the cutoff frequency also depends upon the fingering. Some typical values are given in Table II.

The cutoff frequency predicted by Eq. (1) has been widely and successfully used. Due to approximations in the model Benade used, however, one would expect it to become increasingly imprecise around the cutoff frequency. The transmission line theory used assumes an infinitely long lattice with distributed or continuous components, whereas in a real instrument the tone holes and their separation are of finite size. In the waveguide model the dimensions of the elements in the lattice of open tone holes are always much smaller than the wavelength but this is not the case for the real instrument. Further, the model will not be applicable to sections of bore with a small number of tone holes of different size at higher frequencies. Accordingly we present in the Appendix a derivation of Eq. (1) for tone holes of finite size

TABLE II. Values for the cutoff frequency, calculated using Benade's expression, Eq. (1), and typical values of the relevant parameters. The effective length of the tone hole includes an end effect at each end. For the classical flute, only the tone holes used in the diatonic scale are included in the means. For the modern instrument, all holes except the trill holes are included.

Flute	Bore radius a	Tone hole radius b	Tone hole length t	Effective length t_e	Interhole spacing 2s	Cutoff frequency f_c/Hz
baroque	7.5±1.7	2.8±0.3	5	9.2±1	41±10	1030±300
classical	7.5±1.5	4.3±1.1	5	11.5±1	40±10	1430±500
Boehm	9.5±0	6.7±1.1	2.5	12.6±1.2	27±5	2050±400

and spacing, and some calculations of the impedance and gain of an open tone hole lattice using a simple empirical model.

The cutoff frequency is one of several effects that determine the envelope of the impedance spectra of the flute. For each of the three instruments, the data for all possible simple fingerings (i.e., fingerings that are not cross fingerings) are summarized in Fig. 4. The extrema become successively weaker with increasing frequency, in part because of greater wall losses at high frequency, and in part because the air upstream from the embouchure hole acts as the reservoir of a Helmholtz resonator in parallel with the bore. This is seen most clearly in the impedance spectrum Z_1 for the fingering for the lowest note, in which there are no open tone holes. (The complete impedance spectrum is shown for this fingering only.) For all other simple fingerings, there is a lattice of open tone holes, and the extrema become weak at a frequency near the calculated f_c.

For the baroque and classical flutes, for frequencies above the calculated f_c, the extrema of Z for simple fingerings have an envelope similar to that of Z_1. On the baroque flute, above about 2 kHz, the extrema of simple fingerings tend to cluster near those of Z_1. In other words, for frequencies well above f_c, all fingerings behave approximately as though all holes were closed! This effect is successively less noticeable on the classical and modern flutes, whose tone holes are successively bigger and more numerous, and therefore less negligible, even at high frequency.

C. Standing waves for simple fingerings

To illustrate the effects of cross fingering, we chose one simple and one cross fingering for detailed study. [Measurements for other fingerings are given by Wolfe *et al.* (2001a).] The simple fingering chosen was that for the note G, in which the tone holes or finger holes controlled by the left hand are closed and those for the right hand are open. This is often represented by wind players as XXX-OOO, the characters representing the three largest fingers of the two hands, X being closed and O being open. This choice of fingering allows us to measure the standing waves in both the open and closed parts of the bore, via the finger holes. On all instruments, this is the standard fingering for G4 and G5.

The effect of the cutoff frequency can be seen in both the impedance spectrum and in the standing wave spectra of notes with simple fingerings. Figure 5 shows these spectra for the modern and baroque flutes for the fingering XXX-OOO, used on both instruments for G4 and G5. The impedance spectra are measured at the embouchure hole and include an impedance representing the embouchure radiation impedance, as described above. The standing wave spectra are measured at the tone holes. Only one spectrum measured inside the closed region of the bore is shown, so as not to complicate the figure further.

Below f_c, the flute behaves approximately as an open tube terminated at the position of the first open hole, plus an end correction which increases with frequency because of the greater penetration into the open tone hole lattice by the waves with higher frequency. Below f_c, the impedance minima and the peaks in the standing waves are approximately harmonically related. This is shown in the figures by the arrows, which have been drawn at harmonics of the fundamental frequency. For the baroque flute, the first two resonances fall below f_c. For the modern instrument, the first five resonances fall below f_c. On the modern instrument, the fingering for G4 can be overblown to sound G5, D6, G6, and B6 with intonation error less than the variability among players. The classical flute (data not shown in this figure, but see Figs. 4 and 6) has intermediate cutoff frequency and intermediate behavior: the first four impedance minima are approximately harmonic.

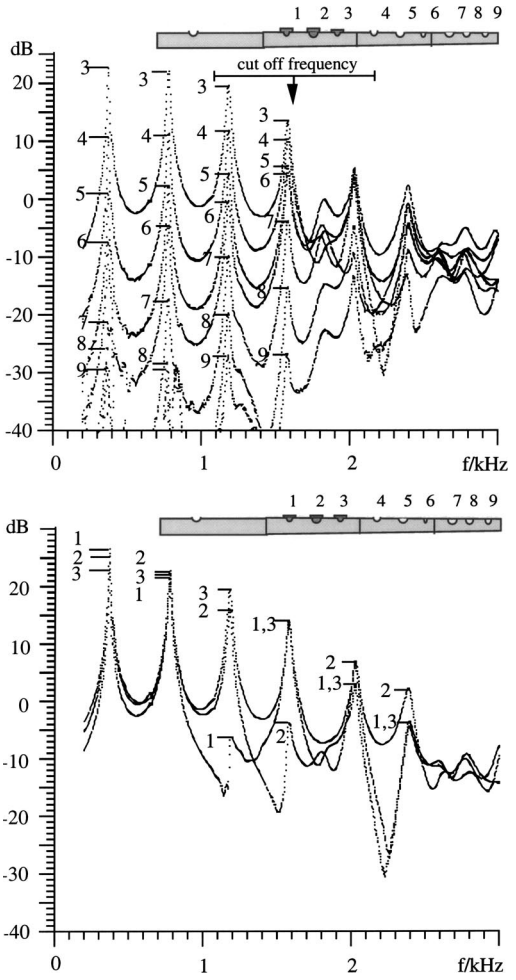


FIG. 6. The sound spectrum in the center of the bore of a classical flute for the fingering XXX-OOO for the notes G4 or G5. Results are expressed as $p_{\text{bore}}/p_{\text{emb}}$, the ratio of measured pressures at positions in the bore to that at the embouchure. Different curves were obtained on the bore axis at the positions of different tone holes, as indicated by the numbers on the curves and the sketch. They therefore represent displacement along the bore, at positions given by Table I. The numbered horizontal dashes are added to show the maxima of the superposed curves.

The increase in cutoff frequency from baroque to classical to modern has important consequences for the sound produced. For most fingerings, the higher f_c increased the number of impedance minima that are in nearly harmonic ratios and thus the number of resonances that interact with harmonics of the air jet to produce the spectrum of the note played. For the same note and dynamic level, the sound spectra of the more recent instruments are richer in higher harmonics [spectra given in Wolfe *et al.* (2001b)]. This

makes them ''brighter'' in timbre, and also considerably louder, because, for most of the range of the flute, the higher harmonics fall in a range where the ear is more sensitive than it is to the fundamental frequency.

At frequencies somewhat higher than f_c, the impedance of the open tone holes becomes sufficiently high that the wave propagates past them all the way down the bore. The resonances in this frequency range are therefore standing waves along the entire length of the instrument. Hence, above about 1.4 kHz for the baroque flute, the standing wave peaks and the impedance minima are approximately equally spaced in frequency, but the spacing is smaller. The average difference in frequency between the last six impedance minima is 278 Hz, which is similar to the frequency of the lowest note of the baroque flute (277 Hz: D4 in baroque tuning with A=415 Hz). For the modern flute at frequencies above about 2.5 kHz, the minima are also more closely spaced, but the resonances are weak at these frequencies (cf. Figs. 3 and 4).

The various features of the standing waves are most clearly seen in the classical flute, which has an intermediate number of tone holes, of intermediate size. Data for the fingering XXX-OOO are shown in Fig. 6. At this stage, we restrict discussion to frequencies below the cutoff, which for this fingering on this instrument is about 1.6 kHz (discussed below).

Figure 6a shows the decay of the standing waves in the lattice of open tone holes. The pressure amplitude falls by several dB from one open hole to the next. This can be analyzed with a model in which the sections of bore are treated as waveguides and the tone holes as discrete elements. Using such a model, and using the typical values in Tables I and II, the attenuation between adjacent open holes is of order 10 dB and depends in a complicated way on frequency (see the Appendix for a calculation). The actual value varies from hole to hole, because of the nonuniform bore and hole size. For instance, the sixth hole is small and the seventh large, so the attenuation from hole 5 to 6 is less than that from hole 6 to 7. For the most distant open holes (8 and 9), the measured attenuation is less strong. The intensity in the bore here is comparable with that of the external sound field. Near and above the cutoff frequency, the spectra are complicated and there is no simple dependence upon position. The simple model fails in this region because the standing waves penetrate further and more closely approach the resonances of the complete tube, as discussed below.

Notice that, at frequencies below the cutoff, the pressure amplitudes of the resonances measured at hole 4 (the first open hole in the lattice) are of approximately equal amplitude. This is because the input acoustic current spectrum has been adjusted to produce equal pressure amplitude at the embouchure hole. At the subsequent open holes (5,6,7), the pressure signal increases with frequency because of greater penetration into the lattice.

In the closed part of the bore [Fig. 6(b)], the behavior below the cutoff frequency has some complications that are explained in terms of the standing waves of the first few resonances. One expects the pressure amplitude to fall monotonically in the last quarter wavelength upstream from the first open hole. This is observed: for the first resonance ($\lambda \cong 870$ mm), the amplitude falls monotonically from holes 1 to 4 (131 mm). For the higher resonances, this length is greater than $\lambda/4$, and so the pressure variation is not monotonic. On this flute, hole 1 can be used as a register hole to produce a note near the third harmonic, D6, and hole 2 can be used as a register hole to produce a note near the fourth harmonic, G6. (A register hole is one that is opened to allow standing waves with nodes near its position, but to disallow lower frequency resonances. It thus selects notes in higher registers, whence the name.) At hole 1 (register hole for the third resonance), the third resonance has a low value of sound pressure and a discontinuity. Similarly at hole 2 (register hole for the fourth resonance), the fourth resonance has a low value of sound pressure and a discontinuity.

Most of the spectra from different positions show strong peaks at the first four harmonics of the note G4, corresponding to standing waves with half wavelengths equal to integral fractions of a length equal to the distance from the embouchure hole to the first open hole plus about 80 mm (i.e., the length of the closed tube, plus end effects for the embouchure and for the tone hole lattice). The frequencies of these peaks correspond, to within 2 Hz, to the minima in the acoustic impedance spectrum measured at the embouchure. The flute can be blown to play notes with these frequencies. Intervening frequencies do not generate strong standing waves with this fingering (and the flute cannot be played at these frequencies without making large changes to the end effects). The good approximation to harmonic ratios may seem surprising, because the end correction in the tone hole lattice is expected to be frequency dependent. The closed air volume upstream from the embouchure and the shape of the bore are responsible for ''correcting'' this frequency dependence, and others related to playing technique (Benade, 1959; 1976; Fletcher and Rossing, 1998).

D. Standing waves for cross fingerings

There is no cross fingering that is equally similar across all three instruments. For the cross fingering, we chose XXX-OXX, which is discussed below. Results for the simple fingering XXX-OOO and the cross fingering XXX-OXX are compared in Fig. 7. The fingering XXX-OXX on the baroque flute plays a sharp F♯4 (for which it is an alternative fingering): i.e., closing the additional holes lowers the pitch by about half of an equal tempered semitone. Overblown, it plays approximately F5, i.e., about two semitones flatter than the XXX-OOO fingering. The note is unstable and difficult to play. On the classical instrument this fingering plays G4 about a quarter of semitone flat, and G5 about half a semitone flatter than the XXX-OOO fingering. The Boehm flute has more tone holes and it also has key linkages that are not present on the other instruments. The fingering XXX-OXX engages a clutch that closes the F♯ key. The cross fingering effect alone lowers the pitch by about 15 and 30 cents in the first and second registers, respectively. (See Table III.) In all flutes, the cross fingering produces notes in the third register in which the open hole operates as a register hole, so considerations of the cross fingering end effect are not relevant.

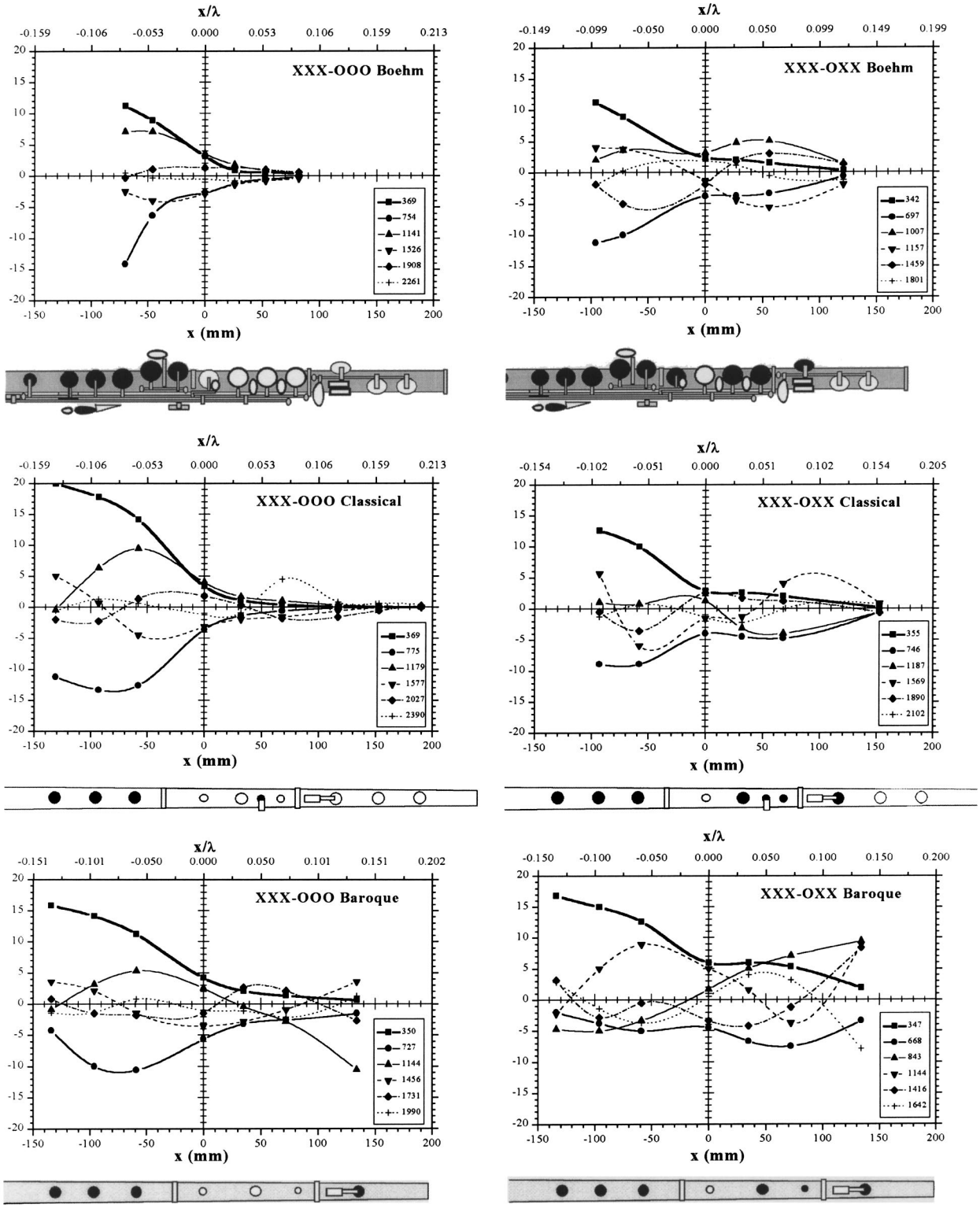


FIG. 7. Acoustic pressure in the bore for simple (left) and cross fingerings. The vertical axis is $p_{\text{bore}}/p_{\text{emb}}$ (linear scale): the ratio of pressure to that measured at the embouchure. Positions are measured with respect to the position of the first open hole and are shown in both mm (lower axis) and in fractions of a wavelength of the fundamental or lowest resonance (upper axis). The figures on the left show the simple fingering XXX-OOO and those on the right show XXX-OXX. The curve with the squares and heaviest line weight is that of resonance with the lowest frequency. The other curves are for next five resonances in increasing frequency. (The frequencies in Hz are given in the inserts.) The sketches of the instruments are drawn to scale in the horizontal direction so that the holes used to make the measurement line up with the data. Lines between data are a guide for the eye only.

The pressure at the peaks of the spectra measured at the tone holes (curves including those in Figs. 5 and 6) is plotted in Fig. 7 as a function of the position of the holes through which they were measured. The first few resonances—those falling below the cutoff frequency for each instrument—are very nearly in a harmonic series (two harmonics for the baroque, four for the classical, and five for the modern). For the simple fingering XXX-OOO, the shape of the fundamental

TABLE III. The pitch change, in cents± standard deviation, due to the cross fingerings shown in Fig. 7. One equal tempered semitone=100 cents.

	Baroque	Classical	Modern
First resonance	−50±20	−25±5	−15±5
Second resonance	−215±20	−55±5	−30±5

wave is simple: it decreases monotonically with distance along the bore, and is very small in the lattice of open tone holes. Attenuation in the open tone hole lattice is greatest in the modern instrument, which has larger holes, and least in the baroque instrument, as expected. For the modern instrument, the higher resonances also behave much as expected. These resonances have nodes in the closed part of the bore, so these curves are not monotonic: for example, the standing wave for the fourth resonance has a node near the hole at −90 mm [See also Fig. 6(b) and the explanation in the penultimate paragraph of section C.]

For the classical and baroque instruments, the greater penetration into the open keyhole lattice results in a complicated pattern, even for the simple fingering. For the baroque instrument, the fifth and higher resonances approximately resemble harmonic standing waves over the full length of the tube [as shown in Fig. 6(b)]. For the baroque flute, the pressure at the last hole downstream is comparable with that in the field outside the instrument.

For the cross fingering, the standing wave of the lowest resonance is substantially terminated by a single open hole in all instruments: the wave falls to a small amplitude at the open hole and its amplitude decreases almost monotonically in the downstream closed section. These curves resemble the sketch given by Benade (1976) to explain the effects of cross fingering.

For the modern flute, the second resonance also falls monotonically beyond the first open hole, but the third, fourth, and fifth all show local maxima inside the downstream tube. The third resonance plays a slightly flat D6 (with a little difficulty), the fourth resonance a sharp F♯6. For these cases, the open hole may be considered to act as a register hole. When the open hole is closed, the fingering is that for D4, whose fourth and fifth harmonics are near D6 and F♯6. In the cross fingering shown, the open hole is between one-quarter and one-fifth of the way along the pipe and thus the third and fourth standing waves of the cross fingering are closely related to the fourth and fifth resonances of the fingering for D4. All of these resonances may be played with the cross fingering, although the third is not stable.

For the classical and baroque flutes, all standing waves except the first increase in amplitude downstream from the open hole, and the third and higher resonances are similar to the fourth and higher harmonics of the D4 fingering: for these notes the open hole is sufficiently close to a pressure node and/or its inertive reactance is sufficiently high that its being open makes relatively little difference to the standing wave.

IV. CONCLUSIONS

The cutoff frequency of the instruments increases, as expected, with increasing tone hole diameter. Above the cut-

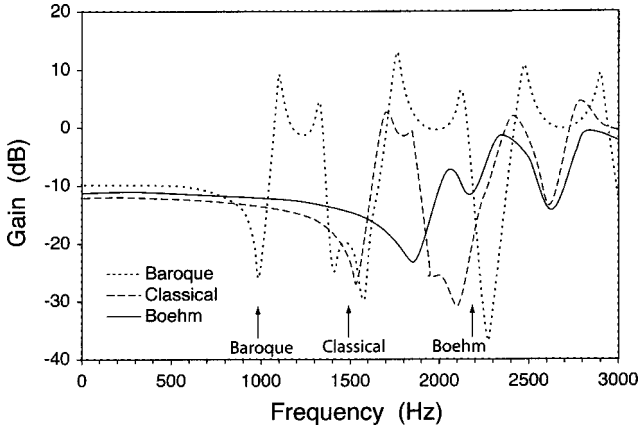


FIG. 8. The theoretical variation of the gain at the start of the tone hole lattice as a function of frequency for three different types of flute. This gain is defined as the ratio of the internal pressure at the second open tone hole to that at the first open tone hole. In this and in Fig. 9, the vertical arrows indicate the average value of the cutoff frequency calculated using dimensions for the first two open tone holes, for each instrument.

off frequency, the standing waves penetrate strongly past the open tone hole. In the case of the instrument with the smallest holes (baroque flute), the standing waves above the cutoff frequency are little affected by the open holes and are close to the expected resonances for the complete length of the instrument, with all tone holes closed. With larger holes, the increasing cutoff frequency gives a larger number of impedance minima that are in approximately harmonic ratio. This is expected to contribute to the production of notes that are brighter in timbre and louder.

Below the cutoff frequency, the standing wave propagates past the first tone hole by an amount that increases with frequency and that decreases with increasing tone hole diameter. For both the cutoff frequency and the attenuation in the tone hole lattice, the observed values are in agreement with those of a simple empirical model. The observed behavior of the standing waves explains the observed effects of cross

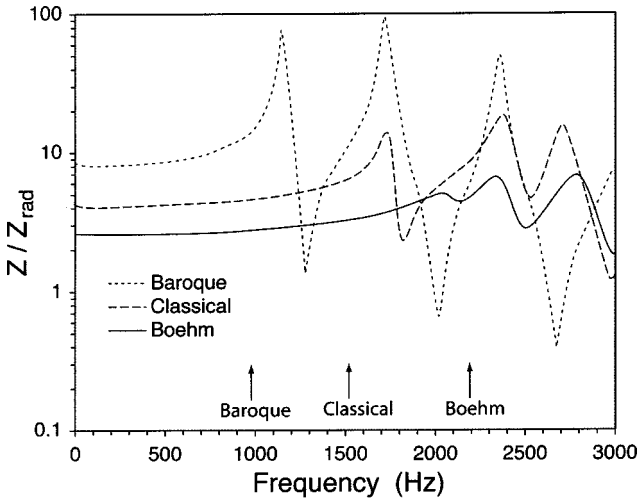


FIG. 9. The theoretical variation of the impedance downstream of the first open tone hole as a function of frequency for three different types of flute. Each curve has been normalized with respect to Z_{rad}, the radiation impedance for an unbaffled aperture with a diameter equal to that of the bore. Thus Z_{rad} is the impedance that would be measured if the flute were physically "cutoff."

fingering on the playing frequency and the impedance spectra for the different instruments and different registers.

ACKNOWLEDGMENTS

We thank Terry McGee, Mark O'Connor, and Matthew Ridley for lending instruments, Attila Stopic and John Tann for technical assistance, and Neville Fletcher for comments on the manuscript. This work was supported by the Australian Research Council.

APPENDIX: THE CHARACTERISTIC FREQUENCY OF AN INFINITE TONE HOLE LATTICE WITH FINITE COMPONENTS

Two successive open tone holes in an infinite, cylindrical bore are separated by a distance L. Looking downstream from one open tone hole, the acoustic impedance is given by the four terminal expression (Fletcher and Rossing, 1998):

$$Z_{\text{in}} = Z_o \frac{Z_L \cos kL + jZ_o \sin kL}{jZ_L \sin kL + Z_o \cos kL},$$

where Z_o is the characteristic impedance of the bore, Z_L is the load impedance present at the next tone hole and other symbols have their usual meaning. Using lower case z for impedances nondimensionalized by dividing by Z_o, this becomes

$$z_{\text{in}} = \frac{z_L + i \tan kL}{jz_L \tan kL + 1}.$$

If the bore has radius a and the tone hole radius b, then the nondimensional radiation impedance z_{hole} at the tone hole is

$$z_{\text{hole}} = j \frac{b^2}{a^2} \tan kt_e \equiv jR \tan kt_e,$$

where $R \equiv b^2/a^2$ and t_e is the effective length of the tone hole when the radiation impedance is considered.

z_L is the impedance of z_{hole} in parallel with the impedance of the bore beyond the hole. As the array is infinite, the latter impedance is z_{in}, so

$$z_{\text{in}} = \frac{z_{\text{in}} z_{\text{hole}} + j(z_{\text{in}} + z_{\text{hole}}) \tan kL}{jz_{\text{in}} z_{\text{hole}} \tan kL + (z_{\text{in}} + z_{\text{hole}})}.$$

Rearrangement gives a quadratic equation whose solution is

$$z_{\text{in}} = j \tan kL \, \frac{1 \pm \sqrt{1 + 4R \tan kt_e / \tan kL - 4R^2 \tan^2 kt_e}}{2(1 - R \tan kt_e \tan kL)}.$$

If $t_e = 0$, the tone hole is a short circuit and $z_{\text{in}} = j \tan kL$, so we reject the solution with the negative sign.

In the cases of interest, $kt_e \ll 1$. kL is on the order of 1 for the characteristic frequencies, so $\tan kL \sim kL$, so

$$z_{\text{in}} \sim jkL \, \frac{1 + \sqrt{1 + 4(a/b)^2 t_e/L - 4(a/b)^4 k^2 t_e^2}}{2(1 - (a/b)^2 k^2 t_e L)}.$$

Because $t_e < L$, the lower characteristic frequency is defined by $(a/b)^2 k^2 t_e L = 1$, which gives

$$f_c \sim \frac{1}{2\pi} \frac{b}{a} \frac{c}{\sqrt{t_e L}} \cong 0.11 \frac{b}{a} \frac{c}{\sqrt{t_e s}},$$

where $s = L/2$. This is the same expression that Benade derives using transmission line theory.

Both the Benade model and the analysis given above fail at high frequencies. The behavior over the whole frequency range of interest may be determined explicitly using four terminal elements to represent the sections of the bore between tone holes. Calculations were made using an empirical model that relates the geometric parameters of the flute to its measured impedance (Botros *et al.*, 2002). Although this model was originally developed for the modern flute, we have also arbitrarily used it for the classical and baroque flutes, altering the bore and tone hole diameters appropriately. For simplicity the effects of energy losses along the bore have not been included in the results presented here. Figure 8 shows the calculated variation of acoustic pressure from one tone hole to the next. Figure 9 shows the calculated frequency dependence of the lattice. In the case of the baroque flute, the acoustic impedance above the cutoff frequency shows a regular series of maxima and minima, rather similar to that of a simple pipe. As observed above, in relation to Figs. 4 and 5, this is because the tone holes behave as though they were closed at these frequencies and so the downstream impedance does indeed approximate that of a simple pipe.

Benade, A. H. (**1959**). "On woodwind instrument bores," J. Acoust. Soc. Am. **31**, 137–146.

Benade, A. H. (**1960**). "On the mathematical theory of woodwind finger holes," J. Acoust. Soc. Am. **32**, 1591–1608.

Benade, A. H. (**1976**). *Fundamentals of Musical Acoustics* (Oxford U.P., New York), pp. 430–501.

Botros, A., Smith, J., and Wolfe, J. (**2002**). "The virtual Boehm flute—a web service that predicts multiphonics, microtones and alternative fingerings," Acoust. Austral. **30**, 61–65.

Coltmann, J. W. (**1979**). "Acoustical analysis of the Boehm flute," J. Acoust. Soc. Am. **65**, 499–506.

Dubos, V., Kergomard, J., Khettabi, A., Dalmont, J-P., Keefe, D. H., and Nederveen, C. J. (**1999**). "Theory of sound propagation in a duct with a branched tube using modal decomposition," Acustica **85**, 153–169.

Fletcher, N. H., and Rossing, T. D. (**1998**). *The Physics of Musical Instruments* (Springer-Verlag, New York), pp. 461–548.

Keefe, D. H. (**1982a**). "Theory of the single woodwind tone hole," J. Acoust. Soc. Am. **72**, 676–687.

Keefe, D. H. (**1982b**). "Experiments on the single woodwind tone hole," J. Acoust. Soc. Am. **72**, 688–699.

Nederveen, C. J. (**1998**). *Acoustical Aspects of Wind Instruments* (Northern Illinois Univ., DeKalb, IL), pp. 45–78.

Nederveen, C. J., Jansen, J. K. M., and van Hassel, R. R. (**1998**). "Corrections for woodwind tone-hole calculations," Acustica **84**, 957–966.

Smith, J. R., Henrich, N., and Wolfe, J. (**1997**). "The acoustic impedance of the Bœhm flute: standard and some non-standard fingerings," Proc. Inst. Acoust. **19**, 315–320.

Strong, W. J., Fletcher, N. H., and Silk, R. K. (**1985**). "Numerical calculation of flute impedances and standing waves," J. Acoust. Soc. Am. **77**, 2166–2172.

Wolfe, J., and Smith, J. (**2001**). Embouchures and end effects in air-jet instruments, Proc. International Congress on Acoustics, Rome, Session 8.06, pp. 8–9.

Wolfe, J., Smith, J., Tann, J., and Fletcher, N. H. (**2001a**). "Acoustics of baroque, classical and modern flutes: a compendium of impedance spectra, sound spectra, sounds and fingerings," JSV+. Electronic publication at http://journals.harcourt-international.com/journals/jsv/supplementary/suppindex.htm

Wolfe, J., Smith, J., Tann, J., and Fletcher, N. H. (**2001b**). "Acoustic impedance of classical and modern flutes," J. Sound Vib. **243**, 127–144.

Measurement and reproduction accuracy of computer-controlled grand pianos

Werner Goebl
Austrian Research Institute for Artificial Intelligence (ÖFAI), Freyung 6/6, 1010 Vienna, Austria and Department of Speech, Music, and Hearing (TMH), Royal Institute of Technology (KTH), Drottning Kristinas väg 31, 10044 Stockholm, Sweden

Roberto Bresin
Department of Speech, Music, and Hearing (TMH), Royal Institute of Technology (KTH), Drottning Kristinas väg 31, 10044 Stockholm, Sweden

(Received 23 February 2003; revised 2 July 2003; accepted 14 July 2003)

The recording and reproducing capabilities of a Yamaha Disklavier grand piano and a Bösendorfer SE290 computer-controlled grand piano were tested, with the goal of examining their reliability for performance research. An experimental setup consisting of accelerometers and a calibrated microphone was used to capture key and hammer movements, as well as the acoustic signal. Five selected keys were played by pianists with two types of touch (''*staccato*'' and ''*legato*''). Timing and dynamic differences between the original performance, the corresponding MIDI file recorded by the computer-controlled pianos, and its reproduction were analyzed. The two devices performed quite differently with respect to timing and dynamic accuracy. The Disklavier's onset capturing was slightly more precise (±10 ms) than its reproduction (−20 to +30 ms); the Bösendorfer performed generally better, but its timing accuracy was slightly less precise for recording (−10 to 3 ms) than for reproduction (±2 ms). Both devices exhibited a systematic (linear) error in recording over time. In the dynamic dimension, the Bösendorfer showed higher consistency over the whole dynamic range, while the Disklavier performed well only in a wide middle range. Neither device was able to capture or reproduce different types of touch. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1605387]

PACS numbers: 43.75.St, 43.75.Mn, 43.75.Tv [NHF]

I. INTRODUCTION

Current research in expressive music performance mainly deals with piano interpretation because obtaining expressive data from a piano performance is easier than, e.g., from string or wind instruments. Pianists are able to control only a few parameters on their instruments. These are the tone[1] onsets and offsets, the intensity (measured as the final hammer velocity), and the movements of the two pedals.[2] Computer-controlled grand pianos are a practical device to pick up and to measure these expressive parameters and—at the same time—provide a natural and familiar setting for pianists in a recording situation. Two systems are most commonly used in performance research: the Yamaha Disklavier (Behne and Wetekam, 1994; Palmer and Holleran, 1994; Repp, 1995, 1996a, b, c, 1997a; Juslin and Madison, 1999; Bresin and Battel, 2000; Timmers *et al.*, 2000; Riley-Butler, 2001, 2002), and the Bösendorfer SE system (Palmer, 1996; Bresin and Widmer, 2000; Goebl, 2001; Widmer, 2001, 2002, 2003). Some studies made use of various kinds of MIDI keyboards which do not provide a natural playing situation to a classical concert pianist because they have a different tactile and acoustic response (e.g., Palmer, 1989; Repp, 1994).

Both the Disklavier and the SE system are integrated systems (Coenen and Schäfer, 1992), which means that they are permanently built into a modern grand piano. They are based on the same underlying principle. That is, to measure and reproduce movements of the piano action, above all the final speed of the hammer before touching the strings. These devices are not designed for scientific purposes and their precise functionality is unknown or not revealed by the companies. Therefore, exploratory studies on their recording and playback precision are necessary in order to examine the validity of the collected data.

Both devices have sensors at the same places in the piano action (Fig. 1). There is a set of shutters mounted on each of the hammer shanks.[3] This shutter interrupts an infrared light beam at two points just before the hammer hits the strings: the first time approximately 5 mm before hammer–string impact, the second time when the hammer crown just starts to contact the strings. These two points in time yield an estimate of the *final hammer velocity* (FHV). In the case of the Disklavier, no further information about how this data is processed was obtainable. On the Bösendorfer, the time difference between these two trip points is called (by definition) *inverse hammer velocity* (IHV) and is stored as such in the internal file format. Since the counter of this infrared beam is operating at 25.6 kHz, the final hammer velocity (in meters per second) is: FHV=128/IHV (Stahnke, 2000; Goebl, 2001, p. 572). The timing of the trip point closer to the strings is taken as the note onset time which has a resolution of 1.25 ms. It seems that the Disklavier uses the same measuring method for hammer velocity and note onset, but as the company does not distribute any more specific details, this is only speculation. The MIDI files of the Disklavier provided

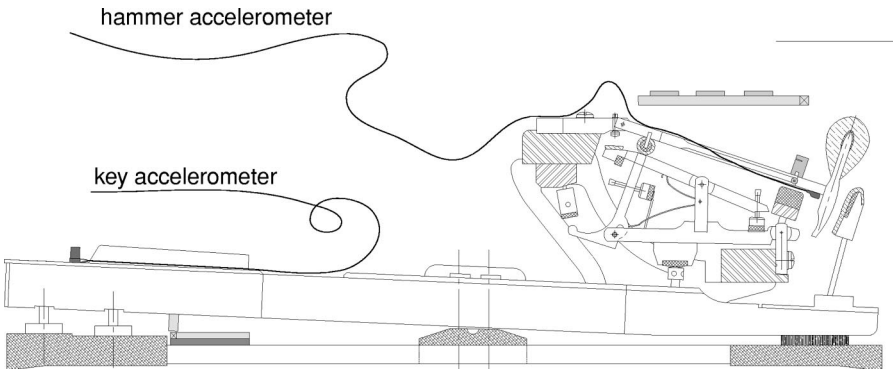


FIG. 1. A Bösendorfer grand piano action with the SE sensors sketched. Additionally, the placement of the two accelerometers are shown. (Figure generated with computer software by the authors. Piano action by Bösendorfer with permission from the company.)

384 MIDI ticks per 512 820 microseconds (as defined in the tempo command in the MIDI file), thus a theoretical timing resolution of 1.34 ms.

A second set of sensors is placed under the keys to measure when the keys are depressed and released. Again, the exact use of this information at the Disklavier cannot be reconstructed, but the Bösendorfer uses this information for releasing the keys correctly (note offsets) and to reproduce silent tones (when the hammer does not reach the strings). The Disklavier used in this study did not reproduce any silent notes at all.

The data picked up by the internal sensors are stored in the Disklavier on an internal floppy drive or externally by using the MIDI out port. The SE system is linked with a special cable plugged into an ISA card of a personal computer running MS DOS. Internal software controls the recording. The information is stored in standard MIDI format on the Disklavier, and in a special file format on the Bösendorfer (each recording comprises a set of three files with the extensions ''.kb'' for keyboard information, ''.lp'' for the loud (right) pedal, and ''.sp'' for the soft (left) pedal). Although the SE file data are encrypted, the content of the files can be listed with the supplied software and used for analysis.

The reproduction is carried out with linear motors (solenoids) placed under the back of each key. The cores of the coils of the Disklavier have a length of approximately 7 cm, whereas those of the SE system are at least double that length or more. Pedal measurement and reproduction is not discussed in the present study.

Only a few studies provide some systematic information about the precise functionality of these devices. Coenen and Schäfer (1992) tested five different reproduction devices (among them a Bösendorfer SE225 and a Yamaha Disklavier grand piano, DG2RE) on various parameters, but their goal was to evaluate their reliability for compositional use; their main focus was therefore on the production mechanism. They determined practical benchmark data like scale speed, note repetition, note density (maximum number of notes which can be played simultaneously), minimum and maximum length of tones, and pedal speed. In their tests, the integrated systems (Disklavier, SE) performed generally more satisfactorily than the systems which are built into an existing piano (Autoklav, Marantz pianocorder). The Bösendorfer, as the most expensive device, had the best results in most of the tasks. Bolzinger (1995) performed some preliminary tests on a Yamaha upright Disklavier (MX-100 A), but his goal was to measure the interdependencies between the pianist's kinematics, performance, and the room acoustics. With his Disklavier, he had the opportunity to play back files and to simultaneously record the movements of the piano with the same device using the MIDI out port. That way, he obtained very easily a production–reproduction matrix of MIDI velocity values, showing a linear reproducing behavior only at MIDI velocity units between approximately 30 and 85 (Bolzinger, 1995, p. 27). On the Disklavier in the present study, this parallel playback and recording was not possible. Maria (1999) developed a complex methodology to perform meticulous tests on a Disklavier (DS6 Pro), but no systematic or quantitative measurements are reported so far.

The focus of this study lies on the recording and reproducing accuracy of two computer-controlled grand pianos with respect to properties of the piano action (hammer–string contact, final hammer velocity), and properties of the sounding piano tone (peak sound-pressure level). In addition to this, we report the correspondence between physical sound properties and their representation as measured by the computer-controlled pianos (MIDI velocity units), in order to provide a benchmark for performance research (see also Palmer and Brown, 1991 and Repp, 1993).

Another issue discussed in the following is the timing behavior of the grand piano action in response to different types of touch and their reproduction by a reproducing piano. Selected keys distributed over the whole range of the keyboard were depressed by pianists with many degrees of force and with two kinds of touch: with the finger resting on the surface of the key (*legato touch*), and with an attack from a certain distance above the keys (*staccato touch*). These different kinds of touch are described in Askenfelt and Jansson (1991).

II. METHOD

A. Material

Two computer-controlled grand pianos were measured in this study.

(i) Yamaha Disklavier grand piano of the Mark II series (DC2IIXG, 173 cm, serial number: 5516392), situated at the Department of Psychology, University of Upp-

sala, Sweden. The Mark II XG series was issued by Yamaha in 1997 (information by Yamaha Germany, Rellingen, personal communication).

(ii) Bösendorfer computer-controlled grand piano (SE290, internal number 290-3, 290 cm), situated at the Bösendorfer company in Vienna, Austria. The *Stahnke Electronics* (SE) system dates back to 1983 (for more information on its development, see Moog and Rhea, 1990), but this particular grand piano was built in 2000. The same system used to be installed in an older grand piano (internal number 19-8974, built in 1986, used, e.g., in Goebl, 2001), but was put into a newer one for reasons of instrumental quality.

Immediately before the experiments, both instruments were tuned, and the piano action and the reproduction unit serviced. In the case of the Disklavier, this procedure was carried out by a specially trained Yamaha piano technician. At the Bösendorfer company, the company's SE technician took care of this work.

B. Equipment

The tested keys were equipped with two accelerometers: one mounted on the key[4] and one on the bottom side of the hammer shank.[5] The accelerometer setting (see Fig. 1) is the same used in Askenfelt and Jansson (1991). Each of the accelerometers was connected with an amplifier[6] with a hardware integrator inside. Thus, their output was velocity in terms of voltage change. A sound-level meter (Ono Sokki LA-210) placed next to the strings of that particular key (approximately 10-cm distance) picked up the sound. The velocities of the key and the hammer as well as the sound were recorded on a multichannel digital audio tape (DAT) recorder (TEAC RD-200 PCM data recorder) with a sampling rate of 10 kHz and a word length of 16 bit. The DAT recordings were transferred onto computer hard disk into multichannel WAV files (with a sampling frequency of 16 kHz).[7] Further evaluation of the recorded data was done in MATLAB programming environment with routines developed for this purpose (by the first author).

C. Calibration

The recordings were preceded by calibration tests in order to be sure about the measured units. The accelerometer amplifiers output ac voltages corresponding to certain measured units (in our case, meters per second) depending on their setting, e.g., 1 V/m/s for the key accelerometer. To calibrate the connection between the TEAC DAT recorder and computer hard disk, different voltages (between -2 and $+2$ V dc) were recorded onto the TEAC recorder and in parallel measured by a volt meter. The recorded dc voltages were transferred to computer hard disk as described above. These values were compared with the values measured by the volt meter. They correlated highly ($R^2=0.9998$), with a factor slightly above 2. The sound recording was calibrated with a 1-kHz test tone produced by a sound-level calibrator.[8]

D. Procedure

Five keys distributed over the whole range of the keyboard were tested: C1 (MIDI note number 24), G2 (43), C4 (60), C5 (72), and G6 (91). The two authors served as pianists to perform the recorded test tones. Each key was hit in as many different dynamic levels (hammer velocities) as possible, in two different kinds of touch: once with the finger resting on the surface of the key (''*legato touch*''), once hitting the key from above (''*staccato touch*''), touching the key already with a certain speed.

Parallel to the accelerometer setting, the grand pianos recorded these test tones with their internal device on computer hard disk (Bösendorfer) or floppy disk (Disklavier). For each of the five keys, both players played in both types of touch 30 to 110 individual tones with interonset time intervals of 1–3 s so that a sufficient amount of data with a large range of different hammer velocities was recorded. Visual feedback of the sound level was provided to the players by the Ono Sokki sound-level meter. Separate MIDI files (or Bösendorfer file triples) were recorded for each key, each type of touch, and for each piano and pianist (5 keys$\times$2 types of touch$\times$2 pianos$\times$2 pianists), containing 964 individual keystrokes for the Yamaha Disklavier and 697 for the Bösendorfer. Immediately after each recording of a particular key and a particular touch by one pianist, the recorded file was reproduced by the grand piano, and the accelerometer data were recorded again onto the multichannel DAT recorder.

This procedure delivered (1) information on timing and dynamics for the *original recording*; (2) the internally stored *MIDI file* of the Disklavier or its correspondent of the SE device; and (3) the precise timing and dynamics for the *reproduction* by the reproducing pianos.

In order to extract discrete data from the hammer and key velocity tracks, several signal processing decisions had to be made.

(1) The *hammer–string contact* was defined as the moment of maximum deceleration of the hammer shank (hammer accelerometer) which corresponded well with the physical onset of the sound, and conceptually with the note onset in the MIDI file. In mathematical terms, the hammer–string contact was the minimum of the first derivative of the measured hammer velocity.

(2) As *hammer velocity*, the maximum hammer velocity (in meters per second) before the hammer–string contact was taken.

(3) An *intensity value* was derived by taking the maximum energy (rms) of the audio signal. The audio channel of each file was calibrated with a 1-kHz pure tone at 94 dB (Brüel & Kjær sound-level calibrator type 4230).

(4) The *MIDI note onset time*, and the *MIDI velocity number* were taken from the MIDI file or the corresponding internal file format of the Bösendorfer.

The onset differences between the original recording and the MIDI file, and those between the original recording and its reproduction were calculated.[9] Since the three measurements (original recording, MIDI file, and reproduction) were not

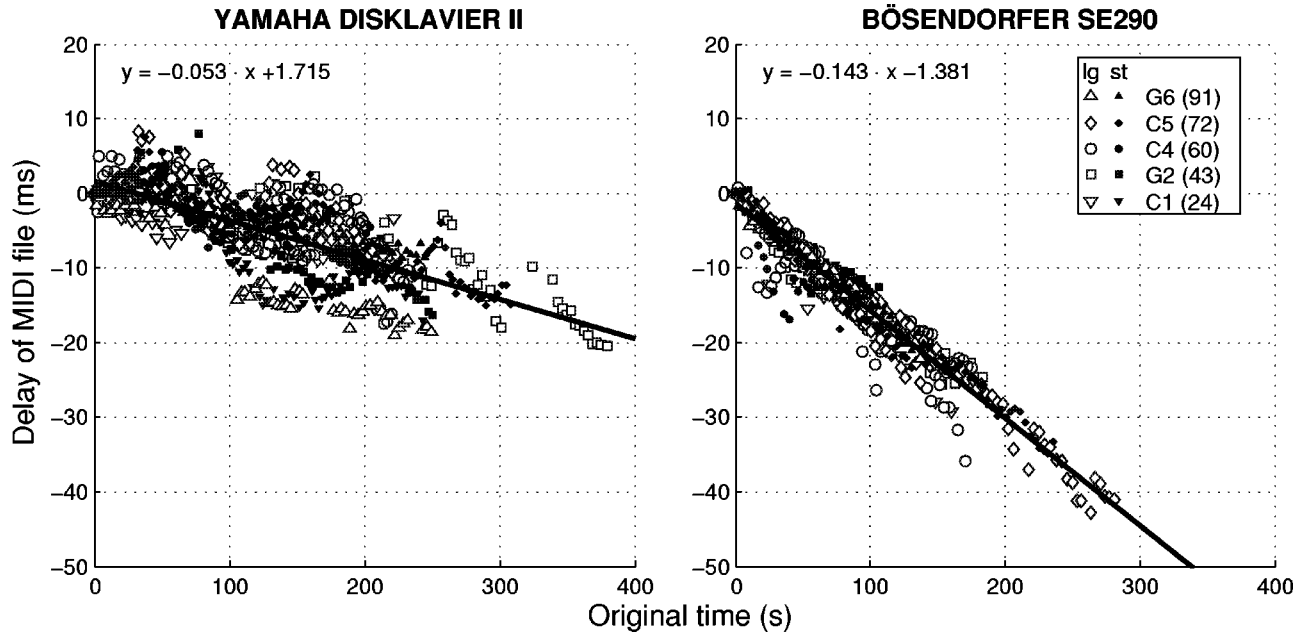


FIG. 2. Timing delays (ms) as a function of recorded time (s) between the original recording and the MIDI file as recorded by the computer-controlled grand pianos for two types of touch: legato (“lg”) and staccato (“st”). Negative values indicate that an onset in the MIDI file was earlier than in the original recording. The straight lines are linear fits of the whole data.

synchronized in time by the measurement procedure, their first attacks were defined as being simultaneous. Care was taken that the first tones always were loud attacks in order to minimize synchronization error, since timing error was smaller the faster (the louder) the attack was. If there was a soft attack at the beginning of a trial, the three files were synchronized by the first occurring louder attack (with hammer velocity over 2 m/s or 77 MIDI velocity units).

III. RESULTS AND DISCUSSION

A. Timing accuracy

In Fig. 2, the note onset delays of the MIDI file in comparison to the original recording are plotted against the recorded time separately for the two pianos.[10] It is evident that both MIDI files showed a constantly decreasing delay over time.

This constant timing error in the MIDI file was larger for the SE system than the Disklavier. The origin of this systematic timing error is unknown, but it is likely that the internal counters of the systems (in the case of the SE system, it is a personal computer) did not operate in exactly the desired frequency, probably due to a rounding error.

This time drift over time was small (0.0053% or 0.014%, respectively) and negligible for performance research (tempo changes of that order are far below just-noticeable differences, cf. Friberg and Sundberg, 1995). But, when such a device has to play in time with, i.e., an audio tape, the synchronization error will already be perceivable after some minutes of performing.

To illustrate the recording accuracy without this systematic error, the residual timing error (the differences between the fitted lines and the data) is plotted in Fig. 3 separately for the two pianos against recorded MIDI velocity.[11] In an earlier conference contribution, a different normalization method was applied on the same data of the Disklavier (see Goebl and Bresin, 2001). The variance was larger for the Disklavier than the SE system (Yamaha mean: 1.4 ms, standard deviation (s.d.): 3.8 ms; Bösendorfer mean: 0.2 ms, s.d.: 2.1 ms), but for both pianos, the residual timing error bore a trend with respect to the loudness of the recorded tones. The Disklavier tended to record softer tones later than louder ones; the SE showed the opposite trend, but to a smaller extent and with much less variation (Fig. 3).

The data in Fig. 3 were approximated by polynomial curves; the formulas are printed there. The R^2 values were different for the two pianos. The Disklavier’s approximation explained barely 40% of the variance, while at the SE system it was about 70%. The Disklavier’s curve fit indicated a larger erroneous trend in recording—in addition to that—it possessed larger variability around that curve.

The timing delays between the original recording and its reproduction are plotted in Fig. 4 separately for the two pianos. The systematic timing error of the recording was not observed, so the display against recorded time (as in Fig. 2) was not required. Evidently, the error in recording was canceled out by the same error in reproduction. The difference between the two systems became most evident in this display. While the reproduced onsets of the Disklavier differed as much as +20 and −28 ms (mean: −0.3 ms, s.d.: 5.5 ms) from the actual played onset, the largest timing error of the SE system rarely exceeded ±3 ms, with a small tendency of soft notes coming up to 5 ms too soon (mean: −0.1 ms, s.d.: 1.3 ms).

Interestingly, the recording accuracy of the SE system was lower than its reproduction accuracy. Obviously, its internal calibration function aimed successfully to absolute precise reproducing capabilities. It could also be that the SE

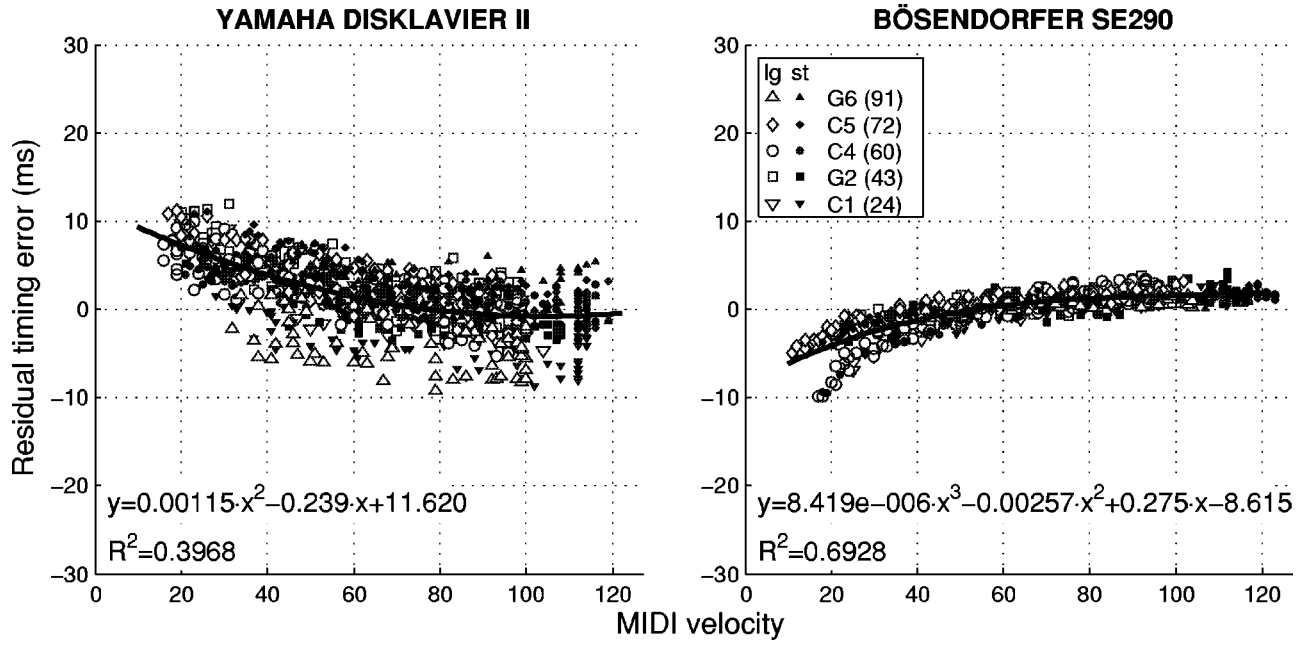


FIG. 3. The residual timing error (ms) between the MIDI file and the original recording as a function of MIDI velocity, as recorded by the computer-controlled pianos. Again, negative values indicate onsets too early in the MIDI data, in comparison to the original file. The trend lines are polynomial functions fitted to the data (as printed in the figures).

takes the first trip point (5 mm before the strings) as being the note onset, but calibrates itself correspondingly to overcome this conceptual mistake. However, this assumption was contradicted by information obtained by the SE's developer, W. Stahnke (Stahnke, 2000; Goebl, 2001).

B. Dynamic accuracy

The second of the investigated parameters is dynamics in terms of the speed of the hammer hitting the strings (m/s) or peak sound-pressure level (dB). We defined the hammer velocity to be the maximum hammer velocity (see above) since it was easy to obtain this value automatically from the recorded hammer track. Usually, this value corresponded very well with the velocity of the hammer when starting to touch the strings (final hammer velocity), but especially for soft notes the maximum hammer speed was larger than the hammer speed at the strings. In this case the time between the escapement (when the hammer loses physical connection to the key, that is, when the jack is catapulted away by the escapement dolly; for more detail see Askenfelt and Jansson, 1990 and Goebl, Bresin, and Galembo, 2003) and hammer–string contact can be as long as 100 ms or more. The actual

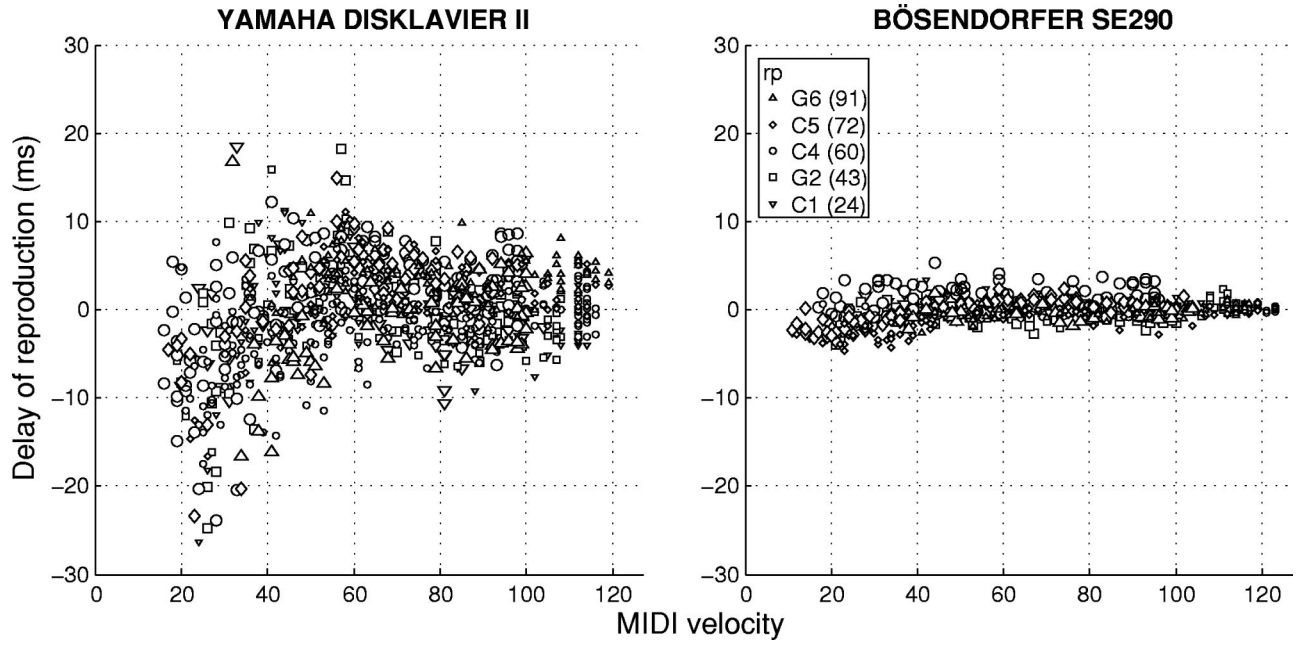


FIG. 4. Timing delays (ms) between the original and its reproduction by the computer-controlled piano. (No systematic trend had to be removed.)

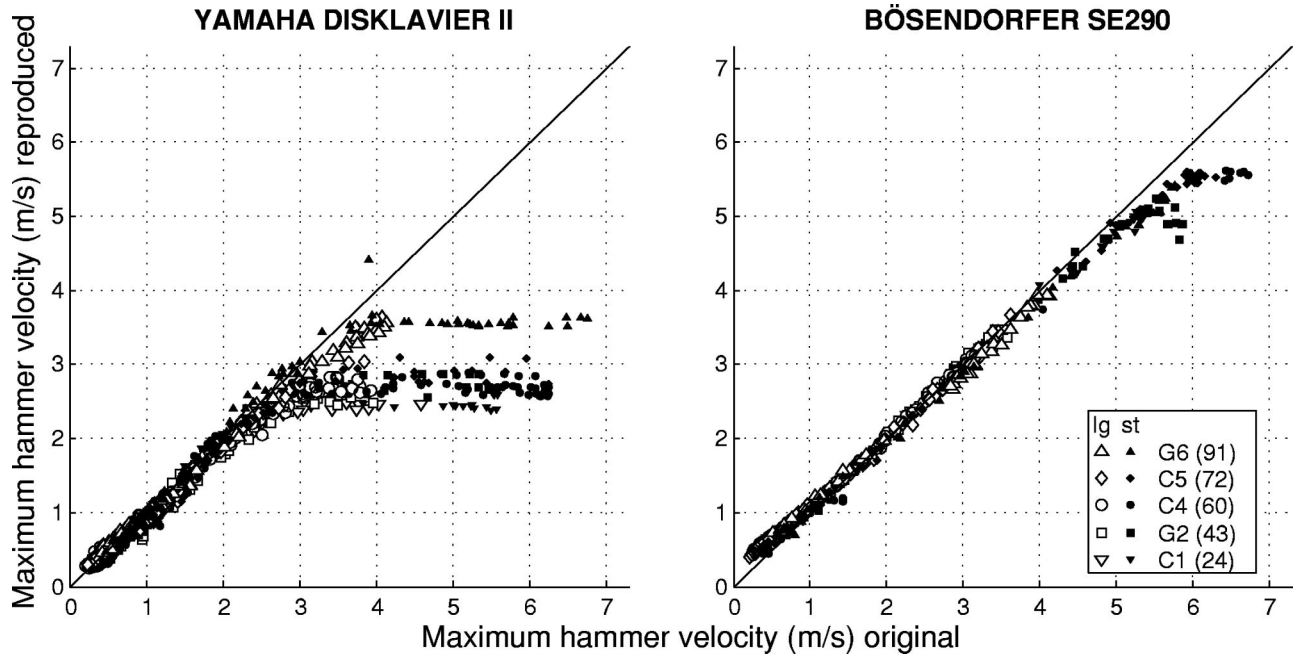


FIG. 5. The maximum hammer velocity (m/s) as played by the pianists (*x* axes) and reproduced by the computer-controlled pianos (*y* axes). (The diagonal line indicates ideal reproduction.)

final hammer velocity was hard to determine from the hammer accelerometer measurements, but the computer-controlled devices measured an average velocity of the last 5 mm of the hammer's travel to the strings (approximately the last 10% of that distance).

In Fig. 5, the reproduced maximum hammer velocity is plotted against the original maximum hammer velocity. It becomes evident that the Disklavier's solenoids were not able to reproduce above a certain hammer speed. This varied slightly between keys, e.g., the G6 (with less hammer mass than hammers at a lower pitch) could be accelerated up to 3.5 m/s, whereas a C1 (with a comparatively heavy hammer) only up to 2.4 m/s. On the SE system, this ceiling effect was not so evident, and there was no obvious effect of pitch as for the Disklavier. Especially in very loud staccato tones, the first impact of the finger hitting the key resulted in a very high-peak hammer velocity which decreases significantly until hammer–string contact. The solenoid was not able to reach this high-peak hammer velocity (and is not programed to do so), but it aimed to reproduce the measured final hammer velocity properly (see also Fig. 8). In this light, the maximum hammer velocity did not seem to be an appropriate measure. Instead, the peak sound-pressure level (dB-SPL) was taken (see Fig. 6).

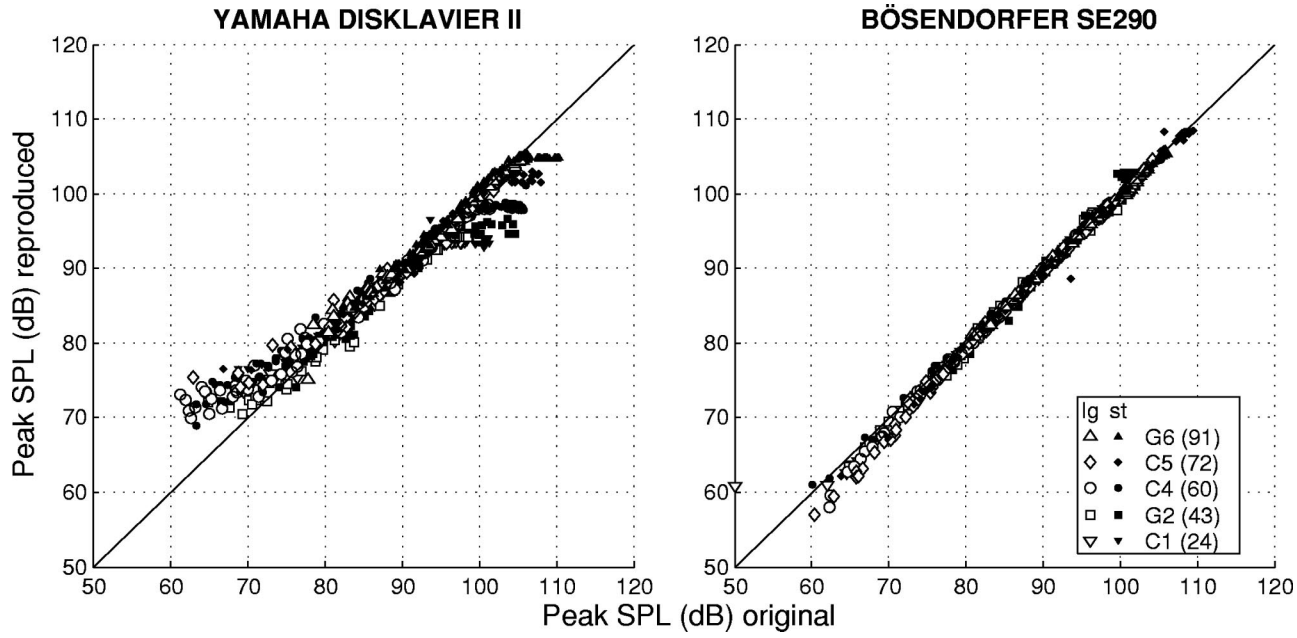


FIG. 6. Peak sound-pressure level (dB) as measured in the tones performed by the pianists (*x* axes) and reproduced by the computer-controlled pianos (*y* axes).

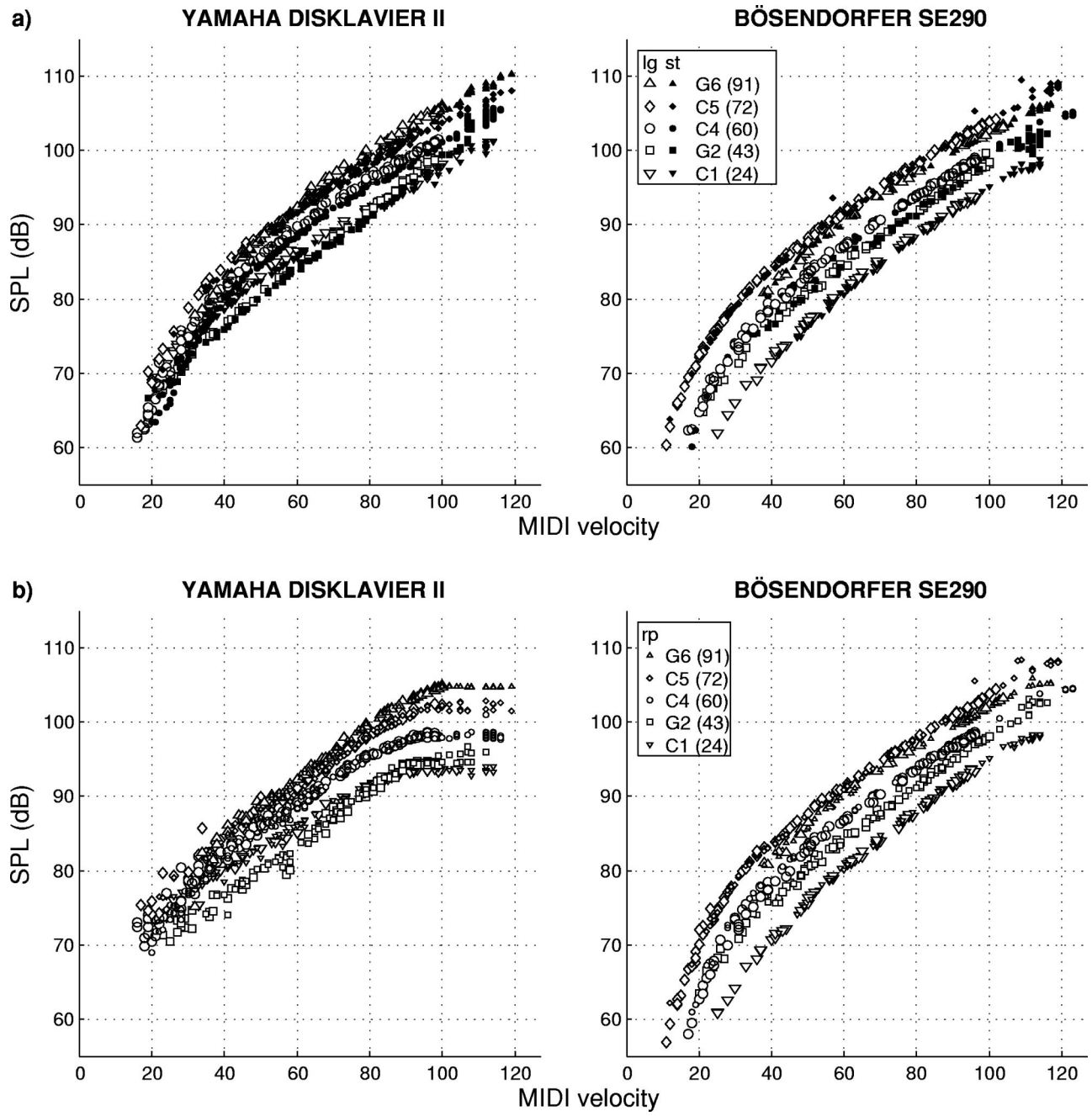


FIG. 7. Peak sound-pressure level (dB) against MIDI velocity as recorded by the computer-controlled pianos. The upper panels show legato touch (''lg''), and staccato touch (''st'') as played by the pianist (a), the lower display the reproduction (''rp'') by the computer-controlled pianos (b).

This display compares acoustic properties of the played tones with their reproduction (peak SPL in dB, Fig. 6). Here, the SE system revealed a much more precise reproducing behavior over the whole dynamic range than the Disklavier. In the latter, the dynamic extremes flattened out, soft tones were played back too loudly, and very loud tones too softly.

In Fig. 7, the relation between MIDI velocity units and peak sound-pressure level is displayed separately for the recording (a) and its reproduction (b). On both instruments, different pitches exhibited a different curve. The higher the pitch, the louder the radiated sound at the same MIDI velocity. The reproduction panel [Fig. 7(b)] reflected the reproducing limitations of the Disklavier already shown in Fig. 6.

C. Two types of touch

Examples of a *legato* keystroke (Disklavier, see Fig. 8) and a *staccato* attack (SE, see Fig. 9) are shown to demonstrate in detail the typical reproducing behavior of the computer-controlled pianos. In these figures, instantaneous key and hammer velocity (first and second row) are plotted above the sound signal (third row). In Fig. 8 on the left side, a *legato* keystroke as played by one of the authors is shown with its smooth acceleration, on the right its reproduction by the Disklavier. The Disklavier hit the key always in a *staccato* manner, with an abrupt acceleration at the beginning of the attack. The parts of the piano action were compressed before their inertia was overcome and the hammer started to

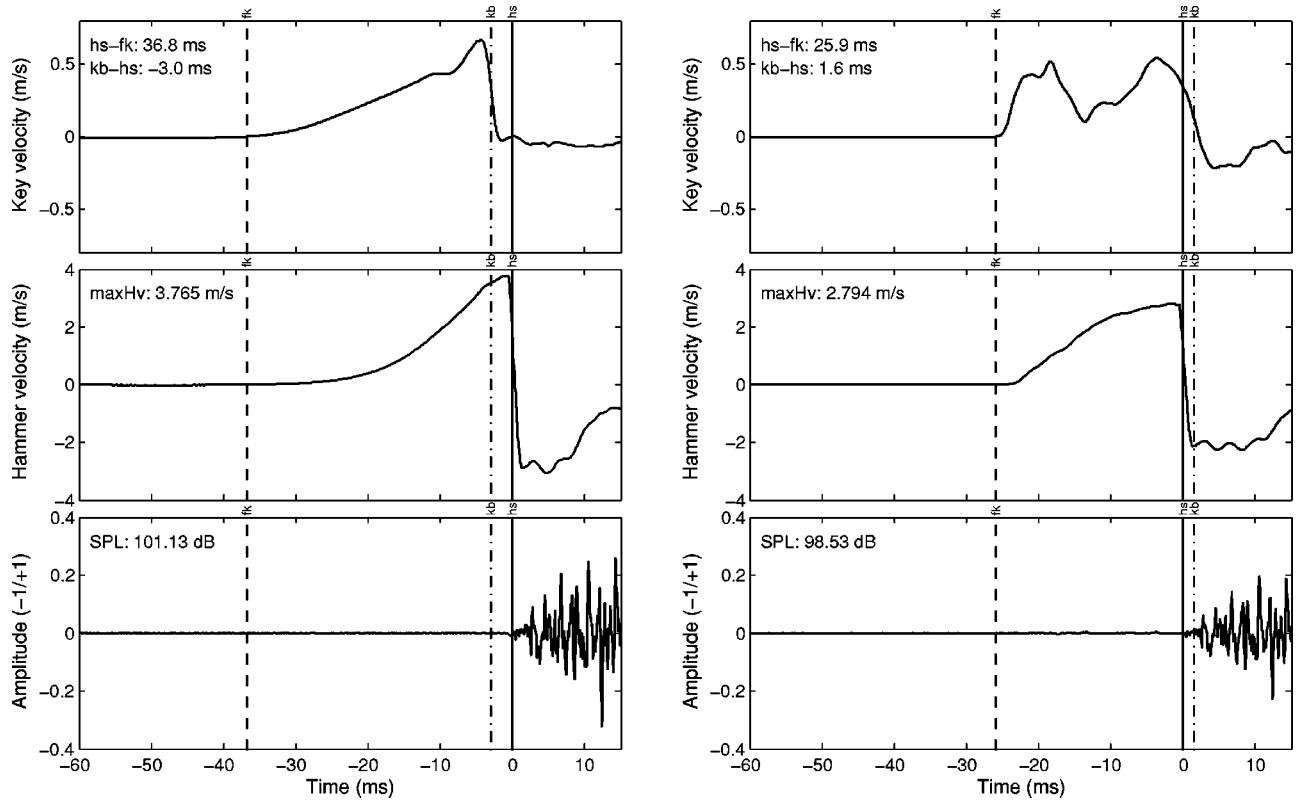


FIG. 8. A *forte* attack (C4, MIDI note number 60) played by one pianist (left panel) ''from the key'' (''*legato touch*''), and its reproduction by the Yamaha Disklavier (right). The upper panels plot key velocity, the middle hammer velocity, the bottom panels the sound signal. The three lines indicate the finger–key contact (start of the key movement, ''fk,'' left dashed line), the key bottom contact (''kb,'' dash-dotted line), and the hammer–string contact (''hs,'' solid line).

move upwards. The solenoid's action resulted in a shorter travel time (the time between finger–key contact (''fk'') and hammer–string contact (''hs'') was 26 ms instead of 37 ms; see Fig. 8, upper panels). The travel time difference between production and reproduction was even larger at very soft keystrokes. This could be one reason why soft notes appeared earlier in the reproduction by the Disklavier than louder notes.

In this particular keystroke, the difference in peak hammer velocity was clearly audible. When the (final) hammer velocities became similar, the two sounds, independently on how they were produced (*legato—staccato*—reproduced) became indistinguishable.[12] We cannot tackle here the controversy as to whether it is only hammer velocity that determines the sound of a single piano tone (White, 1930; Hart, Fuller, and Lusby, 1934; Seashore, 1937) or if there are more influencing factors like various types of noise emerging from the piano action the pianist's interaction with it (Báron and Holló, 1935; Báron, 1958; Podlesak and Lee, 1988; Askenfelt, 1994; Koornhof and van der Walt, 1994).

A very loud staccato attack is plotted in Fig. 9 with the original, human attack on the left, and its reproduction by the Bösendorfer SE on the right. The point of maximum hammer velocity was 5.4 ms before hammer–string contact in the original recording, but only 1.6 ms in the reproduction. Although the reproduced maximum hammer velocity was lower (5.4 m/s instead of 5.8 m/s), the reproduced peak SPL was almost identical with those of the original sound. The human player accelerated the key extremely abruptly so that the hammer reached its highest speed quite some time before hitting the strings and—of course—lost energy at its free flight to the strings. Since the reproducing solenoid cannot accelerate the key in the same abrupt way as the human player, the hammer reached maximum speed later, and—in this example—the machine performed with less energy loss than the human player.

IV. GENERAL DISCUSSION

In this study, we measured the recording and reproducing accuracy of two computer-controlled grand pianos (Yamaha Disklavier, Bösendorfer SE) with an accelerometer setting in order to determine their precision for piano performance research. Both devices showed a systematic timing error over time which was most likely due to a rounding error in the system clock (the internal hardware at the Disklavier, a common personal computer at the SE). This linear error removed, the Bösendorfer had a smaller (residual) timing error than the Disklavier, but both exhibited a certain trend with respect to the loudness of the tones. The Disklavier tended to record soft tones too late, whereas the SE had the tendency to record soft tones too early. But, within these tendencies, the SE was more consistent. At reproduction, the superior performance of the Bösendorfer became more evident: the timing error was smaller than at recording, whereas the Disklavier added some variance in comparison to its recording.

The important point for performance research was the recording accuracy of those systems. Apart from the systematic error that only marginally affected the measured tempo

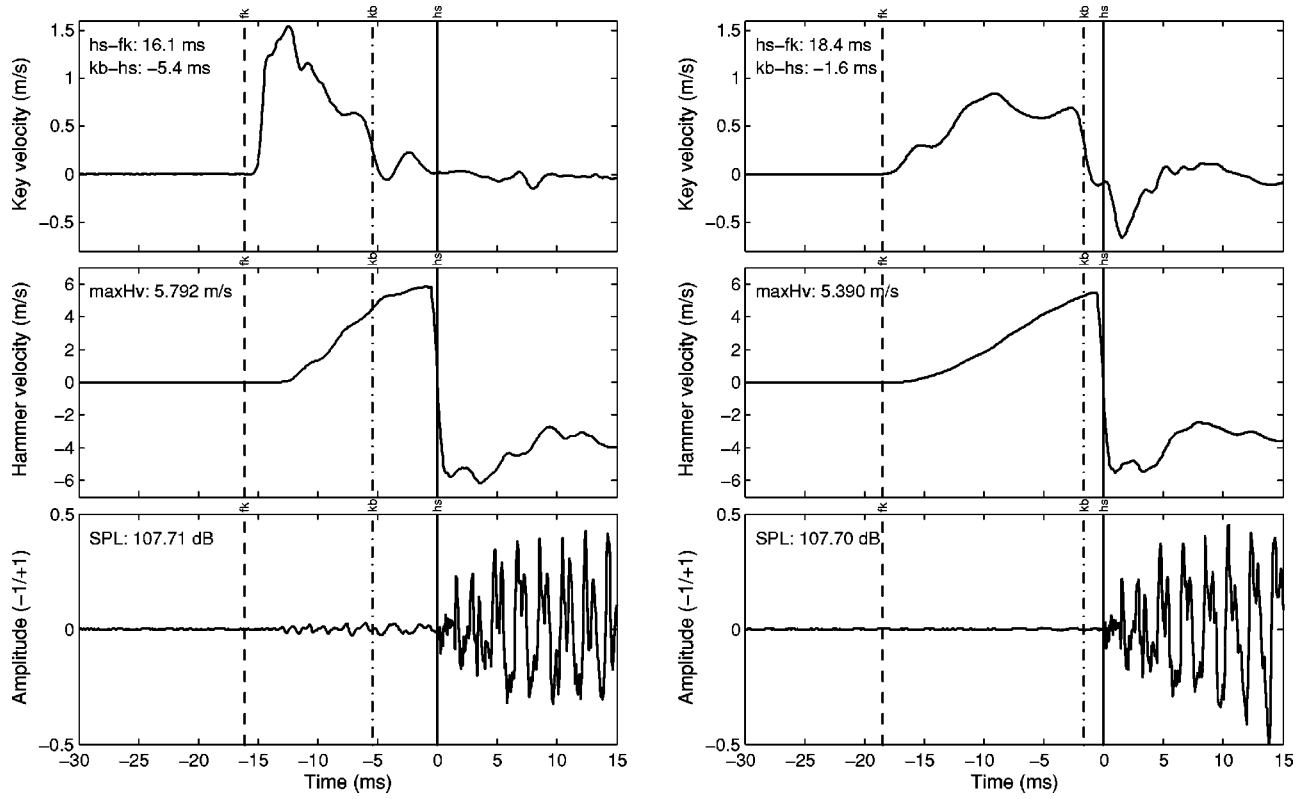


FIG. 9. A *fortissimo* attack (C4, MIDI note number 60) played by one pianist (left panel) from a certain distance above the key (''*staccato touch*''), and its reproduction by the Bösendorfer SE grand piano (right). The upper panels plot key velocity, the middle hammer velocity, the bottom panels the sound signal. The three lines indicate the finger–key contact (start of the key movement, ''fk,'' left dashed line), the key bottom contact (''kb,'' dash-dotted line), and the hammer–string contact (''hs,'' solid line).

value (0.0053% or 0.014%, respectively), the residual timing error (Fig. 3) was considerably large for the Disklavier and smaller for the Bösendorfer. The measurement precision could be improved by subtracting these trends using the polynomial curve approximations as displayed in Fig. 3.

To examine reproducing accuracy in the loudness dimension, we used the maximum hammer velocity and the peak sound-pressure level as measures. Maximum hammer velocity did not correspond to the velocity measures captured by the sensors of the two systems. Considering the peak sound levels of the sounding signal, both devices captured in a similar way, only at reproduction the smaller solenoids of the Disklavier system could not reproduce very loud tones properly. The lower the pitch (and thus the greater the hammer mass), the lower was the maximum sound-pressure level of the Disklavier's reproduction. The reproduction of soft notes was also limited (very soft notes were played back somewhat louder by the Disklavier), because the tested Disklavier prevented very soft tones from being silently reproduced with a minimum velocity matrix, adjustable by the internal control unit. It was also due to this function that the Disklavier was not able to reproduce silent notes, a crucial feature especially for music of the 20th century. The Bösendorfer exhibited linear reproducing behavior over the whole dynamic range (from 60 to 110 dB SPL).

As another, and indeed very important criterion of recording and reproducing capability, we did not investigate the two pedals. (We are talking only of the right and the left pedal of grand pianos, since the middle pedal—the *sostenuto* pedal—only varies the tone length of certain keys depressed during its use, which is recorded and reproduced by simply holding down the corresponding keys at the same time this pedal was depressed.) The use of the right pedal was not investigated extensively up to date (apart from Repp, 1996d, 1997b). We did not have any hypotheses of how pedal recording and reproducing accuracy should be approached. This item remains for future work.

Both the Disklavier and the SE system are based on the same underlying principle: that is, to measure and reproduce movement of the piano action (and the pedals), in particular the final speed of the hammer before touching the strings. This principle is fundamentally different from what a performing artist does when playing expressively. The artist controls finger, hand, and arm movements in order to reproduce a certain mental image of the sound to be produced by continuously listening to the resulting sound and by feeling the hapto-sensory feedback of the keys (Galembo, 1982, 2001). In this way, the performer is able to react to differences in the action, the voicing, the tuning, and the room acoustics, just to mention a few variables that have a certain influence on the radiated sound. On the other hand, a reproducing piano aims to reproduce a certain final hammer velocity independently of whether or not room acoustics, tuning, or voicing changed since the recording. Even if the reproduction takes place on the same piano and immediately after the recording, the tuning might not be the same anymore and the mechanical reproduction, as good as it might be, does not result in an identical sounding performance as

the pianist played it before. This obvious limitation of such devices becomes most evident when a file is played from a different piano or in a different room. Especially, if the damping point (the point of the right pedal where it starts to prevent the strings from freely oscillating) is a different one on another piano, tones in the reproduction will be prolonged (too much pedal) or get cut off (too little pedal) incorrectly.

One possible solution to this problem could be a reproducing device with ''ears,'' in other words, the piano should be able to control its acoustical outcome via a feedback loop through a built-in microphone. If put into a different room, the device could check the room acoustics, its pedal settings, and its current tuning and voicing before the playback starts, much the same as a pianist warming up before a concert. Such a system would require a representation of loudness or timbre other than MIDI velocity, indicating at what relative dynamics a certain note was intended to sound in a pianist's performance.

As the present study was planned to investigate the usefulness of the two devices in question for performance research, we have to consider the obtained results in the light of practical applications. Although the Bösendorfer is the older system, it generally performs better. The disadvantage of the Bösendorfer is its price, around double the price of a grand piano of that size. Moreover, the SE system is not produced anymore, and there are only about 35 exemplars sold around the world, and very few in academic institutions (such as *Ohio State University*, or the *Hochschule für Musik* at Karlsruhe, Germany).[13] On the other hand, the Disklavier is a consumer product, the price level generally cheaper than the Bösendorfer (depending on type of system), and therefore more likely to be obtained by an institution.

The Disklavier measured in this study was certainly not the top model of the Yamaha corporation. Since then, Yamaha issued the Mark III series and the high-end series, called ''Pro'' (e.g., the special ''Pro2000 Disklavier''). The latter series uses an extended MIDI format (with a velocity representation using more than 7 bits), and additional measures like key release velocity to reproduce the way the pianist released a particular key. It can be expected that these newer devices perform significantly better than the tested Mark II grand piano. Since these more sophisticated devices were not available for the authors or too far away from the accelerometer equipment, which was too costly to transport, this has to remain a subject for future investigations.

ACKNOWLEDGMENTS

This study was supported by the European Union (Marie Curie Fellowship, HPMT-GH-00-00119-02, the *Sounding Object* project (SOb), IST-2000-25287, http://www.soundobject.org, and the MOSART IHP network, HPRN-CT-2000-00115) and the START program from the Austrian Federal Ministry for Education, Science, and Culture (Grant No. Y99-INF). The Austrian Research Institute for Artificial Intelligence acknowledges basic financial support from the Austrian Federal Ministry for Education, Science, and Culture. Special thanks are due to Alf Gabrielsson (Department of Psychology, University of Uppsala), who provided a well-maintained Disklavier for experimental use, to the Bösendorfer company, Vienna, for providing the Bösendorfer SE290 grand piano, to the Department of Speech, Music, and Hearing (TMH) of the Royal Institute of Technology (KTH), Stockholm for providing the measurement equipment, and to the Acoustics Research Institute of the Austrian Academy of Sciences for generously providing recording equipment for the test session in Vienna. We are indebted to Tore Persson and Friedrich Lachnit, who maintained and serviced the two reproducing pianos. Finally, many thanks to Anders Askenfelt, Simon Dixon, Alexander Galembo, Erik Jansson, Giampiero Salvi, and two anonymous reviewers for stimulating discussions and helpful comments.

[1]The onset of a sounding tone is very often called ''note onset,'' because of the MIDI world's terminology. In this paper, the terms ''tone'' and ''note'' are used synonymously, since we are not talking about musical notation.

[2]The middle or *sostenuto* pedal only prolongs certain tones and is not counted as an individual expressive parameter.

[3]On the Disklavier, the hammer shutter is mounted closer to the fixed end of the hammer, whereas the SE has its shutter closer to the hammer (as displayed in Fig. 1).

[4]Brüel & Kjær accelerometer type 4393. Mass without cable: 2.4 g; serial number 1190913.

[5]ENDEVCO accelerometer model 22. Mass without cable: 0.14 g; serial number 20845.

[6]Brüel & Kjær charge amplifier type 2635.

[7]Using an analog connection from the TEAC recorder to a multichannel sound card (Producer: Blue Waves, formerly Longhborough Sound Images; model PC/C32 using its four-channel A/D module) on a PC running Windows 2000 operating system.

[8]Brüel & Kjær sound-level calibrator type 4230, test tone: 94 dB, 1 kHz.

[9]$delay_{\mathrm{MIDI}} = t_{\mathrm{MIDI}} - t_{\mathrm{original}}$; $delay_{\mathrm{repro}} = t_{\mathrm{reproduced}} - t_{\mathrm{original}}$.

[10]There were no systematic differences between the two performing pianists, so the data in this and all subsequent figures were not plotted separately for pianists.

[11]For the SE system, the final hammer velocity needs to be mapped to MIDI velocity values by choosing a velocity map. In the present study, a logarithmic map was always used: $\textit{MIDI velocity} = 52 + 25 \log_2(\mathrm{FHV})$.

[12]As informal listening to the material suggests; systematic listening tests will be performed in future work.

[13]The SE system was recently completely re-engineered and was expected to be available commercially at the Bösendorfer company by mid-2002 (Dain, 2002).

Askenfelt, A. (**1994**). ''Observations on the transient components of the piano tone,'' in *SMAC 93: Proceedings of the Stockholm Music Acoustics Conference 28 July–1 August 1993*, edited by A. Friberg, J. Iwarsson, E. V. Jansson, and J. Sundberg (Royal Swedish Academy of Music, Stockholm), Vol. 79, pp. 297–301.

Askenfelt, A., and Jansson, E. V. (**1990**). ''From touch to string vibrations. I. Timing in grand piano action,'' J. Acoust. Soc. Am. **88**(1), 52–63.

Askenfelt, A., and Jansson, E. V. (**1991**). ''From touch to string vibrations. II. The motion of the key and hammer,'' J. Acoust. Soc. Am. **90**(5), 2383–2393.

Báron, J. G. (**1958**). ''Physical basis of piano touch,'' J. Acoust. Soc. Am. **30**(2), 151–152.

Báron, J. G., and Holló, J. (**1935**). ''Kann die Klangfarbe des Klaviers durch die Art des Anschlages beeinflußt werden?'' Z. Sinnesphysiologie **66**(1/2), 23–32.

Behne, K.-E., and Wetekam, B. (**1994**). ''Musikpsychologische Interpretationsforschung: Individualität und Intention,'' in *Musikpsychologie Empirische Forschungen, Ästhetische Experimente*, edited by K. E. Behne, G. Kleinen, and H. d. la Motte-Haber (Noetzel, Wilhelmshaven), Vol. 10, pp. 24–32.

Bolzinger, S. (**1995**). ''Contribution a l'étude de la rétroaction dans la pratique musicale par l'analyse de l'influence des variations d'acoustique de la salle sur le jeu du pianiste,'' Unpublished doctoral thesis, Université Aix-Marseille II, Marseille.

Bresin, R., and Battel, G. U. (**2000**). ''Articulation strategies in expressive piano performance,'' J. New Mus. Res. **29**(3), 211–224.
Bresin, R., and Widmer, G. (**2000**). ''Production of staccato articulation in Mozart sonatas played on a grand piano. Preliminary results,'' TMH-QPSR **2000**(4), 1–6.
Coenen, A., and Schäfer, S. (**1992**). ''Computer-controlled player pianos,'' Comput. Music J. **16**(4), 104–111.
Dain, R. (**2002**). ''The engineering of the concert piano,'' Ingenia **12** (May), 20–39. Published online at http://www.pianosonline.co.uk/.
Friberg, A., and Sundberg, J. (**1995**). ''Time discrimination in a monotonic, isochronous sequence,'' J. Acoust. Soc. Am. **98**(5), 2524–2531.
Galembo, A. (**1982**). ''Quality evaluation of musical instruments,'' (in Russian) Tech. Aesthetics **5**, 16–17.
Galembo, A. (**2001**). ''Perception of musical instrument by performer and listener (with application to the piano),'' in *Proceedings of the International Workshop on Human Supervision and Control in Engineering and Music, 21–24 September 2001* (University of Kassel, Kassel, Germany), pp. 257–266.
Goebl, W. (**2001**). ''Melody lead in piano performance: Expressive device or artifact?'' J. Acoust. Soc. Am. **110**(1), 563–572.
Goebl, W., and Bresin, R. (**2001**). ''Are computer-controlled pianos a reliable tool in music performance research? Recording and reproduction precision of a Yamaha Disklavier grand piano,'' in *Workshop on Current Research Directions in Computer Music, 15–17 November 2001*, edited by C. L. Buyoli and R. Loureiro (Audiovisual Institute, Pompeu Fabra University, Barcelona, Spain), pp. 45–50.
Goebl, W., Bresin, R., and Galembo, A. (**2003**). ''The piano action as the performer's interface: Timing properties, dynamic behaviour, and the performer's possibilities,'' in Proceedings of the Stockholm Music Acoustics Conference, 6–9 August 2003 (SMAC03), edited by R. Bresin (Department of Speech, Music, and Hearing, Royal Institute of Technology, Stockholm, Sweden), Vol. 1, pp. 159–162.
Hart, H. C., Fuller, M. W., and Lusby, W. S. (**1934**). ''A precision study of piano touch and tone,'' J. Acoust. Soc. Am. **6**, 80–94.
Juslin, P. N., and Madison, G. (**1999**). ''The role of timing patterns in recognition of emotional expression from musical performance,'' Music Percept. **17**(2), 197–221.
Koornhof, G. W., and van der Walt, A. J. (**1994**). ''The influence of touch on piano sound,'' in *SMAC 93: Proceedings of the Stockholm Music Acoustics Conference, 28 July–1 August 1993*, edited by A. Friberg, J. Iwarsson, E. V. Jansson, and J. Sundberg (Royal Swedish Academy of Music, Stockholm), Vol. 79, pp. 302–308.
Maria, M. (**1999**). *Unschärfetests mit hybriden Tasteninstrumenten*, Paper presented at the Global Village—Global Brain—Global Music. KlangArt Kongreß 1999, Osnabrück, Germany.
Moog, R. A., and Rhea, T. L. (**1990**). ''Evolution of the keyboard interface: The Bösendorfer 290 SE recording piano and the Moog multiply-touch-sensitive keyboards,'' Comput. Music J. **14**(2), 52–60.
Palmer, C. (**1989**). ''Mapping musical thought to musical performance,'' J. Exp. Psychol. Hum. Percept. Perform. **15**(12), 331–346.
Palmer, C. (**1996**). ''On the assignment of structure in music performance,'' Music Percept. **14**(1), 23–56.
Palmer, C., and Brown, J. C. (**1991**). ''Investigations in the amplitude of sounded piano tones,'' J. Acoust. Soc. Am. **90**(1), 60–66.
Palmer, C., and Holleran, S. (**1994**). ''Harmonic, melodic, and frequency height influences in the perception of multivoiced music,'' Percept. Psychophys. **56**(3), 301–312.
Podlesak, M., and Lee, A. R. (**1988**). ''Dispersion of waves in piano strings,'' J. Acoust. Soc. Am. **83**(1), 305–317.
Repp, B. H. (**1993**). ''Some empirical observations on sound level properties of recorded piano tones,'' J. Acoust. Soc. Am. **93**(2), 1136–1144.
Repp, B. H. (**1994**). ''On determining the basic tempo of an expressive music performance,'' Psychol. Music **22**, 157–167.
Repp, B. H. (**1995**). ''Expressive timing in Schumann's Träumerei: An analysis of performances by graduate student pianists,'' J. Acoust. Soc. Am. **98**(5), 2413–2427.
Repp, B. H. (**1996a**). ''The art of inaccuracy: Why pianists' errors are difficult to hear,'' Music Percept. **14**(2), 161–184.
Repp, B. H. (**1996b**). ''The dynamics of expressive piano performance: Schumann's Träumerei revisited,'' J. Acoust. Soc. Am. **100**(1), 641–650.
Repp, B. H. (**1996c**). ''Patterns of note onset asynchronies in expressive piano performance,'' J. Acoust. Soc. Am. **100**(6), 3917–3932.
Repp, B. H. (**1996d**). ''Pedal timing and tempo in expressive piano performance: A preliminary investigation,'' Psychol. Music **24**(2), 199–221.
Repp, B. H. (**1997a**). ''Acoustics, perception, and production of legato articulation on a computer-controlled grand piano,'' J. Acoust. Soc. Am. **102**(3), 1878–1890.
Repp, B. H. (**1997b**). ''The effect of tempo on pedal timing in piano performance,'' Psychol. Res. **60**(3), 164–172.
Riley-Butler, K. (**2001**). ''Comparative performance analysis through feedback technology,'' *Meeting of the Society for Music Perception and Cognition (SMPC2001), 9–11 August 2001* (Queen's University, Kingston, Ontario, Canada), pp. 27–28.
Riley-Butler, K. (**2002**). ''Teaching expressivity: An aural/visual feedback/replication model,'' *ESCOM 10th Anniversary Conference on Musical Creativity, 5–8 April 2002* (Université de Liège, Liège, Belgium).
Seashore, C. E. (**1937**). ''Piano touch,'' Scientific Monthly, New York **45**, 360–365.
Stahnke, W. (**2000**). Personal communication.
Timmers, R., Ashley, R., Desain, P., and Heijink, H. (**2000**). ''The influence of musical context on tempo rubato,'' J. New Mus. Res. **29**(2), 131–158.
White, W. B. (**1930**). ''The human element in piano tone production,'' J. Acoust. Soc. Am. **1**, 357–367.
Widmer, G. (**2001**). ''Using AI and machine learning to study expressive music performance: Project survey and first report,'' AI Commun. **14**(3), 149–162.
Widmer, G. (**2002**). ''Machine discoveries: A few simple, robust local expression principles,'' J. New Mus. Res. **31**(1), 37–50.
Widmer, G. (**2003**). ''Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries,'' Artif. Intell. **146**(2), 129–148.

Acoustic wave propagation in bovine cancellous bone: Application of the Modified Biot–Attenborough model

Kang Il Lee, Heui-Seol Roh, and Suk Wang Yoon[a)]
Acoustics Research Laboratory and BK21 Physics Research Division, Department of Physics, SungKyunKwan University, Suwon 440-746, Republic of Korea

(Received 25 February 2003; revised 15 July 2003; accepted 28 July 2003)

Acoustic wave propagation in bovine cancellous bone is experimentally and theoretically investigated in the frequency range of 0.5–1 MHz. The phase velocity, attenuation coefficient, and broadband ultrasonic attenuation (BUA) of bovine cancellous bone are measured as functions of frequency and porosity. For theoretical estimation, the Modified Biot–Attenborough (MBA) model is employed with three new phenomenological parameters: the boundary condition, phase velocity, and impedance parameters. The MBA model is based on the idealization of cancellous bone as a nonrigid porous medium with circular cylindrical pores oriented normal to the surface. It is experimentally observed that the phase velocity is approximately nondispersive and the attenuation coefficient linearly increases with frequency. The MBA model predicts a slightly negative dispersion of phase velocity linearly with frequency and the nonlinear relationships of attenuation and BUA with porosity. The experimental results are in good agreement with the theoretical results estimated with the MBA model. It is expected that the MBA model can be usefully employed in the field of clinical bone assessment for the diagnosis of osteoporosis. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1610450]

PACS numbers: 43.80.Cs [FD]

I. INTRODUCTION

Theoretical modeling of acoustic wave propagation in cancellous bone is very important to understand the interaction between acoustic wave and cancellous bone and to improve the quantitative ultrasound (QUS) techniques for the diagnosis of osteoporosis. In order to apply theories such as those developed by Biot[1] and Schoenberg[2] to cancellous bone, attempts have been made by some investigators.[3–8] The Biot theory[1] has attracted most attention with regard to modeling acoustic wave propagation in cancellous bone. The application of the Biot theory to cancellous bone was reviewed by Haire and Langton.[3] The Biot theory accounts for the motion of fluid and solid separately and models the elastic, inertial, and viscous coupling between fluid and solid. The theory predicts the existence of two longitudinal waves: a fast (first kind) wave corresponding to the solid and fluid moving in phase and a slow (second kind) wave corresponding to the solid and fluid moving out of phase. It has shown some success in predicting ultrasound propagation in cancellous bone.[4,5,6] Hosokawa and Otani[5,6] found good agreement between the measured and the calculated propagation speeds, but not for attenuation. This could be because the Biot theory considers the absorption due to only viscous loss at internal interfaces, but the experimental measurements record the signal loss due to other mechanisms as well. Signal loss may include contributions from reflection at the flat surfaces of samples,[9] scattering, absorption of the scattered field, and artifacts of the measurement system such as diffraction[10] and phase cancellation.[11] These factors must be known for the comparison between the theoretical attenuation and the measured signal loss to be meaningful.

Because of the complexity of the Biot theory, Hughes *et al.*[7] suggested using Schoenberg's theory.[2] This theory assumes cancellous bone to be made up of the simple layered structure of alternating bone-marrow plates. Then, the well-established theory of acoustic wave propagation in strata could be applied. The main advantage of Schoenberg's theory compared to the Biot theory is that it requires only two structural parameters: porosity and mean trabeculae thickness. It also predicts two compressional waves for all propagation angles, except for that perpendicular to the plates, where there is only one. The two waves are equivalent to the waves of the first and second kind of the Biot theory. In Schoenberg's theory, the fluid is assumed to be inviscid and so viscous absorption is not considered. Using bovine samples from tibial and femoral epiphyses, Hughes *et al.*[7] found a qualitative agreement between the measured phase velocity and the prediction of Schoenberg's theory. Wear[8] also applied the stratified model to predict the dependence of phase velocity on frequency in trabecular bone. The model performed well for phantoms consisting of alternating layers of polystyrene sheets and water. Moreover, it successfully predicted negative dispersion of phase velocity in human calcaneus samples.

Strelitzke *et al.*[12] proposed that a scattering model based on velocity fluctuations in a binary mixture (marrow fat and cortical matrix) can be used to estimate the ultrasonic attenuation in cancellous bone. Nicholson *et al.*[13] also used this scattering model in cancellous bone to investigate the relationship between broadband ultrasonic attenuation (BUA) and bone mineral density (BMD). The model predicted very similar nonlinear trends to those previously observed experimentally. They also demonstrated that attenuation was dependent on scatterer size in addition to porosity. One of the

a)Electronic mail: swyoon@skku.ac.kr

 0001-4966/2003/114(4)/2284/10/$19.00

potential limitations in this approach is that absorption was not included into the model. Furthermore, the model could not predict velocity relationships. Thus, no single model did not explain all the features of the acoustic behavior of cancellous bone.

Recently, the Modified Biot–Attenborough (MBA) model[14] has been proposed for acoustic wave propagation in a nonrigid porous medium with circular cylindrical pores. Theoretical models for acoustic wave propagation in porous media, such as the Biot theory[1] and Attenborough's approach,[15] have some limitations in applications. Although the Biot theory takes into account the viscous effect as a relatively comprehensive approach for elastic materials, it does not consider the thermal effect. Furthermore, the Biot theory needs a large number of physical parameters that should be measured or estimated, and it does not provide analytic solutions. The theory introduced by Attenborough takes into account both the viscous and the thermal effects in a separate manner, but the pore frame should be considered as a rigid material. The fast wave of the Biot theory is, thus, not included in Attenborough's approach. In our previous work,[14] we proposed the MBA model combining the Biot theory and Attenborough's approach. It includes the thermal effect, which has an analytic solution, and needs fewer input data than those required in the Biot theory. The MBA model in a mixed solid and fluid medium incorporates the fast wave with the slow wave. The MBA model is, thus, a phenomenologically mixed approach based on the strengths of the Biot theory and the separate treatment of Attenborough's approach.

Complicated porous media, such as cancellous bone, may be acoustically modeled using a solid material containing circular cylindrical pores with fluid since the porous medium is regarded as a locally reacting medium. In our previous work,[14] Lucite slabs with circular cylindrical pores filled with water are selected as the preliminary test phantoms for bone since Lucite can be easily manipulated and has similar acoustic impedance to bone. The phase velocity and the transmission coefficient of the Lucite slabs with the circular cylindrical pores were experimentally measured as functions of porosity, pore size, and frequency and also theoretically estimated with the MBA model. The experimental measurements showed good agreement with the theoretical estimates. These results suggest the applicability of the MBA model to the diagnosis of osteoporosis for cancellous bone.

In this study, acoustic wave propagation in bovine cancellous bone was experimentally and theoretically investigated in the frequency range of 0.5–1 MHz. Phase velocity, attenuation coefficient, and BUA of bovine cancellous bone were measured as functions of frequency and porosity. For a theoretical estimation, the MBA model[14] was employed with three new phenomenological parameters: the boundary condition, phase velocity, and impedance parameters. The MBA model is based on the idealization of cancellous bone as a nonrigid porous medium with circular cylindrical pores oriented normal to the surface. In Sec. II we briefly describe the MBA model[14] for acoustic wave propagation in a nonrigid porous medium with circular cylindrical pores.

II. MODIFIED BIOT–ATTENBOROUGH (MBA) MODEL FOR A NONRIGID POROUS MEDIUM

The MBA model[14] is based on Attenborough's approach[15] that separately treats the viscous and the thermal effects. This model requires the introduction of phenomenological, frequency-dependent parameters, without laying the basis for their independent calculation or measurement. Zwikker and Kosten[16] have shown that, at least in the limiting cases of low and high frequency, such independent treatments give the correct results for motion in a viscous, conducting fluid contained within a cylindrical pore, when expressed in terms of a complex density and a complex compressibility. Three phenomenological parameters, one for the boundary condition (s_1), one for the phase velocity (s_2), and one for the impedance (s_3), are introduced to take into account the nonrigid frame in the MBA model. The parameters s_2 and s_3 are, respectively, specified for the fast and the slow waves in the case when the two waves are clearly separated.

A. Sound propagation through a single cylindrical pore

In this study, for simplicity, sound propagation through a single circular cylindrical pore is assumed to be one-dimensional along with the axis of the cylinder. Care was taken in the MBA model to treat the boundary condition at a nonrigid frame which should be different from that at a rigid frame.

The continuity equation for a one-dimensional acoustic wave propagating through a circular cylindrical pore filled with fluid is given by

$$-\rho_0 \frac{\partial \langle v \rangle}{\partial x} = \frac{\partial \rho}{\partial t}, \tag{1}$$

where ρ_0 is the equilibrium density of the fluid and $\langle v \rangle$ is the average particle velocity over the cross section of the circular cylindrical pore for propagation in the x direction of the central axis of the cylindrical pore.

The equation of motion for the same case as the continuity equation is expressed in terms of the frequency-dependent (or complex) density, $\rho_c(\omega)$, by

$$\frac{\partial p}{\partial x} = \rho_c(\omega) \frac{\partial \langle v \rangle}{\partial t}, \tag{2}$$

where p is the acoustic pressure and ω is the angular frequency. The frequency-dependent density can be written as

$$\rho_c(\omega) = \rho_0 [1 - 2(\lambda \sqrt{i})^{-1} T(\lambda \sqrt{i})]^{-1}, \tag{3}$$

where

$$T(\lambda \sqrt{i}) = \frac{J_1(\lambda \sqrt{i})}{J_0(\lambda \sqrt{i})}, \tag{4}$$

$i \equiv \sqrt{-1}$, and J_0 and J_1 are the cylindrical Bessel functions of the zeroth and the first order, respectively. The dimensionless parameter λ is related to the thickness of the viscous boundary layer at a pore wall and is defined by

$$\lambda = a s_1 \left(\frac{\omega}{\nu} \right)^{1/2}, \tag{5}$$

where a is the radius of the circular cylindrical pore, ν is the kinematic viscosity of the fluid, and s_1 is the boundary condition related parameter representing the rigidity of a pore frame. The value of s_1 is determined by the ratio of the density of an ideally rigid frame to that of the (nonrigid) frame considered in the MBA model. As $s_1 \rightarrow 1$, the frame becomes more rigid and, if $s_1 > 1$, it is not rigid. In the present study, the value of s_1 was experimentally set to a constant based on the data for the phase velocity and the transmission coefficient as functions of the frequency.

The effective radius, r_{eff}, of the cylindrical pore may be given as $a s_1$. This is justified since the relative particle velocity at the boundary $r = a$ becomes zero for a rigid wall, but is nonzero at $r = a$ for a nonrigid wall. This suggests that the relative particle velocity for a nonrigid frame becomes zero at an effective radius $r_{\text{eff}} = a s_1$.

Assuming that the fluid within the circular cylindrical pore is a nonviscous conducting fluid, the heat flow may only occur in the transverse direction, and the acoustic pressure, p will be uniform over the cross section of the pore. The thermal equations in this case may be arranged so as to give the frequency-dependent compressibility, $C_c(\omega)$, of fluid within the pore:

$$C_c(\omega) = \frac{1}{\rho_0} \frac{d\rho_c}{dp} = (\rho_0 c_0^2)^{-1} [1 + 2(\gamma - 1) \times (N_{\text{Pr}}^{1/2} \lambda \sqrt{i})^{-1} T(N_{\text{Pr}}^{1/2} \lambda \sqrt{i})], \tag{6}$$

where γ is the specific heat ratio of the fluid, N_{Pr} is the Prandtl number, and c_0 is the equilibrium sound velocity of the fluid.

B. Extension to bulk cylindrical pores for a nonrigid porous medium

Based on the sound propagation through a single circular cylindrical pore, we phenomenologically extend the analysis to the case of a nonrigid porous medium with bulk cylindrical pores. When all the pores are identical, the average particle velocity, $\langle v \rangle$, for the single pore is related to the average particle velocity, u over the unit cross section of the porous medium with the cylindrical pores, and this may be expressed as

$$\langle v \rangle = q u \left[\frac{C_c(\omega) Z_b \omega}{k_b} \right], \tag{7}$$

where q is the tortuosity of the pore, Z_b is the characteristic impedance of the porous medium, and k_b is the propagation constant of the porous medium. The tortuosity factor, q has the particular value $1/\cos\theta$ for a medium containing parallel cylindrical pores inclined at angle θ to the surface normal. The frequency-dependent density of the porous medium, $\rho_b(\omega)$, is phenomenologically given by

$$\rho_b(\omega) = \frac{Z_b k_b}{q \omega}. \tag{8}$$

In this case, the equation of continuity becomes

$$-q \left[\frac{C_c(\omega) Z_b \omega}{k_b} \right] \frac{\partial u}{\partial x} = \frac{1}{\rho_0} \frac{d\rho_b}{dp} \frac{\partial p}{\partial t}. \tag{9}$$

The equation of motion leads to

$$-\frac{\partial p}{\partial x} = \rho_b(\omega) \frac{\partial u}{\partial t}. \tag{10}$$

Combining Eqs. (9) and (10), we obtain the wave equation

$$\frac{\partial^2 p}{\partial x^2} = \frac{1}{c_b^2} \frac{\partial^2 p}{\partial t^2}, \tag{11}$$

where $c_b^2 = \omega^2 / k_b^2$.

The complex propagation constant, k_b, for a nonrigid porous medium may be expressed with a newly introduced parameter s_2 from Eq. (11), with the help of Eqs. (3), (6), and (8), as

$$k_b^2 = q^2 \left[\frac{k_m^2 k_c^2}{(1-\Omega)^{s_2} k_c^2 + \Omega^{s_2} k_m^2} \right], \tag{12}$$

where Ω is the volume porosity of pores, s_2 is a fitting parameter designed to make the phase velocity of the porous medium vary appropriately with the porosity, k_m is the propagation constant of the pore frame, and k_c is the propagation constant of the fluid in the pore and is simply given by

$$k_c^2 = \omega^2 C_c(\omega) \rho_c(\omega). \tag{13}$$

The complex quantity k_b can be rearranged into the real and the imaginary components:

$$k_b = k_{br} + i k_{bi}, \tag{14}$$

where k_{br} and k_{bi} are the real and the imaginary components of k_b, respectively. The phase velocity, V_b, is given by

$$V_b = \frac{\omega}{k_{br}}. \tag{15}$$

The attenuation coefficient, A_b, in units of dB per unit length is then given by

$$A_b = \frac{20}{\ln 10} k_{bi}. \tag{16}$$

The parameter s_2 for the porous medium represents the slope of the phase velocity as a function of the porosity and is determined by fitting a theoretical estimate to the measured phase velocity. The value of s_2 is less than 1 when the fitted sound velocity profile as a function of the porosity is convex in shape, is bigger than 1 when it is concave, and is equal to 1 when it is linear. The phase velocity becomes that of the pore frame at a porosity value of 0 and that of the pore fluid at a porosity value of 1.

The effective characteristic impedance, Z_b, may be written with the parameter s_3 as

$$Z_b = q[-s_3 \Omega^2 + (\rho_c c_c - \rho_m c_m + s_3)\Omega + \rho_m c_m], \tag{17}$$

where ρ_m and c_m are, respectively, the density and the sound velocity of the pore frame, ρ_c and c_c are the density and the sound velocity of the pore fluid, and s_3 is a fitting parameter designed to make the effective impedance of the porous medium vary appropriately with the porosity.

The parameter s_3 represents the slope of the impedance as a function of the porosity and can be phenomenologically determined by fitting the theoretical estimate to the measured transmission or reflection coefficients. It should be noted that the effective impedance becomes that of the frame material at a porosity value of 0 and that of the pore fluid at a porosity value of 1. Equation (17) shows that the effective characteristic impedance, Z_b, increases as the porosity increases in the low-porosity region, where the fast wave is dominant, whereas it decreases as the porosity increases in the high-porosity region, where the slow wave is dominant.

The frequency-dependent effective density, $\rho_b(\omega)$, for a nonrigid porous medium in Eq. (8) can, thus, be written in terms of the parameters s_2 and s_3, by combining Eqs. (12) and (17), as

$$\rho_b(\omega)=\frac{q}{\omega}[-s_3\Omega^2+(\rho_c c_c-\rho_m c_m+s_3)\Omega+\rho_m c_m]\times\left[\frac{k_m^2k_c^2}{(1-\Omega)^{s_2}k_c^2+\Omega^{s_2}k_m^2}\right]^{1/2}. \tag{18}$$

It should be noted that the effective density is not the same as the simple arithmetic mean density $\rho_p=(1-\Omega)\rho_m+\Omega\rho_0$ for a rigid porous medium.

C. Extension of the phenomenological parameters

So far, the parameters s_2 and s_3 have been taken into account for the case in which the fast and the slow waves were mixed. The features of the parameters, however, are somewhat different for each wave, and it is necessary to deal with the parameters independently for each wave when the fast and the slow waves are clearly separated in time. This situation may occur for a porous medium consisting of a very low-density fluid and a highly rigid frame.

From Eq. (12), the complex propagation constant, k_{bf}, for the fast wave may be expressed using the parameter s_{2f} instead of s_2 as

$$k_{bf}^2=q^2\left[\frac{k_m^2k_g^2}{(1-\Omega)^{s_{2f}}k_g^2+\Omega^{s_{2f}}k_m^2}\right], \tag{19}$$

where k_g is the propagation constant of a very low-density fluid medium compared with water. For the extreme cases of $\Omega\to0$ and $\Omega\to1$, $k_{bf}\to k_m$ and $k_{bf}\to k_g$, respectively. As provided in Eqs. (14), (15), and (16), the phase velocity and the attenuation coefficient for the fast wave can be obtained from the real and the imaginary components of k_{bf}, respectively.

The same procedure for the fast wave results in the complex propagation constant k_{bs} for the slow wave expressed by

$$k_{bs}^2=q^2\left[\frac{k_r^2k_c^2}{(1-\Omega)^{s_{2s}}k_c^2+\Omega^{s_{2s}}k_r^2}\right], \tag{20}$$

where k_r is the propagation constant of a very rigid medium compared with a nonrigid medium. For the extreme case of $\Omega\to0$ and $\Omega\to1$, $k_{bs}\to k_r$ and $k_{bs}\to k_c$, respectively. As for the fast wave, the phase velocity and the attenuation coefficient for the slow wave can be obtained from the real and the imaginary components of k_{bs}, respectively.

Using the parameter s_{3f} instead of s_3 in Eq. (17), the effective characteristic impedance, Z_{bf}, for the fast wave can be expressed by

$$Z_{bf}=q[(\rho_g c_g-\rho_m c_m+s_{3f})\Omega+\rho_m c_m], \tag{21}$$

where ρ_m and c_m are the density and the sound velocity of the pore frame, and ρ_g and c_g are the density and the sound velocity of a very low-density fluid medium, respectively. For the extreme value of the porosity, as $\Omega\to0$, $Z_{bf}\to Z_m$ $(=\rho_m c_m)$, and as $\Omega\to1$, $Z_{bf}\to Z_g(=\rho_g c_g)$ assuming that the parameter s_{3f} is negligible compared with Z_g.

The same procedure applies for the effective characteristic impedance, Z_{bs}, for the slow wave, which is given by

$$Z_{bs}=q[(\rho_c c_c-\rho_r c_r+s_{3s})\Omega+\rho_r c_r-s_{3s}], \tag{22}$$

where ρ_c and c_c are the density and the sound velocity of the pore fluid, and ρ_r and c_r are the density and the sound velocity of a very rigid medium, respectively. For the extreme case of the porosity, as $\Omega\to0$, $Z_{bs}\to Z_r(=\rho_r c_r)$, and as $\Omega\to1$, $Z_{bs}\to Z_c(=\rho_c c_c)$ assuming that the parameter s_{3s} is negligible compared with Z_r.

The frequency-dependent effective density, $\rho_{bf}(\omega)$, for the fast wave is given, using Eqs. (19) and (21) together with Eqs. (14) and (15), as

$$\rho_{bf}(\omega)=\frac{q}{\omega}[(\rho_g c_g-\rho_m c_m+s_{3f})\Omega+\rho_m c_m]\times\left[\frac{k_m^2k_g^2}{(1-\Omega)^{s_{2f}}k_g^2+\Omega^{s_{2f}}k_m^2}\right]^{1/2}. \tag{23}$$

The effective density, $\rho_{bs}(\omega)$, for the slow wave is also given, using Eqs. (20) and (22) together with Eqs. (14) and (15), as

$$\rho_{bs}(\omega)=\frac{q}{\omega}[(\rho_c c_c-\rho_r c_r+s_{3s})\Omega+\rho_r c_r-s_{3s}]\times\left[\frac{k_r^2k_c^2}{(1-\Omega)^{s_{2s}}k_c^2+\Omega^{s_{2s}}k_r^2}\right]^{1/2}. \tag{24}$$

III. MATERIALS AND METHODS

A. Bone specimens

Twelve cancellous bone specimens were obtained from the proximal ends of one fresh frozen bovine tibia. The specimens were cut by using a rotary electric saw to make parallel and smooth surfaces without any soft tissue. The surfaces of each specimen were finished with a fine silicon carbide abrasive paper. Care was taken to produce parallel surfaces. The resulting specimens had flat and parallel faces with thicknesses varying from 12 to 18 mm. The opposing planar surfaces were kept as parallel within $\pm0.5°$ by caliper measurements of thickness throughout each specimen. The bone specimens were large enough to cover the receiving transducer face. The specimens were all oriented in the same direction in relation to the bone. Their orientations were chosen so that the ultrasound could pass through the specimens

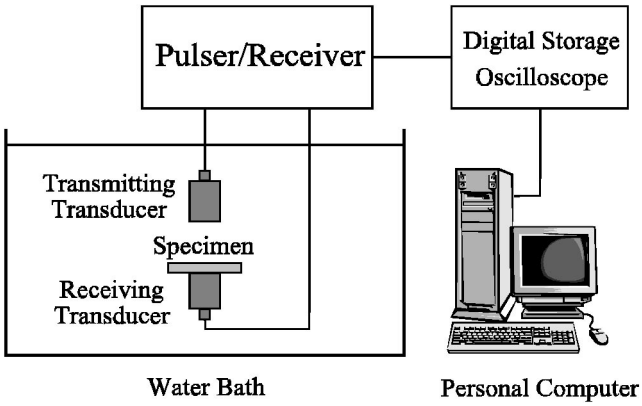


FIG. 1. Schematic diagram of the experimental setup for ultrasonic measurements.

along the mediolateral (ML) direction of the tibia. They are the same orientations with those *in vivo* measurements commonly performed with commercial bone sonometers.

Prior to ultrasonic measurements, the total density ρ of each bone specimen with bone marrow was measured with Archimedes' principle, and then the porosity β was estimated by the following equation:

$$\beta = \frac{\rho_m - \rho}{\rho_m - \rho_0}, \tag{25}$$

where ρ_m is the density of the solid bone and ρ_0 is the density of the bone marrow. The density ρ_m of the solid bone was given as 1960 kg/m^3 by Lang[17] and the density ρ_0 of the bone marrow was given as 930 kg/m^3 by Hosokawa and Otani.[5]

In order to decrease the risk of microorganism contamination of bone specimens, they were defatted by boiling for 1 hour in water. The defatting process was assumed not to significantly affect their acoustic properties because the acoustic properties of defatted trabecular bone have shown just slightly different from those of bone with the marrow left intact.[18] Although boiling may change the bone matrix, correlations between ultrasonic parameters and bone density are largely retained. This is due to the fact that these relationships arise primarily from the interactions of the ultrasonic wave with the bone matrix–pore fluid structure and less from changes in the bone matrix material by itself. To remove air bubbles, the specimens were degassed under vacuum. After removing air bubbles, they were kept at room temperature prior to ultrasonic measurements.

B. Ultrasonic measurements

The schematic diagram of the experimental setup for ultrasonic measurements is shown in Fig. 1. Ultrasonic measurements were performed in a water bath maintained at room temperature between 16 and 17 °C. A 200 MHz computer controlled pulser/receiver (Panametrics 5900PR) was used. One matched pair of coaxially aligned transducers with the center frequency of 1 MHz (Panametrics V303, 0.5 in. diameter) was vertically oriented to make it possible to mount the specimens directly on the face of the receiving transducer. The faces of transducers were separated by 50 mm, a distance greater than the 26.5 mm near-field distance stated by the manufacturer. Received ultrasound signals were acquired using a 500 MHz digital storage oscilloscope (LeCroy LT322) and stored on computer for an off-line analysis. A through-transmission method was used to measure the phase velocity and attenuation coefficient. Using two opposing coaxially aligned transducers, transmitted signals were recorded both with and without the bone specimen in the acoustic path. The bone specimens were large enough to cover the receiving transducer face.

For phase velocity measurements, a fast Fourier transform (FFT) was used to obtain the phase spectra of the transmitted signals with and without the specimen in the water path. The phase velocity $c_p(\omega)$ as a function of frequency can be estimated from the following equation:

$$c_p(\omega) = \frac{c_w}{1 - \dfrac{c_w \Delta\phi(\omega)}{\omega d}}, \tag{26}$$

where ω is the angular frequency, d is the thickness of the specimen, and $\Delta\phi(\omega)$ is the difference in unwrapped phases of the transmitted signals with and without the bone specimen in the acoustic path. The temperature-dependent speed of sound in distilled water c_w is given by[19]

$$c_w = 1402.9 + 4.835 \times T - 0.047016 \times T^2 + 0.00012725 \times T^3, \tag{27}$$

where T is the temperature in °C. The measurements were repeated for the opposite direction in the same specimen. A total of ten measurements of two directions in each specimen were averaged to obtain the mean value of phase velocity.

The attenuation coefficient was determined using the same signals acquired for phase velocity measurements. The signal loss as a function of frequency was obtained by subtracting the logarithm of the power spectrum obtained through the specimen from that of the nonattenuating reference power spectrum obtained through water. The signal loss was divided by the thickness of the specimen to obtain the attenuation coefficient in units of dB/cm. The attenuation slope, i.e., BUA in dB/cm MHz was determined by the linear regression of attenuation over the bandwidth of interest. This parameter is often referred to as ''normalized broadband ultrasonic attenuation'' (nBUA). The usable frequency bandwidth was 0.5–1 MHz. This frequency bandwidth provided good signal to noise ratios.

IV. RESULTS

Figure 2 shows a reference signal only through water and a typical transmitted signal through a bone specimen. The bone signal exhibits the earlier arrival time due to the faster speed of sound in bone than that in water. Figure 3 shows the power spectra corresponding to the signals in Fig. 2. The usable bandwidth was taken to be from 0.5 to 1 MHz. The spectral data within the bandwidth of 0.5–1 MHz are reliable and not subject to the poor signal to noise ratio. The center frequency downshift due to frequency-dependent attenuation is apparent.

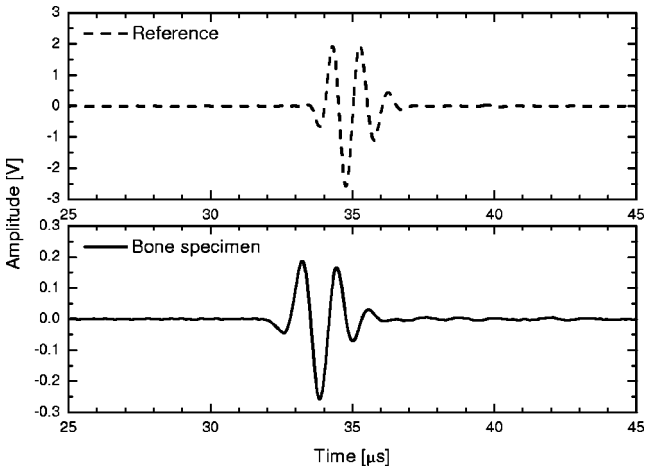


FIG. 2. Reference signal only through water and a typical transmitted signal through a bone specimen.

Table I lists the parameter values of cancellous bone for theoretical estimation with the MBA model. The pore radius a of cancellous bone was assumed to be 0.5 mm, consistent with typical values of trabecular separation in the literature.[5,20,21] On the assumption that all of the pores are identical, the tortuosity factor q was taken to be 1. This is because the pores in cancellous bone were supposed to be oriented normal to the surface. Therefore, the MBA model is based on the idealization of cancellous bone as a nonrigid porous medium with circular cylindrical pores oriented normal to the surface. The kinematic viscosity and Prandtl number of the bone marrow, ν and N_{Pr}, were assumed to be 1×10^{-6} m^2/s and 7, respectively, which are equivalent in values to water. The density $\rho_0=930$ kg/m^3 and the sound speed $c_0=1450$ m/s of the bone marrow, and the sound speed $c_m=3500$ m/s of the solid bone were taken from work by Hosokawa and Otani.[5] The density ρ_m of the solid bone was given as 1960 kg/m^3 by Lang,[17] and is equivalent to the value of bovine cortical bone measured by Hosokawa and Otani.[5] The other required parameters for the MBA model are the boundary condition (s_1), the phase velocity (s_2), and the impedance (s_3) parameters, which will be specified later.

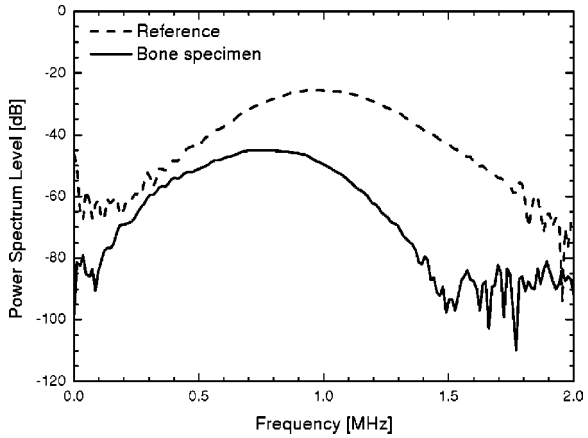


FIG. 3. The power spectrum corresponding to the signals in Fig. 2. The usable bandwidth was taken to be from 0.5 to 1 MHz. The spectral data within the bandwidth of 0.5–1 MHz are reliable and not subject to the poor signal to noise ratio.

TABLE I. Parameter values of cancellous bone for a theoretical estimation with the MBA model.

Parameter	Value
Radius of pore, a	0.5 mm
Tortuosity factor, q	1
Kinematic viscosity of bone marrow, ν	1×10^{-6} m^2/s
Prandtl number of bone marrow, N_{Pr}	7
Density of bone marrow, ρ_0	930 kg/m^3
Sound speed of bone marrow, c_0	1450 m/s
Density of solid bone, ρ_m	1960 kg/m^3
Sound speed of solid bone, c_m	3500 m/s
Boundary condition parameter, s_1	1.5
Phase velocity parameter, s_2	1.5
Impedance parameter, s_3	1.3485×10^6 Pa·s/m

Figure 4 shows the experimental and theoretical phase velocities as functions of frequency for the three cancellous bone specimens of porosity $\beta=81.9\%$, 78.9%, and 74.4%. Plotted error bars represent the standard deviation in terms of reproducibility. The theoretical estimation was performed in terms of the total wave mixed (or overlapped) with the fast and the slow waves using the MBA model. The experimental results show good agreement with the theoretical results estimated with boundary condition parameter $s_1=1.5$ and phase velocity parameter $s_2=1.5$. As shown in Fig. 4, although the experimental phase velocities in bovine cancellous bone specimens can be approximated as nondispersive in the frequency range of 0.5–1 MHz, the MBA model predicts a slightly negative dispersion linearly with frequency in this frequency range.

The experimental and theoretical phase velocities at 1 MHz are shown as functions of porosity in Fig. 5. The theoretical phase velocity varies from 3500 to 1450 m/s as the porosity increases. This is very similar to the propagation speed of the fast wave in cancellous bone with trabeculae aligned perpendicular to the direction of propagation as reported previously by Hosokawa and Otani.[6] They experimentally measured the phase velocities of the fast and the

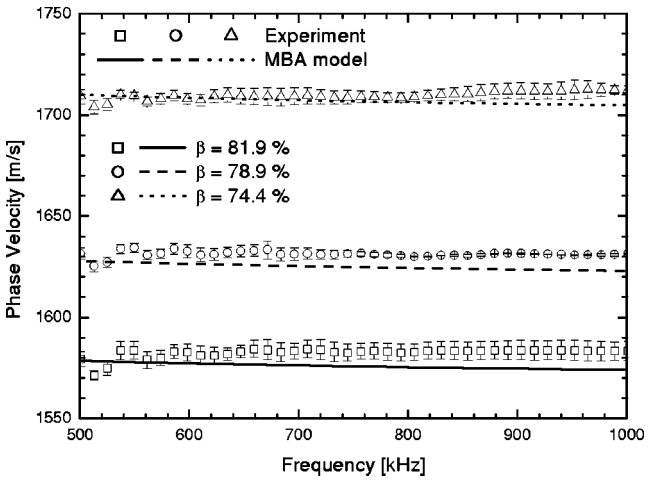


FIG. 4. Experimental and theoretical phase velocities as functions of frequency for the three cancellous bone specimens of porosity $\beta=81.9\%$, 78.9%, and 74.4%. Although the experimental phase velocity in bovine cancellous bone specimens can be approximated as nondispersive in the frequency range of 0.5–1 MHz, the MBA model predicts a slightly negative dispersion linearly with frequency in this frequency range.

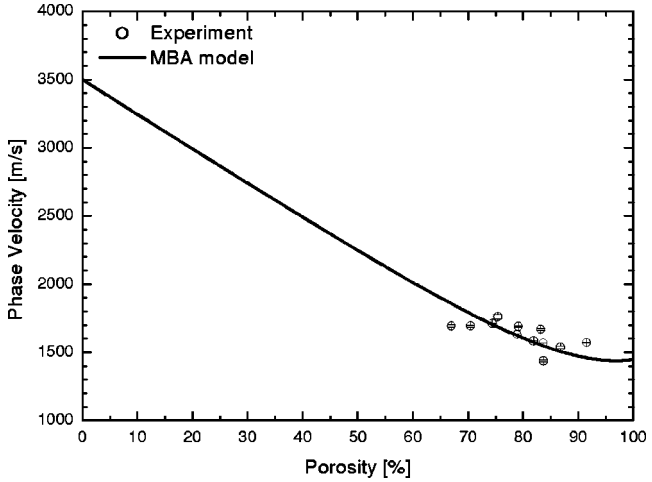


FIG. 5. Experimental and theoretical phase velocities as functions of porosity at 1 MHz. The theoretical phase velocity varies from 3500 to 1450 m/s as the porosity increases.

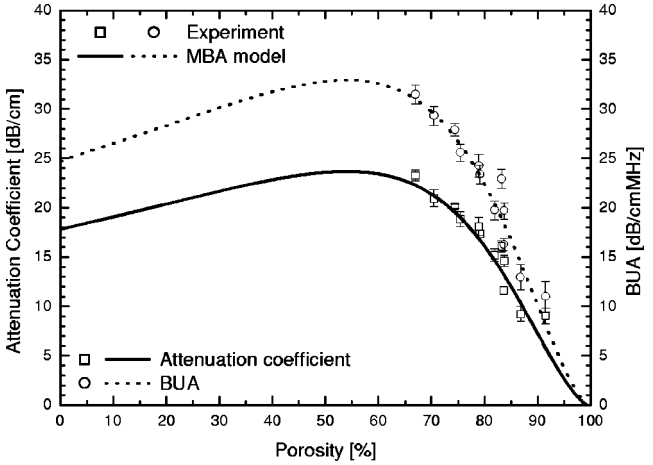


FIG. 7. Experimental and theoretical attenuation coefficients at 1 MHz and BUA values as functions of porosity. As the porosity increases, the theoretical attenuation coefficient and BUA slowly increase and rapidly decrease thereafter, reaching maxima at the porosities of approximately 55%–60%.

slow waves as functions of the propagation angle to the trabecular alignment, and theoretically estimated them with the Biot theory. No data for the slow wave was obtained for propagation in the perpendicular direction because the fast waveform completely overlaps the slow waveform.

Figure 6 shows the experimental and theoretical attenuation coefficients as functions of frequency for the three cancellous bone specimens of porosity $\beta=81.9\%$, 78.9%, and 74.4%. The experimental results are in good agreement with the theoretical results estimated with the impedance parameter $s_3=\rho_0 c_0=1.3485\times10^6$ Pa·s/m. In Fig. 6, it can be found that the experimental attenuation coefficients linearly increase with frequency in the frequency range of 0.5–1 MHz. The frequency variation of attenuation coefficient for the solid bone has been theoretically modeled with a simple linear relationship between attenuation and frequency, $\alpha(f)=\alpha_0+\alpha_1 f=-80+2.85\times10^{-4}f$ over the frequency range of 0.5–1 MHz. The constant α_0 has been introduced in order to account for the transmission loss and to avoid any assumption about the frequency variation of attenuation outside the measured frequency bandwidth.

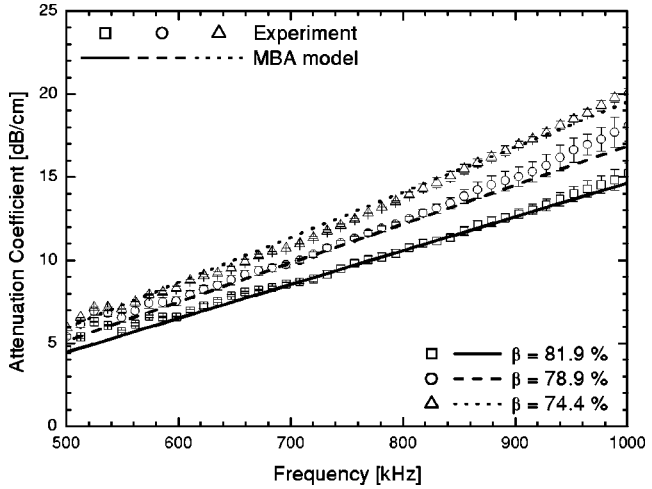


FIG. 6. Experimental and theoretical attenuation coefficients as functions of frequency for the three cancellous bone specimens of porosity $\beta=81.9\%$, 78.9%, and 74.4%. The experimental attenuation coefficients linearly increase with frequency in the frequency range of 0.5–1 MHz.

The experimental and theoretical attenuation coefficients at 1 MHz and BUA values are shown as functions of porosity in Fig. 7. The experimental results agree well with the theoretical results estimated with the MBA model. As the porosity increases, the theoretical attenuation coefficient slowly increases and rapidly decreases thereafter, reaching maxima at the porosities of approximately 55–60%. A similar nonlinear trend can be observed from the relationship between theoretical BUA and porosity.

V. DISCUSSION

Previous studies[3–6] have applied the Biot theory to investigate ultrasonic properties in cancellous bone. It was reported that both fast and slow longitudinal waves were clearly identified when the acoustic wave propagated in bovine femoral cancellous bone.[5,6] This could only occur in the parallel direction of ultrasonic propagation to the trabeculae, not in the perpendicular direction. In the present study, the relative orientation between the ultrasonic beam and the bone specimens was the ML direction of the bovine tibia where the trabeculae are aligned in the perpendicular direction to propagation at normal incidence. This is the same orientation with that *in vivo* measurements commonly performed with commercial bone sonometers. As shown in Fig. 2, since the fast and the slow waves were completely overlapped, only one longitudinal wave was observed for propagation in the ML direction. It is one of the reasons why there is no evidence of two longitudinal waves *in vivo*. In this study, therefore, the theoretical estimation was performed in terms of the total wave mixed (or overlapped) with the fast and the slow waves using the MBA model.

Phase velocity in cancellous bone has been investigated by a number of groups in various frequency ranges.[8,22–27] In human calcaneal cancellous bone, significant negative dispersion was observed by four different groups.[8,22–25] Moreover, a cancellous bone mimicking material also shows negative dispersion, but only at high porosities. The negative

dispersion in human calcaneus has yet to be explained. Recently, Wear[8] predicted the negative dispersion of phase velocity in trabecular bone using the stratified model. However, this negative dispersion is contrary to Kramers–Kronig (K–K) dispersion relations.[28] The K–K relations predict that if BUA is positive, then dispersion should be positive too. The violation of the K–K relationships suggests that there is something distinctly abnormal about wave propagation in low-density cancellous bone.

Fry and Barger[26] measured phase velocity increase with frequency in skull cancellous bone between 0.3 and 2.0 MHz. Positive velocity dispersion was also measured in one cancellous bone specimen by Pal *et al.*[27] Hosokawa and Otani[5] observed nondispersive phase velocities of fast and slow waves in bovine femoral cancellous bone between 0.5 and 5.0 MHz, and predicted them using the Biot theory. Either negative or positive velocity dispersion of cancellous bone was found in the frequency range 0.2 to 0.6 MHz by Droin *et al.*[24] In the present study, the experimental phase velocities in bovine cancellous bone specimens with relatively high bone mineral densities were found to be approximately nondispersive between 0.5 and 1.0 MHz, while the MBA model predicted a slightly negative dispersion linearly with frequency in this frequency range as shown in Fig. 4. The velocity dispersion predicted with the MBA model is somewhat lower than reported measurements for dispersion in human calcaneus.[8,22–25] The negative dispersion may be only associated with much lower density and velocity in human calcaneal cancellous bone. The denser bovine cancellous bone than human cancellous bone may cause ultrasonic absorption and scattering in fundamentally different mechanisms. There should be further studies in this subject.

Hosokawa and Otani[5,6] estimated the propagation speed of the fast and the slow waves propagating through bovine cancellous bone samples with trabeculae parallel and perpendicular directions using the Biot theory. The geometrical structure of the trabecular frame determines the parameter n which is the exponent of the power law for the elastic moduli of cancellous bone.[4] The concept of the parameter n introduced by Williams[4] is similar to that of the phase velocity parameters in the MBA model. The value of n was estimated in order to fit the experimental data of the speed of the fast wave as a function of bone volume fraction. They obtained the value of $n=2.14$ for propagation in the perpendicular direction to the trabeculae and $n=1.46$ in the parallel direction. In the MBA model, the phase velocities are determined by phase velocity parameters s_2, s_{2f}, and s_{2s}, which represent the slopes of the phase velocities of a porous medium as functions of porosity. The phase velocity parameter can be estimated by fitting the theoretical calculation to the experimental data of the phase velocity as a function of porosity. This resulted in the value of $s_2=1.5$ for the total wave in the perpendicular direction of propagation to the trabeculae (ML direction). The theoretical results of the phase velocity estimated with $s_2=1.5$ showed good agreement with the experimental results as shown in Fig. 5.

In the case of cancellous bone with trabeculae aligned parallel to the direction of propagation, the phase velocities of the fast wave and the slow wave can be theoretically estimated with $s_{2f}=0.6$ and $s_{2s}=0.5$, respectively, referring to the experimental data of Hosokawa and Otani.[5] Figure 8 shows the theoretical phase velocities estimated by the MBA model with $s_2=1.5$ for the total wave mixed with the fast and the slow waves, $s_{2f}=0.6$ for the fast wave, and $s_{2s}=0.5$ for the slow wave, respectively, as functions of porosity at 1 MHz. The trends clearly correspond to predictions from the Biot theory by Hosokawa and Otani.[5,6] Accordingly, the MBA model is applicable to the prediction of the phase velocity in a nonrigid porous medium with anisotropic structure such as cancellous bone by taking into account the phase velocity parameter in the direction of propagation.

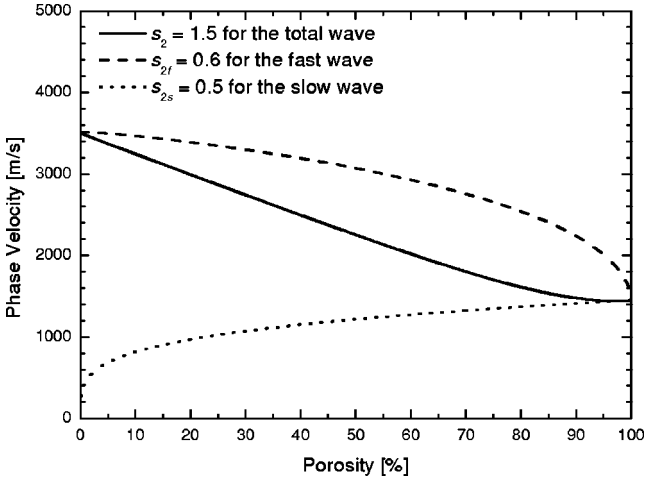


FIG. 8. Theoretical phase velocities estimated using the MBA model with $s_2=1.5$ for the total wave mixed with the fast and the slow waves, $s_{2f}=0.6$ for the fast wave, and $s_{2s}=0.5$ for the slow wave, respectively, as functions of porosity at 1 MHz.

In this study, the attenuation coefficients were theoretically modeled with a simple linear dependence on frequency, $\alpha(f)=\alpha_0+\alpha_1 f$ over the frequency range of 0.5–1 MHz. It is reported in many studies that attenuation shows a linear dependence on frequency in the frequency range of 0.2–1 MHz. The frequency-dependent attenuation in human calcaneus *in vitro* from 0.2 to 1.7 MHz has been reported by Chaffai *et al.*[29] and Wear.[18] The linear dependence of attenuation on frequency appears to extend up to at least 1.7 MHz for human calcaneus *in vitro*. The attenuation in a defatted proximal bovine tibia was studied *in vitro* up to 2.0 MHz by Hoffmeister *et al.*[30] and Lee *et al.*[31] The reliability of the estimates of frequency dependence increases with the frequency bandwidth used for measurements. However, the limiting factor to these measurements is the significant signal to noise ratio problem in increasing the frequency bandwidth due to the high attenuation of cancellous bone. At the present time no satisfactory explanation exists for the variation of attenuation with frequency. However, Wear[32] has reported markedly different frequency dependences for calcaneal attenuation and backscatter coefficients. He developed a scattering model in which the trabeculae are modeled as long thin cylinders with a diameter small compared with the ultrasonic wavelength, and obtained the experimental results in good agreement with the theoretical predictions in which scattering should lead to the frequency-dependent attenuation varying as f^3. Since attenuation is the result of both

absorption and scattering, this result suggests that absorption is dominant and that scattering is comparatively small.[33]

The MBA model for a nonrigid porous medium predicts the nonlinear relationship between BUA and porosity with maxima at the porosities of approximately 55–60%. Clarke *et al.*[34] demonstrated that attenuation increases in the low porosity region; it passes through a peak at roughly 50%, and it decreases in the high porosity region. Several experimental studies[35–40] also demonstrated that BUA is the profoundly nonlinear function of porosity in cancellous bone and bone phantom materials. It suggests that this nonlinear behavior is the general feature of fluid-saturated porous media. The role of scattering and absorption in the attenuation of ultrasound provides a qualitative explanation for the nonlinear relationship of BUA with density. Han *et al.*[36] and Serpe *et al.*[37] suggested that attenuation in low density cancellous bone and in cortical bone may be primarily due to absorption, compared with scattering important only in dense cancellous bone. Nonlinear trends in BUA rising to maxima at the porosities of 65–70% were reported by Nicholson *et al.*[13] using a simple scattering model in cancellous bone. This scattering model provides strong evidence for the important role of scattering in cancellous bone. It is demanded to develop a more detailed understanding for the scattering mechanism in cancellous bone.

VI. CONCLUSIONS

Acoustic wave propagation in bovine cancellous bone is experimentally and theoretically investigated in the frequency range of 0.5–1 MHz. The phase velocity, attenuation coefficient, and BUA of cancellous bone are measured as functions of frequency and porosity. For a theoretical estimation, the MBA model is employed with three new phenomenological parameters: the boundary condition, phase velocity, and impedance parameters. The relative orientation between the ultrasonic beam and the bone specimen is the ML direction of the bovine tibia. Since the fast and the slow waves are completely overlapped in this direction and are observed as if one longitudinal wave propagates, the theoretical estimation is performed in terms of the total wave mixed with the fast and the slow waves using the MBA model. It is experimentally observed that the phase velocity is approximately nondispersive and the attenuation coefficient linearly increases with frequency. The MBA model predicts a slightly negative dispersion of phase velocity linearly with frequency and the nonlinear relationships of attenuation and BUA with porosity. The experimental results are in good agreement with the theoretical results estimated with the MBA model.

It is expected that the MBA model can be usefully employed in the field of clinical bone assessment for the diagnosis of osteoporosis. Future studies will be required to address some of the limitations of the present study. The most significant limitation involves the fact that all measurements were performed on specimens with relatively high bone mineral densities compared with human cancellous bone. The denser bovine cancellous bone than human cancellous bone may cause ultrasonic absorption and scattering in fundamentally different mechanisms. Repeating this study on specimens with much lower density, such as osteoporotic bone, by using a similar *in vitro* procedure would be valuable.

ACKNOWLEDGMENTS

This work was supported by a Korea Research Foundation grant (KRF-2000-015-DP0178) and the BK21 Program of the Ministry of Education.

[1] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated solid. II. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[2] M. Schoenberg, "Wave propagation in alternating solids and fluid layers," Wave Motion **6**, 303–320 (1984).
[3] T. J. Haire and C. M. Langton, "Biot theory: a review of its application to ultrasound propagation through cancellous bone," Bone **24**, 291–295 (1999).
[4] J. L. Williams, "Ultrasonic wave propagation in cancellous and cortical bone: prediction of some experimental results by Biot's theory," J. Acoust. Soc. Am. **91**, 1106–1112 (1992).
[5] A. Hosokawa and T. Otani, "Ultrasonic wave propagation in bovine cancellous bone," J. Acoust. Soc. Am. **101**, 558–562 (1997).
[6] A. Hosokawa and T. Otani, "Acoustic anisotropy in bovine cancellous bone," J. Acoust. Soc. Am. **103**, 2718–2722 (1998).
[7] E. R. Hughes, T. G. Leighton, G. W. Petley, and P. R. White, "Ultrasonic propagation in cancellous bone: a new stratified model," Ultrasound Med. Biol. **25**, 811–821 (1999).
[8] K. A. Wear, "A stratified model to predict dispersion in trabecular bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 1079–1083 (2001).
[9] K. Y. Wu, Q. Xue, and L. Adler, "Reflection and transmission of elastic waves from a fluid-saturated porous solid boundary," J. Acoust. Soc. Am. **87**, 2349–2358 (1990).
[10] W. Xu and J. J. Kaufman, "Diffraction correction methods for insertion ultrasound attenuation estimation," IEEE Trans. Biomed. Eng. **40**, 563–570 (1993).
[11] G. W. Petley, P. A. Robins, and J. D. Aindow, "Broadband ultrasonic attenuation: are current measurement techniques inherently inaccurate?," Br. J. Radiol. **68**, 1212–1214 (1995).
[12] R. Strelitzki, P. H. Nicholson, V. Paech, "A model for ultrasonic scattering in cancellous bone based on velocity fluctuations in a binary mixture," Physiol. Meas **19**, 189–196 (1998).
[13] P. H. F. Nicholson, R. Strelitzki, R. O. Cleveland, and M. L. Bouxsein, "Scattering of ultrasound in cancellous bone: predictions from a theoretical model," J. Biomech. **33**, 503–506 (2000).
[14] H. S. Roh, K. I. Lee, and S. W. Yoon, "Acoustic characteristics of a non-rigid porous medium with circular cylindrical pores," J. Korean Phys. Soc. **43**, 55–65 (2003).
[15] K. Attenborough, "Acoustic characteristics of rigid fibrous absorbents and granular materials," J. Acoust. Soc. Am. **73**, 785–799 (1983).
[16] C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, Amsterdam, 1949).
[17] S. B. Lang, "Ultrasonic method for measuring elastic coefficients of bone and results on fresh and dried bovine bone," IEEE Trans. Biomed. Eng. **BME-17**, 101–105 (1970).
[18] K. A. Wear, "Ultrasonic attenuation in human calcaneus from 0.2 to 1.7 MHz," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 602–608 (2001).
[19] G. W. C. Kaye and T. H. Laby, *Table of Physical and Chemical Constants* (Longman, London, 1973).
[20] C. C. Gluer, C. Y. Wu, M. Jeregas, S. A. Goldstein, and H. K. Genant, "Three quantitative ultrasound parameters reflect bone orientation: an *in vitro* study," Calcif. Tissue Int. **55**, 46–52 (1994).
[21] D. Hans, M. E. Arlot, A. M. Schott, J. P. Roux, P. O. Kotzki, and P. J. Meunier, "Do ultrasound measurements on the os calcis reflect more the bone microarchitecture than the bone mass? A two-dimensional histomorphometric study," Bone **16**, 295–300 (1995).
[22] R. Strelitzki and J. A. Evans, "On the measurement of the velocity of ultrasound in the os calcis using short pulses," Eur. J. Ultrasound **4**, 205–213 (1996).
[23] P. H. F. Nicholson, C. G. Lowet, C. M. Langton, J. Dequeker, and G. Van der Perre, "Comparison of time-domain and frequency-domain approaches to ultrasonic velocity measurements in trabecular bone," Phys. Med. Biol. **41**, 2421–2435 (1996).

[24] P. Droin, G. Berger, and P. Laugier, ''Velocity dispersion of acoustic waves in cancellous bone,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 581–592 (1998).
[25] K. A. Wear, ''Measurements of phase velocity and group velocity in human calcaneus,'' Ultrasound Med. Biol. **26**, 641–646 (2000).
[26] F. J. Fry and J. E. Barger, ''Acoustical properties of the human skull,'' J. Acoust. Soc. Am. **63**, 1576–1589 (1978).
[27] S. Pal, S. Saha, and G. N. Reddy, ''Frequency dependence of ultrasonic characteristics of cancellous bone,'' in *Biomedical Engineering* (Pergamon, London, 1982).
[28] M. O'Donnell, E. T. Jaynes, and J. G. Miller, ''Kramers–Kronig relationship between ultrasonic attenuation and phase velocity,'' J. Acoust. Soc. Am. **69**, 696–701 (1981).
[29] S. Chaffai, F. Padilla, G. Berger, and P. Laugier, ''In vitro measurement of the frequency-dependent attenuation in cancellous bone between 0.2 and 2 MHz,'' J. Acoust. Soc. Am. **108**, 1281–1289 (2000).
[30] B. K. Hoffmeister, S. A. Whitten, and J. Y. Rho, ''Low-megahertz ultrasonic properties of Bovine cancellous bone,'' Bone **26**, 635–642 (2000).
[31] K. I. Lee, H. S. Roh, and S. W. Yoon, ''Correlations between acoustic properties and bone density in bovine cancellous bone from 0.5 to 2 MHz,'' J. Acoust. Soc. Am. **113**, 2933–2938 (2003).
[32] K. A. Wear, ''Frequency dependence of ultrasonic backscatter from human trabecular bone: theory and experiment,'' J. Acoust. Soc. Am. **106**, 3659–3664 (1999).
[33] K. A. Wear, ''Anisotropy of ultrasonic backscatter and attenuation from human calcaneus: implications for relative roles of absorption and scattering in determining attenuation,'' J. Acoust. Soc. Am. **107**, 3474–3479 (2000).
[34] A. J. Clarke, J. A. Evans, J. G. Truscott, R. Milner, and M. A. Smith, ''A phantom for quantitative ultrasound of trabecular bone,'' Phys. Med. Biol. **39**, 1677–1687 (1994).
[35] J. A. Evans and M. B. Tavakoli, ''Ultrasonic attenuation and velocity in bone,'' Phys. Med. Biol. **35**, 1387–1396 (1990).
[36] S. Han, J. Y. Rho, and I. Ziv, ''Ultrasound velocity and broadband attenuation over a wide range of bone mineral density,'' Osteoporosis Int. **6**, 291–296 (1996).
[37] L. Serpe and J. Y. Rho, ''The nonlinear transition period of broadband ultrasound attenuation as bone density varies,'' J. Biomech. **29**, 963–966 (1996).
[38] J. M. Alves, W. Xu, D. Lin, R. S. Stiffert, J. T. Ryaby, and J. J. Kaufman, ''Ultrasonic assessment of human and bovine trabecular bone: a comparison study,'' IEEE Trans. Biomed. Eng. **43**, 249–258 (1996).
[39] R. Strelitzki, J. A. Evans, and A. J. Clarke, ''The influence of porosity and pore size on the ultrasonic properties of bone investigated using a phantom material,'' Osteoporosis Int. **7**, 370–375 (1997).
[40] J. Toyras, H. Kroger, and J. S. Jurvelin, ''Bone properties as estimated by mineral density, ultrasound attenuation, and velocity,'' Bone **25**, 725–731 (1999).

PROGRAM OF

The 146th Meeting of the Acoustical Society of America

Austin, Texas • Renaissance Austin Hotel • 10–14 November 2003

NOTE: All Journal articles and Letters to the Editor are peer reviewed before publication. Program abstracts, however, are not reviewed before publication, since we are prohibited by time and schedule.

MONDAY MORNING, 10 NOVEMBER 2003 PECOS ROOM, 8:55 A.M. TO 12:00 NOON

Session 1aAO

Acoustical Oceanography: Developments in Multibeam Sonar for Water-Column Measurements I

Kenneth G. Foote, Chair
Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Chair's Introduction—8:55

Invited Papers

9:00

1aAO1. Application of multibeam sonar in marine ecology and fisheries research: New fields and limitations. Francois M. Gerlotto, Patrice A. Brehmer (IRD Ctr., Halieutique Mediterraneen et Tropical, Ave. Jean Monnet, B.P. 171, 34203 Sete Cedex, France, gerlotto@ird.fr), Paul G. Fernandes, David G. Reid, Philip Copland (Marine Lab., Aberdeen AB11 9DB, Scotland, UK), Stratis Georgakarakos (Inst. of Marine Biol. of Crete, Heraklion, Greece), and Jorge Paramo (INPESCA, Talcahuano, Chile)

Multibeam sonars have been used since the mid 90s for three dimensions and a dynamic observation of fish schools and shoals in their environment. A 455 kHz Reson Seabat 6012, with 60 beams of 1.5×22 deg, was used. It has allowed the quantification of school avoidance in acoustic surveys; the three-dimensional description of school structure and position in the water column; the noninvasive study of the real schooling behavior of fish in the wild; the observation and quantification of fish in relation to survey gear; and the predation on mussels in aquaculture by fish schools; the fish school distribution in relation to ecological factors, even in very shallow water. These observations showed that a multibeam sonar can be a useful tool for improving the quality of stock assessment surveys and studying the behavioral ecology of important commercial species. To date all these applications are essentially observational. The next stage in the development of these systems will be to produce quantitative biomass estimates. To achieve this, calibration systems and side aspect TS measurements for the fish will be required. The paper presents the current uses of this instrument described above, and discuss the research requirements for use in fisheries science.

9:25

1aAO2. *In situ* fish school studies enabled by multi-beam sonar. Ole Arve Misund (Inst. of Marine Res., P.O. Box 1870, N-5817 Bergen, Norway)

Since multi-beam sonar became available to fisheries research applications in the seventies, much insight has been accumulated on metric aspects of fish schools as geometry, packing density, and school biomass. Behavioral aspects such as swimming speed, migration and avoidance of vessels and fishing gears have also been a focus. In later years, ecosystem aspects of fish school studies such as spatial distribution, influence of physical features and interactions with predators and prey have been quantified. An update is given on how the knowledge on these various aspects of fish schools have been collected by use of multi-beam sonar. Prospects for further insight in the role of fish schools in marine ecosystems are considered in the light of the technological development of multi-beam sonar that is now taking place. Both the development of recording instruments and post-processing systems are considered.

9:50

1aAO3. Three-dimensional acoustic backscatter measurements in the water column with hull-mounted and towed multibeam echo-sounders. Christian de Moustier (Ctr. for Coastal & Ocean Mapping, Univ. of New Hampshire, 24 Colovos Rd., Durham, NH 03824), Timothy C. Gallaudet (USS KITTY HAWK (CV-63), FPO AP 96634-2770), and Daniel S. Brogan (Ctr. for Coastal & Ocean Mapping, Univ. of New Hampshire, Durham, NH 03824)

Multibeam echo-sounders have been used very effectively to map the seafloor for nearly 30 years but their use in water column measurements has received little attention. This is partly due to the usually narrow-band, single-frequency operation of bottom mapping sonars which is not well suited to water column applications, and to the 40–60 dB higher receiver sensitivity and gain needed for acoustic backscatter measurements in the water column, leading to echo saturation in seafloor mapping applications [de Moustier and Alexandrou, J. Acoust. Soc. Am. Suppl. 1 **83**, S48 (1988)]. Recent work with towed toroidal multibeam sonars [Gallaudet and de Moustier, J. Acoust. Soc. Am. **112**, 489–503 (2002)] has shown the potential benefits of 3-D water column acoustic backscatter measurements obtained with such sonars. However care must be taken to avoid multipath boundary interferences received through the sidelobes of beams pointed in the ocean volume that can significantly bias the measurements. Ocean volume reverberation measurements recorded with the original 12-kHz SeaBeam hull-mounted swath bathymetry sonar, with towed multibeam volume search sonars, are used to illustrate the spatial and temporal resolution capabilities and limitations of such measurements. [Work supported by NRL/ONR.]

10:15–10:30 Break

Contributed Papers

10:30

1aAO4. Hawaiian spinner dolphins aggregate midwater food resources through cooperative foraging. Kelly Benoit-Bird and Whitlow Au (Hawaii Inst. of Marine Biol., Univ. of Hawaii, P.O. Box 1106, Kailua, HI 96734, benoit@hawaii.edu)

To test the hypothesis that spinner dolphins in Hawaii may actively aggregate their prey through cooperative foraging, a 200-kHz multi-beam sonar (Simrad MS2000) was used to observe 323 groups of spinner dolphins foraging within a midwater, micronekton sound-scattering layer off Oahu. Strong cooperation was observed in groups of 8–14 pairs of spinner dolphins. The dolphin group size was highest at midnight when the density of prey was highest and was significantly higher in Makua Beach, where the prey density was higher, than Electric Beach, where the prey density was lower. Cooperative groups of dolphins aggregated their food resources, apparently using their preys' avoidance behavior to create distinct, high-density patches in the prey. Prey aggregation was strongly stereo-typed, regardless of the distribution of the scattering layer. Dolphins swam around the edge of a 28–40 m diameter circle at least 5 times, concentrating prey within this area before pairs of dolphins on opposite sides of the circle swapped positions in the circle, swimming through the high density prey 'donut' they had formed. The hypothesis that nocturnal animals aggregate prey in midwater could not have been tested without the three-dimensional information on prey distribution and dolphin geometry provided by the multi-beam.

10:45

1aAO5. Schooling in pelagic fish: An approach in three dimensions using vertical scanning multibeam sonar. Jorge Paramo, Franois Gerlotto, and Ciro Oyarzun (Instituto de Investigacin Pesquera, Casilla 350, Talcahuano, Chile; Institut de Recherche pour le Developpement, Univ. Concepcin, jparamo@inpesca.cl)

In fish resource assessment, it is very important to know about the behavior and the form in that schools are occupying the space. The present paper describes the morphology and internal structure in pelagic schools using vertical scanning multibeam sonar. This sonar provides with data making it possible to reconstruct the school in three-dimensions. Acoustic surveys were carried out in Venezuela, Senegal, and Mexico. Geometrical (position of the school), morphological, density and internal structure parameters were extracted from a total of 668 schools: 257 from Mexico, 343 from Venezuela, and 68 from Senegal. The analysis of fish schools avoidance effect caused by vessels showed two modes in school frequency distribution in the case of Venezuela and Senegal schools and three modes for Mexico schools. This frequency distribution was decreasing as a vessel distance increase. In general, the schools show an amoeboid shape, which cannot be simply described with some geometric volume (e.g., spheres or ellipsoids). Also, it shows a contagious internal structure, i.e., fishes are not positioned at random inside the school. Another interesting characteristic is the presence of empty spaces inside the school that are called vacuoles.

11:00

1aAO6. Inferring the vertical turbulent diffusion coefficient from backscatter measurements with a multibeam sonar. Thomas Weber, David Bradley, R. Lee Culver, and Anthony Lyons (Appl. Res. Lab. and the Grad. Prog. in Acoust., P.O. Box 30, State College, PA 16804, tcw141@psu.edu)

A 240 kHz multibeam sonar was deployed on a sub-surface mooring in order to collect acoustic backscatter measurements from the bubbles in a ships wake in August, 2002. The sonar was located at a depth of 30 m and oriented so that its swath of beams was orthogonal to the ships track. Measurements proportional to the acoustic scattering cross-section per unit volume were made and used to estimate the vertical turbulent diffusion coefficient in the ships wake using the same methodology that Thorpe used with a single-beam echosounder underneath breaking waves [J. Phys. Ocean. **14**, 855–863 (1984)]. The wake was viewed as a distributed source of bubbles that are transported downward from the surface by turbulence. A steady state representation of the bubble field was used, where it was assumed that the bubble population exponentially decayed with depth, and that a bubble conservation equation could be written that only accounted for turbulent diffusion, buoyancy, and gas dissolution. Estimates of the vertical turbulent diffusion coefficient in the wake of a research vessel will be presented. [Work supported by ONR Code 321US. First author supported by a National Defense Science and Engineering Graduate Fellowhip.]

11:15

1aAO7. Water-column measurements of hydrothermal vent flow and particulate concentration using multibeam sonar. Christopher D. Jones (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, cjones@apl.washington.edu)

High-frequency multibeam systems enable water-column studies of fluid flow from deep ocean hydrothermal vent systems. Backscatter from high-temperature particulate-laden vent fluid (black smokers) is used to obtain three-dimensional images of the plume's temporal and spatial struc-

ture. Backscattering intensity may provide a means to remotely estimate plume particulate concentration and turbulent micro-structure created by temperature and density fluctuations in the plume fluid, providing denser sampling in space and time than conventional point measurements. Acoustic Doppler measurements can provide estimates of the entrainment and transport of water, particulate matter, and larvae by hydrothermal systems, which is of interest in physical, chemical, and biological oceanography. Toward the goal of providing an acoustic means of inverting for such oceanographic parameters with a multibeam system, several mechanisms and models of volume scattering from the water column will be discussed. The results from field experiments with a calibrated 200-kHz multibeam system (Simrad SM2000) will be presented, as well as future developments and applications of mutlibeam systems to water-column scattering. [Work supported by NSF.]

11:30

1aAO8. Echoview as a multibeam sonar data processing and analysis toolkit for fisheries research. Bart Buelens, Tim Pauly, and Ian Higginbottom (SonarData Pty. Ltd., GPO Box 1387, Hobart 7001, Tasmania, Australia)

Echoview is a hydroacoustic data analysis software package, widely used in the fisheries research and stock assessment communities. Originally developed to handle a variety of single-beam sonar data formats, Echoview has been extended to support multibeam data. Multibeam data logging, lossless compression, and real time beamforming and display are some of the software's core features. Multibeam data has an additional dimension compared to single-beam data, and a 3D data viewer has been developed providing 3D visualizations of the seabed and fish schools detected by built-in algorithms. Since the multibeam module is just one of many software modules of the Echoview package, data from other sources such as single-beam sonar systems and current profilers can be combined and analyzed together with the multibeam data. The combination of coincident fish density estimates from calibrated single-beam backscatter data with school volume estimates from multibeam data will represent a significant improvement in stock assessment methods. Ongoing research and development will make it possible for Echoview to follow and even set new trends in multibeam water-column data analysis for fisheries research. Features under development include calibration, vessel motion compensation, improved feature detection, and enhanced and animated 3D displays.

11:45

1aAO9. 3D visualization of marine organism distribution and movement in water column using virtual reality techniques. Marek Moszynski, Andrzej Stepnowski, and Krzysztof Bikonis (Gdansk Univ. of Technol., Narutowicza 11/12, 80-952 Gdansk, Poland)

Virtual reality techniques mainly developed for computer graphics had found numerous applications in various scientific areas like chemistry, astronomy, geography, etc. The paper proposes application of the virtual reality techniques in hydroacoustics for 3D visualization of data acquired by split-beam or multi-beam echosounders. Both types of echosounders allow for nonambiguous localization of targets in the water column and their imaging. As the multi-beam systems also delivers bathymetric information, in addition to imaging of the spatial distribution, it is possible to create complete virtual scene of sampled volume. Using builtin properties of Virtual Reality Modeling Language (VRML) it was possible to animate movement of fishes and other marine organisms in the water column. As the language is supported by Internet technology, its textlike description allows for easy rendering by most popular WWW browsers.

MONDAY MORNING, 10 NOVEMBER 2003 SABINE ROOM, 8:30 TO 11:30 A.M.

Session 1aUW

Underwater Acoustics: Modeling of Scattering and Reverberation

Kevin D. LePage, Chair
Naval Research Laboratory, Code 7144, 4555 Overlook Avenue Southwest, Washington, D.C. 20375

Contributed Papers

8:30

1aUW1. A hybrid model for the three-dimensional scattering from objects in underwater waveguides. Mario Zampolli, David S. Burnett, Finn B. Jensen (NATO Undersea Res. Ctr., Viale San Bartolomeo 400, 19138 La Spezia, Italy, zampolli@saclantc.nato.int), Henrik Schmidt (MIT, Cambridge, MA 02139), and John B. BlottmanIII (Naval Undersea Warfare Ctr., 1176 Howell St., Newport, RI 02841)

The scattering from objects in underwater waveguides is a multi-scale problem, involving both near-field effects in the vicinity of the scatterer as well as long-range propagation through the waveguide. To solve this problem, 3-D Finite-Element STructural Acoustics software developed at SACLANTCEN (FESTA) and an underwater waveguide propagation model based on wavenumber integration developed at MIT (3-D OASES), are coupled into a hybrid model. In a three-step method, the propagation model is used to compute the incident acoustic field in the vicinity of the target, which may be floating, proud, partially buried or buried in the sediment. The incident field data is subsequently passed as an input to the finite-element tool to compute the target-scattered acoustic nearfield. In the final step, the scattered field is propagated through the waveguide by OASES. A second method of coupling between the two models is based on the characterization of the target scattering via spherical harmonic basis responses. The advantage of the second method is that the finite-element computations need to be performed only once for each frequency, regardless of the incident field. Results for different targets with multistatic source–receiver configurations and with focused acoustic incident fields are presented.

8:45

1aUW2. Separating volume and roughness components of scattering from the seafloor using spatial coherence filtering. Anatoliy N. Ivakin, Darrell R. Jackson, and Dajun Tang (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

Seabed scattering is usually associated with two mechanisms, interface roughness and volume heterogeneity, which can operate at the same time. In this paper, a method of separately identifying these two mechanisms is

presented, involving measurements of spatial coherence of the field with a vertical linear array located near the source. Analytical expressions are derived for the spatial coherence of interface and volume scattering and used to illustrate the potential for their separating and quantifying. It is shown that the correlation scales for the volume and roughness components can be significantly different, and thus spatial filtering can be used for their separating. A Monte Carlo simulation technique based upon the point scatterer model is also developed that makes it possible to generate random time series for the scattered field. An algorithm for data processing is proposed to estimate spatial correlation as an average over realizations (from separate pings) of the measured time series. The simulations illustrate the ability of the algorithm to distinguish and separate volume and surface components of scattering and also verify the analytical expressions for spatial coherence. It is shown that the convergence to the theoretical prediction for both volume and surface components of scattering indeed takes place, but longer pulses require a larger number of measurements (realizations). [Work supported by ONR Code 321OA.]

9:00

1aUW3. Analytic expressions for time-domain forward propagation through a waveguide with random inhomogeneities including causality and dispersion relations. Tianrun Chen, Purnima Ratilal, and Nicholas C. Makris (MIT, Cambridge, MA 02139)

Analytic expressions in the time domain are derived for the mean forward field propagated through a waveguide containing random volume and surface inhomogeneities. The results are expressed in terms of the modal attenuation and dispersion coefficients in a waveguide [P. Ratilal and N. C. Makris, **112**, 2403 (2002)]. Simulations in a shallow-water waveguide containing bubbles for a finite-time duration pulse show that the scattering leads to additional time delay and distortion of the signal waveform. Stationary phase approximations are also applied to represent the time-domain field in terms of the modal group velocities that are smaller than those in a waveguide without inhomogeneities. We show that the dispersion and attentuation effects cannot be explained by heuristic results based on an effective medium sound speed computed using bulk moduli and volume fractions since it does not account for scattering from inhomogeneities in the medium. The time-domain expressions for the forward propagated field obey causality and are consistent with Kramer–Kronig relations when used in their range of validity.

9:15

1aUW4. Range-dependent reverberation modeling with the parabolic equation. Purnima Ratilal, Sunwoong Lee, and Nicholas C. Makris (MIT, Cambridge, MA 02139)

Propagation and scattering are generally convolved in a waveguide. For scatterers small compared to the wavelength, however, the waveguide Green functions to and from the scatterer can be separated from the free space scattering amplitude [Ratilal *et al.*, J. Acoust. Soc. Am. **112**, 1797–1816 (2002)]. In this case, waveguide scattering can be modeled using efficient and accurate range-dependent methods such as the parabolic equation (PE) for the waveguide Green function. This approach is implemented for scattering from small particles in the water column and seabed at arbitrary bistatic orientations. Scenarios investigated involving targets of large impedance contrast include scattering from bubbles and fish schools. When the Rayleigh–Born single scatter approximation is valid, PE-based range-dependent scattering can also be used to scatter from elemental water-column and seabed inhomogeneities. This approach is used to model bistatic reverberation from the water column and seabed including returns arising from both diffuse and discrete scattering processes. The former is associated with smoothly decaying reverberation, the latter with clutter. Also, the possibility that internal waves may lead to discrete seafloor scattering returns by focusing energy on the ocean bottom is also investigated with this model.

9:30

1aUW5. Effects of refraction on the prediction of bistatic reverberation in range dependent shallow water waveguides. Kevin D. LePage (NRL Code 7144, 4555 Overlook Ave. SW, Washington, DC 20375) and Chris Harrison (SACLANT Undersea Res. Ctr., Vle. San Bartolomeo 400, 19138 La Spezia, Italy)

The BiStaR bistatic reverberation model has been benchmarked against analytic formulas for bistatic reverberation in isovelocity range dependent waveguides. Studies conducted using BiStaR to evaluate the effects of introducing refraction into the benchmark scenarios indicate that downward refraction can have a tremendous effect on the time-angle evolution of mono- and bistatic reverberation, especially in shoaling directions. Here recent advances in analytic formulas for propagation and reverberation in refractive range dependent waveguides are exercised to evaluate analytically the predicted changes in RL caused by the different propagation characteristics. The analytic results are compared with the numerical results and conclusions are drawn regarding the propagation regimes and mechanisms that cause the strong deviations from the isovelocity predictions. [Work supported by ONR and SACLANTCEN.]

9:45

1aUW6. Kirchhoff approximation for backscattering by partially illuminated circular cylinders: Two-dimensional case. Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

The Kirchhoff approximation is widely used for estimating reflection contributions to the backscattering by rigid surfaces [W. G. Neubauer, J. Acoust. Soc. Am. **35**, 279–285 (1963)]. Backscattering of a plane wave by a right-circular cylinder has been approximated this way using Struve and Bessel functions [G. C. Gaunaurd, IEEE J. Ocean Eng. **10**, 213–230 (1985)]. For various situations of interest, the cylinder may be partially illuminated such that the illumination is abruptly truncated along lines parallel to the cylinder's axis. This configuration does not appear to have been analyzed previously. This two-dimensional scattering problem is reformulated by an application of the plane-wave expansion theorem to a phase factor. The scattering is expressed in terms of an infinite series of terms containing Bessel functions and elementary trigonometric integrals. An evaluation of the integrals and numerical evaluation of the series for the fully illuminated case reduces to the aforementioned Struve and Bessel function result. The partial-illumination case exhibits Fresnel zone behavior and illustrates the limitations of quadratic phase approximations. [Work supported by ONR.]

10:00–10:15 Break

10:15

1aUW7. Uniform approximation method for backscattering of sound from a target in a caustic field. Ben Dzikowicz and Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99163)

In certain cases, underwater backscattering of sound by targets in a caustic field associated with reflection from curved surfaces takes the form of a diffraction integral over the product of a wavefront amplitude and a phase factor. The phase factor can be expressed using the relative return

times of a transient pulse (B. Dzikowicz, Ph.D. dissertation, Washington State University, Pullman, WA, 2003). In a homogeneous, isotropic medium, the variation in ray return angles can be related to geometrically predicted ray amplitudes at the receiver. These geometrically predicted amplitudes as a function of target position can be recovered from the diffraction integral by using the results of a uniform approximation [C. Chester, B. Friedman, and F. Ursell, Proc. Cambridge Philos. Soc. **53**, 599–611 (1957)]. The result is that for targets near an Airy caustic, the coefficient of the Airy function amplitude factor depends on the derivatives of ray return angles. The improvement using this technique for two merging rays (a target at an Airy caustic) versus a simpler model (where the aforementioned coefficient is constant) will be discussed and compared to experiment for the single-bounce echo. [Work supported by ONR.]

10:30

1aUW8. Scattering calculations for a sphere ensonified by a fully depth sampling TRM in a shallow water waveguide. Kevin D. LePage and B. Edward McDonald (NRL Code 7140, 4555 Overlook Ave. SW, Washington, DC 20375)

One potential application of a Time Reversal Mirror (TRM) is as an acoustic tripwire [Kuperman *et al.*, IEEE J. Oceanic Eng (submitted)] where the focus of the TRM at the probe source range and depth is degraded by the presence of an object in the waveguide. Here an efficient technique for approximating the time domain scattered field of a rigid or elastic sphere in a shallow water waveguide is presented based on the Ingenito model [J. Acoust. Soc. Am. **82**, 2051–2059 (1987)], a WKB decomposition of normal modes into up- and down-going plane waves at the scatterer depth, and the narrowband approximation for the time domain signal broadcast from a fully depth sampling TRM. Results show that the degradation of the TRM focusing near the focal region is as expected a strong function of the scatter size and frequency, as well as the usual parameter of the critical angle in the bottom. Calculations in the nearfield of the scatterer show shadowing for spheres above $ka=1$ of all the pseudo plane wave components of the TRM field, which yields an interesting evolution of the forward scattered field as a function of time, angle, and scatterer depth. [Work supported by ONR.]

10:45

1aUW9. Space–time variations of the acoustic field scattered from a penetrable isovelocity wedge. C. Feuillade (Naval Res. Lab., Stennis Space Center, MS 39529-5004), D. Chu (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), and C. S. Clay (Univ. of Wisconsin–Madison, Madison, WI 53706)

The formalism of D. Chu [J. Acoust. Soc. Am. **86**, 1883–1896 (1989)] is used to perform an investigation of the spatial and time domain variations of the scattered field of a penetrable isovelocity wedge. The direct, reflected, and diffracted field components are considered, and the relationships between them are investigated. The density contrast surface is an appropriate model for many water–seafloor interfaces and, since diffractions often dominate scattering from rough surfaces, an emphasis is placed on achieving an improved description and understanding of these waves. Particular emphasis is placed on a detailed study of the wave phenomena and transition behavior around the forward scattering shadow boundary, where the causal relationship between the direct incident field and the diffracted field is specifically examined. At the apex the two fields coincide, and their propagational effects here, and also above and below the geometrical shadow zone, lead to complicated temporal and azimuthal variations of the field. The origin and behavior of diffractions associated with specular reflections of the source in the faces of the wedge is also studied. [Work supported by ONR.]

11:00

1aUW10. Statistical characterization of active sonar reverberation using extreme value theory. Brian La Cour (Appl. Res. Labs., Univ. of Texas, 10000 Burnet Rd., Austin, TX 78758)

The statistics of reverberation in active sonar are characterized by non-Rayleigh amplitudes in the normalized matched filter output. Unaccounted for, this property can lead to high false alarm rates in fixed-threshold detectors. A new approach to modeling threshold-crossing statistics based on extreme value theory is proposed which uses the generalized Pareto distribution as the unique, asymptotic model of the tail distribution, valid at large thresholds. Methods of parameter estimation are discussed and applied to active sonar reverberation collected on a hull-mounted sonar system. The statistics of reverberation in active sonar are found to generally have a power-law behavior in the tails with a shape parameter that is persistent in time and bandwidth dependent. The threshold needed for accurate parameter estimation is generally found to be well below that of typical fixed-threshold detectors.

11:15

1aUW11. Pseudo-hybrid PDF estimators for active sonar classification. James M. Gelb, Mark A. Donnell, and Brian R. La Cour (Appl. Res. Lab., Signal Phys. Div., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

A class of multivariate PDF estimators for use in feature-based target classification in active sonar is presented. The goal is to develop a method that (1) preserves feature identity, (2) captures complex data structures and tails using non-parametric estimates, (3) utilizes parametric models for higher-order inter-feature dependencies, and (4) provides a systematic enhancement to classifier performance given limited target data. The estimators are labeled pseudo-hybrid because (1) they are not true PDFs but they have sensible limiting behavior and (2) they combine non-parametric and parametric models. The simplest form is: $P_{\mathrm{PH}}(\mathbf{f})=\Pi_{i=1}^{N}p_i(f_i)\times M_N(\mathbf{f})/\Pi_{i=1}^{N}M_i(f_i)$, where $\mathbf{f}$ is the feature set of N features, p's are the marginal distributions, and M_N is a parametric model with marginals M_i. The denominator was chosen so that the overall expression reduces to the correct PDF in the limits that the p's are independent or M_N happens to be the correct model. The efficacy of this PDF estimator to mimic the sampled, test distributions will be demonstrated. Other pseudo-hybrid PDF estimators will also be discussed.

Session 1pAA

Architectural Acoustics and Speech Communication: Designing for Speech—Both Intelligibility and Privacy

Kenneth P. Roy, Chair
Innovation Center, Armstrong World Industries, 2500 Columbia Avenue, Lancaster, Pennsylvania 17604

Chair's Introduction—1:00

Invited Papers

1:05

1pAA1. Measurement of speech privacy in closed offices. Kenneth P. Roy (Armstrong Innovation Ctr., 2500 Columbia Ave., Lancaster, PA 17604)

Two closed offices of different sizes were constructed using a demountable wall system as a mockup for a planned new courthouse facility. The occupants of the existing fixed construction facility were concerned about the possible impact of the new wall system on the speech privacy performance. Field measurements were made to determine the speech privacy performance in terms of the PI rating. Issues relating to the application of ASTM E 1130 for closed office measurements will be discussed.

1:25

1pAA2. Classroom acoustic evaluation and field study. Kenneth P. Roy and Micaelina Campos (Armstrong Innovation Ctr., 2500 Columbia Ave., Lancaster, PA 17604)

The Saguaros Corazones de Alameda school in Santiago, Chile had a problem—bad acoustics. Both the background noise levels and reverberation time were much too high. The children had a difficult time understanding the teacher such that their attention was less than exemplary. Teachers had a significant absentee rate due to vocal strain issues. A sound reinforcement system was being considered—but architectural acoustic treatments were substituted with very good results. See the video and hear the results that good architectural design can make.

1:45

1pAA3. Sound system coverage and intelligibility using AFR devices. Peter Mapp (Peter Mapp Assoc., Colchester C03 4JZ, UK, petermapp@btinternet.com)

AFR loudspeaker technology (audio full-range) combines the characteristics of distributed mode loudspeakers at mid and high frequencies with conventional loudspeaker piston action at low frequencies in a single driver. The combination retains the well-established advantages of the DML in respect of wide dispersion and improved interaction with reflective boundaries. A comparison study has been made using conventional 6-in. ceiling loudspeakers and AFR devices of comparable size in a typical open plan office situation. It is shown that the AFR devices exhibited noticeably improved coverage, clarity, and speech intelligibility. However, the use of traditional intelligibility measures, such as STI did not always reflect this subjectively observed trend. The use of STI or STIPa under high signal to noise conditions is questioned.

2:05

1pAA4. Frequency response and systematic errors in STI measurements. Peter Mapp (Peter Mapp Assoc., Colchester CO3 4JZ, UK, petermapp@btinternet.com)

The paper examines two fundamental problems when measuring and verifying the STI of sound systems. These relate to the overall accuracy of the measurement platform in use and the fact that STI does not correctly account for system frequency response and spectral aberrations under certain conditions. As has previously been reported by the author, STI appears to significantly overestimate the intelligibility of limited bandwith sound systems under reverberant, high S/N ratio conditions. Further evidence and word score comparisons for a number of different frequency response conditions, measured under several reverberant situations are

presented. The results indicate that errors of up to 30 comparisons with low-reverberant field and anechoic conditions are also made. It is also shown that the measurement platform itself and sound source characteristics can affect the resultant error. Tests, carried out under highly controlled conditions, are reported that enable the causes of the potential errors to be established. The implications of reported limitations in STI and STIPa, when assessing the potential intelligibility of a sound system or acoustics of a space, are discussed.

2:25

1pAA5. Auralization studies on talker comfort. Jessica Shearer and Rendell R. Torres (Prog. in Arch. Acoust., Rensselaer Polytechnic Inst., Greene Bldg., 110 8th St., Troy, NY 12180-3590)

Although much research has focused on determining optimal acoustical environments for students in classrooms, relatively little has addressed the classroom as an acoustical workspace for teachers, who may suffer from stress and vocal strain due in part to poor acoustical environments. Although the primary problem is typically the background noise level (whether due to ventilation or students), it is also interesting to study systematically how controlling early reflections may improve the audible ''room response'' at the teacher's speaking location without inordinately increasing the reverberant level of the background noise. Moreover, the room response at the talker's position may help reduce the perceived need to strain the voice, as long as the reflections are not so delayed as to be disturbing. In this study approximately ten configurations of absorptive and reflective surfaces in a ''typical-sized'' classroom are auralized in real time. For each room condition, subjects rate the ''talker comfort'' in terms of perceived loudness of their speech, possible disturbance from echoes or increased background noise, and other factors. The primary descriptive physical parameter is essentially the relative amplitude and delay of clusters of early reflections, which are not always well characterized by the classical room-acoustics descriptors. Initial results of the modeling and subject testing will be presented.

2:45

1pAA6. Predicting speech intelligibility in noise for hearing-critical jobs. Sigfrid D. Soli (House Ear Inst., 2100 W. 3rd St., Los Angeles, CA 90057), Chantal Laroche, and Christian Giguere (Univ. of Ottawa, 451 Smyth Rd., Ottawa, ON K1H 8M, Canada)

Many jobs require auditory abilities such as speech communication, sound localization, and sound detection. An employee for whom these abilities are impaired may constitute a safety risk for himself or herself, for fellow workers, and possibly for the general public. A number of methods have been used to predict these abilities from diagnostic measures of hearing (e.g., the pure-tone audiogram); however, these methods have not proved to be sufficiently accurate for predicting performance in the noise environments where hearing-critical jobs are performed. We have taken an alternative and potentially more accurate approach. A direct measure of speech intelligibility in noise, the Hearing in Noise Test (HINT), is instead used to screen individuals. The screening criteria are validated by establishing the empirical relationship between the HINT score and the auditory abilities of the individual, as measured in laboratory recreations of real-world workplace noise environments. The psychometric properties of the HINT enable screening of individuals with an acceptable amount of error. In this presentation, we will describe the predictive model and report the results of field measurements and laboratory studies used to provide empirical validation of the model. [Work supported by Fisheries and Oceans Canada.]

3:05–3:15 Break

3:15

1pAA7. Performance comparison of above ceiling and through ceiling sound masking systems. Carl Rosenberg, Christopher Savereid, and Thomas Horrall (Acentech, 33 Moulton St., Cambridge, MA 02138, crosenberg@acentech.com)

Sound masking systems are an important tool for providing speech privacy in offices and similar spaces. These systems typically use an array of loudspeakers radiating masking sound above an acoustical tile ceiling. The ceiling acts to diffuse the sound laterally in order to minimize localization on loudspeakers by the listeners below. The diffusion also helps keep the speech-to-noise ratio relatively uniform spatially, which may help minimize variations in privacy conditions. Unfortunately, the ceiling material, the plenum, and the space below are rarely homogeneous acoustically. There are usually openings in the ceiling for HVAC return air and lighting fixture ventilation, and different parts of the ceiling often use a variety of materials and elevations for aesthetic reasons. Large ducts and structural elements may compartmentalize the plenum. These variations pose substantial acoustical challenges for the designer and additional costs for the owner, not all of which can be satisfactorily resolved. Recognizing these issues, several manufacturers have recently introduced masking systems that might be characterized as direct field systems. These systems use masking loudspeakers that radiate directly into the occupied space rather than into a plenum. This paper contrasts the typical performance and other characteristics of the two basic types of systems.

3:35

1pAA8. Instrumentation for measuring speech privacy in rooms. Thomas Horrall, Rein Pirn, and Ben Markham (Acentech, 33 Moulton St., Cambridge, MA 02138)

Federal legislation pertaining to oral privacy in healthcare and financial services industries has increased the need for a convenient and economical way to document speech privacy conditions in offices, medical examination rooms, and certain other workspaces. This legislation is embodied in the Health Insurance Portability and Accountability Act (HIPAA) and Gramm-Leach-Bliley Act

(GLBA). Both laws require that reasonable measures be put in place to safeguard the oral privacy of patients and clients. While techniques for privacy documentation are known within the acoustical consulting community, it is unlikely that community alone has the capacity to provide the surveys needed to evaluate acoustical conditions and demonstrate compliance with the legislation. A portable computer with integrated soundboard and a suitable amplified loudspeaker and test microphone are all that are needed to perform *in situ* measurements of articulation index or other accepted indices of speech privacy. Along with modest training, such instrumentation allows technicians to survey a large number of sites economically. Cost-effective components are shown that can meet the requirements for testing in most common environments where oral privacy is likely to be required. Example cases are presented to demonstrate the feasibility of such instrumentation.

3:55

1pAA9. A new measure for assessment of architectural speech security. Bradford N. Gover and John S. Bradley (Inst. for Res. in Construction, Natl. Res. Council, Ottawa, ON K1A 0R6, Canada, brad.gover@nrc-cnrc.gc.ca)

A new measure has been developed to indicate the audibility and intelligibility of speech transmitted through walls from adjacent rooms. The new measure is a frequency-weighted signal-to-noise ratio. It is shown to be a more accurate predictor of subjective ratings than the Articulation Index, the Speech Intelligibility Index, and simple A-weighted signal-to-noise ratios. Listening tests using English sentences were conducted to measure the fraction of words intelligible to acute-hearing, native-English-speaking listeners under a range of acoustical conditions. Each subject listened to 500 sentences, each of which had been filtered to represent passage through some type of wall construction. The level of the speech, as well as the level and spectrum of background ventilation-type noise was varied. The conditions in the test ranged from those where all subjects were able to correctly identify all the words in a sentence (i.e., very poor security), through to those where all subjects were unable to even recognize the presence of a speaking voice (i.e., excellent security). The tests, analysis, and derivation of the measure will be discussed, as will the directions of ongoing work in the area.

4:15

1pAA10. Correlation of subjective and objective measures of speech intelligibility. Erica E. Bowden, Lily M. Wang, and Milena S. Palahanska (Architectural Eng. Prog., Univ. of Nebraska–Lincoln, Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182)

Currently there are a number of objective evaluation methods used to quantify the speech intelligibility in a built environment, including the Speech Transmission Index (STI), Rapid Speech Transmission Index (RASTI), Articulation Index (AI), and the Percentage Articulation Loss of Consonants (%ALcons). Many of these have been used for years; however, questions remain about their accuracy in predicting the acoustics of a space. Current widely used software programs can quickly evaluate STI, RASTI, and %ALcons from a measured impulse response. This project compares subjective human performance on modified rhyme and phonetically balanced word tests with objective results calculated from impulse response measurements in four different spaces. The results of these tests aid in understanding performance of various methods of speech intelligibility evaluation. [Work supported by the Univ. of Nebraska Center for Building Integration.] For Speech Communication Best Student Paper Award.

Contributed Papers

4:35

1pAA11. Evaluation of sound field systems in elementary school classrooms. Michelle C. Vigeant and Kelly Kruger (AB Infrastructure, 6950 113 St., Edmonton, AB T6H 5V7, Canada)

Our primary purpose in this study was to determine the relevant ergonomic issues associated with daily use of sound field systems in elementary school classrooms, in order to develop a purchasing guideline and technical specification. The secondary purpose was to evaluate these systems to identify if one or more acoustical parameters could be used to determine the quality and effectiveness of a system. Six sound field systems, with varying numbers and types of speakers, were chosen as a cross-section of available systems on the market. Six representative classrooms, currently in use, were selected based on a range of reverberation times and background noise levels. All systems were installed for two weeks in each classroom. Student speech intelligibility (SI) tests using phonetically balanced word lists were conducted, as well as teacher interviews. The acoustical parameters measured were clarity ratios C50 and C80, speech transmission indices STI and R(rapid)STI, sound pressure level (SPL) uniformity and frequency response. An improvement in SI was found for all systems. Only SPL uniformity and frequency response were found to be useful distinguishing performance parameters between systems. Ergonomic design aspects of sound field systems had a significant influence on the acceptance and usage in the classroom.

4:50

1pAA12. Analysis of practical sound transmission class (STC) measurements performed under a variety of field conditions. James Conroy (ENSCO, Inc., APA Div., 5400 Port Royal Rd., Springfield, VA 22151)

Sound transmission class (STC) measurements were made using a simplified version of the procedure presented in ASTM E336-96 and the results compared to STC values found in the literature. STC is used to provide a single number speech privacy rating for a building partition. It is calculated by measuring the transmission loss of a partition and applying a curve fitting procedure. While the standards provide a highly detailed data collection and processing procedure, they give limited insight into the type

of response and deviation that might be expected from measurements made along a partition or when comparing results from similar partitions under different field conditions. Likewise, while there is discussion in the literature on STC, with some attempt to relate values to human perception and provide ratings for common partitions there is a generally limited discussion involving data collected under real conditions. The partitions in three different office rooms were tested and the results described. In particular, the effect of speaker type and location, source type, room geometry, microphone placement, the spatial variation and repeatability of measurements on a particular partition, and a comparison of results from similar partitions under different field conditions were analyzed and discussed in detail.

MONDAY AFTERNOON, 10 NOVEMBER 2003 PECOS ROOM, 1:30 TO 3:15 P.M.

Session 1pAO

Acoustical Oceanography: Developments in Multibeam Sonar for Water-Column Measurements II

Kenneth G. Foote, Chair

Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Contributed Papers

1:30

1pAO1. Acoustic near field effects for the Simrad–Mesotech SM 2000 multibeam sonar. Norman A. Cochrane (Ocean Sci. Div., Dept. of Fisheries and Oceans, Bedford Inst. of Oceanogr., P.O. Box 1006, Dartmouth, NS B2Y 4A2, Canada) and Gary D. Melvin (St. Andrews Biological Station, St. Andrews, NB E5B 2L9, Canada)

Targets quantitatively observed with a multibeam sonar based on a truncated circular arc transducer geometry, focused at infinity, display anomalous apparent target strengths at short measurement ranges. Three near field mechanisms, affecting both the sonar transmit and receive responses, are identified: (1) Anomalous inter-elemental differential phase shifts, (2) inter-elemental differential spreading losses, (3) shifts in elemental viewing aspect. With reference to a commercial 200 kHz Simrad–Mesotech SM 2000 multibeam sonar, the relative importances of the three mechanisms are numerically examined at a 1-m target range. Combined effects in both transmit and in receive are further evaluated for target ranges between 0.5 and 20 m. For the instrument considered, near field effects are significant at ranges of 5 m and less and are characterized by systematically reduced target amplitudes after the application of ''normal'' sonar time variable gain. Systematic broadening of combined transmit–receive response patterns at short target ranges result in target defocusing in sonar imaging applications. Experimentally measured responses are presented for comparison with theory. [Work supported by the DFO National Hydroacoustic Program and generous access to the Defence Research & Development Canada—Atlantic acoustic calibration facilities.]

1:45

1pAO2. Development of a sea-well facility for calibrating multibeam sonar. Kenneth G. Foote, Terence R. Hammar, Stephen P. Liberatore, and Kenneth W. Doherty (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

In order to calibrate multibeam sonars for use in quantifying water-column backscattering, a sea-well facility has been developed on Iselin Dock at the Woods Hole Oceanographic Institution. The suite of equipment and calibration instruments is described. This includes systems for mounting and rotating transducer arrays, a number of precision targets, systems for target suspension and positioning, and an external sonar-triggering unit. [Work supported by NSF Grant No. OCE-0002664.]

2:00

1pAO3. Development of a fresh-water tank facility for calibrating multibeam sonar. Kenneth C. Baldwin, Larry Mayer, Andrew McLeod (C-COM, Chase Ocean Eng. Lab, Univ. of New Hampshire, Durham, NH 03824), Kenneth G. Foote, Dezhang Chu (Woods Hole Oceanogr. Inst., Woods Hole, MA), Jonathan Beaudoin (Univ. of NB, Fredericton, NB E3B 5A3, Canada), and Tom Weber (Penn State Univ., University Park, PA 16802)

Multibeam sonars are being used increasingly to image fish. To realize their quantitative potential for measuring the numerical density of fish and other aquatic organisms, it is essential that they be calibrated. This can be done by the use of standard targets or reference hydrophones. The calibration of narrow beam acoustic arrays requires precision angular positioning of the transducer under test. This precision is defined as 0.1 deg of angular position control. This degree of control is achievable with the use of a precision rotary table typically used in CNC machining. This presentation describes: system specifications and the LABVIEW program used to control and coordinate position and acoustic data acquisition, the initial evaluation of the rotary table for repeatability and possible backlash, and representative acoustic measurements made with multibeam sonars using the new system. [Work supported by NSF Contract No. OCE 0002842.]

2:15

1pAO4. Protocols for calibrating multibeam sonar. Kenneth G. Foote, Dezhang Chu (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), Kenneth C. Baldwin, Larry A. Mayer, Andrew McLeod (Univ. of New Hampshire, Durham, NH 03824), Lawrence C. Hufnagle, Jr. (Northwest Fisheries Sci. Ctr., Seattle, WA 98112), J. Michael Jech, and William Michaels (Northeast Fisheries Sci. Ctr., Woods Hole, MA 02543)

The use of multibeam sonar for quantifying scatterers in the water column requires calibration. Principal elements of this include measurement of the directional characteristics of the transducer array or arrays, the determination of dynamic range, the definition of the system sensitivity, and the verification of linearity and time-varied-gain functions. Protocols developed or proposed in the course of a three-year project are reviewed. [Work supported by NSF Grant No. OCE-0002664.]

2:30

1pAO5. Calibration trials with multibeam sonars. Dezhang Chu, Kenneth G. Foote (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), Kenneth C. Baldwin, Larry A. Mayer, Andrew McLeod (Univ. of New Hampshire, Durham, NH 03824), Lawrence C. Hufnagle, Jr. (Northwest Fisheries Sci. Ctr., Seattle, WA 98112), J. Michael Jech, and William Michaels (Northeast Fisheries Sci. Ctr., Woods Hole, MA 02543)

A series of calibration trials have been performed with several multibeam sonars by means of the standard-target method. These have included multiple units of the Simrad SM2000 Multibeam Echo Sounder operating at 90 or 200 kHz, with external transmitter in each case. The principal measurements have been of the full two-dimensional directivity characteristics of the main lobes. Issues of sensitivity, linearity, dynamic range, and near-field effects have also been studied. [Work supported by NSF Grant No. OCE-0002664.]

2:45

1pAO6. A new multibeam echo sounder/sonar for fishery research applications. Lars Nonboe Andersen, Sverre Berg, Erik Stenersen, Ole Bernt Gammelsaeter, and Even Borte Lunde (Simrad AS, P.O. Box 111, N-3191 Horten, Norway, fish_research@simrad.com)

Fisheries scientists have for many years been requesting a calibrated multibeam echo sounder/sonar specially designed for fishery research applications. Simrad AS has, in cooperation with IFREMER, France, agreed on specifications for a multibeam echo sounder and with IMR, Norway for a multibeam sonar, and contracts were signed for development of such systems in January 2003. The systems have 800 transmitting and receiving channels with similar hardware, but different software, and are characterized by narrow beams, low-sidelobe levels, and operate in the frequency range 70–120 kHz. The echo sounder is designed for high operating flexibility, with 1 to 47 beams of approximately 2°, covering a maximum sector of 60°. In addition, normal split beam mode on 70 and 120 kHz with 7° beams for comparison with standard system is available. The sonar will be mounted on a drop keel, looking horizontally, covering a horizontal sector of ±30°, and a vertical sector of 45°. Total number of beams is 500, 25 beams horizontally with a resolution of ~3°, and 20 beams vertically with a resolution of ~4°. Both systems are designed for accurate fish-stock assessment and fish-behavior studies.

3:00

1pAO7. Midwater acoustic modeling for multibeam sonar simulation. Bart Buelens, Ray Williams, Arthur Sale (School of Computing, Univ. of Tasmania, Sandy Bay Campus, Hobart 7005, Tasmania, Australia, bart@sonardata.com), and Tim Pauly (Sonardata Pty. Ltd., Tasmania, Australia)

Simulation and modeling software has been developed to generate synthetic midwater multibeam data. Essentially, the simulator can be considered as a virtual test tank. In order to develop multibeam data analysis methods for fisheries research, it is essential to have a variety of test data sets available, which are ground truthed, georeferenced and corrected for vessel motion. Since equipment and ship time are expensive and data quality not always guaranteed, the simulator provides an effective alternative. The seabed and any objects in the water column such as fish and fish schools can be defined in a 3-dimensional space. A specification for a generic linear array multibeam sonar and its position in space and time can be chosen. The acoustic model implements the technique of acoustic ray-tracing to obtain the pressure at the transducer face, which is converted to individual samples by modeling the working of a digital multibeam system. Beamforming is performed on the fly, and both raw and beamformed complex data sets are generated. Current research on model validation, calibration and analysis techniques will be presented along with an outline of planned future research.

MONDAY AFTERNOON, 10 NOVEMBER 2003 SAN ANTONIO ROOM, 1:30 TO 5:15 P.M.

Session 1pED

Education in Acoustics and Musical Acoustics: Neat Acoustics Websites and Software for Teaching Musical Acoustics

Uwe J. Hansen, Cochair
Department of Physics, Indiana State University, Terre Haute, Indiana 47809

Thomas D. Rossing, Cochair
Physics Department, University of Illinois, DeKalb, Illinois 60115

Invited Papers

1:30

1pED1. Acoustics and vibration animations: A surprisingly successful website. Daniel Russell (Sci. & Mathematics Dept., Kettering Univ., 1700 W. Third Ave., Flint, MI 48532-4898)

For the past 8 years the author has been creating mathematically and physically correct computer animations for use in teaching acoustics to advanced undergraduate engineering and science majors [D. Russell, J. Acoust. Soc. Am. **106**, 2197 (1999)]. Compiling these animations, along with supporting text, on a web site (http://www.kettering.edu/~drussell/demos.html) has resulted in a surprising response from students, teachers, and professionals who have found the animations useful for their own presentations, study, or courses. Unsolicited recognition in the form of web awards and coverage in magazines adds weight to the usefulness of well-made animations for conveying difficult concepts to a wide audience. This presentation will showcase as many of the animations as possible,

especially newer ones, including: oscillation, wave superposition, reflection and refraction, particle motion for various wave types, radiation from simple and complex sources, room modes, waves on strings and membranes, bending waves and boundary conditions, and the vibration of guitars and baseball bats. If time permits, some of the methods used to create the animations will be shared.

2:00

1pED2. Musical acoustics demonstrations. P. L. Hoekje (Dept. of Phys. and Astron., Baldwin-Wallace College, 275 Eastland Rd., Berea, OH 44017)

The ASA Musical Acoustics Demonstrations website (trial version at http://www.bw.edu/~phoekje) includes sound files, video clips, program code listings, and other material for demonstrations related to musical acoustics. Many of the sound demonstrations may be experienced either as expositions, in which the phenomena are explained before they are presented, or as experiments, in which the explanation comes after listeners have had the opportunity to draw their own conclusions. Suggestions are provided for apparatus construction and classroom experiments, as well as for building simple musical instruments. Software is recommended if it is available free and compatible with multiple personal computer operating systems. For example, Audacity (http://audacity.sourceforce.net) is a sound file editor and analyzer that can be used to visually represent sounds and manipulate them. Source files are included for the synthesized sound examples, which were created in Csound (http://csounds.com), so that interested users may create their own variations. Source code is also included for visual demonstrations created in Visual Python and Python (http://www.python.org), an efficient, high level programming language. Suggestions, criticisms, and contributions are always welcome! [Work supported by ASA and Baldwin-Wallace College.]

2:30

1pED3. Teaching acoustics online. Andrew Morrison and Thomas D. Rossing (Dept. of Phys., Northern Illinois Univ., DeKalb, IL 60115)

We teach an introductory course in musical acoustics using a Blackboard. Students in this course can access audio and video materials as well as printed materials on our course website. All homework is submitted online, as are tests and examinations. The students also have the opportunity to use synchronous and asynchronous chat rooms to discuss the course with each other or with the instructors.

3:00–3:15 Break

3:15

1pED4. ASA Education in Acoustics website as new portal. E. Carr Everbach (Eng. Dept., Swarthmore College, 500 College Ave., Swarthmore, PA 19081)

The ad-hoc On Line Education Committee of the Acoustical Society of America was charged in 2002 with revitalizing the Education in Acoustics website. A web portal, demonstrated at the Nashville, TN, meeting in April, 2003, consists of a fanciful view of Lord Rayleigh's drawing room. Items within the room are ''hot'' and lead the visitor to web links whose content is related to the technical committee areas and other areas of interest. The links are configurable via a web (MySQL) database interface so that they can be manipulated without programming. Current efforts include populating the portal with links and adding active content in topic areas that need to be strengthened.

3:45

1pED5. Software tools for developing an acoustics multimedia CD-ROM. Todd W. Bigelow and Paul A. Wheeler (Utah State Univ., 4120 Old Main Hill, Logan, UT 84322 big@cc.usu.edu)

A multimedia CD-ROM was developed to accompany the textbook, *Science of Sound*, by Tom Rossing. This paper discusses the multimedia elements included in the CD-ROM and the various software packages used to create them. PowerPoint presentations with an audio-track background were converted to web pages using Impatica. Animations of acoustic examples and quizzes were developed using Flash by Macromedia. Vegas Video and Sound Forge by Sonic Foundry were used for editing video and audio clips while Cleaner by Discreet was used to compress the clips for use over the internet. Math tutorials were presented as whiteboard presentations using Hitachis Starboard to create the graphics and TechSmiths Camtasia Studio to record the presentations. The CD-ROM is in a web-page format created with Macromedias Dreamweaver. All of these elements are integrated into a single course supplement that can be viewed by any computer with a web browser.

Contributed Papers

4:15

1pED6. Numerical modeling of the acoustic guitar. Antoine Chaigne (ENSTA-UME, Chemin de la Hunière, 91761 Palaiseau cedex, France), Grégoire Derveaux, Patrick Joly, and Eliane Bécache (INRIA, 78153 Le Chesnay cedex, France)

An interactive DVD has been created, based on a numerical model of the acoustic guitar. In a first chapter, the retained physical model is described and illustrated, from the pluck to the 3D radiation field. The second chapter is devoted to the presentation of the numerical tools used for solving the equations of the model. Numerical simulations of plate vibrations and radiated sound pressure are shown in the third chapter. A number of simulated sounds are presented and analyzed in the fourth chapter. In addition, the DVD includes a discussion between a guitar maker, an acoustician, a guitar player and a mathematician. This discussion is entitled ''towards a common language.'' Its aim is to show the interest of simulations with respect to complementary professional approaches of the instrument. This DVD received the Henri Poincaré Prize from the 8th Research

Film Festival of Nancy (June 2003), sponsored by the CNRS, in the category ''Documents for the scientific community and illustrations of the research for teaching purpose.''

4:30

1pED7. "Lurching waves" in DPGraph. R. Dean Ayers and Nader Inan (Dept. of Phys. and Astron., California State Univ.–Long Beach, 1250 Bellflower Blvd., Long Beach, CA 90840-3901, rdayers@csulb.edu)

Introductory treatments of waves usually emphasize undamped traveling waves and ideal standing waves with perfect nodes. Those are just special cases from a larger class of waves in which the crests perform a characteristic ''lurching'' or ''galloping'' motion. The variation of a terminal reflection coefficient and the constant for damping in propagation generates a continuum of more realistic behaviors that connect the special, simple cases. Attempts to develop this larger class verbally and mathematically might seem abstract and complicated, but the use of kinetic computer graphics in an interactive mode makes their introduction straightforward. Preliminary observations and explorations with these images can then lead naturally to a mathematical treatment at a level appropriate for the audience. Software from DPGraph has been particularly convenient for the development of the figures. The fact that programming must be done using analytic expressions and no iterations is a valuable constraint; it forces the user to stay close to fundamentals in the physics and mathematics. Exploratory studies then encourage the programmer to ask analytic questions that might not have been considered otherwise. Several representative figures will be presented. [Work supported by the Paul S. Veneklasen Research Foundation and the CSULB Scholarly and Creative Activities Committee.]

4:45

1pED8. Acoustics textbooks of the future. R. Dean Ayers (Dept. of Phys. and Astron., California State Univ.–Long Beach, 1250 Bellflower Blvd., Long Beach, CA 90840-3901, rdayers@csulb.edu)

There is likely to be a time when electronic textbooks will have become the norm. If these are essentially free, there will be some distinct advantages for both students and teachers. (Imagine a truly free marketplace of ideas, unaffected by the ploys of textbook publishers.) Electronic resources provide flexibility for adaptation to the organization and level of treatment suitable for particular groups of students. Internal links can be used to remove digressions, fine details, and advanced treatments from the direct line of sight of unsophisticated readers, making it easier to focus on the main ideas. Background materials can be linked to highlighted words, for the benefit of both the students who need them and those who do not. Links among chapters and sections could encourage encyclopedic browsing and the development of a cross-linked structure of understanding. Multimedia resources have an obvious relevance to acoustics in the form of audio examples. Kinetic, interactive figures could play an important role by encouraging the development of graphical literacy, coupled with the more widely recognized verbal and mathematical literacies needed for scientific work. MERLOT and other existing models for the development of electronic textbooks will be discussed. [Work supported by the Paul S. Veneklasen Research Foundation.]

5:00

1pED9. Using acoustics to lure high school students into a career in science. Andrew Piacsek (Dept. of Phys., Central Washington Univ., 400 E. 8th Ave., Ellensburg, WA 98926-7422, piacsek@cwu.edu)

As part of a National Science Foundation STEP grant to stimulate recruitment and improve retention of students in science, mathematics, and technology, several science faculty at Central Washington University developed a 2-week summer program for high school students entering their senior year. From June 23 to July 2, 2003, 13 high school juniors from the primarily agricultural region of central Washington lived on the CWU campus and spent their days working on college-level science projects with CWU faculty. Each week, the students worked in groups of three or four with a faculty mentor from one of five disciplines: chemistry, biology, geology, physics, and computer science. The week-long physics project consisted of a series of simple experiments designed to build an understanding of standing waves and resonance. By the end of the week, students were testing hypotheses relating the resonance frequency of a wineglass to its thickness, height, and opening diameter. They also measured Q for several glasses and chose the one with the highest Q to break with sound from a loudspeaker. Although the students were in the lab for an average of 4.5 h per day, they showed unflagging enthusiasm.

MONDAY AFTERNOON, 10 NOVEMBER 2003 SABINE ROOM, 1:30 TO 5:00 P.M.

Session 1pUW

Underwater Acoustics: Scattering and Reverberation: Experimental Observations

Thomas W. Yudichak, Chair
Applied Research Laboratories, University of Texas, P.O. Box 8029, Austin, Texas 78713

Contributed Papers

1:30

1pUW1. Multistatic broadband sonar (MSBA): Principles and measurements. William M. Carey (Dept. of Aerosp. and Mech. Eng., College of Eng., Boston Univ., Boston, MA 02215) and Peter G. Cable (BBN Technologies, Mystic, CT 06355)

Shallow-water (SW) coastal areas with depths between 40 and 200 m with bottom sediment layers ($<$25 m) formed by deposition and sea level variations are known to have frequency-dependent sound transmission and reverberation characteristics especially under the adverse conditions of downward refraction. Coastal ocean-dynamics complicates these environments by coupling the surface, volume, and bottom characteristics and can dramatically affect the temporal and spatial coherency of the signal. The broad-band signal and processing smear these frequency-dependent effects as well as signal excess fluctuations. Furthermore, recent experiments have shown that despite these complexities, explosive sources coupled with directional receivers have the capability to detect, localize, and classify sonar targets in the presence of clutter most likely due to bottom features, fish schools, and sunken objects. Active search sonar operations in SW are typically reverberation limited and experiments have shown that MSBA sonar when employed in bistatic source-receiver geometries can be effective. This paper reviews the principles underlying the MSBA sonar in

shallow water and presents results from several experiments that demonstrated the utility of MSBA. The results are discussed in terms of a classical sonar model with basic statistical considerations to define limits on the detection, classification, and localization with multiple receivers.

1:45

1pUW2. Rough surface scattering from an elastic scale model of an ocean bottom. Raymond J. Soukup, Robert F. Gragg (Naval Res. Lab., Washington, DC 20375), Robert W. Wiley (The Univ. of Houston, TX), and Burcin Inanli (Thelen Assoc., Erlanger, KY)

Monostatic and bistatic scattering strength measurements with a rough PVC surface were collected during two experiments in an acoustic tank facility at the Allied Geophysical Laboratories in the University of Houston. The PVC surface was analogous to limestone ocean bottoms in its two-dimensional power-law roughness spectrum and its large dependence of scattering strength on the roughness parameters. The experiments represent an initial effort to use physical models with ground-truth measurements of roughness and compressional/shear speeds and attenuations to verify the predicted effects of interface scattering models, e.g., the small-slope model developed at the Naval Research Laboratory for elastic bottoms. Comparisons between the small-slope model, perturbation theory, and the observed data are shown for the various geometries using acoustic transmissions in the 100–400 kHz band. The success in obtaining a good model-data fit is shown to be directly related to the ensonification of an area that represents a sufficient statistical sample of the roughness. Plans for a series of tank experiments with physical models for verifying predictions of rough surface scattering theories and elastic PE are described. [Work supported by ONR.]

2:00

1pUW3. Azimuthal variability of acoustic backscatter from 3 to 8 kHz on the New Jersey continental shelf. Paul C. Hines, Darcy J. MacDougald, and John C. Osler (Defence R&D Canada Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada)

The acoustic backscattering strength of the seabed has been demonstrated to be one of the key inputs required in sonar performance prediction (SPP) models. Extending range independent SPP models to full 3-dimensional—or even the simpler N×2-dimensional—models requires measurements of the azimuthal variability of the backscattering strength. DRDC Atlantic's Wide Band Sonar (WBS) system, which consists of a parametric transmitter and a superdirective receiver, is ideally suited to make these measurements. In the spring of 2003 the system was used to measure the azimuthal variability of acoustic backscatter at a site referred to as the New Jersey Strataform. Data were collected at 3, 5, and 8 kHz at grazing angles spanning 10–40. Interpretation of the data set is enhanced with swath bathymetry measurements made during the experiment as well as 3-D seismic profiles made by other researchers [C. S. Duncan, J. A. Goff, J. A. Austin, Jr., and C. S. Fulthorpe, Marine Geology **170**, 395–421 (2000)]. In this paper the experimental geometry is described and the backscatter measurements are presented.

2:15

1pUW4. Modeling of shallow water reverberation data. T. W. Yudichak, D. P. Knobles (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713-8029), and P. G. Cable (BBN Technologies, Mystic, CT 06355-3641)

A broadband model of seabed reverberation in a shallow water waveguide is used to analyze data measured in the Area Characterization Test I (ACT I). Given a source spectrum, a mean geoacoustic profile for the waveguide, and statistics characterizing the scattering sources (including fluctuations in seabed sound speed and density as well as seabed interface roughness) the model simulates time series measured in a monostatic geometry. It accomplishes this by solving the integral form of the Helmholtz equation in a random medium in the Born (or single scattering) approximation using the ORCA normal mode model to compute the unperturbed Green's functions. Statistics of the scattering sources are inferred from comparisons between modeled time series and experimental data. Also, the effects of uncertainties in the source spectrum are considered, and the validity of the Born approximation in this context is discussed. [Work supported by ONR.]

2:30

1pUW5. Statistics of high-resolution seafloor backscatter and the detection of spatial changes. Jerald W. Caruthers, Ralph R. Goodman (Dept. of Marine Sci., Univ. of Southern Mississippi, 1020 Balch Blvd., Stennis Space Center, MS 39529), and Natalia A. Sidorovskaia (Univ. of Louisiana at Lafayette, Lafayette, LA 70505-4210)

The dependence of probability density functions on patch size for high-resolution backscattering, change detection techniques, and implications for scattering models for high-resolution systems are discussed. Data used in these analyses were collected off Panama City using NRL's dual-angle, multiple-frequency towed sonar array and using USM's dual-frequency, side-scan sonar (SSS) system. Frequencies used in the NRL system discussed here are 200, 225, 250, and 275 kHz, and aspects of that work are discussed [Caruthers *et al.*, NRL/MR/7184-00-8256 (2000)]. Scattering patch sizes of the NRL system were of the order of a square meter. The SSS operated at the dual frequencies of 150 and 300 kHz, and its patch sizes were nominally a half meter cross-range by a few centimeters downrange. The SSS is capable of digitizing the four signal waveforms at a 1-MHz rate and 12 bits. Each pulse was six cycles long. The characteristics of this unique SSS system are discussed in Caruthers and Fisher [USM/DMS-2003-01]. Techniques are discussed for determining change detection and varying the scattering patch sizes through coherent combinations of the SSS waveforms. The returns for each of these systems appear to be distinctly different from Rayleigh and suggest point-scatter models are appropriate for these high-resolution data. The false alarm probabilities of the observed distributions are discussed and implications addressed. [Work supported by NRL, SPAWAR, and ONR.]

2:45

1pUW6. Transfer of oceanographic and bottom variability into shallow water propagation and reverberation uncertainty. Kevin D. LePage (Naval Res. Lab. Code 7144, 4555 Overlook Ave. SW, Washington, DC 20375), Charles W. Holland (Penn State Univ., State College, PA 16804), and John A. Goff (Univ. of Texas, Austin, TX 78759)

Often an oceanographic or geoacoustic description of a particular environment is not available on a fine enough scale to capture all the variability that is known to exist at the site. However in cases where resources are available to characterize a particular piece of the environment at the scales of acoustic importance, these measurements may be used to predict the uncertainty of propagation and reverberation in nearby regions as long as the bottom and oceanographic processes are homogeneous over the length scale covering the two sites. In these cases stochastic treatments of the environmental variability at fine scales may be employed and a resulting uncertainty on acoustic propagation and reverberation predicted. Here the uncertainty of acoustic propagation and reverberation due to fine scale sound speed profile variability associated with internal waves and seafloor heterogeneity is evaluated using Monte-Carlo techniques. Results show that for the New Jersey and Eel river STRATAFORM sites seafloor vari-

ability leads to predictions of reverberation uncertainty of less than 1 dB at 4 kHz, while oceanographic variability due to internal waves is predicted to cause about 2 dB of uncertainty in the same frequency regime. [Work supported by ONR Capturing Uncertainty DRI.]

3:00–3:15 Break

3:15

1pUW7. Experimental demonstration of adaptive reverberation nulling using a time reversal mirror. H. C. Song, W. A. Kuperman, W. S. Hodgkiss, P. Roux, T. Akal (Scripps Inst. of Oceanogr., Univ. of California–San Diego, La Jolla, CA 92093-0238), and M. Stevenson (NATO SACLANT Undersea Res. Ctr., La Spezia 19138, Italy)

Boundary reverberation (e.g., due to bottom roughness) can serve as a surrogate probe source in time reversal. A time-gated portion of the reverberation is then refocused to the bottom interface at the corresponding range [Lingevitch *et al.*, J. Acoust. Soc. Am. **111**, 2609–2614 (2002)]. In this paper, reverberation nulling is investigated to enhance active target detection. The basic idea is to minimize the acoustic energy incident on the corresponding interface by applying a weight vector on the time reversal mirror which is in the orthogonal subspace to the focusing vector. The feasibility of the reverberation nulling is demonstrated in a recent time reversal experiment at 850 and 3500 Hz.

3:30

1pUW8. Reverberation measurements in the ocean with a source-receiver array. Joseph F. Lingevitch, David M. Fromm, Charlie F. Gaumond, David C. Calvo, John S. Perkins, Edward L. Kunz, and Y. E. Kim (Naval Res. Lab., Washington, DC 20375)

During our time reversal experiment (TREX-03) conducted in May 2003 off the New Jersey coast, a 64 channel midfrequency source-receiver array (SRA) was deployed to measure reverberation and test the focusing capabilities of an SRA in the ocean without explicitly introducing a known probe source. Previous simulations [J. Acoust. Soc. Am. **111**, 2609–2614 (2002)] and laboratory experiments [J. Acoust. Soc. Am. **113**, 3155–3160 (2003)] have shown that the monostatic reverberation signal due to a rough interface can be decomposed to yield SRA source functions that will focus energy at the interface. The range of the focus is controlled by time windowing the reverberation signal. Additionally, measured environmental data (sound-speed profiles and geo-acoustic parameters) were used to estimate the SRA source weights for focusing at targets that were deployed in the water column and on the bottom. [Work supported by ONR.]

3:45

1pUW9. Comparison of modeled and measured sonar returns from calibrated targets in the Main Acoustic Clutter Experiment 2003. Yi-san Lai, Sunwoong Lee, Purnima Ratilal, and Nicholas C. Makris (MIT, 77 Massachusetts Ave., Cambridge, MA 02139, makris@mit.edu)

In a shallow-water environment the sound field scattered by a target can be complicated by waveguide effects such as multimodal propagation and dispersion. It has been shown that relying on the sonar equation in a shallow-water waveguide can lead to errors in estimating a target's scattering properties [Ratilal, Lai, and Makris, J. Acoust. Soc. Am. **112**, 1797–1816 (2002)]. Furthermore, applying conventional beamforming and matched filtering for target localization can yield errors in range and bearing estimation. To quantify these errors, an analytic expression for active sonar returns after beamforming and matched filtering is derived from waveguide scattering theory. This expression is used to model the beamformer and matched filter output from calibrated targets deployed in the Main Acoustic Clutter Experiment 2003. Modeled and measured sonar returns are compared from several calibrated targets, and the errors in target localization and scattering amplitude estimation are quantified in a realistic active sonar scenario. Accurate locations of the source, targets, and receiver array were measured by GPS during the experiment.

4:00

1pUW10. Bistatic scattering from flush and partially buried targets in shallow water. Irena Veljkovic, Henrik Schmidt (MIT, 77 Massachusetts Ave., Cambridge, MA 02139), and Alessandra Tesei (SACLANT Undersea Res. Ctr., 19138 San Bartolomeo (SP), Italy)

GOATS'98 experiment had as one of its objectives to explore the possibility of vastly increasing the coverage rate of shallow water mine countermeasures (MCM), especially for buried targets using subcritical seabed penetration. The evanescent incident field combined with 3D bistatic configuration brought about a rich and intricate set of physical phenomena occurring in the target scattering and radiation process. Resonance of manmade elastic targets is a pivotal concept distinguishing targets from rocks or other clutter. Nevertheless, burial in sediment results in a variety of modifications to the scattered response caused by different physical mechanisms, geometric constraints, and intrinsic sediment properties. It is the goal of this work to identify, analyze, and explain the fundamental effects of the outer sediment and the proximity of the seabed interface on the scattering from elastic shells insonified at subcritical incident angles in a bistatic configuration. Using time-frequency and spectral analysis, a new, more comprehensive, understanding of GOATS'98 experimental data was obtained distinguishing the effects of the acoustics environment from the resonant signature of buried elastic targets. To achieve this, and to further investigate the details of the scattering process a new target scattering module of OASES-3D modeling framework was used. [Work supported by ONR.]

4:15

1pUW11. Automated buried object classification with a broadband autonomous underwater vehicle sonar system. Monica Montanari, Joseph R. Edwards, and Henrik Schmidt (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

The classification of buried objects in the littoral ocean is a challenging task for the underwater community. The high attenuation rate of the seabed imposes the use of a low frequency sonar system, which leads to an increased aperture length requirement to achieve high resolution. A simple classification algorithm is proposed in which a broadband source insonifies the target area while multiple receiver vehicles process the target scattering strengths over various bistatic angles to reconstruct a planar cut of the scattered field of the target. The advantages of this technique are the low rate of information exchange between vehicles, the simplicity of the processing algorithm, and less demanding processing requirements compared to the image-based techniques. The algorithm is based on multistatic measurements of the target spatial acoustic signatures, which permits a parameterization of the identification problem to a decreased number of degrees of freedom, with the consequence of decreased processing and implementation difficulty. The use of a broadband sonar system splits the processing in parallel tasks, leading to an overall classification improvement. The GOATS experiment series is used to demonstrate the classification performance of the algorithm for various targets and conditions. The results compare favorably with the state-of-the-art imaging techniques.

4:30

1pUW12. Very-low-frequency scattering experiments from proud targets in a littoral environment using a 55-m rail. Harry J. Simpson (Naval Res. Lab., Code 7136, 4555 Overlook Ave., Washington, DC 20375-5320, harry.simpson@nrl.navy.mil), Carl K. Frederickson (Univ. of Central Arkansas, Conway, AR 72035-0001), Erik C. Porse, Brian H. Houston (Naval Res. Lab., Code 7136, Washington, DC 20375-5320), Larry A. Kraus, Alain R. Berdoz, Philip A. Frank, and Steve W. Liskey (SFA, Inc., Largo, MD 20774-5322)

Scattering measurements from proud targets were conducted in the littoral environment of St. Andrews bay. A 55-m long rail was assembled 2.3 m above the sandy bottom to position transducers using an attached robot. The rail provided a straight and level structure that repeatably scanned the target field. This system was used in both a bistatic and monostatic configuration to produce synthetic array measurements of the target field. The bistatic experiments used a stationary source at the center of the rail to insonify the target field while the receiver was moved to synthesize a 55-m array. The monostatic experiments had both a source and receiver mounted together on the robot, and both were moved together to scan the target field. All the experiments were conducted using a very-low-frequency band (1 to 12 kHz). The results of these experiments will be presented along with signal processing of the target field to study the physical responses of the individual targets to the low-frequency acoustic insonification. [Work supported by ONR.]

4:45

1pUW13. Hybrid volume scattering model: Application of the Biot theory to the fluid-bottom model. Keiichi Ohkawa and Tokuo Yamamoto (Geoacoustic Lab., Appl. Marine Phys. Div. RSMAS, Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149)

The volume backscattering model, which was derived by Yamamoto (1996) applying the Born approximation to the Helmholtz equation, is improved by replacing transmission coefficients at the fluid–fluid interface to those at the fluid–porous seabed interface. The spectra of velocity and density fluctuations in sediments are assumed to be the same. The sea bottom is hereby treated as an isotropic porous medium, in which sound speeds have frequency dependence. Frequency dependent reflection and transmission coefficients at the fluid–porous seabed interface for a compressional wave incident from fluid upon porous media and vice versa are easily calculated applying the mode's decoupling method to Biot's body waves without considering behaviors of inhomogeneous waves in porous media. In this problem the reflection and transmission coefficients in terms of displacement potentials are computed. The hybrid model is applied to medium to high frequency bottom backscattering from the sandy bottom. In soft sediments, however, the difference between the hybrid and the fluid–fluid bottom models is small. It is concluded that the hybrid model can be applied to volume backscattering from any kind of unconsolidated sediments throughout the frequency range. [Work supported by ONR Code 321OA.]

NOTE: Separate registration required. See page A28

MONDAY EVENING, 10 NOVEMBER 2003 WEDGWOOD ROOM, 7:00 TO 9:00 P.M.

Session 1eID

Interdisciplinary: Tutorial Lecture: Classroom Acoustics

Yves H. Berthelot, Chair

School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

Chair's Introduction—7:00

Invited Paper

7:05

1eID1. Children's need for favorable acoustics in schools. Peggy B. Nelson (Dept. of Commun. Disord., Univ. of Minnesota, Minneapolis, MN 55455, nelso477@umn.edu)

Children continue to improve their understanding of speech in noise and reverberation throughout childhood and adolescence. They do not typically achieve adult performance levels until their late teenage years. As a result, schools that are designed to be acoustically adequate for adult understanding may be insufficient for full understanding by young children. In addition, children with hearing loss, those with attention problems, and those learning in a non-native language require even more favorable signal-to-noise ratios. This tutorial will review the literature gathered by the ANSI/ASA working group on classroom acoustics that shaped the recommendations of the working group. Special topics will include speech perception data from typically developing infants and children, from children with hearing loss, and from adults and children listening in a non-native language. In addition, the tutorial will overview recommendations contained within ANSI standard 12.60-2002: Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools. The discussion will also include issues related to designing quiet classrooms and working with local schools and professionals.

Session 2aAA

Architectural Acoustics: Small Rooms—Big Challenges

Alexander U. Case, Chair
Fermata Audio and Acoustics, P.O. Box 1161, Portsmouth, New Hampshire 03802

Invited Papers

8:00

2aAA1. Maximum performance synergy: A new approach to recording studio control room design. Jeff D. Szymanski (Auralex Acoust., Inc., 8851 N. Hague Rd., Indianapolis, IN 46256, consulting@auralex.com)

Popular recording studio control room designs include LEDE(tm), RFZ(tm), and nonenvironment rooms. The common goal of all of these is to create an accurate acoustical environment that does not distort or otherwise color audio reproduction. Also common to these designs is the frequent need to have multiple ancillary recording rooms, often adjacent to the main control room, where group members perform. This approach, where group members are physically separated from one another, can lead to lack of ensemble in the finished recordings. New twists on old acoustical treatment techniques have been implemented at a studio in Nashville, Tennessee, which minimize the need for multiple ancillary recording rooms, thus creating an environment where talent, producer and recording professionals can all occupy the same space for maximum performance synergy. Semi-separated performance areas are designed around a central, critical listening area. The techniques and equipment required to achieve this separation are reviewed, as are advantages and disadvantages to this new control room design approach.

8:25

2aAA2. Control room design: An analytic perspective. Douglas Jones (Columbia College Chicago, 600 S. Michigan, Chicago, IL 60605)

The recording studio control room is a special case of small room. In this paper the subjective attributes of a control room and their objective counterparts will be explored, discussing the analytic tools that might be used to evaluate and predict the subjective performance. A number of areas where more research is needed will also be identified.

8:50

2aAA3. Modeling sound fields in small rooms. Sung Yoon and Rendell R. Torres (Prog. in Arch. Acoust., Rensselaer Polytechnic Inst., Greene Bldg., 110 8th St., Troy, NY 12180-3590)

In computationally modeling and auralizing acoustics of small rooms, geometrical-acoustics methods do not apply at lower frequencies, and models must account for the wave-nature of sound fields, especially modal effects and scattering in the early impulse response. Common approaches include finite-difference methods, finite-element methods, boundary-element methods, edge-diffraction models, boss models, among others. This investigation describes different time- frequency-domain techniques for computing the sound field in small rooms and introduces new methods as well, including fast algorithms based on cellular automata for calculating scattering from edge discontinuities and other boundary conditions. The latter approach will be described in relation to the pioneering work on the lattice gas time domain (LGTD) approach by Sudo and Sparrow [AIAA J. **33** (1995)] and how it can be applied to modeling the typical timbral and temporal (reverberant) characteristics of small room acoustics. [Work supported by Rensselaer Polytechnic Institute.]

9:15

2aAA4. Controlling loudspeakers in rooms. Jan Abildgaard Pedersen and Poul Praestgaard (Acoust. Res., Bang & Olufsen, Denmark)

How to control the perceived timbre of a loudspeaker in different listening positions, using different loudspeaker positions in different listening rooms is a central problem, when designing and using loudspeakers is discussed. Loudspeaker directivity has been found to be one of the critical parameters in the solution to this problem. This paper presents three hypotheses for the optimal loudspeaker directivity and a novel realization of that directivity, i.e., the Acoustic Lense (ALT). The dominant problem at low frequencies has been found to relate to the fact, that the acoustic power output of loudspeakers is highly dependent on the loudspeaker position and on the acoustic properties of the listening room. This paper also presents a novel system for adapting a loudspeaker to its position and to the acoustic properties of the listening room, i.e., Adaptive Bass Control (ABC).

9:40

2aAA5. Variable acoustics in small rooms for music. Yasushi Shimizu (Adv. System Development Ctr., Yamaha, Japan, Hamamatsu 430-8650, Japan)

Small rooms for music are frequently discussed to consider variable acoustics with different program in practice room, rehearsal room as well as audio listening room. In practice room major criteria in acoustics is usually discussed how to determine reverberation with less than 1.0 s. RT for different use such as practice for different musical instrument. And on audio listening room, acoustical effect on sound reproduced by speakers is important especially for reproducibility of 3D audio representation and usually surface material behind stereophonic speakers gave significant effect on the accuracy of sound localization with the sound played by speakers. Rehearsal room acoustics is always discussed compared to the stage acoustics in primary concert hall and some electro-acoustic technique is discussed how to simulate stage acoustics. This paper presents examples of acoustical design for different practice rooms, rehearsal room and audio listening room. A different RT setting for each room with different program is discussed as well as different configuration of surface material especially in audio listening room. Finally rehearsal room with electro-acoustics system to reproduce stage acoustics of specific concert hall is discussed and measurement result is presented.

10:05–10:20 Break

10:20

2aAA6. Low frequency evaluation and treatment of small rooms. Peter D'Antonio (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774)

At low frequencies, the acoustical coupling of the listener and loudspeakers with reflections from the room's boundary surfaces and its modal pressure distribution can cause significant acoustical distortion. For rectangular room's, software programs exist to predict the magnitude of these effects. However, there is no substitute for experimental measurements. When evaluating small rooms, it is often desirable to isolate the modal effects from the speaker-boundary effects, so that appropriate corrective measures can be applied. A MLS measurement procedure will be presented. After the room dimensions and listener/loudspeaker positions are optimized, one can apply dedicated low frequency absorbers to further control low frequency problems. The characterization of low frequency absorbers will be described, using a 7-ton, 22 long impedance tube.

Contributed Papers

10:45

2aAA7. Improvement of low frequency sound variability in a small enclosure by use of a source with frequency-independent radiated power. Alexandra Loubeau and Jiri Tichy (Grad. Prog. in Acoust., The Pennsylvania State Univ., P.O. Box 30, State College, PA 16804, aloubeau@psu.edu)

The irregularity in the low frequency response encountered in small rooms adversely affects the transmission of sound from a source to a receiver. Variations in the frequency response, which are dependent on source position, receiver position, and room absorption, result in perceived inconsistencies in sound quality throughout the room. This computational study examines the effect of constant sound power radiation on the sound field variability with frequency in a small enclosure. Effects of the parameters listed above are investigated. Frequencies of interest are 20–80 Hz. Source positions include the room corner, typical subwoofer locations, and typical full-range loudspeaker locations. Grids of receiver positions are selected around the center of the room, forward of center, back of center, and to the side of center. Room absorption values explored correspond to reverberation times of 0.5 s, 1 s, and 2 s. The benefit of constant sound power radiation is shown using the metric of standard deviation. Depending on the location of the source and receiver, the standard deviation of the frequency response decreases by approximately 2–3 dB with constant sound power radiation. [Research supported by Penn State ARL.]

11:00

2aAA8. Real-time virtual room acoustic simulation. James P. Carneal, Jan Johnson, Troge Johnson, and Marty Johnson (Vib. and Acoust. Labs., Dept. of Mech. Eng., VPI&SU, Blacksburg, VA 24061, jcarneal@vt.edu)

A realistic virtual room acoustic simulation has been implemented on a PC-based computer in near real-time. Room acoustics are calculated by the image source method using realistic absorption coefficients for a variety of realistic surfaces and programmed in MATLAB. The resulting impulse response filters are then applied in near real-time using fast convolution DSP techniques using data being read from a CD-ROM. The system was implemented in a virtual acoustic room facility. Optimizations have been performed to retain the realistic virtual room effect while minimizing computations through limited psycho-acoustic testing. In general, realistic anechoic to reverberant virtual rooms have been re-created with six 8192 coefficient filters. To provide realistic simulations, special care must be taken to accurately reproduce the low frequency acoustics. Since the virtual room acoustic facility was not totally anechoic (as are most anechoic chambers), inverse filters were applied to compensate for over-amplified acoustics at frequencies below 350 Hz.

11:15

2aAA9. Acoustical evaluation of preschool classrooms. Wonyoung Yang and Murray Hodgson (School of Occupational & Environ. Hygiene, Univ. of British Columbia, 2206 East Mall, 3rd Fl., Vancouver, BC V6T 1Z3, Canada)

An investigation was made of the acoustical environments in the Berwick Preschool, Vancouver, in response to complaints by the teachers. Reverberation times (RT), background noise levels (BNL), and in-class sound levels (Leq) were measured for acoustical evaluation in the classrooms. With respect to the measured RT and BNL, none of the classrooms in the preschool were acceptable according to the criteria relevant to this study. A questionnaire was administered to the teachers to assess their subjective responses to the acoustical and nonacoustical environments of the classrooms. Teachers agreed that the nonacoustical environments in the classrooms were fair, but that the acoustical environments had problems. Eight different classroom configurations were simulated to improve the acoustical environments, using the CATT room acoustical simulation program. When the surface absorption was increased, both the RT and speech levels decreased. RASTI was dependent on the volumes of the classrooms when the background noise levels were high; however, it depended on the total absorption of the classrooms when the background noise levels were low. Ceiling heights are critical as well. It is recommended that decreasing the volume of the classrooms is effective. Sound absorptive materials should be added to the walls or ceiling.

11:30

2aAA10. Estimating reverberation time in classrooms by the power method. Richard D. Godfrey (Owens Corning Sci. & Technol., 2790 Columbus Rd., Granville, OH 43023)

The Sabin equation relates reverberation time to room volume and absorption. If an acoustic source of known power is placed in a room, the absorption is also related to the power per bandwidth, volume, and pressure. Combining the Sabin equation with the power equation and solving for reverberation time yields a relationship for reverberation time as a function of volume, pressure, and the power of the source. This approach has been reported to deviate from the decay method at low frequency. In a laboratory reverberation room, it yielded reasonable results in the frequency range of interest in classrooms. Reasonable results were also achieved in real classrooms in the 500- to 2000-Hz frequency range. Based on these findings, it may be possible to make acceptable reverberation time measurements in less time with less sophisticated instrumentation than the decay method.

11:45

2aAA11. A survey of the acoustical quality of seventeen libraries at Princeton University. Benjamin Markham (Acentech, Inc., 33 Moulton St., Cambridge, MA 02134, bmarkham@acentech.com)

The purpose of this study was to identify objective acoustic measures that correlate with the subjective responses of students and administrators to libraries at Princeton University. The motivation for this study was to determine what is necessary in order to provide a comfortable acoustic environment for users of a new science library to be built on campus. On 31 March 2003, Acentech, Incorporated evaluated 17 library spaces and interviewed a number of students and librarians at Princeton. Based on the results of the survey, the author proposes that a comfortable acoustic environment in a library is an environment that provides freedom from distraction; in other words, casual conversation and other noises in the library will not distract users reading or studying in the library. In order to provide such an environment, a library must have (1) appropriate levels of background sound, (2) a physical barrier between noise-producing and noise-sensitive sections, and (3) sufficient sound absorbing material in the space. Measured quantitative metrics support these conclusions.

TUESDAY MORNING, 11 NOVEMBER 2003 SABINE ROOM, 8:25 A.M. TO 12:00 NOON

Session 2aAO

Acoustical Oceanography and Underwater Acoustics: Aubrey L. Anderson Memorial Session on Acoustics of Gas-Bearing Sediments

Anthony P. Lyons, Cochair
Applied Research Laboratory, The Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804-0030

Michael D. Richardson, Cochair
Naval Research Laboratory, Code 7430, Stennis Space Center, Mississippi 39529

Chair's Introduction—8:25

Invited Papers

8:30

2aAO1. The nature and location of gassy sediment sections in the continental shelf and slope in the northwestern Gulf of Mexico. William Bryant (Dept. of Oceanogr., Texas A&M Univ., College Station, TX 77843-3146)

In the northwestern Gulf of Mexico continental shelf and upper slope gassy sediments are a pervasive phenomena and an important consideration relative to engineering and acoustic activities on the sea floor. An examination of seismic data from over a thousand M.M.S. geohazard reports and core logs of 1,670 foundation boreholes drilled to an average subbottom depth of 125 m on the continental shelf and upper slope in the northwestern Gulf, has revealed that gassy sediment sections are most abundant near the Mississippi River Delta, in the sediment fill of buried stream channels that were eroded during the early and late Wisconsinan, and in Miocene and Plio–Pleistocene depocenters on the continental shelf and upper slope. Out of the 1,670 bore holes examined 1,158 (68%) contained indications of gassy sediments most of which is of biogenic origin. Large patches of gassy sediments exist, some exceeding 10 km in size but most are less than 500 m. The examination of 500 piston cores, up to 40 meters in length, taken on the mid and lower continental slope areas were almost void of gassy sediments as the result of the halokeiesis of allocthonous salt.

8:50

2aAO2. A global survey of the distribution of free gas in marine sediments. Peter Fleischer (Naval Oceanogr. Office, Stennis Space Center, MS 39522-5001), Tim Orsi (Planning Systems, Inc., Slidell, LA 70458), and Michael Richardson (Naval Res. Lab., Stennis Space Center, MS 39529-5004)

Following the work of Aubrey Anderson in the Gulf of Mexico, we have attempted to quantify the global distribution of free gas in shallow marine sediments, and have identified and indexed over one hundred documented cases in the scientific and engineering literature. Our survey confirms previous assumptions, primarily that gas bubbles are ubiquitous in the organic-rich muds of coastal waters and shallow adjacent seas. Acoustic turbidity as recorded during seismo-acoustic surveys is the most frequently cited evidence used to infer the presence of seafloor gas. Biogenic methane predominates within these shallow subbottom deposits. The survey also

reveals significant imbalances in the geographic distribution of studies, which might be addressed in the future by accessing proprietary data or local studies with limited distribution. Because of their global prevalence, growing interest in gassy marine sediments is understandable as their presence has profound scientific, engineering and environmental implications.

9:10

2aAO3. Marine gas hydrate: Fabric, quantification and free gas content. Examples from Hydrate Ridge–Cascadia margin. Friedrich Abegg (GEOMAR Res. Ctr. for Marine Geosciences, Wischhofstr. 1-3, 24148 Kiel, Germany, fabegg@geomar.de), Johannes Freitag (Alfred Wegener Inst. for Polar and Marine Res., Columbusstr., 27568 Bremerhaven, Germany), and Gerhard Bohrmann (Univ. Bremen, Am Fallturm 1, 28334 Bremen, Germany)

A recent advance in the investigation of internal gas hydrate structure and in small scale quantification is based on a combination of a new coring device and computerized x-ray tomography (CT). This approach has been chosen because gas hydrate is only stable within a special (low) temperature and (high) pressure field. A MultiAutoclaveCorer, developed by the Technical University of Berlin, is in principle similar to a piston corer and has the size of a multiple corer. During the deployments on Hydrate Ridge, two pressure vessels filled with seafloor samples could be recovered. They were CT-investigated after the cruise in a medical clinic close to San Francisco. As a result, a 3-D-dataset of the cores is available which allows to quantify the components' gas hydrate, sediment and free gas and also shows the distribution and orientation of the gas bubbles. One of the pressure vessels showed a distinct gas hydrate horizon where the gas hydrate content reached close to 50 vol%. Within this horizon there was a free gas content of 2.4 vol%. The preferential bubble orientation, compared to free gas in soft marine sediments, is horizontal, not vertical, which is an indicator for the mechanism of gas hydrate formation.

9:30

2aAO4. Data-model comparisons for scattering by a bubbly sediment. Darrell R. Jackson and Dajun Tang (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

As part of the Coastal Benthic Boundary Layer Program, seafloor backscattering measurements were made at 40 kHz in Eckernfoerde Bay, Germany, in areas of high concentration of gaseous methane. Although it had long been known that methane bubbles have a profound influence on acoustic scattering, this was the first time that contemporaneous measurements were made of acoustic scattering and bubble size distributions. The latter measurements were part of a collaboration between Aubrey Anderson's group and NRL-SSC and made detailed model-data comparisons possible. Acoustic interferometry was used to conclude that the dominant scattering layer was about 1 m below the sediment–water interface. A single scattering model accounted for the angular dependence and level of scattering, however, the model gave unphysical results at very small grazing angles. A multiple scattering model has been developed to remedy this defect. [Work sponsored by ONR.]

9:50

2aAO5. Ups and downs of a gas horizon under environmental control. Thomas F. Wever (Forschungsanstalt der Bundeswehr für Wasserschall und Geophysik, Klausdorfer Weg 2-24, 24148 Kiel, Germany, Wever@fwg-kiel.de)

Repeated measurements with a sub-bottom profiler along reference tracks in Eckernförde Bay (Western Baltic) led to the recognition that acoustic turbidity (the ''gas horizon'') within the muddy sediments changed its depth. Soon a dependence on the annual temperature cycle with a 3–4 month delay became obvious. In addition, short-term variations correlated with atmospheric pressure changes. For a further investigation, a tower equipped with an echosounder, pressure, and temperature probes at different sediment depths was deployed for four years in the center of the bay. Data were recorded every hour. The data show the migration of the annual temperature peak and low through the sediment and its long-term influence on acoustic turbidity depth. Short-term (''high-frequency'') variations can be correlated with changes in total effective pressure. Repeated investigations of the total methane content in sediment cores during different times of the year and the analysis of the cores with the use of computer tomography under the guidance of Aubrey L. Anderson proved a constant total methane content. This led to the explanation that the environmental factors pressure and temperature cause variations of gas solubility that move up and down and thus control the existence or absence of free gas bubbles within the sediment.

10:10–10:30 Break

10:30

2aAO6. A biological source of bubbles in sandy marine sediments. D. V. Holliday, Charles F. GreenlawIII (BAE Systems, 4669 Murphy Canyon Rd., San Diego, CA 92123, van.holliday@baesystems.com), David Thistle (Florida State Univ., Tallahassee, FL 32306), and Jan E. B. Rines (Univ. of Rhode Island, Narragansett, RI 02882)

Gas in sediments, even in small quantities, will modify the propagation and scattering of sound. Shallow water littoral environments are often sufficiently well lit by sunlight to support healthy populations of benthic and epibenthic marine microalgae. Photosynthesis in marine algae produces oxygen. Oxygen saturation levels as high as 600% have been measured in the pore water of a sandy sediment at 1 mm depth, decreasing to 100% saturation at ca. 4.5 mm. Bioturbation and physical processes routinely mix materials at the surface of the seabed, including algae, to depths of at least a few cm. Mixing times and depths vary with the sediment

type and the species and abundance of organisms present, but time scales of minutes to hours are common. While light is rapidly attenuated with depth in sand, measurements show that it can penetrate to depths of a few mm. Physical and biological mechanisms are suggested which could produce gas bubbles in oxygen saturated pore water. Laboratory measurements of sound scattering from marine algae on a sand surface suggest a possible method for *in situ* bubble detection in shallow marine environments. [Work supported by ONR.]

10:50

2aAO7. Frequency-dependent acoustic properties of gassy marine sediments. Angus I. Best (Challenger Div. for Seafloor Processes, Southampton Oceanogr. Ctr., European Way, Southampton SO14 3ZH, UK, aib@soc.soton.ac.uk), Michael D. J. Tuffin, Justin K. Dix, and Jonathan M. Bull (School of Ocean and Earth Sci., Southampton Oceanogr. Ctr., Southampton SO14 3ZH, UK)

Acoustic velocity and attenuation were measured during two *in-situ* experiments in gassy intertidal muds in Southampton Water, United Kingdom. The horizontal transmission results gave frequency-independent velocity (1431 m/s) and attenuation (4 dB/m) over the frequency range 600 to 3000 Hz, representative of the soft (non-gassy) muds shallower than about 1 m. The results from a vertical transmission experiment straddling the top of the gassy zone (about 1 m depth) showed strong frequency-dependent velocity and attenuation over 600 to 3000 Hz. They showed velocity and attenuation maxima predicted by the Anderson and Hampton model, associated with gas bubble resonance. Moreover, attenuation maxima shifted in frequency with water depth over a tidal cycle that was monitored, suggesting variations in gas bubble size with hydrostatic pressure. X-ray CT images on a sealed core from the site revealed vertically-aligned, centimeter-scale, gas-filled cracks in the muddy sediments. Ultrasonic (300 to 700 kHz) velocities and attenuations were higher in the gassy zone than in the nongassy parts of the core. Overall, the results give a fascinating insight into the acoustical behavior of gassy sediments that could be used to extract sediment physical properties information from seabed acoustic reflection data. [Work supported by NERC].

11:10

2aAO8. Determining gas bubble morphology and size distribution in mud using CT imagery. Kevin B. Briggs and Allen H. Reed (Seafloor Sci. Branch, Naval Res. Lab., Stennis Space Center, MS 39529-5004)

Sampling for the measurement and description of methane gas bubbles *in situ* was accomplished using the method of Abegg and Anderson [Mar. Geol. **137**, 137–147 (1997)]. Sediment cores from East Bay, off the mouth of the Mississippi River, were collected by divers and placed into aluminum pressure transfer chambers while on the seafloor. With the cores at seafloor pressure within the chambers, they were transferred to an x-ray computed tomography (CT) scanner where high-resolution images were made of the sediment within the cores. Data, in the format of series of cross-sectional images of x-ray attenuation reconstructed in 3-D, were evaluated in terms of spatial distribution, sizes, and shapes of bubbles. CT imagery was obtained from a GE LS medical CT scanner at a local hospital and the Naval Research Laboratory's new HD-500 industrial CT scanner specifically designed for core sample imaging. The medical scanner provided images of stationary cores at 625-μm intervals with a rotating x-ray source and was able to resolve bubbles down to 625 μm in diameter. The industrial scanner provided images of rotating cores at 25-μm intervals with a stationary source and was able to resolve bubbles down to 10 μm in diameter.

Contributed Papers

11:30

2aAO9. Gas bubbles in marine mud—How small are they? Allen H. Reed and Kevin B. Briggs (Seafloor Sci. Branch, Naval Res. Lab., Stennis Space Center, MS 39529)

Free gas in marine mud poses a challenging problem in the realm of ocean acoustics as it readily attenuates (i.e., scatters or absorbs) energy, such that objects lying below the gassy sediment are acoustically masked. Gas-laden sediments were located in 10- to 120-m water depth adjacent to the South Pass of the Mississippi River in East Bay using a 12-kHz transducer and the Acoustic Sediment Classification System. Several cores were collected in this region for physical property measurements. Some of the cores were x-rayed on medical and industrial computed tomography (CT) scanners. Volumetric CT images were used to locate gas bubbles, determine shapes and sizes to within the limits of the CT resolution. Free gas in the East Bay sediments was relegated to worm tubes as well as isolated pockets as was the case in Eckernförde Bay sediments [Abegg and Anderson, Mar. Geol. **137**, 137–147 (1997)]. The primary significance of the present work is that gas bubbles have been determined to exist in the tens of μm size range, which is significantly smaller than the smallest bubbles that were previously resolved with medical CT ($\sim$440 μm) with NRL's HD-500 micro-CT System. [Work supported by ONR and NRL.]

11:45

2aAO10. Sensitivity of bubble resonance to environmental inputs using the Anderson model. Warren Wood and Michael Richardson (Naval Res. Lab., Code 7430, Stennis Space Center, MS 39529)

The Anderson model relates the expected sound speed and attenuation (both a function of bubble resonance) through bubbly sediment to physical properties of the gas (including shape and temperature) as well as the anelastic properties of the surrounding sediment. This mathematical formulation allows relatively straightforward assessment of the sensitivity of the sound speed and attenuation to each of the approximately 20 inputs to the model. Our sensitivity analysis was performed at several locations in model space that correspond to realizations of frequently encountered, realistic, bubbly sediments. We first determine a baseline multi-frequency acoustic response from a starting sediment and gas type, then compare the baseline response to responses from a perturbed model. Coupled parameters are perturbed simultaneously (by differing percentages) such that at no time are we simulating the response of an unrealistic sediment. Results suggest a generally greater sensitivity to gas and pore fluid properties than to sediment grain, or frame properties.

Session 2aBB

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Ultrasound Contrast Agents

Constantin-C. Coussios, Cochair
Department of Aerospace and Mechanical Engineering, Boston University, Boston, Massachusetts 02215

Preston S. Wilson, Cochair
Mechanical Engineering Department, University of Texas at Austin, Austin, Texas 78712-0292

Invited Papers

8:30

2aBB1. Methods for blood flow measurements using ultrasound contrast agents. J. Brian Fowlkes (Dept. of Radiol., Univ. of Michigan, Kresge III R3315, Ann Arbor, MI 48109-0553)

Blood flow measurements using ultrasound contrast agents are being investigated for myocardial perfusion and more recently in other organ systems. The methods are based largely on the relative increase in echogenicity due to the concentration of bubbles present in the ultrasound beam. In the simplest form, regional differences in blood volume can be inferred but the possibility exists to extract perfusion from the transit of contrast agent through tissue. Perfusion measurements rely on determining the flux of blood through a tissue volume and as such require knowledge of the fractional blood volume (FBV), i.e., ml blood/g tissue and the rate of exchange, commonly measured as the mean transit time (MTT). This presentation will discuss methods of determining each of these values and their combination to estimate tissue perfusion. Underlying principles of indicator-dilution theory will be provided in the context of ultrasound contrast agents. Current methods for determining MTT will include imaging of the intravenous bolus, in-plane contrast disruption with interval and real-time contrast recovery imaging, and control of contrast agent flow using arterial disruption (contrast interruption). The advantages and limitations of the methods will be examined along with current applications. [Work supported in part by NIH.]

8:55

2aBB2. A review of theoretical models for microbubble contrast agents. Charles C. Church and Xinmai Yang (Natl. Ctr. for Phys. Acoust., Univ. of Mississippi, University, MS 38677)

Theoretical investigation of the acoustic responses of albumin-encapsulated microspheres began nearly fifteen years ago when Albunex, the first agent approved for clinical use in the U.S., was still in development. Since that time, the number of potential ultrasound contrast agents has grown considerably. These agents utilize a variety of both shell materials (e.g., proteins, synthetic polymers, surfactants, lipids) and encapsulated gases (e.g., air, sulfur hexafluoride, perfluorocarbons). The shell may be very thick or vanishingly thin. A thorough understanding of the interaction between ultrasound pulses and contrast microbubbles is essential for the successful clinical application of a particular agent. In this talk, models developed by de Jong *et al.*, Church, Ye, Allen *et al.*, Khismatullin *et al.*, and others will be reviewed. The basis of these theories is a free bubble model supplemented by the effect of the encapsulating shell. The differences among these models lie primarily in their treatment of the encapsulating layer and, to some extent, the surrounding medium. The nature of various contrast agents will be discussed, and appropriate models for each will be described. Comparisons among models will include predictions of clinically significant acoustic responses (resonance frequency, scattering strength, nonlinearity, etc.).

9:20

2aBB3. Acoustical and optical monitoring of contrast microbubbles. Jingfeng Guan, Wen-shiang Chen, and Thomas Matula (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

We have undertaken a series of studies to better understand the response of ultrasound contrast microbubbles to pulsed ultrasound. Toward this end we have used acoustical and optical scattering techniques. The acoustic interrogation of shell-disrupted microbubbles immediately following HIFU was used to monitor microbubble (or fragmented daughter) dissolution. Results were compared to calculations, assuming a simple Gaussian distribution of fragmented microbubbles and the dissolution characteristics of a mixed-gas system (such as a perfluorocarbon gas bubble in air-saturated water). Good results were obtained by considering a multi-modal bubble size distribution. That is, the HIFU pulse created a distribution of smaller bubbles. In order to follow the instantaneous motion of a single microbubble during pulsed ultrasound, light scattering was employed. For this experiment, a diagnostic ultrasound system was used to force the microbubble into oscillation. Light scattered off the microbubble was collected by a photodetector. A simple bubble dynamics model was used to fit the data. At low pressures, consecutive data segments were averaged to increase the signal/noise. At higher pressures, the bubble (size or shell) properties required systematic adjustment in order to obtain a good fit, suggesting that the bubble evolved with each pulse.

9:45

2aBB4. Ultrasound contrast agents and their use in monitoring therapy. Katherine Ferrara, Paul Dayton, Michaelann Shortencarrier, and Dustin Kruse (Dept. of Biomed. Eng., Univ. of California–Davis, Davis, CA 95616)

The shell of ultrasound contrast agents can be modified to include a molecular targeting ligand, and the properties of the agent with and without molecular targeting can be used to monitor changes produced by a therapy. We have investigated the use of ligands targeted to an integrin expressed in cancer, whose expression correlates with tumor grade. Acoustic studies illustrate a 3- to 20-fold increase in echo amplitude from integrin-expressing cells exposed to the targeted contrast agent, as compared to controls, and depending on cell type, stimulation, and targeting ligand. Changes in integrin expression with therapy may be important in future studies. We have also developed a system to quantify small changes in vascular parameters due to effects of new anti-angiogenic drugs using the intrinsic properties of contrast agents. Regions containing intravascular contrast agents are identified using a strategy that combines subharmonic and phase inversion imaging. As predicted by a Rayleigh–Plesset analysis, this strategy can successfully detect flow over a range of transmission frequencies from 4–6 MHz. We demonstrate that regions of viable tumor as small as 1 mm, as verified by histology, can be detected and show similar morphology to images acquired with computed tomography (CT).

10:10–10:30 Break

Contributed Papers

10:30

2aBB5. Effect of contrast agent on the incidence of ultrasound-induced lung hemorrhage in rats. William D. O'Brien, Jr. (Dept. of Elec. and Computer Eng., Univ. of Illinois, 405 N. Mathews, Urbana, IL 61801), Douglas G. Simpson (Univ. of Illinois, Champaign, IL 61820), Leon A. Frizzell (Univ. of Illinois, Urbana, IL 61801), and James F. Zachary (Univ. of Illinois, Urbana, IL 61802)

The objective is to test further the hypothesis that if inertial cavitation in the vasculature of the lung is the physical mechanism responsible for ultrasound-induced lung hemorrhage, then the addition of cavitation nuclei to the blood will enhance the occurrence of lung hemorrhage. A factorial design was used to study the effects of two types of injected agents (IA; 0.25 mL per rat of saline or Optison given intravenously) and two levels of pulsed ultrasound exposure (UE; *in situ* peak rarefactional pressures of 2.74 and 5.86 MPa) on the incidence and size of lung lesions. Ten 10-to-11-week-old Sprague–Dawley rats were exposed at each of the four combinations of IA and UE at 3.1 MHz for 10 s (1-kHz PRF, 1.2-microns PD). Rats administered contrast agent prior to exposure did not have an increase in lesion occurrence or size compared to rats that received saline with no contrast agent. These results provide further evidence that the mechanism of lung hemorrhage is not inertial cavitation. These findings are consistent with another group's results from another species (mouse) showing no increase in the area of lung hemorrhage using a different contrast agent (Albunex) when exposed to pulsed ultrasound. [Work supported by NIH Grant No. R01EB02641.]

10:45

2aBB6. Role of ultrasound contrast agent models in applications and design. John S. Allen, Shin Yoshizawa, Yukio Kaneko, and Yoichiro Matsumoto (Dept. of Mech. Eng., Univ. of Tokyo, Hongo 731, Bunkyo-ku, Tokyo 113, Japan, allenj@fel.t.u-tokyo.ac.jp)

The earliest ultrasound contrast agents were constructed with an air core surrounded by a protein shell. However, contrast agents are currently being produced and developed with a variety of internal gases and shell materials. The agent's shell has been constructed with lipid and polymer materials. These developments have been driven more by empirical experimental fittings rather than a rigorous study of material properties and a construction process that involves both theory and experiment. Furthermore, for diagnostic applications the optimization of the resonance behavior as a function of shell material and thickness is desired. Also the role of the shell in destruction remains an outstanding issue. Recently developed contrast agent models based on hyperelastic materials are further developed to investigate both more extensive strain energy functions and different types of damping from the shell materials. The role of viscous and viscoelastic damping is highlighted. From these radial equations, corresponding versions can be derived in the membrane limit. These models may be used with previously developed bubble dynamics codes, which have a comprehensive formulation for the internal phenomena, and for the analysis of recent experimental results involving therapeutic applications.

11:00

2aBB7. Acoustical correction to the bubble dynamics equation for elastic media. Stanislav Y. Emelianov (Dept. of Biomed. Eng., Univ. of Texas, Austin, TX 78712-1084), Mark F. Hamilton, Yurii A. Ilinskii, G. Douglas Meegan, and Evgenia A. Zabolotskaya (Univ. of Texas, Austin, TX 78713-8029)

Nonlinear oscillations of microbubbles are being investigated extensively for innovative medical and industrial applications. The Rayleigh–Plesset equation successfully describes the dynamics of gas bubbles in liquids and viscoelastic media. To model bubble dynamics in weakly compressible elastic media such as soft tissue, effects of shear stress should be taken into account. These effects were included recently by the authors in a model for bubble oscillations that differs from the Rayleigh–Plesset equation by a single term associated with nonlinear elasticity of the medium surrounding the bubble. Here, we augment this dynamical equation to account for acoustic radiation and thus describe nonlinear bubble oscillations in slightly compressible elastic media. The bubble is assumed to be spherical and the surrounding elastic medium to be isotropic. The analysis is performed in Eulerian coordinates, and attention is devoted to terms associated with radiation loss. Our analysis builds on results obtained by Prosperetti and Lezzi [J. Fluid Mech. **168**, 457 (1986)] for a bubble in liquid. The stress tensor for the elastic medium is expressed through invariants of the Green deformation tensor in spherical coordinates. Explicit expressions for components of the stress tensor are obtained using Mooney's potential function. [Work supported by IR&D at ARL:UT.]

11:15

2aBB8. Forced linear oscillations of microbubbles in blood capillaries. Elisabetta Sassaroli and Kullervo Hynynen (Dept. of Radiol., BWH, Harvard Med. School, Boston, MA)

A numerical investigation of the forced oscillations of ultrasound contrast agents in blood capillaries are investigated by means of a simplified model based on the assumption of small amplitude oscillations of the microbubbles acoustically driven. The natural frequencies of oscillations, the thermal and viscous damping coefficients, the amplitude resonances, as well as the average power absorbed by the system, bubble plus vessel, are computed for different kinds of contrast agent bubbles, containing air, octafluropropane, and perflurobutane. It is found that viscous damping is much larger than thermal damping and that the natural frequencies depend much more on the system parameters than the nature of the gas contained in the bubble. Our motivation for this study lies in the possibility of using bubbles as vectors for directed drug delivery and gene transfection into living cells.

11:30

2aBB9. An interfacial rheological model of ultrasound contrast agents: Theory and experiments. Kausik Sarkar, Dhiman Chatterjee (Mech. Eng., Univ. of Delaware, Newark, DE 19716, sarkar@me.udel.edu), William T. Shi, and Flemming Forsberg (Thomas Jefferson Univ., Philadelphia, PA 19107)

A new interface model with zero thickness and intrinsic surface rheological parameters has been developed for the encapsulation of a microbubble contrast agent. A modified Rayleigh–Plesset equation with these parameters (surface tension γ and surface dilatational viscosity κ^s) was obtained. The rationale for the zero-thickness interface lies in the anisotropy and the inhomogeneity in the structure of the encapsulation. The parameters of the model were determined using in vitro attenuation data for various contrast agents (Albunex, Optison, Sonazoid, and Quantison). The linearized dynamics was used to match data from a controlled set of experiments performed at smaller pressure amplitudes (<0.1 MPa). The model was investigated for its ability to predict the nonlinear response at higher amplitudes (>1.0 MPa) for two test cases: Optison (Mallinckrodt, St. Louis, MO) with $\gamma=0.9$ N/m, $\kappa^s=0.08$ msP and Sonazoid (Amersham Health, Oslo, Norway) with $\gamma=0.6$ N/m, $\kappa^s=0.01$ msP. In contrast to existing models, the new model captured the distinct characteristics of subharmonic emissions—initiation and rapid growth—for these agents. The present model seems to be a reliable descriptor of contrast agent behaviors, and therefore, may be used to characterize different agents and to design next generation agents by tuning the model parameters. [Work supported by DOD.]

11:45

2aBB10. Parameters affecting the detection of microbubbles in blood. Constantin-C. Coussios, Paolo Zanetti, and Ronald A. Roy (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

The ability to detect small gas bubbles in blood depends on the relative magnitude of the acoustic power backscattered from the microbubbles (''signal'') to the power backscattered from the red blood cells (''noise''). Erythrocytes are weak, Rayleigh scatterers, and therefore the backscattering coefficient (BSC) of blood increases as the fourth power of frequency throughout the diagnostic frequency range. Microbubbles, on the other hand, are either resonant or super-resonant in the range 10–30 MHz. Above resonance, their total scattering cross-section remains constant with increasing frequency and the directivity pattern of the scattered wave changes significantly. Therefore, increasing the detection frequency may lead to a reduction in signal-to-noise ratio. An active cavitation detector (ACD) was utilized to observe the gradual obscuring of a steel target in blood with increasing frequency, and to measure the BSC of suspensions of Optison® microspheres in blood, as a function of microsphere concentration, hematocrit and frequency in the range 10–30 MHz. The experimental results were compared with theoretical predictions of the BSC of Optison® and blood, in order to determine whether the presence of tightly packed red blood cells affects the acoustic response of the microbubbles. [Work supported by the ASA, the US Army, and the NSF.]

TUESDAY MORNING, 11 NOVEMBER 2003 PECOS ROOM, 8:00 A.M. TO 12:00 NOON

Session 2aEA

Engineering Acoustics: New Developments

Kim C. Benjamin, Chair
NAVSEA Division Newport, 1176 Howell Street, Newport, Rhode Island 02841

Contributed Papers

8:00

2aEA1. Broadband sonar considerations for small underwater vehicle applications. Kim C. Benjamin (NAVSEA Div. Newport, 1176 Howell St., Newport, RI 02841)

As underwater vehicles become more prevalent, so too does the design and integration of compact broadband sonar arrays. Rather than rely on conventional tonpilz technology, where the bandwidth is governed by the width of the transducer's mechanical resonance, designers must consider other transducer technologies that are better suited to small vehicle packaging constraints. These constraints include operational ruggedness, light weight, conformability, and low cost. This talk advocates the use of 1–3 piezocomposite and discusses the rationale behind such a selection. In going to a wideband material such as a 1–3 piezocomposite, with typical mechanical quality factors (Qm) around 2, the selection of where to place the resonance frequency differs from that of its tonpilz counterpart. For the tonpilz array, the sonar operational bandwidth is totally governed by the resonance response of the array element and is typically limited to approximately its 3-dB or half power points. Relaxor-based ferroelectric materials such as single crystal PMN-PT, which exhibit extremely large electromechanical coupling coefficients, cannot attain 3-dB bandwidths of a decade or more when configured in a tonpilz design. This presentation will discuss a 1–3 piezocomposite-based approach that places mechanical resonance near the upper band edge of an operational bandwidth of 1 decade (10–100 kHz).

8:15

2aEA2. The comb waveform as an efficient method for wideband measurements. Walter H. Boober, Gorham Lau, Kim C. Benjamin (NAVSEA Div. Newport, 1176 Howell St., Newport, RI 02841), and Kenneth M. Walsh (K&M Eng. Ltd., Middletown, RI 02843)

An efficient acoustic calibration technique based on a uniformly-weighted comb waveform is presented. The method takes advantage of the linear, time invariant nature of the measurement configuration and the comb's wide bandwidth to capture all spectral components of interest for a device under test in a single ping. Results comparing single ping comb measurements with conventionally obtained CW measurements are presented. The examples given illustrate the accuracy and utility of this technique for the calibration of broadband systems.

8:30

2aEA3. *In situ* calibration of sonar arrays. L. D. Luker and S. E. Forsythe (Naval Undersea Warfare Ctr. Div. Newport, USRD Code 216, Bldg. 1171B, Rm. 205, 1176 Howell St., Newport, RI 02841, lukerld@npt.nuwc.navy.mil)

The transmitting and receiving properties of the channels of sonar arrays can change with time resulting in a degradation of the array's performance. Fortunately, the degradation in performance can be minimized, perhaps even eliminated, if the changes in a channel's transmitting or receiving properties are compensated for in the array's beamformer elec-

tronics. However, this requires up-to-date knowledge of the acoustic performance of each of the array's channels. This paper describes a procedure for the *in situ* calibration of sonar arrays when the vessel they are installed on is in open water. It can be used to determine changes in the electroacoustic performance of the projecting and receiving channels of the array. The method used is based on a procedure for *in situ* comparison calibration of transducers [A. L. Van Buren, ''Procedure for the *in situ* calibration of sonar transducers,'' J. Acoust. Soc. Am. **90**, 48–52 (1991)] that uses sound-propagation factors measured when the vessel is first deployed to account for the influence of the vessel's structure. Results are presented that show comparisons of the measured degradation of numerous channels in a planar array using an independent acoustic measurement and the *in situ* method. [Work supported by ONR.]

8:45

2aEA4. Shear mode properties of single crystal ferroelectrics. E. A. McLaughlin and H. C. Robinson (NUWC, Code 2132, 1176 Howell St., Newport, RI 02841, mclaughlinea@npt.nuwc.navy.mil)

Single crystal ferroelectrics or piezocrystals were recently introduced into the electroactive materials community. The 33-mode electromechanical coupling factor of piezocrystals is typically greater than 0.90, which is significantly larger than typical values for piezoelectric ceramics (0.62–0.74). For sonar projector applications this large k33 has been responsible for more than doubling the bandwidth of active sonar arrays over what is currently achievable with ceramics. Last year a crystal grower produced a cut of lead magnesium niobate-lead titanate (PMN-PT) single crystal with piezoelectric shear coefficient values of 7000 pm/V and shear coupling factors of 0.97. (For PZT5H, d15 is 730 pm/V.) This piezocrystal d15 coefficient implies significantly improved sensitivity and signal-to-noise ratio for accelerometers and hydrophones, while the high coupling promises bandwidth increases greater than those realized in 33-mode projectors using piezocrystals. This research studies the shear-mode behavior of PMN-PT piezocrystals for use in sensors and projectors. By measuring the response of the materials to high and low level electrical bias and excitation fields, frequency, and temperature, the materials' effective material properties as a function of these operational variables were determined. [Work sponsored by ONR and NUWC ILIR.]

9:00

2aEA5. Frequency shift of a rotating-imbalance vibration source caused by radiative damping in the surrounding medium. Stephen R. Novascone (Idaho Natl. Eng. Lab., 225 N. Fremont, Idaho Falls, ID 83415-2110), Michael J. Anderson (Univ. of Idaho, Moscow, ID 83844-0902), David M. Weinberg (Idaho Natl. Eng. Lab., Idaho Falls, ID 83415-2110), and Jack H. Cole (Univ. of Arkansas)

The motion of vibrating bodies in a surrounding fluid is often used to infer the transport properties of the fluid. A new sensor configuration is presented that consists of a rotating imbalance source radiating into an unbounded fluid medium. Under these circumstances, the reaction of the fluid medium onto the vibration source includes a steady state torque that opposes the applied torque required to sustain the rotating imbalance. This reaction torque causes a shift in frequency of the vibration source. The frequency shift is related to the density of the surrounding fluid medium and vibration source characteristics. A description of measurements taken with a rotating-imbalance source located in unbounded water and air is provided. The total mass, eccentricity, and length of the source were 4.1 kg, $3.28(10^{-4})$ kg m, and 0.432 m, respectively. Motive torque to drive the imbalance was provided by a permanent-magnet dc motor. For an applied dc voltage that caused the source to operate at a nominal frequency near 150 Hz, a frequency shift of approximately 11 Hz was observed when the source was moved from air to water. Experimentally measured frequency shifts compared favorably with predictions provided by a nonlinear steady state model of the source and surrounding medium.

9:15

2aEA6. Amplitude and frequency experimental field measurements of a rotating-imbalance seismic source associated with changes in lithology surrounding a borehole. Stephen R. Novascone (Idaho Natl. Eng. Lab., 2525 N. Fremont, Idaho Falls, ID 83415-2110), Michael J. Anderson (Univ. of Idaho, Moscow, ID 83844-0902), David M. Weinberg (Idaho Natl. Eng. Lab., Idaho Falls, ID 83415), and Jack H. Cole (Univ. of Arkansas)

Field measurements of the vibration amplitude of a rotating-imbalance seismic source in a liquid-filled borehole are described. The borehole was a cased oil well that had been characterized by gamma-ray cement bond and compensated neutron litho-density/gamma-ray logs. The well logs indicated an abrupt transition from shale to limestone at a depth of 2638 ft. The vibration amplitude and frequency of a rotating-imbalance seismic source was measured versus applied voltage as the source was raised from 2654 to 2618 ft through the shale–limestone transition. It was observed that the vibration amplitude changed by approximately 10% in magnitude and the frequency changed approximately 15% as the source passed the shale–limestone transition. The measurements were compared to predictions provided by a two-dimensional analytical model of a rotating-imbalance source located in a liquid-filled bore hole. It was observed that the sensitivity of the experimentally measured vibration amplitude of the seismic source to the properties of the surrounding geologic media was an order of magnitude greater than that predicted by the two-dimensional analytical model.

9:30

2aEA7. Development of a stress gradient enhanced piezoelectric actuator composite (GEPAC) with integrated ultrasonic NDE capablity. Antoine Bechet, Yves Berthelot, and Christopher Lynch (Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

Stress gradient enhanced piezoelectric composites (GEPACs) are unimorph based curved actuators that can be primarily used as low-frequency actuators embedded within an aircraft skin. The actuator is made of a thin PZT ceramic bonded between two composite layers with different coefficients of thermal expansion. Actuators have been manufactured with segmented upper electrodes so that they can also be used at ultrasonic frequencies to monitor continuously the integrity of the actuator itself (cracks), or of the actuator/skin assembly (debonding). This study will focus on the detection of debonding at the interface between the piezoelectric layer and the outer layers. The input signal is a one cycle burst at 1 MHz. The time domain signals recorded experimentally are compared with predictions from a finite element model (ABAQUS explicit). It is shown that the presence of a debonding defect can convert a symmetric pulse into a primarily antisymmetric wave. A simple correlation technique is used to estimate the size of the debonding defect. The ultrasonic NDE operation superposed to the low frequency actuation is demonstrated. [Work supported by the AFOSR.]

9:45

2aEA8. Acoustics of automotive catalytic converter assemblies. Nolan S. Dickey, Ahmet Selamet (Dept. of Mech. Eng., Ohio State Univ., 930 Kinnear Rd., Columbus, OH 43212-1443, dickey.21@osu.edu), Steve J. Parks, Kevin V. Tallio, Keith D. Miazgowicz, and Paul M. Radavich (Ford Motor Co., Dearborn, MI 48121)

In an automotive exhaust system, the purpose of the catalytic converter is to reduce pollutant emissions. However, catalytic converters also affect the engine and exhaust system breathing characteristics; they increase backpressure, affect exhaust system acoustic characteristics, and contribute to exhaust manifold tuning. Thus, radiated sound models should include catalytic converters since they can affect both the source characteristics and the exhaust system acoustic behavior. A typical catalytic converter assembly employs a ceramic substrate to carry the catalytically active noble metals. The substrate has numerous parallel tubes and is mounted in a housing with swelling mat or wire mesh around its periphery. Seals at the ends of the substrate can be used to help force flow

through the substrate and/or protect the mat material. Typically, catalytic converter studies only consider sound propagation in the small capillary tubes of the substrate. Investigations of the acoustic characteristics of entire catalytic converter assemblies (housing, substrate, seals, and mat) do not appear to be available. This work experimentally investigates the acoustic behavior of catalytic converter assemblies and the contributions of the separate components to sound attenuation. Experimental findings are interpreted with respect to available techniques for modeling sound propagation in ceramic substrates.

10:00–10:15 Break

10:15

2aEA9. Whistle suppression at a duct–sidebranch interface by boundary layer perturbation. Andrew Madden, Ahmet Selamet (Ctr. for Automotive Res., The Ohio State Univ., 930 Kinnear Rd., Columbus, OH 43212, selamet.1@osu.edu), and Kevin V. Tallio (Ford Motor Co., Dearborn, MI 48124)

Whistles are generated when shear layer instabilities in a main duct couple with acoustic resonances in a sidebranch. Such flow-acoustic coupling phenomenon is suppressed by perturbing, thus thickening the boundary layer near the interface of the adjoining ducts in an experimental setup, consisting of a closed sidebranch connected to a main duct. The boundary layer is perturbed by inserting pins that protrude into the main duct immediately upstream of the sidebranch duct. The pin height, diameter, and spacing are varied. The sound pressure is measured at the inlet of the main duct using a sound level meter, followed by a spectral analysis to determine the whistle frequencies. Pressure is also measured at the end of the sidebranch by a piezoelectric transducer to quantify the resonant amplitudes. The large-magnitude whistles produced by flow-acoustic coupling are suppressed in nearly all cases by the use of pins which provide, at times, more than 10 dB reduction at peak amplitudes. Furthermore, the introduction of pins into the main flow does not increase the background noise in the duct.

10:30

2aEA10. Special short wave finite elements for flow acoustics. Pablo Gamallo and Jeremy Astley (ISVR, Univ. of Southampton, Highfield, Southampton SO17 1BJ, UK, rja@isvr.soton.ac.uk)

Short wave problems of practical interest in acoustics include the modeling of sound attenuation in the ducted regions of turbofan aircraft engines where the convective and diffractive effects of the mean flow are significant. Also, in this case the acoustic wavelength is generally much smaller than the length scale of the scattering geometry and smaller than the characteristic length scale for variations in the mean flow. The Partition of Unity Method (PUM) [J. M. Melenk and I. Babuška, Comput. Methods Appl. Mech. Eng. **139**, 289–314 (1996)] is a proved efficient numerical method for solving short wave problems in the absence of flow (Helmholtz equation) [Lagrouche *et al.*, Int. J. Numer. Methods Eng. **54**, 1501–1533 (2002)]. The PUM approach is based on the use of a discrete set plane of wave as a local basis for the spatial discretization. When flow is present the wave number of each locally defined plane wave becomes dependent upon the magnitude and direction of the mean flow. The implementation of such scheme for the harmonic acoustic propagation within an irrotational mean flow is proposed in this work. One- and two-dimensional model problems are presented to demonstrate the effectiveness of the approach. [Work supported by EPSRC.]

10:45

2aEA11. Noise control for a ChamberCore cylindrical structure using long T-shaped acoustic resonators. Deyu Li and Jeffrey S. Vipperman (Dept. of Mech. Eng., Univ. of Pittsburgh, 531 Benedum Hall, Pittsburgh, PA 15261, jsv@pitt.edu)

The Air Force Research Laboratory, Space Vehicles Directorate has developed a new advanced composite launch vehicle fairing (referred to as ''ChamberCore''). The ChamberCore is sandwich-type structure fabricated from multi-layered composite face sheets separated by channels that form passive acoustic chambers. These acoustic chambers have a potential to create an acoustic resonator network that can be used to attenuate noise inside the closed ChamberCore cylindrical structure. In this study, first, the feasibility of using cylindrical Helmholtz resonators to control noise in a mock-scale ChamberCore composite cylinder is investigated. The targeted frequencies for noise control are the first four acoustic cavity resonances of the ChamberCore cylinder. The optimal position of the Helmholtz resonators for controlling each targeted cavity mode is discussed, and the effects of resonator spacing on noise attenuation are also experimentally evaluated. Next, six long T-shaped acoustic resonators are designed and constructed within the acoustic chambers of the structure and investigated. Several tests are conducted to evaluate the noise control ability of the resonators in the ChamberCore cylinder. Reductions ranging from 3.2 to 6.0 dB were observed in the overall mean-square noise reduction spectrum at the targeted inner cavity resonance frequencies. [Work supported by AFRL/DV.]

11:00

2aEA12. Design and resonant frequency prediction for long T-shaped acoustic resonators. Deyu Li and Jeffrey S. Vipperman (Dept. of Mech. Eng., Univ. of Pittsburgh, 531 Benedum Hall, Pittsburgh, PA 15261, jsv@pitt.edu)

The use of acoustic resonators is an effective way to control cavity resonances in small enclosures. One popular device is the long, T-shaped acoustic resonator which consists of three branches. Two branches (referred to as ''Branch-1'' and ''Branch-2'') are co-axial and both have one open end and one closed end, and the third branch (referred to as ''Branch-3'') is perpendicular to the co-axis and has two open ends. In practical cavity noise control, the optimal position of Branch-3, i.e., the length of Branch-1 or Branch-2 is determined by the mode shape of the controlled cavity mode, and the length of Branch-3 is typically chosen to be as short as possible to minimize the occupied space of the resonator. If the cross-sectional areas are given, the only design parameter is the length of Branch-1 or Branch-2. In this study, three new models are developed to calculate the end corrections for the three branches. The novel theory is also used to design long T-shaped acoustic resonators for control. In addition to the fundamental acoustic resonator mode, higher frequency resonances of the resonator can be used to target a specific cavity mode for control. Several tests are conducted to experimentally evaluate and validate these new models.

11:15

2aEA13. Infrasonic windscreen. Qamar A. Shams, B. Scott Sealey (M.S. 236, NASA Langley Res. Ctr., Hampton, VA 23681), Allan J. Zuckerwar (NASA Langley Res. Ctr., Hampton, VA 23681), and Laura M. Bott (Univ. of Toledo, Toledo, OH 43606)

Infrasonic windscreens, designed for service at frequencies below 20 Hz, were fabricated from a variety of materials having a low acoustic impedance, and tested in a small wind tunnel against four specifications: (1) attenuation of wind-generated sound (at a free-stream wind speed of 9.4 m/s), (2) the transmission of low-frequency sound from a known source (Janis subwoofer), (3) spectrum of sound generated from trailing vortices, and (4) water absorption (to determine the suitability for all-weather service). Windscreen materials included three woods (pine, cedar, and balsa), polyurethane foam, and Space Shuttle tile material. The windscreen outside diameter ranged from 0.0254 to 0.1016 m (1 to 4 inches), and wall thickness from 0.003175 to 0.01905 m (1/8 to 3/4 inch). A windscreen made of polyurethane foam revealed a wind noise attenuation of 10–20 dB from 0.7–25 Hz, transmission coefficient near unity from 10–25 Hz, and a spectral peak of 23 Hz due to vortex-generated sound. Data will be presented for a variety of windscreens.

11:30

2aEA14. A rumbling index development for sound quality analysis of a vehicle. Sang-Kwon Lee and Byung-Soo Kim (Dept. of Mech. Eng., Inha Univ., 253 Yonghyun Dong, Inchon)

Rumbling sound is one of the most important interior sounds of a passenger car, according to the conventional research. The evaluation of sound quality has been focused on the reduction of the A-weighted sound pressure level. However, an A-weighted sound pressure level cannot give the whole story about the sound quality based on human sensibility. The rumble index is required for the evaluation of the sound quality of a vehicle. In order to make a rumble index, the artificial neural network has been used. The sound metrics for the interior sounds of vehicles are used for the input of the neural network. The subjective ratio for the interior sounds of vehicles is used for the target of artificial neural network to train the weights of the neuron. For the train of the artificial neural network, 150 signals for rumbling sound are synthesized. The interior sounds of 16 vehicles are tested by using the trained artificial neural network. The output of artificial neural network becomes the rumbling index and has 90% correlation with the subjective results for interior sounds of 16 vehicles. [Work supported by BK21.]

11:45

2aEA15. Inversion for acoustic impedance using artificial neural network algorithm. Gee-Pinn Too and Suzi Hwang (Dept. of System and Naval Mechatronic Eng., Natl. Cheng Kung Univ., Tainan, Taiwan, z8008070@email.ncku.edu.tw)

A new approach for measuring acoustic impedance is developed by using Artificial Neural Network (ANN) algorithm. Instead of using impedance tube, a rectangular room or a box is simulated with known boundary conditions at some boundaries and a unknown acoustic impedance at one side of the wall. The training data basis for the ANN algorithm is evaluated by Similar Source Method which was developed earlier by Too (1999) for the estimation of interior and exterior sound field. The training data basis is constructed by evaluating of acoustic pressure at a field point with various acoustic impedance conditions at one side of the wall. The simulation result indicates that the prediction of acoustic impedance is very accurate with error percentage under 1%. Also, one point field measurement in the present approach provides a straightforward and easy evaluation than that in the two points measurement of the impedance tube approach.

TUESDAY MORNING, 11 NOVEMBER 2003 SAN ANTONIO ROOM, 8:30 TO 11:35 A.M.

Session 2aMU

Musical Acoustics and Physical Acoustics: Wind Instruments and Nonlinearity I

Peter L. Hoekje, Chair
Department of Physics and Astronomy, Baldwin-Wallace College, 275 Eastland Road, Berea, Ohio 44017

Chair's Introduction—8:30

Invited Papers

8:35

2aMU1. Linear and nonlinear behavior of human and artificial lip reeds. Murray Campbell and Orlando Richards (School of Phys., Univ. of Edinburgh, Edinburgh EH9 3JZ, UK)

In a musical instrument of the lip reed aerophone class, the flow of air from the player's lungs into the resonating air column is modulated by the periodic opening and closing of the pressure-controlled valve formed by the player's lips. The nature of the operation of this valve has been the subject of considerable study in recent years. Since the pressure–flow relationship is strongly nonlinear, the behavior of the coupled system of lips and air column can only be modeled using the methods of nonlinear dynamics. Extensive studies of artificial lip reeds, in which the lips are simulated by water-filled latex tubes, have shown them to be capable of reproducing musically important features of human playing, including the lipping of notes both below and above an acoustic resonance of the air column. Measurements of the linear response of artificial reeds have guided the development of more realistic models of the lip reed, while studies of both real and artificial lips using a high-speed digital camera have shed fresh light on the nature of the lip motion at the large amplitudes typical of loud playing. [Work supported by EPSRC.]

9:05

2aMU2. Propagation of nonlinear acoustic waves and shocks in variable cross-section resonators. Bart Lipkens (Mech. Eng. Dept., Western New England College, 1215 Wilbraham Rd., Springfield, MA 01119)

Relative peak acoustic pressures in wind instruments such as oboes, trumpets, and trombones can reach levels of 10%. At these elevated levels nonlinear acoustic effects have to be taken into account to accurately represent the acoustic wave propagation in these wind instruments. A short survey is presented of some of the past work in nonlinear resonant oscillations in tubes. Included in this discussion is the effect of shaping the resonator on nonlinear wave propagation. A one-dimensional model for the analysis of nonlinear acoustic waves in a variable cross-section resonator is introduced [Ilinski *et al.*, J. Acoust. Soc. Am. **104**, 2664–2674 (1998)]. The resonator is assumed to be of an arbitrary axisymmetric shape. The model equations are derived from the fundamental gas dynamic equations for an ideal gas. The effects of bulk viscous absorption and absorption in the viscous boundary layer are included. Nonlinear spectral equations are developed which are integrated numerically to obtain the solution. The model has been modified to include a radiation impedance and the excitation of a harmonic spectrum. The effect of resonator shape, radiation condition, and harmonic excitation on nonlinear wave propagation is discussed.

9:35

2aMU3. Localized and cumulative nonlinearity in wind instruments. Joel Gilbert (Laboratoire d'Acoustique de l'Universit du Maine—UMR CNRS 6613, Ave. Olivier Messiaen, 72085 Le Mans, Cedex 9, France)

Nonlinearities are very common in wind instruments. A crucial one localized at the input of the wind instrument is responsible for the sound production mechanism. As an illustration, some recent measurements done at clarinet mouthpieces will be shown. Some other localized nonlinear effects take place at the open tube ends. They imply extra losses whose amount depends on the internal geometry of the termination. They control the sound extinction phenomena. It will be shown how the playing range of a clarinet-like instrument is determined by these extra losses. Besides localized nonlinearity, cumulative nonlinearity effects are present as well. The cumulative nonlinear propagation phenomena along the tube of brass instruments can lead to shock waves obtained when the player is playing very loudly with a ''brassy sound.''

10:05

2aMU4. Interaction of pipe tone and edge tone in organ pipe oscillation. A. W. Nolle (Dept. of Phys., Univ. of Texas, Austin, TX 78712, nolle@mail.utexas.edu)

Inharmonic partials in an organ pipe tone can occur in clusters, at uniform frequency separation. The primary object of this discussion is to demonstrate that these clusters arise in a nonlinear process: Edge tone oscillation is modulated at the fundamental frequency of the pipe tone. The mechanism is demonstrated by a laboratory analog, in which periodic deflection of a jet producing an edge tone is forced by a transverse acoustic flow, and also by a numerical analog. Additional examples show that the simultaneous occurrence of a pipe tone and an unmodulated edge tone can occur at low amplitudes of oscillation, where the nonlinear interaction is unimportant.

10:35

2aMU5. Feathering collisions in beating reed simulation. Tamara Smyth (Ctr. for Computer Res. in Music and Acoust., Dept. of Music, Stanford Univ., Stanford, CA 94305-8180), Jonathan S. Abel (Univeral Audio, Inc., Santa Cruz, CA 95063-3818), and Julius O. SmithIII (Stanford Univ., Stanford, CA 94305-8180)

Pressure controlled valves are the primary sound production mechanisms for woodwind and brass musical instruments, as well as for many bioacoustic vocal systems such as the larynx and syrinx (the vocal organ in birds). During sound production, air flow sets a reed or membrane into motion creating a variable height in the valve channel and, potentially, periodically closing the channel completely. Depending on how this event is handled, an abrupt termination of air flow between open and closed states can cause undesirable discontinuities and inaccuracies in a discrete-time simulation—particularly at relatively low audio sampling rates. A solution was developed by re-examining the behavior of the differential equation governing volume flow through a pressure-controlled valve, paying particular attention to this rather troublesome transition. A closed-form solution for the time evolution of volume flow is given and used to derive an update for volume flow. The result is a smoother, more accurate, and nearly alias-free transition from open to closed. ''Feathered collisions'' of this nature can refine the sound quality produced by the numerical simulation of beating reeds, such as in clarinets, at typical audio sampling rates.

Contributed Papers

11:05

2aMU6. Error metrics for predicting discrimination of original and spectrally altered musical instrument sounds. James W. Beauchamp (School of Music and Dept. of Elec. & Computer Eng., Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801, jwbeauch@uiuc.edu) and Andrew Horner (Hong Kong Univ. of Sci. & Technol., Clear Water Bay, Kowloon, Hong Kong)

The correspondence of various error metrics to human discrimination data was investigated. Time-varying harmonic amplitude data were obtained from spectral analysis of eight musical instrument sounds (bassoon, clarinet, flute, horn, oboe, saxophone, trumpet, and violin). The data were altered using fixed random multipliers on the harmonic amplitudes, and the sounds were additively resynthesized with estimated average spectral errors ranging from 1% to 50%. Listeners attempted to discriminate the randomly altered sounds from reference sounds resynthesized from the original data. Then, various error metrics were used to calculate the spectral differences between the original and altered sounds, and the R^2 correspondence between the error metrics and the discrimination data was measured. A relative-amplitude spectral error metric gave the best correspondence to average subject discrimination data, capturing over 90% of the variation relative to a Fourth-order regression curve, although other formulas gave similar results. Error metrics which used a small number of representative analysis frames gave results which compared favorably to using all frames of the analysis.

11:20

2aMU7. The rams horn in western history. David Lubman (David Lubman & Assoc., 14301 Middletown Ln., Westminster, CA 92683)

The shofar or rams horn—one of the most ancient of surviving aerophones—may have originated with early Neolithic herders. The shofar is mentioned frequently and importantly in the Hebrew bible and in later biblical and post-biblical literature. Despite its long history, contemporary ritual uses, and profound symbolic significance to western religion, no documentation of shofar acoustical properties was found. Since ancient times, shepherds of many cultures have fashioned sound instruments from the horns of herd animals for practical and musical uses. Shepherd horns of other cultures exhibit an evolution of form and technology (e.g., the inclusion of finger holes). The shofar is unique in having retained its primitive form. It is suggested that after centuries of practical use, the shofar became emblematic of the shepherd culture. Ritual use then developed, which froze its form. A modern ritual rams horn played by an experienced blower was examined. This rather short horn was determined to have a source strength of 92 dB (A) at 1 m, a fundamental frequency near 420 Hz, and maximum power output between 1.2 and 1.8 kHz. Sample sounds and detection range estimates are provided.

Session 2aNS

Noise and Architectural Acoustics: Acoustical Design Issues for the Health Care Industry

Kerrie G. Standlee, Cochair
Daly-Standlee and Associates, 4900 Southwest Griffith Drive, Suite 216, Beaverton, Oregon 97005

Jack B. Evans, Cochair
Engineering, Vibration and Acoustics, JEAcoustics, 5806 Mesa Drive, Suite 380, Austin, Texas 78731-3742

Chair's Introduction—9:00

Invited Papers

9:05

2aNS1. Acoustical criteria for hospital patient rooms: Resolving competing requirements. Bennett M. Brooks (Brooks Acoust. Corp., 27 Hartford Turnpike, Vernon, CT 06066, brooks@brooks-acoustics.com)

The acoustical criteria for patient rooms in hospitals, nursing homes, and rehabilitation facilities may be based on several needs. One important requirement is that noise levels in the room be conducive to restful sleep. Also, caregivers must have easy auditory and visual access to the patients, and be able to hear vital sign monitor alarms. This often means that patient rooms are located near central nurse stations and that patient room doors are left open. Further, the recently published federal privacy standards developed by the U.S. Department of Health and Human Services (HSS) under the Health Insurance Portability and Accountability Act (HIPAA) require that ''appropriate physical safeguards'' be put in place to protect the confidentiality of patient health information. The simultaneous and competing requirements for speech privacy, caregiver access, and good sleeping conditions present a serious acoustical challenge to health care facility designers. Specific facility design issues and potential solution strategies are presented.

9:25

2aNS2. Noise levels, spectra, and operational function of an occupied newborn intensive care unit built to meet recommended permissible noise criteria. M. Kathleen Philbin (Children's Regional Hospital at Cooper Univ. Hospital, One Cooper Plaza, Dorrance 755, Camden, NJ 08103, philbika@umdnj.edu) and Jack B. Evans (JEAcoust., Inc., Austin, TX 78731)

A group of clinical experts developed recommended permissible noise criteria for newly constructed or renovated hospital nurseries [Philbin *et al.*, J. Perinatol. **19**, 559–563 (2000); R. White, *ibid.* **23**, S1–22 (2003)]. These criteria are based principally on research regarding wake-up thresholds for term newborns and speech interference levels for adults. These criteria are: The overall continuous A-weighted, slow response, sound level at any bed or patient care area shall not exceed: (1) an hourly Leq of 50 dB, (2) an hourly L10 of 55 dB, and (3) a 1-s Lmax of 70 dB. A new hospital building was designed to meet these criteria by using specific acoustical criteria for the structure and space arrangement [J. B. Evans and M. K. Philbin, J. Perinatol. **20**, S105–S112 (2000)]. Acoustical criteria for sound isolation, background NC, structural vibration, and reverberation will be presented along with space arrangements that ensure staff efficiency, clinical safety, and family privacy. Post-occupancy measurements of sound levels and spectra along with photographs of a nursery in operation will be presented to illustrate how an ICU can have a quiet, highly functioning intensive care environment while meeting the operational goals and acoustical criteria.

9:45

2aNS3. Floor vibration evaluations for medical facilities. Chad N. Himmel (JEAcoust., 5806 Mesa Dr., Ste. 380, Austin, TX 78731, himmel@jeacoustics.com)

The structural floor design for new medical facilities is often selected early in the design phase and in renovation projects, the floor structure already exists. Because the floor structure can often have an influence on the location of vibration sensitive medical equipment and facilities, it is becoming necessary to identify the best locations for equipment and facilities early in the design process. Even though specific criteria for vibration-sensitive uses and equipment may not always be available early in the design phase, it should be possible to determine compatible floor structures for planned vibration-sensitive uses by comparing conceptual layouts with generic floor vibration criteria. Relatively simple evaluations of planned uses and generic criteria, combined with on-site vibration and

noise measurements early in design phase, can significantly reduce future design problems and expense. Concepts of evaluation procedures and analyses will be presented in this paper. Generic floor vibration criteria and appropriate parameters to control resonant floor vibration and noise will be discussed for typical medical facilities and medical research facilities. Physical, economic, and logistical limitations that affect implementation will be discussed through case studies.

10:05–10:20 Break

10:20

2aNS4. Structure-borne sound from magnetic resonance imaging systems. Eric E. Ungar and Jeffrey A. Zapfe (Acentech, Inc., 33 Moulton St., Cambridge, MA 02138)

Magnetic resonance imaging (MRI) systems are known to produce a considerable amount of audible noise. The recent tendency to install such systems on above-grade floors has led to increasing concerns about structure-borne noise transmission from the MRI to adjacent occupied areas. This paper presents the results of a study in which structure-borne noise forces produced by two operational MRI systems were determined via measurement of the floor vibrations induced by the systems and of the impedance of their supporting floors. Forces with known spectra were applied to the floors of planned MRI suites in a hospital extension and the corresponding noise in adjacent areas was measured. Similarly, airborne noise was introduced in the planned suites and the related noise in adjacent areas was measured. The results then were scaled to correspond to the measured MRI forces and airborne noise. It was found that in areas below the planned MRI installations structure-borne noise would predominate, unless it is mitigated. Structure-borne noise isolation of MRI systems, whose environments must meet stringent vibration criteria, is discussed briefly.

10:40

2aNS5. The MRI: A noise source of concern in the health care industry. Kerrie G. Standlee and Joseph C. Begin (Daly-Standlee & Assoc., Inc., 4900 SW Griffith Dr., Ste. 216, Beaverton, OR 97005, jbegin@acoustechgroup.com)

Two recent trends in the development and use of magnetic resonance imaging (MRI) equipment have created challenges for acoustical engineers: (1) the trend toward more powerful MRI machines with greater magnetic field strengths, and (2) the tendency of health care facilities to locate these machines, which were previously located in basements or on grade, on upper floors adjacent to (and in some cases above) other critical use areas. For newer, 3-T MRI machines, sound levels well over 100 dBA in the examination room are common. Along with these trends, some equipment manufacturers are now providing design recommendations to address the issues of airborne and structure-borne noise within hospitals and clinics. In addition, MRI manufacturers sometimes have strict requirements for acceptable levels of building vibration from other sources, to prevent potential image quality problems. This paper discusses experience gained during the course of addressing MRI-generated noise on several projects. Data for airborne sound levels measured inside MRI rooms and adjacent rooms and vibration levels measured below MRI units will be presented.

11:00

2aNS6. MRI structure borne noise, in and out. K. Anthony Hoover (Cavanaugh Tocci Assoc., 327 F Boston Post Rd., Sudbury, MA 01776, thoover@cavtocci.com)

MRI equipment typically requires very low levels of ambient floor vibration. However, this same equipment often generates levels of vibration that are sufficient to cause complaints of excessive MRI noise in nonadjacent rooms. This paper will review two projects. The first investigates levels of floor vibrations caused by flyovers of medivac helicopters. The second investigates a structural floor slab designed to satisfy MRI ambient-vibration criteria, but, as predicted, allowed for excessive structure borne MRI noise. Difficulties with standard vibration isolation of MRI equipment will be discussed.

11:20

2aNS7. Environmental noise issues associated with medical facility HVAC equipment. Ted N. Carnes, Howard K. Pelton, and Daniel W. Saenz (Pelton Marsh Kinsella, 1420 W. Mockingbird Ln., Ste. 400, Dallas, TX 75247, PeltonHK@c-b.com)

Medical facilities comprise a variety of types of buildings from research facilities to hospitals to professional office buildings to name but three. Many new hospitals are being located in close proximity to residential housing. Furthermore, as hospitals expand the buildings are closer to residential communities. Thus, the community noise aspects of these facilities are of even more importance than ever. This paper will examine how community noise intrusions can occur without proper planning, and how this can be avoided by knowing the community noise criteria, working closely with the design team to assist them with the proper planning and noise control design. Case histories will be presented that will include a hospital expansion with large rooftop equipment, new cooling towers and how these can be modeled as part of the overall hospital campus model to determine the community noise impact. In addition, a new chiller plant expansion adjacent to the residential areas will be examined, and how this can be modeled to show the various aspects of the community noise impact.

Session 2aPA

Physical Acoustics: Thermoacoustics and Resonators

Gregory W. Swift, Chair
Condensed Matter and Thermal Physics Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Contributed Papers

9:00

2aPA1. Regenerator-based thermoacoustic refrigerator for ice cream storage applications. Matthew E. Poese, Robert W. M. Smith, and Steven L. Garrett (Penn State Appl. Res. Lab., P.O. Box 30, State College, PA 16804, poese@psu.edu)

A regenerator-based chiller has been built in the "bellows bounce" style [J. Acoust. Soc. Am. **112**, 15 (2002)] to replace the vapor compression system in an ice cream sales cabinet. It utilizes a 6-in.-diam metal bellows to form a compliant cavity that contains the dynamic pressure oscillation ($>$50 kPa). The stiffness of the gas trapped in the bellows is resonated against the mass of the bellows-cap and the mass of a moving-magnet linear motor which is capable of high ($>$85%) electro-acoustic efficiency. A second resonator, operated well below its natural frequency, uses the gas stiffness of a 1-l volume nested within the bellows and the inertia of an ordinary loudspeaker cone to create the pressure difference across the regenerator that drives gas flow that is in-phase with pressure. The mass of the cone can be adjusted to vary the multiplication factor that is typically 5%–10% greater than the dynamic pressure within the bellows. The loudspeaker cone suffers none of the hydrodynamic losses associated with an acoustic inertance and eliminates problems with dc gas flow in the energy feedback path. The cold heat exchanger forms one surface of the pressure vessel permitting direct contact with any thermal load. [Work supported by Ben and Jerry's Homemade.]

9:15

2aPA2. The performance of a high-frequency thermoacoustic-Stirling engine. Kevin J. Bastyr and Robert M. Keolian (Penn State Univ., P.O. Box 30, State College, PA 16804)

A thermoacoustic-Stirling engine that operates at 400 Hz with a working fluid of 1-MPa helium is constructed. For proper acoustic phasing in this engine's regenerator, an acoustic power feedback path exists in the form of an annulus surrounding the regenerator. This feedback path is obtained by suspending an insulated, stainless steel sleeve containing a wire mesh regenerator, which is flanked by two heat exchangers, a short distance from one end of the larger diameter resonator. The ambient heat exchanger is a shell and tube exchanger, while the hot heater consists of nichrome ribbon wound on an aluminum silicate frame. Gedeon streaming is prevented by a diaphragm covering the end of the stainless steel sleeve adjacent to the ambient heat exchanger. A variable acoustic load provides a convenient means of testing this engine at various hot heater temperatures, while operating at different acoustic pressure amplitudes effects the acoustic power generated by the engine. [Work supported by ONR.]

9:30

2aPA3. Resonant self-pumped thermoacoustic heat transfer. Greg Swift and Scott Backhaus (Condensed Matter and Thermal Phys. Group, Los Alamos Natl. Lab., Los Alamos, NM 87545, swift@lanl.gov)

As a thermoacoustic engine or refrigerator is scaled up to higher power, it generally needs heat exchangers of a larger cross-sectional area. Constrained to lie adjacent to a stack or regenerator, large heat exchangers must use intricate geometries to interweave the thermoacoustic working gas and an external fluid (e.g., ambient cooling water or hot combustion gas) to bring them into intimate thermal contact. Such heat exchangers have many thermoacoustic passages in parallel, often entailing many parts and many joints that must be leak tight. Resonant self-pumped thermoacoustic heat transfer is a new alternative in which the thermoacoustic-to-external heat-transfer surface area is in the form of one long tube (or a few) extending far from the associated stack or regenerator. With a resonant sound wave in the tube and with diode-like nonlinear flow elements near the wave's velocity maxima, a strong time-averaged circulation of the working gas carries the heat. This concept will be described and preliminary experimental results will be presented. [Work supported by DOE's Office of Science and by Praxair, Inc.]

9:45

2aPA4. Driver parameter characterization method for a thermoacoustic cooler. Insu Paek, Luc Mongeau, and James E. Braun (Ray W. Herrick Labs., Purdue Univ., West Lafayette, IN 47907)

The efficiency of transducers used in electromechanically driven standing wave thermoacoustic coolers is governed by six parameters. The maximum efficiency is obtained around the frequency at which the total electrical input reactance is zero. The parameters in a linear circuit representation of the dynamic behavior of the driver are often frequency and amplitude dependent. This nonlinear response is due to the large acoustic power delivered and the resonant nature of the acoustic load. A method was developed to determine quickly and accurately driver equivalent parameters from independent measurements of the input voltage and current, as well as the piston acceleration and pressure. A system identification technique was used. Transfer function measurements were performed for a few driving frequencies close to the optimal operating condition of the thermoacoustic cooler. The data were used to calculate driver parameters off-line. Parameter values obtained for low piston force and displacement amplitudes were found to be significantly different than their values for larger amplitudes. The method can be used to find the optimal driver parameters for target operating conditions of a particular thermoacoustic cooler.

10:00

2aPA5. Heat transfer of heat exchangers in a standing wave thermoacoustic cooler. Insu Paek, James E. Braun, and Luc Mongeau (Ray W. Herrick Labs., Purdue Univ., West Lafayette, IN 47907)

Standing wave thermoacoustic coolers require heat exchangers to move heat in and out of the working gas. Within most finned-tube type heat exchangers, heat is transferred from the oscillating gas to a liquid flowing through the tubes. Experiments were performed to evaluate the thermal performance of heat exchangers immersed in a standing wave thermoacoustic cooler. The gas-side convection heat transfer coefficient was calculated from the measured data, and compared to steady flow heat transfer model predictions. A simple relation between the Reynolds number in steady flow and the time dependent acoustic Reynolds number was obtained. The results suggest that substituting this relation into known steady flow correlations may yield accurate predictions of the heat transfer coefficient in oscillating flow.

10:15

2aPA6. Minor losses in high-amplitude acoustic resonators with varying changes in cross section. Andrew J. Doller and Anthony A. Atchley (Grad. Prog. in Acoust., Penn State Univ., University Park, PA 16802, adoller@psu.edu)

In a previous paper [J. Acoust. Soc. Am. **112** (2003)], measurements of time-averaged pressure at the closed end of a high-amplitude resonator having a non-uniform cross section were presented. The motivation for this continued work is to better understand minor losses in high-amplitude oscillatory flows through changes in resonator cross section. The resonators consist of two sections of straight brass pipe of different diameters and joined through either step-like or conical couplers. They are driven at the end of the large diameter pipe with a rigid piston. The end of the small tube is rigidly terminated. In the current paper, data obtained from an improved experimental apparatus will be presented. The improvements consist of better temperature control and the ability to measure the power delivered to the resonator by the piston. Subtraction of the linear acoustic boundary-layer losses from the measured input power provides an upper estimate on the power dissipated through other mechanisms. The dependence of the time-averaged pressure and power dissipation on drive amplitude, junction type, and junction location in the standing wave will be discussed. [Research supported by ONR.]

10:30

2aPA7. Numerical simulation of hydrodynamic and thermal effects around a 2-D stack plate in a thermoacoustic refrigerator. David Marx and Philippe Blanc-Benon (Ecole Centrale de Lyon, Ctr. Acoustique, LMFA UMR CNRS 5509, 69134 Ecully Cedex, France)

Two-dimensional Navier–Stokes equations for an unsteady and compressible flow are solved around a 2-D stack plate of a thermoacoustic refrigerator. The computational domain is a resonator slice including the resonator end but not the acoustic source. The standing wave is sustained in the domain using the characteristics method. The same methods were already used for the simulation of the coupling between stack plates and heat exchangers (AIAA Paper No. 2003-3150). The oscillating flow above the 2-D plate induces a periodic vortex creation at the ends of the plate. Numerical results will be compared to PIV measurements. The effect of the acoustic Mach number and geometrical parameters are studied. The temperature difference between the ends of the plate is also investigated and compared to linear theory. Some deviations can be attributed to the difference between the mean temperature gradient in the plate and the mean temperature gradient in the fluid. A streaming flow is also observed above the plate. [Work supported by DGA.]

10:45

2aPA8. Open cycle traveling wave thermoacoustics. Nathan T. Weiland and Ben T. Zinn (Schools of Mech. and Aerosp. Eng., Georgia Inst. of Technol., Atlanta, GA 30332, gte852f@mail.gatech.edu)

The concepts and basic operation of open cycle traveling wave thermoacoustic devices are discussed for the case of a simple thermoacoustic engine, where a steady flow of hot gas replaces the hot heat exchanger as the means for supplying heat to the engine. It has been determined that a large mean temperature difference must exist where the mean flow first meets the regenerator for the engine to produce any acoustic power, due to the joining conditions between the ideally isothermal environment in the regenerator and the isentropic environment of the adjoining open duct that contains the mean flow. The physics behind this temperature difference is discussed, as well as its effects on the various energy fluxes in the thermoacoustic engine. In particular, it is shown that the acoustic power output and thermal efficiency of the engine can be maximized with respect to the magnitude of this temperature difference.

11:00

2aPA9. Investigation of harmonic generation in thermoacoustic engines. James E. Parker, Mark F. Hamilton, and Yurii A. Ilinskii (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713-8029)

One factor limiting the efficiency of thermoacoustic engines and refrigerators is the generation of harmonics. Not only do thermoviscous losses increase in relation to the generation of higher frequencies, but the stack is not designed to optimize thermoacoustic processes at these frequencies. We describe here a semi-analytical investigation of nonlinear effects associated with harmonic generation. The approach is based on a regular perturbation expansion of the acoustical quantities, combined with the assumptions and methodology underlying the linear theory developed by Rott. At first order, Rott's equations for the pressure and particle velocity are obtained and solved numerically. The second-order system consists of Rott's equations for the second harmonic driven by a forcing function containing products of the first-order solutions. The goal is to solve the third-order system, given by Rott's equations for the fundamental driven by products of the first- and second-order solutions, to determine the nonlinear correction to the fundamental component. The investigation was not completed at the time of writing, and we therefore present here the intermediate results for second-harmonic generation. These results provide partial explanations of how harmonic generation influences thermoacoustics. [Work supported by ONR and ARL:UT IR&D.]

11:15

2aPA10. Multiple existence of steady state solutions of acoustic streaming in a long two-dimensional rectangular box. Takeru Yano (Div. of Mech. Sci., Grad. School of Eng., Hokkaido Univ., Sapporo 060-8628, Japan, yano@mech-me.eng.hokudai.ac.jp)

The large amplitude standing wave excited in a resonator induces acoustic streaming of Rayleigh type outside the acoustic boundary layer on the wall of the resonator. The streaming motion with large Reynolds number is examined numerically in relatively long two-dimensional rectangular boxes. The two-dimensional incompressible Navier–Stokes equations with no external force are used as the governing equations for the streaming velocity. The steady velocity component at the outer edge of the acoustic boundary layer, which induces Rayleigh type streaming, is employed as the boundary condition for the Navier–Stokes equations. By using a finite-difference method, the existence of multiple steady state solutions is demonstrated. Some of the numerically obtained flow patterns bear a resemblance to those obtained experimentally. It may be remarked that all the flow patterns induced by the fifth mode standing wave in a relatively long box can be regarded as combinations of those induced by the second mode standing wave in a short box.

11:30

2aPA11. Update on the simulation of streaming in tapered resonators. Brian C. Tuttle and Victor W. Sparrow (Grad. Prog. in Acoust., Penn State Univ., 217 Appl. Sci. Bldg., University Park, PA 16802)

In a previous report [J. Acoust. Soc. Am. **110**, 2652 (2001)] the authors described a plan to simulate nonlinear acoustic streaming by way of a MacCormack finite-difference scheme running on a cluster-type multiprocessor computer. The object was to investigate how the tapering of a resonator tube affects the acoustic streaming. The experiments of Olson and Swift predict that a temperature-dependent viscosity in conjunction with tube geometry can minimize Rayleigh streaming for a particular taper angle. Recent progress in the present research includes a careful perturbation analysis of the model equations to incorporate additional nonlinear terms that are important to simulate flow near a boundary. A modified computational grid necessary to accommodate the tapered domain as well as the refined boundary layer is discussed. The method of driving the tube and issues arising in the simulation starting conditions are also described. [Work supported in part by ONR.]

Session 2aPP

Psychological and Physiological Acoustics: A Potpourri for the Informative

W. Jay Dowling, Chair
Program in Cognitive Sciences, University of Texas–Dallas, Richardson, Texas 75083-0688

Chair's Introduction—9:00

Contributed Papers

9:05

2aPP1. Stimulus-driven and knowledge-driven processes in attention to warbles. W. Jay Dowling (Prog. in Cognit. Sci., Univ. of Texas–Dallas, Richardson, TX 75083-0688, jdowling@utdallas.edu) and Barbara Tillmann (Univ. Claude Bernard Lyon I, Lyon, France)

Listeners identified warbles differing in amplitude-modulation rate (3–10 Hz). And measured RT while listeners maintained above 90% correct responses. After a practice session listeners identified target warbles following stimulus-driven or knowledge-driven cues. The stimulus-driven cue was a 250-ms ''beep'' at the target pitch (valid) or another pitch (invalid); the knowledge-driven cue was a midrange ''melody'' pointing to the target pitch (always valid). A 500-ms target warble followed the cue after delays of 0–500 ms (250–750 ms SOA). The listener pressed a key to indicate ''slow'' or ''fast.'' RTs were shortest at the briefest delay. In contrast to results from a memory task, RTs here were much shorter, and we found no evidence for IOR or attentional blink. Listeners began generating responses while the target was still sounding. Invalid ''beeps'' slowed responses at the briefest (but not the longer) delays; adding a valid ''beep'' to the valid ''melody'' did not speed responses.

9:20

2aPP2. Interpreting electrically evoked emissions using a finite-element model of the cochlea. Niranjan V. Deo (Dept. of Mech. Eng., Univ. of Michigan, Ann Arbor, MI 48109, ndeo@umich.edu), Karl Grosh (Univ. of Michigan, Ann Arbor, MI 48109), and Anand Parthasarathi (Bose Corp., Framingham, MA)

Electrically evoked otoacoustic emissions (EEOAEs) are used to investigate *in vivo* cochlear electromechanical function. Electrical stimulation through bipolar electrodes placed very close to the basilar membrane (in the scala vestibuli and scala tympani) gives rise to a narrow frequency range of EEOAEs, limited to around 20 kHz when the electrodes are placed near the 18-kHz best frequency place. Model predictions using a three-dimensional inviscid fluid model in conjunction with a middle ear model [S. Puria and J. B. Allen, J. Acoust. Soc. Am. **104**, 3463–3481 (1998)] and a simple model for outer hair cell activity [S. Neely and D. Kim, J. Acoust. Soc. Am. **94**, 137–146 (1993)] are used to interpret the experimental results. To estimate effect of viscosity, model results are compared with those obtained for a viscous fluid. The models are solved using a 2.5-D finite-element formulation. Predictions show that the high frequency limit of the excitation is determined by the spatial extent of the current stimulus. The global peaks in the EEOAE spectra are interpreted as constructive interference between electrically evoked backward traveling waves and forward traveling waves reflected from the stapes. Steady state response predictions of the model are presented.

9:35

2aPP3. Global electro-structural-acoustic filtering in the cochlea. Sripriya Ramamoorthy and Karl Grosh (Dept. of Mech. Eng., Univ. of Michigan, Ann Arbor, MI 48105)

The active filtering in the cochlea is now believed to be due to the piezo-motility of the Outer Hair Cell (OHC), and the modulation of stereocilia resistance. Physiologically, the OHC resting potential is known to be modulated by the K+ ions in the cochlear fluid entering or leaving the OHC via gated channels, in response to structural motion, which in turn couples the electrical potential in the cochlear fluids to OHCs piezomotility. However, the electrotonic space-constant of the ionic fluid in the cochlea is not negligible compared to the structural-acoustic wavelength for all frequencies, thus limiting the currently existing models that adopt local structural-electrical interaction. The electrical wave interacts globally with the structural-acoustic wave. This paper proposes a three-dimensional global electro-structural-acoustic linear finite element model to analyze the coupled behavior in the active cochlea. The OHC is modeled as a piezoelectric bar. A cable model is used for the electrical voltages, and a micromechanical model based on Dallos (2002) is proposed. Modeling is maintained as physiological as possible. The active response of the cochlea to acoustic input, Electrically Evoked Oto-Acoustic Emissions, efferent feedback, and the potential of IHC in response to acoustic or electrical input are studied.

9:50

2aPP4. A nonlinear cochlear model with the outer hair cell piezoelectric activity. Xiaoai Jiang and Karl Grosh (Dept. of Mech. Eng., Univ. of Michigan, 2350 Hayward, Ann Arbor, MI 48109, jiangxa@umich.edu, grosh@umich.edu)

In this paper we present a simple cochlear model which captures the most important aspect of nonlinearity in the cochlea—the nonlinearity caused by the piezoelectric-like activity of outer hair cells and the variable conductance of the outer hair cell stereocilia. A one-dimensional long-wave model is built to simulate the dynamic response of the fluid-loaded basilar membrane. The basilar membrane is simulated as isolated linear oscillators along the cochlear length, and its motion is coupled with the fluid pressure and the nonlinear force produced by the outer hair cells. As the basilar membrane moves, the fluid shears stereocilia, and the resulting ion flow changes the transmembrane potential of the outer hair cells and subsequently their length, leading to further movement of the basilar membrane. The piezoelectric-like activity of the outer hair cell is simulated by a current source, and stereocilia motion is modeled as a varying conductance that changes as the basilar membrane moves. A solution in the time domain will be presented. [Work supported by NIH.]

10:05

2aPP5. Effect of loudness on the haptic force-feedback perception in virtual environments. M. Ercan Altinsoy, Jens Blauert (Inst. of Commun. Acoust., Ruhr Univ. Bochum, D-44780 Bochum, Germany), and Richard H. Y. So (Hong Kong Univ. of Sci. & Technol.)

Due to increasing usage of the haptic modality in virtual reality applications, the perceptual aspects of the interaction between haptic and auditory modalities is becoming more important. In our daily life, force-feedback which we perceive in our hands after beating a drum as well as the loudness of the drum sound give us the required information about how much force we have applied when playing a drum. In this study, psychophysical experiments were conducted to investigate the effect of loudness on the haptic force-feedback perception (strength) by playing a virtual drum. Subjects were presented (1) only haptic force-feedback information, (2) only auditory information, and (3) auditory and haptic in-

formation together. In each experiment, subjects were asked how much force they had applied when playing the virtual drum. The results show that the magnitude of strength increases with increasing loudness in spite of no change in force-feedback as generated by a virtual drum and applied to the subject's hand. These results contribute to our knowledge on how to effectively combine haptic and auditory information in virtual environments. [Work supported by IGSN, Ruhr University Bochum.]

10:20–10:35 Break

10:35

2aPP6. Effect of a non-target sound event on motion perception of a target sound event. Tsuruyo Nishida (Dept. of Elec. and Computer Eng., Gifu Natl. College of Technol., 2236-2 Kamimakuwa, Shinsei-cho, Motosu-gun, Gifu 501-0461, Japan) and Kazuhiko Kakehi (Grad. School of Information Sci./CIAIR, Nagoya Univ., Furo-cho, Chikusa, Nagoya 464-8601, Japan)

Most psychoacoustics studies on sound localization and motion perception have been conducted for only one sound event. In this report, one- or two-sound events were presented and the direction and motion perception were investigated using two-sound phenomena to produce motion perception; the first was apparent motion and the second was a synthesized sound image. There were no significant differences in the correct response percentage for both two-sound phenomena. The correct response percentage for the direction of motion was significantly worse for two-sound events than that for one-sound event, even if the two-sound events are separately perceived in terms of frequency, space domain, and temporal order. The analysis of wrong responses showed that the perception depends on IEOIs and the combination of target and non-target moving conditions (e.g., moving or still). Furthermore, it was clarified with signal detection theory that the motion perception of target was strongly affected by the non-target conditions. It means that the process of perceiving the direction and motion of the target might be different than that of perceiving the sound feature.

10:50

2aPP7. Review and comparison of methods using spectral characteristics for the purposes of CASA, Computational Auditory Scene Analysis. Jerry Gregoire (Montana State Univ., 622 Cobleigh Hall, Bozeman, MT 59717, jgregoire@ece.montana.edu)

The brain has a remarkable ability to separate signals from a mixture. Although much progress has been made, the goal of replicating this has eluded us. Bregman (90) suggests that the brain divides this analysis into low or primitive, and high or schema processes. Schema processes require memory and subsequent interpretation. Conversely, primitive processes operate only on the innate properties of the incoming mixture and the sources that created it. The major parameters are intra aural time differences, spectrum, onset and offset, phase, and amplitude of each source. In this paper I review and compare previously proposed methods to analyze the frequency spectrum. This includes estimation of the fundamental frequency and partials produced by each source in the mixture.

11:05

2aPP8. Exposure smart hearing protector for reducing noise-induced hearing loss. Ed Nykaza and Tom Frank (Dept. of Acoust., Penn State, University Park, PA 16802, etn106@psu.edu)

The Exposure Smart Hearing Protector (ESHP) is a new device that can be used for measuring noise exposure levels (NELs) and the prevention of noise induced hearing loss (NIHL). The ESHP consists of two microphones, located in a right and left earplug, that are connected to a dosimeter. In practice, the user wears the ESHP. When the noise level exceeds a safe dose a warning light comes on. The user then inserts the earplugs. If the earplugs are correctly inserted and the noise level in the user's earcanal is below a safe level the warning lights go off. As a result, the ESHP measures the user's total daily noise exposure (unprotected and protected). To increase the efficiency of using the ESHP for preventing NIHL, the user downloads the information stored in the ESHP via a scanner into user friendly-software. The software can be used not only to record a user's daily NELs, but more importantly to determine if the user needs intervention because the NELs exceed a safe level. The purpose of this poster session is to demonstrate the ESHP and software, and to report the results of a pilot study. [Work supported by NIOSH/CDC Grant No. U60/CCU 315855.]

11:20

2aPP9. Subjective assessment for the number of channel signals to realize sound field based on wavefield synthesis. Toshiyuki Kimura (Res. Fellow of the Japan Society for the Promotion of Sci./CIAIR, Nagoya Univ., Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan), Kazuhiko Kakehi, Kazuya Takeda, and Fumitada Itakura (Nagoya Univ., Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan)

It is very important to study the affect of the number of channel signals in the sound field reproduction technique based on wavefield synthesis. Though there are many studies of objective affect such as wavefront accuracy, there are not as many studies of the subjective affect such as sound field perception. Therefore, the subjective assessment was designed to evaluate the effect of the number of channel signals, which were synthesized by convolving a sound source to room transfer functions of free field, on the directional perception. In the low-reverberation room, a circle loudspeaker array with a radius of 2 m was set on the horizontal plane. The subjects evaluated the directional perception of the sound image reproduced at distances of 3 and 4 m by playing simultaneous channel signals from the loudspeakers set at intervals of 10, 15, 20, 30, and 45 degrees, respectively. As a result, it was subjectively confirmed that 24 channel signals were enough to reproduce the directional perception of 5 degrees accuracy if the control area was the circle of radius 2 m.

11:35

2aPP10. Coupled volume/double slope subjective listening test. Michael Ermann and Rebecca Stuecker (Virginia Tech Dept. of Architecture, 201 Cowgill Hall, Blacksburg, VA 24061)

Can experienced listeners of music discern a double-sloped decay from a Sabine decay? Do they prefer the double slope? Concert hall designers use coupled-volumes and their signature double-slope sound decay in an effort to reconcile the inversely related qualities of reverberance and clarity. A simulated space, based on an actual built coupled-volume hall, was conceived in the room acoustics software CATT-Acoustic. Variations in the aperture sizes that sonically expose the main hall to the coupled volume generated both classic Sabine decays and double-sloped decays. The impulse responses generated were convolved with the same anechoic musical recording, grouped in pairs, and played for an opportunity-sample of 21 volunteers from the Architectural Acoustics section of the 145th meeting of the Acoustical Society of America in Nashville. Participants listened to the 11 recorded pairs over headphones and were asked to determine (1) if the two recordings sounded different, (2) which recording was more likely to have a double slope or had a more dramatic double slope, and (3) which of the two recordings they prefer.

11:50

2aPP11. Probabilistic neurotransmitter release for dynamic synapse neural networks and its application on pure tone recognition. Alireza A. Dibazar, Hassan H. Namarvar, Sageev George, and Theodore W. Berger (BME Dept., Univ. of Southern California, OHE 500, Los Angeles, CA 90089-1451)

In this paper, we propose a probabilistic neurotransmitter release for dynamic synapses neural network (DSNNs). The capabilities of DSNNs have already been investigated in the processing of spatio-temporal patterns of action potentials [J.-S. Liaw and T. W. Berger (1996)]. The deterministic model of synapse is substituted by probabilistic Markov model. The action potentials generated by auditory system are the inputs of the model. The probability of neurotransmitter release is then estimated from

the model. In general, the aim of this study is to present a robust pure tone recognition system based on DSNN. To generate action potential from music tone we have employed pulse code modulation method. The action potentials are plugged into the DSNN for recognition purpose. Our simulation results showed that the DSNN based on probabilistic release has 4% better recognition performance with respect to the deterministic model [A. A. Dibazar *et al.*, SFN 2002]. [Work supported in part by DARPA-CVS, NASA, and ONR.]

TUESDAY MORNING, 11 NOVEMBER 2003 NUECES ROOM, 9:00 A.M. TO 12:00 NOON

Session 2aSA

Structural Acoustics and Vibration: Sound/Structure Interaction

Jeffrey S. Vipperman, Chair
Department of Mechanical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261

Contributed Papers

9:00

2aSA1. Computation of scattering by high-aspect ratio spheroids using a new wide-angle on-surface radiation condition approach. David C. Calvo (Naval Res. Lab., Code 7145, 4555 Overlook Ave. SW, Washington, DC 20375)

In a previous study [Calvo *et al.*, IEEE Trans. Antennas Propag. (2003)], a wide-angle on-surface radiation condition (OSRC) was introduced to compute scattering by hard and soft elliptical cylinders. This OSRC was based on a rational approximation of the exact Dirichlet-to-Neumann map for a circular cylinder. As a result of the wide-angle accuracy, exceptional results were obtained for scattering by hard elliptical cylinders when a plane wave was incident end-on with the target. In this talk, the form of the rational approximation is economized and the OSRC is extended to the case of three-dimensional convex bodies (spheroids). A comparison of scattered fields using the OSRC and the BEM methods shows excellent agreement. Owing to the tridiagonal form of the differentiation matrices appearing in the rational approximation, computational costs scale only linearly with frequency making this technique ideal for mid-to-high frequency target scattering computations for orientations where diffraction effects may be important.

9:15

2aSA2. Acoustic diffraction by a disk in a viscous medium. Anthony M. J. Davis (Univ. of Alabama, Tuscaloosa, AL 35487) and Raymond J. Nagem (Boston Univ., Boston, MA 02115)

The diffraction of a time-harmonic acoustic plane wave by a circular disk in a viscous fluid medium is considered by using the linearized equations of viscous fluid flow and the no-slip condition on the rigid disk. Sets of dual integral equations for the fluid velocity and pressure are derived for an arbitrary disk radius and an arbitrary angle of incidence. The dual integral equations are solved by an analytic reduction to sets of linear algebraic equations. In the cases of normal or tangential incidence, numerical results are presented for the fluid velocity in the plane of the disk and the scattered acoustic disturbance in the far field.

9:30

2aSA3. A 3-D hp-adaptive finite-element code for modeling acoustic scattering from elastic structures. David S. Burnett and Mario Zampolli (NATO Undersea Res. Ctr., Viale San Bartolomeo 400, 19138 La Spezia, Italy, burnett@saclantc.nato.int)

The Centre has developed a 3-D hp-adaptive finite-element structural acoustics code, FESTA, for modeling acoustic scattering from underwater elastic structures. The project is using a unique commercial software system for developing customized application codes. To develop a code, the researcher derives the Galerkin residual equations and then links a self-written code with a vendor-written application-independent library of routines, which creates an application code with a commercial-quality GUI. For FESTA the entire fluid-structure domain is treated as a single continuum, which is modeled using only one type of finite element: a fluid-solid element. The wave equations for both media are derived from the same underlying equations of continuum mechanics, and then combined into a single wave equation, from which the fluid–solid finite element is derived. The code employs 3-D continuum mechanics throughout the computational domain. Thin structural components, such as plates and shells, are modeled with 3-D physics rather than plate or shell theories. The code is therefore fully 3-D, i.e., in both physics and geometry. This paper will describe the theoretical development and present several scattering and propagation models analyzed by FESTA.

9:45

2aSA4. Computing fluid-coupled resonance frequencies, mode-shapes, and damping loss factors using the singular value decomposition. John B. Fahnline (Appl. Res. Lab, Penn State Univ., Appl. Sci. Bldg., University Park, PA 16801, jbf@wt.arl.psu.edu)

In many acoustic design problems, it would be useful to be able to compute fluid-coupled resonance frequencies, mode shapes, and their associated damping levels. Unfortunately, conventional eigenvalue solution procedures are either computationally-inefficient, unreliable, or have limited applicability. Sophisticated methods for identifying modal parameters using the singular value decomposition have recently emerged in the area of experimental modal analysis, where the available data typically consists of velocity to force transfer function data as a function of frequency for several drive point locations. Here, these techniques are shown to be even more effective for coupled finite element/boundary element solutions because full matrices of transfer function data can be computed as a function of frequency. This allows the modes to be completely separated from each other, such that the modal parameters can be identified using simple methods designed for single degree of freedom systems. Several benchmark example problems are solved numerically including a baffled circular plate, an unbaffled rectangular plate, and a spring-mounted piston coupled to fluid within a rigid-walled pipe.

10:00

2aSA5. System identification and energy-based active noise control performance. Benjamin M. Faber and Scott D. Sommerfeldt (Brigham Young Univ., N283 ESC, Provo, UT 84602, bmf3@email.byu.edu)

An active noise control system, which minimizes acoustic energy density, can often become unstable due to significant changes in the acoustical environment in which it operates. Typically, the control system of interest, which is based on a modified version of the filtered-x LMS algorithm, is operated with a fixed set of system identification filter coefficients. These coefficients are determined off-line, prior to execution of the main control

algorithm. The performance of the control system depends on the quality of the system model obtained through the system identification algorithm. Changes in the system transfer function and impulse response, due to various factors, including temperature, boundary conditions, and sensor/actuator location, have been investigated. Results will be presented, which give indications regarding what types of changes to the system have the greatest effect on the transfer function and impulse response, and how sensitive the control system is to those effects. The effects of changing sampling rates and system identification filter lengths on the quality of the system model have also been investigated and will be discussed.

10:15–10:30 Break

10:30

2aSA6. Acoustic control in a tractor cabin using two optimally designed Helmholtz resonators. Patricia L. Driesch (United Technologies Res. Ctr., E. Hartford, CT 06108) and Gary H. Koopmann (Penn State Univ., University Park, PA 16801)

A virtual design methodology is developed to minimize the noise in enclosures with optimally designed, passive, 20 acoustic absorbers (Helmholtz resonators). A series expansion of eigenfunctions is used to represent the acoustic=20 absorbers as external volume velocities, eliminating the need for a solution of large matrix eigenvalue problems. A determination of this type (efficient model/reevaluation approach) significantly increases the design possibilities when optimization techniques are implemented. As a full-scale demonstration, the acoustic response from 90–190 Hz of a tractor cabin was investigated. The lowest cabin mode proposes a significant challenge to a noise control engineer since its anti-node is located near the head of the operator and often generates unacceptable sound-pressure levels. Exploiting the low-frequency capability of Helmholtz resonators, lumped parameter models of these resonators were coupled to the enclosure via an experimentally determined acoustic model of the tractor cabin. The virtual design methodology uses gradient optimization techniques as a post-processor for the modeling and analysis of the unmodified acoustic interior to determine optimal resonator characteristics. Using two optimally designed Helmholtz resonators, potential energy was experimentally reduced by 3.4 and 10.3 dB at 117 and 167 Hz, respectively.

10:45

2aSA7. Micromachined variable impedance acoustic waveguides for fluid-structure waves. Robert D. White and Karl Grosh (Dept. of Mech. Eng., Univ. of Michigan, Ann Arbor, MI 48109, grosh@umich.edu)

Micromachined fluid filled variable impedance waveguides intended to mimic the mechanics of the passive cochlea have been fabricated and experimentally examined. These devices consist of a fluid filled chamber of dimension 37 mm by 6.25 mm by 0.1 mm fabricated on a pyrex wafer. One side of this chamber is constrained by a 3-cm-long, exponentially tapered, 0.3-μm thick tensioned membrane made of low-pressure chemical vapor deposited (LPCVD) silicon nitride. The membranes vary in width from 170 μm to 1.9 mm, and are fabricated using deep reactive ion etching (DRIE). The devices are filled with silicone oil or de-ionized water. Experimental tests demonstrate acoustically excited traveling fluid-structure waves with wave numbers well below the free-plane wave wave number, and with phase accumulations as high as 25 rad (at 6.5 kHz) in the 3 cm length of the device. Comparison of experimental measurements with both numerical (finite-element) and asymptotic (Liouville–Green) models assist in understanding the results. A thin-layer finite-element approximation is developed to take advantage of the device aspect ratio and increase computational efficiency. Viscous fluid effects are included in the models to accurately capture damping phenomena. [Work supported by ONR and NSF.]

11:00

2aSA8. Acoustic signature of an AUV. Joe Cuschieri and Susan Frandsen (Ctr. for Acoust. and Vib., Dept. of Ocean Eng., Florida Atlantic Univ., 101 N. Beach Rd., FL 33004)

In this presentation further results from an experimental analysis of the acoustic signature of an Ocean Explorer class AUV are presented. The results to be presented include, data collected in an open water environment with the AUV operating at cruising speed, measurements performed in a reverberant test tank on an AUV model, and measurements in the reverberant test tank of an AUV under typical operating conditions. Different operating conditions and different mountings for the main propulsion and control module (podule) inside the AUV are considered. Inside the podule are the propulsion motor and the motors for the control surfaces with penetrations for the main propulsion shaft and the shafts for the control surfaces. Considered in the measurements is the influence of the propeller, the mounting of the podule, and the covering the podule with a compliant layer. It is shown that for the type of AUV considered here, the type of mounting of the podule is not very significant and that significant energy is transferred through the water trapped in between the podule and the AUV hull. Furthermore, the propeller has a significant influence on the acoustic signature.

11:15

2aSA9. Correlation of acoustic and vibration signatures of faulty bearings. Nagarjuna Jillella, Corrinne Darvennes, and Sally Pardue (Mech. Eng., Brown Hall, Tennessee Technolog. Univ., Cookeville, TN 38505)

The use of acoustic based measurements for machinery health monitoring of rolling element bearings is explored. Acoustic data was taken with several types of microphones placed very close to the bearings, to minimize the effect of background noise. Accelerometer measurements were also made on the system. A frequency spectrum analysis of acoustic and vibration data was done. The experimentation was performed on a Machine Diagnostics Trainer unit. The shaft is driven by a motor and supported by two bearings that can be easily replaced to incorporate known defects. Three bearings were used to make measurements: a good bearing, a bearing with an inner race defect, and a bearing with an outer race defect. Frequency spectrums and rms values will be presented showing the effects of the transducer type and location, and of the bearing defects.

11:30

2aSA10. Computational and experimental techniques for structural acoustics. Timothy F. Walsh (Dept. of Computational Solid Mech. and Structural Dynam., Sandia Natl. Labs., P.O. Box 5800, M.S. 0835, Albuquerque, NM 87185), Hartono Sumali, and Jeffrey L. Dohner (Sandia Natl. Labs., Albuquerque, NM 87185)

In this paper a discussion of various computational and experimental methods under development at Sandia for evaluating structural acoustic interactions will be presented. The driving applications range from the microscale to the macroscale. An overview will be given of Salinas, a massively parallel structural dynamics code that has been developed at Sandia, and its embedded structural acoustic capabilities. Salinas can perform coupled structural acoustic simulations of the interior and exterior of large-scale structures as well as Micro-Electro-Mechanical Systems (MEMS). An ongoing research effort in Salinas for including viscosity in the acoustic computations will be discussed. Viscosity is typically ignored in commercial acoustics codes. From the experimental side an acoustic wave tube will be described, which is to be used for validating structural acoustic computations and other acoustic experiments of interest. Also, acoustic experimental facilities for MEMS applications will be described. [Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company for the United States Department of Energy's National Nuclear Security Administration under Contract No. DE-AC04-94AL85000.]

11:45

2aSA11. Computer methods in vibrational ecology problems. Samuil A. Rybak (N. N. Andreev Acoust. Inst., Moscow 117036, Russia), Sergey A. Makhortykh (Inst. of Mathematical Problems of Biol. RAS, Pushchino Moscow reg. 142290, Russia), Stanislav A. Kostarev (Lab. of Acoust. Vib. Tunnel Assoc., Moscow 107217, Russia), and Aleksandra R. Gatina (Moscow State Univ., Vorobievy Gory, Moscow, Russia)

In the paper formulations of direct vibrational ecology problems are described. Linearly spaced source of acoustic waves in the solid media is considered (e.g., tunnel in soil, simulated by the thin elastic shell). In this case the shell is exited by the inner force changing in time and spatial variables. The proposed computer prognosis method is realized in VibLab software of the Russian Tunneling Association for two versions of algorithms: in homogeneous and stratified two-dimensional media. Three general stratification types are considered: two waveguides (subsurface and waveguide with the axis on the depth H) and stratification with linearly increased wave speed. On the basis of these approximations the amendments, respectively, liquid soil case were calculated. A vibrational field in the ground is described by cylinder waves; modulation of this field type is assessed by a simple final elements scheme. The aspects of application of the proposed algorithms in inverse problems of media parameters determined are also discussed in the report. [Work supported by the Russian Foundation for Basic Research Grants No. 01-02-16127 and No. 02-02-17143.]

TUESDAY MORNING, 11 NOVEMBER 2003

RIO GRANDE EXHIBIT HALL, 9:00 A.M. TO 12:00 NOON

Session 2aSC

Speech Communication: Speech Perception in Normal and Impaired Hearing (Poster Session)

Judith L. Lauter, Chair
Department of Human Services, Stephen F. Austin University, Nacogdoches, Texas 75962

Contributed Papers

All posters will be on display from 9:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 9:00 a.m. to 10:30 a.m. and contributors of even-numbered papers will be at their posters from 10:30 a.m. to 12:00 noon.

2aSC1. "The perceptual bases of speaker identity" revisited. William D. Voiers (Dynastat, Inc., 2704 Rio Grande, Austin, TX 78705, bvoiers@aol.com)

A series of experiments begun 40 years ago [W. D. Voiers, J. Acoust. Soc. Am. **36**, 1065–1073 (1964)] was concerned with identifying the perceived voice traits (PVTs) on which human recognition of voices depends. It culminated with the development of a voice taxonomy based on 20 PVTs and a set of highly reliable rating scales for classifying voices with respect to those PVTs. The development of a perceptual voice taxonomy was motivated by the need for a practical method of evaluating speaker recognizability in voice communication systems. The Diagnostic Speaker Recognition Test (DSRT) evaluates the effects of systems on speaker recognizability as reflected in changes in the inter-listener reliability of voice ratings on the 20 PVTs. The DSRT thus provides a qualitative, as well as quantitative, evaluation of the effects of a system on speaker recognizability. A fringe benefit of this project is PVT rating data for a sample of 680 voices. [Work partially supported by USAFRL.]

2aSC2. Children's discrimination of vowel sequences. Jeffry A. Coady, Keith R. Kluender, and Julia Evans (Dept. of Psych., Univ. of Wisconsin–Madison, 1202 W. Johnson St., Madison, WI 53706)

Children's ability to discriminate sequences of steady-state vowels was investigated. Vowels (as in "beet," "bat," "bought," and "boot") were synthesized at durations of 40, 80, 160, 320, 640, and 1280 ms. Four different vowel sequences were created by concatenating different orders of vowels for each duration, separated by 10-ms intervening silence. Thus, sequences differed in vowel order and duration (rate). Sequences were 12 s in duration, with amplitude ramped linearly over the first and last 2 s. Sequence pairs included both same (identical sequences) and different trials (sequences with vowels in different orders). Sequences with vowel of equal duration were presented on individual trials. Children aged 7;0 to 10;6 listened to pairs of sequences (with 100 ms between sequences) and responded whether sequences sounded the same or different. Results indicate that children are best able to discriminate sequences of intermediate-duration vowels, typical of conversational speaking rate. Children were less accurate with both shorter and longer vowels. Results are discussed in terms of auditory processing (shortest vowels) and memory (longest vowels). [Research supported by NIDCD DC-05263, DC-04072, and DC-005650.]

2aSC3. A comparison of speech recognition training programs for cochlear implant users: A simulation study. Marie E. McCabe (Dept. of Psych., Dyer Hall, ML 0376, Univ. of Cincinnati, Cincinnati, OH 45229-0376, mccabeme@email.uc.edu) and C.-Y. Peter Chiu (Univ. of Cincinnati, Cincinnati, OH 45229-0376)

The present simulaton study compared two training programs with very different design features to explore how each might improve the ability of listeners with normal hearing to recognize speech generated by a cochlear implant simulator. The first program, which focused training on specific areas of difficulty for individual patients across multiple levels of linguistic content (e.g., vowels, consonants, words, and sentences), was modeled after a standard program prescribed by one of the US manufacturers of cochlear implants. The second program consisted of exposure to multiple sentences with feedback regardless of subjects' performance level, and had been used in previous studies from this laboratory. All speech materials were reduced spectrally to simulate an 8-channel CIS cochlear implant processor with a "6mm frequency upshift" [Fu and Shannon, J. Acoust. Soc. Am. **105**, 1889 (1999)]. Test sessions were administered to all subjects to assess recognition of sentences, consonants (/aCa/), and vowels (in /hVd/ and /bVt/ contexts) pre- and post-training. In

a subset of subjects, a crossover design, in which subjects were trained first with one program and then with the other, was employed. Results will be discussed both in terms of theory and practice of therapeutic programs for cochlear implant users.

2aSC4. Auditory color constancy. Keith R. Kluender (Dept. of Psych., Univ. of Wisconsin, Madison, WI 53706) and Michael Kiefte (School of Human Commun. Disord., Dalhousie, Halifax, NS, Canada)

It is both true and efficient that sensorineural systems respond to change and little else. Perceptual systems do not record absolute level be it loudness, pitch, brightness, or color. This fact has been demonstrated in every sensory domain. For example, the visual system is remarkable at maintaining color constancy over widely varying illumination such as sunlight and varieties of artificial light (incandescent, fluorescent, etc.) for which spectra reflected from objects differ dramatically. Results will be reported for a series of experiments demonstrating how auditory systems similarly compensate for reliable characteristics of spectral shape in acoustic signals. Specifically, listeners' perception of vowel sounds, characterized by both local (e.g., formants) and broad (e.g., tilt) spectral composition, changes radically depending upon reliable spectral composition of precursor signals. These experiments have been conducted using a variety of precursor signals consisting of meaningful and time-reversed vocoded sentences, as well as novel nonspeech precursors consisting of multiple filter poles modulating sinusoidally across a source spectrum with specific local and broad spectral characteristics. Constancy across widely varying spectral compositions shares much in common with visual color constancy. However, auditory spectral constancy appears to be more effective than visual constancy in compensating for local spectral fluctuations. [Work supported by NIDCD DC-04072.]

2aSC5. Processing of speech and non-speech stimuli in children with specific language impairment. Madhavi L. Basu and Aimee M. Surprenant (Purdue Univ., West Lafayette, IN 47907)

Specific Language Impairment (SLI) is a developmental language disorder in which children demonstrate varying degrees of difficulties in acquiring a spoken language. One possible underlying cause is that children with SLI have deficits in processing sounds that are of short duration or when they are presented rapidly. Studies so far have compared their performance on speech and nonspeech sounds of unequal complexity. Hence, it is still unclear whether the deficit is specific to the perception of speech sounds or whether it more generally affects the auditory function. The current study aims to answer this question by comparing the performance of children with SLI on speech and nonspeech sounds synthesized from sine-wave stimuli. The children will be tested using the classic categorical perception paradigm that includes both the identification and discrimination of stimuli along a continuum. If there is a deficit in the performance on both speech and nonspeech tasks, it will show that these children have a deficit in processing complex sounds. Poor performance on only the speech sounds will indicate that the deficit is more related to language. The findings will offer insights into the exact nature of the speech perception deficits in children with SLI. [Work supported by ASHF.]

2aSC6. A novel probabilistic framework for event-based speech recognition. Amit Juneja and Carol Espy-Wilson (Dept. of Elec. and Computer Eng., Univ. of Maryland, AVW Bldg. ECE, College Park, MD 20742, juneja@glue.umd.edu)

One of the reasons for unsatisfactory performance of the state-of-the-art automatic speech recognition (ASR) systems is the inferior acoustic modeling of low-level acoustic-phonetic information in the speech signal. An acoustic-phonetic approach to ASR, on the other hand, explicitly targets linguistic information in the speech signal, but such a system for continuous speech recognition (CSR) is not known to exist. A probabilistic and statistical framework for CSR based on the idea of the representation of speech sounds by bundles of binary valued articulatory phonetic features is proposed. Multiple probabilistic sequences of linguistically motivated landmarks are obtained using binary classifiers of manner phonetic features—*syllabic*, *sonorant* and *continuant*—and the knowledge-based acoustic parameters (APs) that are acoustic correlates of those features. The landmarks are then used for the extraction of knowledge-based APs for source and place phonetic features and their binary classification. Probabilistic landmark sequences are constrained using manner class language models for isolated or connected word recognition. The proposed method could overcome the disadvantages encountered by the early acoustic-phonetic knowledge-based systems that led the ASR community to switch to systems highly dependent on statistical pattern analysis methods and probabilistic language or grammar models.

2aSC7. Effects of lag-time on dichotic word perception in children. Sheryl S. Shoemaker, Edward L. Goshorn, and Shannon Jette (Speech Dept., Louisiana Tech Univ., P.O. Box 3165 Tech Station, Ruston, LA 71272, egoshorn@ltparts.latech.edu)

Dichotic word perception is an essential component in audiological test batteries that are designed to identify auditory processing disorders (APD). Lag-time is the time difference between onsets of dichotic stimuli. The contribution of lag-time to perception of dichotic speech is not fully known and there is a lack of data available on children. Studies by Black (1955) and Berlin *et al.* (1973) found intermittent effects of lag-time on perception of dichotic speech in adults. This project examined the effects of 3 lag-times (0, 150, and 300 ms) on perception of NU-CHCIPS [Elliot and Katz (1980)] words presented dichotically at 50 dB HL through ER-3A insert phones. The sequence of lag-times and ear order was balanced. Subjects were six children age 7 to 10 years who had normal hearing and normal findings on standardized APD tests. Lag-times of 0, 150, and 300 ms yielded mean error rates of 1.3, 1.5, and 1.7 ms, respectively. Mean right ear and left ear error rates were 1.4 and 1.6, respectively. A two-way ANOVA showed no significant ($p>0.05$) effect for lag-time, ear preference, or their interaction. Implications for findings and development of APD treatment materials based on lag-time data will be presented.

2aSC8. Representational specificity of within-category phonetic variation in the mental lexicon. Min Ju and Paul A. Luce (Dept. of Psych., Univ. at Buffalo, Amherst, NY 14260, mju@buffalo.edu)

This study examines (1) whether within-category phonetic variation in voice onset time (VOT) is encoded in long-term memory and has consequences for subsequent word recognition and, if so, (2) whether such effects are greater in words with voiced counterparts (pat/bat) than those without (cow/*gow), given that VOT information is more critical for lexical discrimination in the former. Two long-term repetition priming experiments were conducted using words containing word-initial voiceless stops varying in VOT. Reaction times to a lexical decision were compared between the same and different VOT conditions in words with or without voiced counterparts. If veridical representations of each episode are preserved in memory, variation in VOT should have demonstrable effects on the magnitude of priming. However, if within-category variation is discarded and form-based representations are abstract, the variation in VOT should not mediate priming. The implications of these results for the specificity and abstractness of phonetic representations in long-term memory will be discussed.

2aSC9. Grammatical judgments and phonetic reality: A study of internal constraints in phonological variation. Richard File and Manuel Diaz-Campos (Indiana Univ., Ballantine Hall 844, 1020 E. Kirkwood Ave., Bloomington, IN 47405)

Previous investigations analyzing phonological variation in Spanish have pointed out that phonetic context is an important factor for predicting different variants of /s/. In many studies (Cedergren, 1973; Poplack, 1979, 1980, 1986; Ranson, 1993; Samper Padilla, 1990; Terrell, 1977, 1978,

1979, 1986, etc.), it is claimed that /s/ aspiration is more likely to happen word-internally, preceding a consonant, while elision is more likely to happen word-finally preceding a pause. The methodology used in those studies relies on grammatical judgment for performing the analysis of phonetic context, which could present a perceptual bias that could misrepresent the reality of the realization of the segments. The present investigation examines the phonetic cues that might have an effect in the perception of syllable-final /s/ including the duration of the preceding vowel as well as the following context. Preliminary results show that when the preceding vowel had a shorter duration, retention was easily perceived, whereas when the preceding vowel was longer in duration, the perception of aspiration was favored. The results suggest that for a segment with the same degree of aspiration, deletion was perceived before a pause more often than before a consonant, especially when the preceding vowel was shorter in duration.

2aSC10. Do listeners perceive coarticulatory differences for normal talkers and neurologically impaired talkers? Kris Tjaden, Joan Sussman, and Ya-ju Yu (Dept. of Communicative Disord. & Sci., Univ. at Buffalo, 122 Cary Hall, 3435 Main St., Buffalo, NY 14214-3005, tjaden@acsu.buffalo.edu.)

Speech production studies suggest that the extent of coarticulation varies across talkers. The extent of coarticulation may even vary for repetitions of the same utterance produced by a single talker. Thus, an utterance or talker could be characterized by an average degree of coarticulation, a high degree of coarticulation, or a low degree of coarticulation. The perceptual consequences, if any, of this coarticulatory variability are not well understood, particularly for individuals with speech motor control disorders. The current study examined whether the listeners' speed and accuracy of vowel identification for naturally-produced CV sequences varied depending on the extent of anticipatory vowel coarticulation. Speakers and speech tokens included those characterized by an average degree of anticipatory coarticulation, a high degree of coarticulation, and a low degree of coarticulation. Healthy talkers, speakers with multiple sclerosis, and speakers with Parkinson's disease produced the stimuli. Consonants in CV syllables beginning with /s/, /k/, or /t/ followed by the vowels /i/ or /u/ were excised and presented to listeners for identification of the following vowel. While identification accuracy may be high across speakers and stimuli, the speed of response should vary with extent of coarticulation, if listeners are sensitive to coarticulatory cues in the acoustic speech stream.

2aSC11. Effects of frequency shifts and visual gender information on vowel category judgments. Catherine Glidden and Peter F. Assmann (School of Behavioral and Brain Sci., Univ. of Texas–Dallas, Box 830688, Richardson, TX 75083)

Visual morphing techniques were used together with a high-quality vocoder to study the audiovisual contribution of talker gender to the identification of frequency-shifted vowels. A nine-step continuum ranging from ''bit'' to ''bet'' was constructed from natural recorded syllables spoken by an adult female talker. Upward and downward frequency shifts in spectral envelope (scale factors of 0.85 and 1.0) were applied in combination with shifts in fundamental frequency, $F0$ (scale factors of 0.5 and 1.0). Downward frequency shifts generally resulted in malelike voices whereas upward shifts were perceived as femalelike. Two separate nine-step visual continua from ''bit'' to ''bet'' were also constructed, one from a male face and the other a female face, each producing the end-point words. Each step along the two visual continua was paired with the corresponding step on the acoustic continuum, creating natural audiovisual utterances. Category boundary shifts were found for both acoustic cues ($F0$ and formant frequency shifts) and visual cues (visual gender). The visual gender effect was larger when acoustic and visual information were matched appropriately. These results suggest that visual information provided by the speech signal plays an important supplemental role in talker normalization.

2aSC12. Part Ia: Spatial separation on McGurk effect applying three-dimensional sounds. Klaus A. J. Riederer (Lab. of Acoust. and Signal Processing, Helsinki Univ. of Technol., Otakaari 5 A, P.O. Box 3000, FIN-02015 TKK, Finland, klaus.riederer@hut.fi)

The dependence of *sound direction* on the *McGurk effect* [McGurk and McDonald, Nature (London) **264**, 746–748 (1976)] is less known. Jones and Munhall [Canadian Acoust. **25**, 13–19 (1997)] concluded with no spatial separation dependence, applying 30° horizontally spaced loudspeakers. Current *dual* study investigated the full 360° horizontal space applying *head-related transfer functions* (HRTFs) from a Cortex dummy head [Riederer, J. Audio Eng. Soc. (Abstracts) **46**, 1036 (1998), preprint 4846]. Dry acoustic /ipi/ and /iti/ recorded from a professional speaker were convolved with HRTFs, measured at azimuths 0°, ±40°, ±90°, ±130°, and 180°, headphones (Sennheiser HD580) equalized. DVcam-recorded visual /ipi/, /iti/ (and black screen) were randomly presented synchronously with the 3-D sounds using Presentation 0.20 [http://nbs.neuro-bs.com]. Totally 1024 incongruent audiovisual stimuli were perceived by eight 20–30-year-old normal hearing ($\leqslant$20 dBHL) native subjects (2 female) as follows. Visual /ipi/ + auditory /iti/: /ipi/ 59.96%, /iti/ 15.63%, and /ipti/ 24.02%; visual /iti/ + auditory /ipi/: 66.02%, 22.07%, and 11.52%, respectfully. No significant dependence of spatial separation was found for the McGurk effect, except for *reaction times*. The obtained fusions were atypically weak, probably because visual /iti/ was less pronounced than visual /ipi/. [Work supported by Graduate School of Electronics, Telecommunication and Automation.]

2aSC13. Shifts in the perceived voicing boundary of bilingual listeners. Adrian Garcia-Sierra, Craig A Champlin (Univ. of Texas–Austin, 1 University Station A1100, Austin, TX 78712, gasa@austin.utexas.edu), and Maritza Rivera-Gaxiola (Univ. of Washington, Seattle, WA 98105)

Three groups of listeners, American-English monolinguals, Mexican-Spanish monolinguals, and Mexican-Spanish bilinguals, identified synthetic /ga/–/ka/ syllables varying in VOT. In order to induce a language set in the bilingual listeners, they were asked to produce syllables and phrases in either English or Spanish prior to the identification task. Bilinguals identification of the VOT stimuli varied depending on the language set condition, with the voicing boundary in each case shifting in the direction of values characteristic of monolingual listeners. This effect of language set replicates earlier findings [L. Elman, R. Diehl, and S. E. Buchwald, J. Acoust. Soc. Am. **62**, 971–974 (1977)]. Bilinguals velar consonant productions showed a corresponding shift in VOT values across the two language conditions. A Mismatch Negativity paradigm will be used in an attempt to record preattentive discrimination of the language-set-induced phonemic boundary shift in bilingual listeners. [Work supported by the Department of Communication Sciences and Disorders Univ. of Texas–Austin.]

2aSC14. Shifting perceptions of age in voice. Rahul Shrivastav, Harry Hollien, W. S. Brown, Jr., Howard B. Rothman (Inst. for Adv. Study of the Commun. Proc., Univ. of Florida, Gainesville, FL 32611), and James D. Harnsberger (Univ. of Florida, Gainesville, FL 32611)

A series of experiments have been carried out in order to identify the acoustical and perceptual correlates of the aging voice. The initial phase of the program was to identify those voice parameters which signal a person's age; the second phase was to systematically shift these parameters in

order to determine if a parallel change in perceived age would occur. This second study focused on temporal characteristics related to voice. In this instance, standard speech samples for 16 males aged 70–90 years were contrasted with those of 14 males aged 20–33 years. The features studied included the following: (1) sentence duration, (2) word duration, (3) diphthong duration, (4) consonant-vowel ratios, (5) number of pauses and (6) pause duration. Significant differences were found for all relationships. Subsequently, a preliminary study was carried out where the voices were synthesized and the temporal parameters for the two groups shifted toward each other. The preliminary data suggest that such modifications lead especially to the idea that the voices of older individuals actually were those of younger men.

2aSC15. Structural equation modeling for estimating the identification accuracy and detection time latency of English monosyllabic words. Sumiko Takayanagi, Lynne E. Bernstein, and Edward T. Auer, Jr. (Dept. of Commun. Neurosci., House Ear Inst., Los Angeles, CA 90057, stakayanagi@mailhouse.hei.org)

Structural equation modeling (SEM) was used to examine the statistical structure among sets of experiential (word age of acquisition and subjective familiarity) and lexical similarity (lexical equivalence class size and neighborhood density) variables for word identification and reaction time latency tasks. Stimuli were 240 vocoded monosyllabic English words with reduced intelligibility and altered similarity relationships. Participants detected a target word following a prime and on every trial reported the prime. The identification accuracy was estimated by words and phonemes correct, and detection latency was estimated by trimmed and harmonic mean RTs. A parsimonious SEM was chosen in terms of the chi-square and model fit indices that determine whether the models adequately described the particular associations of variables/interfactor relationships. The variable/factor error variances were constrained to be uncorrelated with each other in order to evaluate effects independently. A bootstrapping technique indicated that the regression weights of the top-down and bottom-up factors were small, but they were significant in the model. The variance accounted for (VAF) by the model was 7.1% for identification accuracy, and 5.2% for RT latency. The model also indicated that RT latency was highly influenced by prime identification accuracy (15% VAF). [Work supported by NIH/NIDCD00695.]

2aSC16. The effects of auditory and visual vowel training on speech reading performance. Carolyn Richie and Diane Kewley-Port (Dept. of Speech and Hearing Sci., Indiana Univ., IN)

Speech reading, the use of visual cues to understand speech, may provide a substantial benefit for normal-hearing listeners in noisy environments and for hearing-impaired listeners in everyday communication. However, there exists great individual variability in speech reading ability, and studies have shown that only a modest improvement in speech reading ability is achieved with training. The purpose of this investigation was to determine the effects of a novel approach to speech reading training on word and sentence identification tasks. In contrast to previous research, which involved training on consonant recognition, this study focused on vowels. Two groups of normal-hearing adults participated in auditory–visual (AV) conditions with added background noise. The first group of listeners received training on the recognition of 14 English vowels in isolated words, while the second group of listeners received no training. All listeners performed speech reading pre- and post-tests, on words and sentences. Results are discussed in terms of differences between groups, dependent upon whether training was administered, and a comparison is made between this and other speech reading training methods. Finally, the potential benefit of this vowel-based speech reading training method for the rehabilitation for hearing-impaired listeners is discussed. [Work supported by NIHDCD-02229.]

2aSC17. Two-C but not two-V: Segment similarity in learning an artificial lexicon. Sarah C. Creel, Richard N. Aslin, and Michael K. Tanenhaus (Dept. of Brain and Cognit. Sci., Univ. of Rochester, Rochester, NY 14627, screel@bcs.rochester.edu)

The role of segment similarity (C1_C2_or _V1_V2) in a word learning task was assessed using an artificial lexicon in a referential context. Learning consisted of 480 trials in which S's heard one of 40 CVCV nonsense strings, accompanied by an unfamiliar picture. In testing, participants heard the direction "Click on the [nonsense word]," and chose one of four pictures that matched the test item. On some trials, target lexical items (pibo) appeared with foils that contained matched consonants (pabu) or matched vowels (diko). There were higher rates of confusion errors to the matched-consonant items than to non-matched items, but no significant elevation in errors to matched-vowel items. A second experiment examined the role of differences in informativeness between C's and V's by inverting the numbers of C and V types (first experiment: 10 C, 5 V; second experiment: 5 C, 10 V). This made the consonants less predictive of word identity (more words contained the same consonants), and made the vowels more predictive of word identity. The matched-consonant effect remained undiminished while no corresponding matched-vowel effect emerged, ruling out a segment-informativeness explanation. Other accounts based on the syllable positions and confusability patterns of consonants are being explored.

2aSC18. An investigation of perceptual tolerance limits of stop constriction regions along the vocal tract. Kang Li and Brad Story (Dept. of Speech and Hearing Sci., Univ. of Arizona, P.O. Box 210071, 1131 E. 2nd St., Tucson, AZ 85721-0071, lik@u.arizona.edu)

In this study we explore the perceptual categories of constriction regions along the vocal tract for voiced stop consonants /b/, /d/, and /g/. The three consonants were imbedded in consonant-vowel (CV) syllables. Using an area function model, a series of stop-vowel syllables that differed only in the place of constriction at even spaces was constructed. A second series was generated by eliminating the third formant of each syllable in the first series. It was predicted that the continuum of place of constriction would result in the discontinuity of acoustic representation and the categorical perception. The predicted perception of each synthesized syllable was based on their spectrographic shapes. The results showed that the perception of the constriction continuum was categorical in both series. The boundaries of actually perceived regions agree reasonably well with the predicted ones. It turned out that the perception of the midway stimuli between two categories was ambiguous instead of being perceived as a member of one category or the other. The Distinctive Region Model based on the sensitivity function is discussed to explore a possible explanation for the ambiguity.

2aSC19. Otoacoustic emissions measured in children diagnosed with attention-deficit/hyperactivity disorder. Dennis McFadden, J. Gregory Westhafer, Edward G. Pasanen, David M. Tucker, and Caryn L. Carlson (Dept. of Psych., Univ. of Texas, Austin, TX 78712-0187, mcfadden@psy.utexas.edu)

Attention-deficit/hyperactivity disorder (ADHD) is generally acknowledged to be more prevalent in males than in females. Further, some precursors to ADHD appear early in life. Together these facts suggest that ADHD may be influenced by androgenic mechanisms operating early in development. This reasoning raises the question of whether the otoacoustic emissions (OAEs) of children with ADHD are masculinized. Click-evoked OAEs were measured for one click level in 8 boys and 3 girls diagnosed as ADHD/Combined, in 11 males and 5 females diagnosed as ADHD/Inattentive (IA), and in 17 male and 18 female controls. The ages

of these samples ranged between 7 and 15. As in adults, the CEOAEs of the control males were weaker than those of the control females. Further, the CEOAEs of the ADHD/IA males were weaker than in the control males (a hypermasculinization) and the CEOAEs of the IA females were weaker than in the female controls (a masculinization). The CEOAEs of the Combined groups were slightly stronger (feminized) than those of the control males and females. One interpretation is that the IA subgroup of ADHD boys and girls (but not the Combined subgroup) was exposed to higher-than-normal levels of androgens sometime early in development. [Work supported by NIDCD.]

2aSC20. The time course of indexical specificity effects in the perception of spoken words. Conor T. McLennan and Paul A. Luce (Dept. of Psych. and Ctr. for Cognit. Sci., Univ. at Buffalo, 245 Park Hall, Buffalo, NY 14260, mclennan@buffalo.edu)

This research investigates the time-course of indexical specificity effects in spoken word recognition by examining the circumstances under which the variability in the speaking rate affects the participant's perception of spoken words. Previous research has demonstrated that variability has both representational and processing consequences. The current research examines one of the conditions expected to influence the extent to which indexical variability plays a role in spoken word recognition, namely the time-course of processing. Based on our past work, it was hypothesized that indexical specificity effects associated with the speaking rate would only affect later stages of processing in spoken word recognition. The results confirm this hypothesis: Specificity effects are only in evidence when processing is relatively slow. [Research supported (in part) by Research Grant No. R01 DC 0265801 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.]

2aSC21. AXS and SOM: A new statistical approach for treating within-subject, time-varying, multivariate data collected using the AXS Test Battery. Judith L. Lauter and Chris Ninness (Dept. of Human Services, Stephen F. Austin State Univ., Box 13019 SFA Station, Nacogdoches, TX 75962, jlauter@sfasu.edu)

The Auditory Cross-Section (AXS) Test Battery [J. L. Lauter, Behav. Res. Methods Instrum. Comput. **32**, 180–190 (2000)], described in presentations to ASA in 2002 and 2003, is designed to document dynamic relations linking the cortex, brainstem, and body periphery (whether physics, physiology, or behavior) on an individually-specific basis. Data collections using the battery typically employ a within-subject, time-varying, multivariate design, yet conventional group statistics do not provide satisfactory means of treating such data. We have recently developed an approach based on Kohonens (2001) Self-Organizing Maps (SOM) algorithm, which categorizes time-varying profiles across variables, either within- or between-subjects. The treatment entails three steps: (1) z-score transformation of all raw data; (2) employing the SOM to sort the time-varying profiles into groups; and (3) deriving an estimate of the bounds for the Bayes error rate. Our three-step procedure will be briefly described and illustrated with data from a recent study combining otoacoustic emissions, auditory brainstem responses, and cortical qEEG.

2aSC22. Some effects of intonation contour on sentence intelligibility. James M. Hillenbrand (Speech Pathol. and Audiol., Western Michigan Univ., Kalamazoo, MI 49008)

This experiment was designed to measure the effects of pitch movement on sentence intelligibility. A source-filter synthesizer was used to generate three synthetic versions of 60 sentences drawn from the TIMIT multi-talker speech database: (1) an original pitch (OP) condition in which the fundamental frequency ($F0$) contour matched that of the original utterance, (2) a monotone pitch (MP) condition in which $F0$ was held constant at the median value measured from the original utterance, and (3) an inverted pitch (IP) condition in which the $F0$ contour was reflected around the median $F0$ value (i.e., pitch rises were changed to pitch drops, and vice versa). Results from 30 listeners showed a small but statistically reliable drop in intelligibility from the OP condition to either the MP or IP condition, with no difference between the MP and IP conditions. A second group of 22 listeners was tested on the same task, but with overall sentence intelligibility reduced by running all signals through a 2-kHz low-pass filter. As with the unfiltered signals, intelligibility was reduced for the MP and IP conditions relative to OP; however, the decrements in intelligibility were somewhat larger for the filtered signals, and inverting pitch caused a larger intelligibility decrement than flattening pitch.

2aSC23. Perceptual rate normalization in naturally produced bilabial stops. Kyoko Nagao and Kenneth de Jong (Dept. of Linguist., Indiana Univ.–Bloomington, 406 Memorial Hall, Bloomington, IN 47405, knagao@indiana.edu)

The perception of voicing categories is affected by the speaking rate, so that listeners' category boundaries on a VOT continuum shift to a lower value when the syllable duration decreases (Miller and Volaitis, 1989; Volaitis and Miller, 1992). Previous rate normalization effects have been found using computer-generated stimuli. This study examines the effect of speech rate on voicing categorization in naturally produced speech. Four native speakers of American English repeated syllables (/bi/ and /pi/) at increasing rates in time with a metronome. Three-syllable stimuli were spliced from the repetitive speech. These stimuli contained natural decreases in VOT with faster speech rates. Besides, this rate effect on VOT was larger for /p/ than /b/, so that VOT values for /b/ and /p/ overlapped at the fastest rates. Eighteen native listeners of American English were presented with 168 stimuli and asked to identify the consonant. Perceptual category boundaries occur at VOT values 15 ms shorter than the values reported for synthesized stimuli. This difference may be due to the extraordinarily wide range of VOT values in previous studies. The values found in the current study closely match the actual division point for /b/ and /p/. The underlying mechanism of perceptual normalization will be discussed.

2aSC24. The Ganong paradigm: Converging evidence supporting initial phoneme weighting. Erik C. Tracy and Mark A. Pitt (Dept. of Psych., The Ohio State Univ., 1885 Neil Ave. Mall, Columbus, OH 43210)

In the present experiment we investigate whether the initial phoneme is given more weight in word recognition [W. D. Marslen-Wilson and A. Welsh, Cognit. Psych. **10**, 29–63 (1978)] or if all phonemes in a word are weighted equally [C. M. Connine, D. G. Blasko, and D. Titone, J. Mem. Lang. **32**, 193–210 (1993)]. Using the Ganong paradigm [W. F. Ganong, JEP:HPP. **6**, 110–125 (1980)], participants were instructed to categorize a final ambiguous fricative in the target items, which included both words and pseudowords. Pseudowords were created by changing either the initial or a medial phoneme within the words. For example, the word diminish was altered to create the pseudowords timinish and dimimish. In addition, initial and medial phonemes were altered by either one or three distinctive features. The differences in the labeling of the final ambiguous fricative in the target items led to the conclusion that the initial phoneme is weighted more heavily. [Work supported by NIDCD.]

Session 2aSP

Signal Processing in Acoustics and Noise: Signal Processing for Aircraft Noise

Joe W. Posey, Chair
NASA Langley Research Center, Mail Stop 461, Hampton, Virginia 23681

Chair's Introduction—8:00

Invited Papers

8:05

2aSP1. Beamforming for aircraft noise measurements. Robert P. Dougherty (OptiNav, Inc., 10914 NE 18 St., Bellevue, WA 98004, rpd@optinav.com)

Phased array beamforming for aircraft noise source location has a long history, including early work on jet noise, wind tunnel measurements, and flyover testing. In the last 10 years, advancements in sparse 2-D and 3-D arrays, wind tunnel test techniques, and computer power have made phased array measurements almost common. Large aerospace companies and national research institutes have an advantage in access to major facilities and hundreds of measurement microphones, but universities and even consulting companies can perform tests with electret microphones and PC data acquisition systems. The type of testing remains a blend of science and art. A complex noise source is approximated by a mathematical model, and the microphones are deployed to evaluate the parameters of the model. For example, the simplest, but often the best, approach is to assume a distribution of mutually incoherent monopoles. This leads to an imaging process analogous to photography. Other models include coherent distributions of multipoles or duct modes. It is sometimes important to simulate the results that would have been obtained from single microphone measurements of part of the airplane in an ideal environment, had such measurements been feasible.

8:50

2aSP2. The rotating rake fan mode measurement system. Daniel Sutliff (NASA Glenn Res. Ctr., MS 54-3, 21000 Brookpark Rd., Cleveland, OH 44212)

An experimental measurement system was developed and implemented by the NASA Glenn Research Center in the 1990s. The system is a continuously rotating radial rake immersed into the duct. This rotating rake provides a complete map of the acoustic duct modes present in a ducted fan and has been used on a variety of test articles: from a low-speed, concept test rig to a full-scale production turbofan engine. The rotating rake has been critical in developing and evaluating a number of noise reduction concepts as well as providing experimental databases for verification of several aero-acoustic codes. This paper will describe the physical theory (Sofrin) and the analytical techniques (Moore) upon which the rotating rake is based will be described. Data processing and analysis as well as implementation issues will be discussed. Several Rotating Rake systems have been custom built for 3 facilities. In order of complexity of the turbo machinery test article, these are (1) the Advanced Noise Control Fan, (2) various 22-inch fan rigs in the NASA Glenn 9×15 wind tunnel, and (3) a full scale turbofan, the Honeywell TFE-731-60. Descriptions and measurement achievements of these systems will be provided (Heidelberg, Sutliff).

9:20

2aSP3. Novel error sensing microphone arrays for active control of turbofan rotor/stator tones. Bruce E. Walker, Alan S. Hersh (Hersh Walker Acoust., 780 Lakefield Rd., Unit G, Westlake Village, CA 91361), Edward J. Rice (E. J. Rice Consulting, Westlake, OH), and Daniel L. Sutliff (NASA Glenn Res. Ctr., Cleveland, OH)

Active control of turbofan rotor/stator interaction tones is complicated by the simultaneous presence of multiple duct propagation modes. In-duct error sensing microphone arrays that can adequately resolve these modes typically require duct lengths that are incompatible with modern compact engine design. Two alternative approaches have been investigated. For inlet noise, an external linear array of microphones was positioned in the near/far radiation field transition region and weighted to provide error signals resolved either by duct mode or by radiation angle. For the exhaust, radially spaced microphones have been placed on duct bifurcation panels to provide supplemental radial-mode resolution. The concepts were tested in combination with an adaptive segmented liner in a static duct and as part of an active stator-vane system in the ANCF research facility at NASA/Glenn Research Center. [Work sponsored by NASA/Langley Research Center.]

9:50

2aSP4. Methodologies for duct liner impedance eduction. Tony L. Parrott, Michael G. Jones (Structural Acoust. Branch, MS 463, 2 N. Dryden St., NASA Langley Res. Ctr., Hampton, VA 23681-2199, t.l.parrott@larc.nasa.gov), Willie R. Watson (NASA Langley Res. Ctr., Hampton, VA 23681-2199), and Charles D. Smith (Lockheed Martin Eng. and Sci. Co., NASA Langely Res. Ctr., Hampton, VA 23681-2199)

Methodologies and techniques are reviewed that are currently employed at Langley Research Center to educe (from primitive measurements) the impedance of acoustically absorbing liner structures. These structures are of interest for suppressing noise emission from aircraft engine nacelles. The accuracy and precision of the primitive measurements and their impact on the educed impedance of liners when exposed to high speed grazing flows is of special interest for aircraft engine nacelle applications. The test setups range from the classical standing wave tube for which the primitive measurement is a complex transfer function between two judiciously chosen locations, to an elaborate grazing flow duct arrangement (the Langley Grazing Incidence Tube) for which the primitive measurements (acoustic pressure and phase) are compromised by increased flow noise contamination due to high speed grazing flows up to a Mach number of 0.5. Results of different techniques/methodologies are compared on the basis of how the primitive measurements are processed and mapped into impedance spectra for different test setups. The results are compared on the basis of bias and precision errors that are specific to the impedance eduction methodology employed.

10:20–10:35 Break

10:35

2aSP5. *In situ* evaluation of aircraft interior noise reduction technologies. Jacob Klos and Daniel L. Palumbo (Structural Acoust. Branch, M.S. 463, NASA Langley Res. Ctr., Hampton, VA 23681, j.klos@larc.nasa.gov)

In order to quantify the performance of interior noise treatments under flight conditions, it is desirable to evaluate the noise reduction due to treatment of a limited portion of an aircraft fuselage. However, radiation from the untreated areas of the fuselage can corrupt an intensity measurement in front of the treated area. In the past, this problem of corrupting noise has been solved by acoustically isolating the treated area from the rest of the fuselage. In this presentation, a method to evaluate the performance of an acoustic treatment applied to an aircraft fuselage *in situ* using correlation analysis is documented. The insertion loss of the acoustic treatments is estimated from the ratio of the intensity, correlated to reference transducers, measured with and without the treatment applied. The formulation is presented for both single and multiple reference transducers. Several experimental studies and numerical simulations have been conducted, and the results are documented. Through these case studies, it is demonstrated that this method can be used to evaluate the insertion loss of fuselage treatments without having to acoustically isolate the treated area.

10:55

2aSP6. Signal processing for aircraft flyover noise synthesis and propagation. Stephen A. Rizzi, Brenda M. Sullivan (Structural Acoust. Branch, NASA Langley Res. Ctr., Hampton, VA 23681, Stephen.A.Rizzi@nasa.gov), and Bryan A. Cook (AuSIM, Inc., Mountain View, CA 94043)

Subjective assessments of low noise aircraft flight operations require time histories of acoustic pressure at listener positions. Synthesized sound has an advantage over recordings by allowing the examination of proposed aircraft, flight procedures, and other conditions or configurations for which recordings are unavailable. A two-stage process for synthesizing flyover noise at listener positions on the ground is presented. The first stage entails synthesizing time histories at the flying source. Rizzi and Sullivan [J. Acoust. Soc. Am. **113**, 2245 (2003)] developed an approach for synthesizing sound from broadband sources (e.g., jet noise) based upon predicted 1/3-octave band source spectra. Further developments in source synthesis signal processing are presented here, including the introduction of temporal fluctuations and the synthesis of tone-dominated sources (e.g., fan noise). The second stage entails propagation of the synthesized sound from the flying source to the listener. Its signal processing aspects are also presented and include a dynamic absolute delay of sound and location (producing accurate Doppler shift and emission position), spreading loss, atmospheric attenuation, and binaural filtering. Whereas the first stage prediction analysis and synthesis are computed *a priori*, the second stage is performed in real-time, allowing creation of an immersive test environment.

Contributed Papers

11:15

2aSP7. Near-field acoustical holography in enclosed spaces. Courtney Burroughs (Appl. Res. Lab., Penn State Univ., State College, PA 16804)

In enclosed spaces, the locations of surfaces from which most of the noise radiation occurs can be difficult to determine in the presence of reverberation and contributions from other sources. By using microphones configured in an array that confirms to the surface of interest, it should be possible to map the acoustic field inside the enclosed space produced by the surface under the array using nearfield acoustical holography processing on the unsteady pressures measured by the microphones in the conformal array. Such an acoustic mapping could therefore be used to isolate contributions to the interior noise make by different surfaces. Because enclosed spaces rarely have simple geometries, it is often necessary to measure the propagation function (i.e., the Green's function) for the enclosed space. To explore the sensitivity of mappings to measurement and physical parameters, a mathematical model of the acoustic field inside a capped cylindrical shell was developed. Measurements at locations on a conformal array are simulated, maps of the interior acoustic fields developed by applying numerical nearfield acoustical holography algorithms on the simulated measurements and resulting acoustic mappings compared to acoustic fields predicted directly from known surface vibration distributions.

11:30

2aSP8. Supersonic naval missile sounds over San Nicolas Island. Charles R. Greene, Jr., Robert G. Norman (Greeneridge Sci., Inc., 4512 Via Huerto, Santa Barbara, CA 93110), Meike Holst (LGL Ltd. Environ. Res. Assoc., King City, ON L7B 1A6, Canada), and Charles I. Malme (Eng. & Sci. Services, Hingham, MA 02043)

Vandals and other missiles are launched occasionally from San Nicolas Island, CA, during Naval exercises and tests. Pinnipeds on the island beaches are exposed to the flight sounds, some of which are sonic booms from directly overhead. Environmental concerns led the Navy to support acoustic studies of the missile sounds at the beaches. The results show flat-weighted sound pressures from Vandals as high as 150 dB *re*: 20 μPa(peak) [140 dB *re*: 20 μPa(rms)] at a near-vertical distance of 400 m. Other flat-weighted pressures from Vandals were as low as 107 dB *re*: 20 μPa(peak) [95 dB *re*: 20 μPa(rms)] at a beach 3.9 km horizontally behind the launcher. Pulse durations and sound exposure levels were also measured. One-third octave band sound exposure levels were measured. All parameters (except one-third octave band levels) were also measured with *A* weighting. Other missiles measured include Tomahawk cruise missiles, Rolling Airframe Missile, Advanced Gun System, Terrier, and the Supersonic Sea-Skimming Target. [Work supported by U.S. Navy.]

11:45

2aSP9. Multiresolution analysis of noise generated by aircrafts during landing/take-off operations. Luigi Maxmilian Caligiuri and Adolfo Sabato (Univ. of Calabria, via P. Bucci, 87036 Rende (CS), Italy)

The evaluation of noise impact generated by aircrafts is usually realized by means of indicators such as Lden and/or by indexes obtained opportunely combining the values of single event levels relative to the events occurred within the measurement time interval. In any case is fundamental, to ensure the validity of result, to properly find and characterize the noise events produced by aircrafts landing/take-off operations, in particular distinguish them from that produced by other sources. Since the traditional analysis techniques are generally inadequate to describe these nonstationary signals, its necessary to employ time-frequency or time-scale based techniques. In this paper it will be shown the application of multiresolution time-scale and time-frequency analysis techniques, based on Wavelet (WT) and smoothed Wigner Ville transforms (SWVT), to noise signals related to aircrafts landing/take-off operations, also comparing the result with those obtained using a Windowed Fast Fourier Transform (WFFT). The effects of changing the parameters related to the smoothing WVT windows and to the WT base functions atoms will be investigated, showing that an adequate choose of such parameters will allow to us to properly analyze and characterize these noise events in order to quantify their impact.

TUESDAY AFTERNOON, 11 NOVEMBER 2003 BRAZOS AUDITORIUM, 1:00 TO 5:00 P.M.

Session 2pAA

Architectural Acoustics and Musical Acoustics: Electroacoustic Enhancement System in Rooms for Music

Yasushi Shimizu, Chair
200-41 Timitsuka, Hamamatsu 432-8002, Japan

Chair's Introduction—1:00

Invited Papers

1:10

2pAA1. Electronic architecture: An acoustician's approach. J. Christopher Jaffe (Jaffe Holden Acoust., 114A Washington St., Norwalk, CT)

Electronic architecture is a process for modifying the sound fields of an auditorium, concert hall, or theatre to recreate the sound fields of outstanding examples of individual performance spaces. This paper discuss how acousticians develop proper acoustic environments for specific program use in physical acoustic halls and how these design goals are incorporated into electronically enhanced facilities.

1:35

2pAA2. Control of early and late energy in rooms with the variable room acoustics system. Mark Poletti (Industrial Res. Ltd., P.O. Box 31-310, Lower Hutt, New Zealand, m.poletti@irl.cri.nz) and Steven Ellison (Level Control Systems, Sierra Madre, CA 91024)

Many electroacoustic systems seek to offer control of subjective sound quality using either regeneration of sound or by minimizing feedback to produce in-line sound control. In-line systems provide control of early energy from the stage but do not respond globally to all sound sources. Regenerative systems offer global enhancement of late energy but are less able to influence early energy. The Variable Room Acoustics System is a hybrid approach which contains a regenerative system for reverberation enhancement and an in-line system for early energy control. Both systems are optimized to provide maximum power gain. The principles of the system are outlined and results from recent installations are presented.

2:00

2pAA3. Active field control (AFC) -electro-acoustic enhancement system using acoustical feedback control. Hideo Miyazaki, Takayuki Watanabe, Shinji Kishinaga, and Fukushi Kawakami (Adv. System Development Ctr., Yamaha Corp., 10-1 Nakazawa-cho Hamamatsu, Japan)

AFC is an electro-acoustic enhancement system using FIR filters to optimize auditory impressions, such as liveness, loudness, and spaciousness. This system has been under development at Yamaha Corporation for more than 15 years and has been installed in approximately 50 venues in Japan to date. AFC utilizes feedback control techniques for recreation of reverberation from the physical reverberation of the room. In order to prevent coloration problems caused by a closed loop condition, two types of time-varying control techniques are implemented in the AFC system to ensure smooth loop gain and a sufficient margin in frequency characteristics to prevent instability. Those are: (a) EMR (electric microphone rotator) -smoothing frequency responses between microphones and speakers by changing the combinations of inputs and outputs periodically; (b) fluctuating-FIR -smoothing frequency responses of FIR filters and preventing coloration problems caused by fixed FIR filters, by moving each FIR tap periodically on time axis with a different phase and time period. In this paper, these techniques are summarized. A block diagram of AFC using new equipment named AFC1, which has been developed at Yamaha Corporation and released recently in the US, is also presented.

2:25

2pAA4. Various applications of Active Field Control (AFC). Takayuki Watanabe, Hideo Miyazaki, Shinji Kishinaga, and Fukushi Kawakami (Adv. System Development Ctr., Yamaha Corp., 10-1 Nakazawa-cho, Hamamatsu, Japan, watanabe@yarl.yamaha.co.jp)

AFC is an electro-acoustic enhancement system, which has been under development at Yamaha Corporation. In this paper, several types of various AFC applications are discussed, while referring to representative projects for each application in Japan. (1) Realization of acoustics in a huge hall to classical music program, e.g., Tokyo International Forum. This venue is a multipurpose hall with approximately 5000 seats. AFC achieves loudness and reverberance equivalent to those of a hall with 2500 seats or fewer. (2) Optimization of acoustics for a variety of programs, e.g., Arkas Sasebo. AFC is used to create the optimum acoustics for each program, such as reverberance for classical concerts, acoustical support for opera singers, uniformity throughout the hall from the stage to under-balcony area, etc. (3) Control of room shape acoustical effect, e.g., Osaka Central Public Hall: In this renovation project, preservation of historically important architecture in the original form is required. AFC is installed to vary only the acoustical environment without architectural changes. (4) Assistance with crowd enthusiasm for sports entertainment, e.g., Tokyo Metropolitan Gymnasium. In this venue, which is designed as a very absorptive space for speech intelligibility, AFC is installed to enhance the atmosphere of live sports entertainment.

2:50

2pAA5. Active acoustics for music rehearsal rooms. Ronald R. Freiheit (555 Park Dr., Owatonna, MN 55060, ron.freiheit@wengercorp.com)

The use of virtual acoustics has the ability to provide a new level of rehearsal experience for the musician. By integrating the signal processing of an active acoustic system (with time variant-gain before feedback) into a relatively small rehearsal room, musicians can now benefit from the experience of rehearsing in multiple acoustic environments including those of the actual performance venue in which they will perform. To effectively communicate the various acoustics environments, the musicians must be immersed in the sound field of the active acoustics without being able to discern source locations of the speakers. The system must also be cable of supporting the dynamic range of the musicians without presenting artifacts of its own such as system noise or audible distortion. An installation of such a system will be provided as a case study describing the challenges that were overcome for a successful implementation including areas such as adequate sound isolation, background noise levels and system security. The paper will also briefly discuss programming methodologies for the system. Anecdotal responses from musicians who have used the active acoustic rehearsal room and some unexpected issues will also be covered.

3:15–3:30 Break

3:30

2pAA6. Acoustical enhancement systems: Design criteria and evaluation of room acoustical parameters based on *in situ* measurements. Bjorn van Munster and Wim Prinssen (Systems for Improved Acoust. Performance, Runmolen 3, 5404 KP Uden, The Netherlands, b.v.munster@pbri.nl)

Acoustic enhancement systems have evolved significantly during the years. Where the early systems only aimed to increase the reverberation time in a hall, nowadays the increase of the reverberation time is only one of the features of such a system. Contrary and additionally to passive acoustics, an acoustic enhancement system enables a designer or acoustical consultant to change the acoustical characteristics of a hall in a more flexible way. Due to the sophisticated convolution processes and layout of such a system besides the reverberation time also, e.g., speech intelligibility and spaciousness can be improved or special effects can be added to shows. In this paper the applications of an enhancement system in general will be outlined in more detail. Furthermore, design criteria will be given which can be formulated for the installation of such a system. These criteria can be used to evaluate proposed designs, but also to estimate the required provisions to be included in the planning of a system installation. Besides, the paper describes the results of *in situ* measurements of one such system (SIAP) whereby the increase of the reverberation time is evaluated with respect to certain important room acoustical parameters, i.e., reverberant level, lateral efficiency and clarity.

3:55

2pAA7. An electronic enhancement system for Silva Concert Hall. Timothy E. Gulsrud (Kirkegaard Assoc., 954 Pearl St., Boulder, CO 80302, tgulsrud@kirkegaard.com), Scott D. Pfeiffer, and Frans H. H. Swarte (Kirkegaard Assoc., Chicago, IL 60607)

An electronic enhancement system and new orchestra shell have been designed for and installed in Silva Concert Hall, a 2500-seat multipurpose hall at the Hult Center for the Performing Arts in Eugene, Oregon. The enhancement system provided by Acoustic Control Systems B.V. features early reflection and reverberation modules in the house and a foldback system for performers on stage. The design considerations, interaction with the new shell, tuning, and in-hall measurements of the system are discussed, including reactions by musicians, administrators, and critical listeners.

Contributed Paper

4:20

2pAA8. Experimental evaluation of the omnidirectional behavior of platonic polyhedron loudspeakers. Sarah Rollins, Timothy Leishman, and Gordon Dix (Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602)

Many architectural acoustics measurements require the use of an omnidirectional source. For several years, the source predominantly used for such applications has been the dodecahedron loudspeaker with small in-phase drivers mounted in each face. While other platonic polyhedron loudspeakers (PPLs) have not been used as frequently, they also produce nearly omnidirectional fields over limited bandwidths. Above cutoff frequencies specific to their geometries, all PPLs depart from ideal omnidirectional behavior, with varying degrees of directivity. While these cutoff frequencies are typically higher for higher-order polyhedra, they commonly fall within the bandwidths of standard measurements. The five types of PPLs have been constructed and measured to gain greater insight into their omnidirectional behaviors. Their frequency responses were taken at 2664 points over a sphere (5-deg polar and azimuthal angle increments) in an anechoic chamber. The measurements were then processed to produce directivity balloon plots. However, to better compare directivities and find the source consistently producing the most omnidirectional field over a useful bandwidth, a frequency-dependent standard deviation formula was implemented. Average values of the standard deviation parameter produce figures of merit that further characterize omnidirectionality. [Research supported by funding from the NSF REU program.]

4:35–5:00

Panel Discussion

TUESDAY AFTERNOON, 11 NOVEMBER 2003 PECOS ROOM, 1:00 TO 5:00 P.M.

Session 2pAO

Acoustical Oceanography: Geoacoustic Inversion

Gopu R. Potty, Chair
Department of Ocean Engineering, University of Rhode Island, Narragansett, Rhode Island 02881

Contributed Papers

1:00

2pAO1. Geoacoustic inversion of broadband data in the Florida Straits. N. Ross Chapman and Yongmin Jiang (School of Earth and Ocean Sci., Univ. of Victoria, P.O. Box 3055, Victoria, BC V8W 3P6, Canada)

Acoustic propagation experiments have been carried out in the Florida Straits with a multi-frequency broadband source that transmitted m-sequence pulses over a range of 10 km to a sparse-filled vertical line array. This paper presents results of matched field inversions of the acoustic field data at low frequencies to estimate geoacoustic model parameters for the experimental site. Two approaches were taken for the inversions. The first was a conventional matched field inversion using multi-frequency data centered at 200 and at 400 Hz from the vertical array. The second approach was designed to model the low frequency waveform at a single hydrophone. For the very long range experimental geometry, the waveform was modeled in terms of modes. Each inversion was cast as an optimization problem using the adaptive simplex simulated annealing algorithm. The inversions provide a comparison between approaches that take advantage of the spatial coherence in one case, and the time coherence in the received signal in the other case. Both inversions give similar results for the parameters of a simple geoacoustic bottom model, and the sensitivities and relative uncertainties of the model parameters are consistent for the two approaches. Notably, the inversions are sensitive to compressional wave attenuation.

1:15

2pAO2. Attenuation inversions in the East China Sea. Gopu Potty and James Miller (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882)

Data from the ASIAEX East China Sea (ECS) experiment is used to estimate sediment attenuation coefficients as a function of frequency and depth. The ECS experiment offers a number of independent measures of sediment parameters using seismic and chirp surveys, gravity and piston cores and historic data. Modal amplitude ratios for modes 1 to 3 are used to obtain the modal attenuation coefficient and attenuation profiles. These inversions use broad-band data from Wide Band Sources (WBS) in the frequency range 20–100 Hz. The modes are detected, identified and their amplitudes measured using a time–frequency wavelet analysis. A joint inversion for modal attenuation coefficients, water depth, source and receiver depths, range and Empirical Orthogonal Functions in the water column is performed. The results will be compared with core data and frequency and depth dependence will be examined. [Work supported by ONR.]

1:30

2pAO3. Geoacoustic inversion results of low- to high-frequency source tow data from the ASIAEX East China experiment. Chen-Fen Huang and William S. Hodgkiss (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238)

During the 2001 ASIAEX East China Sea experiment, source tow data was collected by a 16-element, 75-m aperture, autonomously recording vertical line array in approximately 105-m-deep water. Transmissions from two similar 6-km-long tracks were transmitted. In the first, CW tonals at 95, 195, 295, 395, 805, 850, and 905 Hz were transmitted from a J-15 transducer at a nominal depth of 46 m. In the second, CW tonals at 1.6, 2.4, 3.5, and 4.4 kHz were transmitted from an ITC-2015 transducer at a nominal depth of 49 m. An environmental model in the study area initially was derived by matched-field geoacoustic inversion using the low-frequency track. In this study, the high-frequency data from the second track are included in the parameter assessment. A multiple-stage geoacoustic inversion scheme is applied. Inversion results for seafloor geoacoustic parameters from these transmissions will be presented. [Work supported by ONR.]

1:45

2pAO4. Acoustic inversions from an explosive source of opportunity in the ASIAEX SCS experiment. Ying-Tsong Lin (Natl. Taiwan Univ., No. 1, Sec. 4, Roosevelt Rd., Taipei 106, Taiwan), James F. Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), Nicholas P. Chotiros (Office of Naval Res., Arlington, VA 22217), Altan Turgut (Naval Res. Lab., Washington, DC), Ching-Sang Chiu (Naval Postgrad. School, Monterey, CA), and Steven G. Schock (Boca Raton, FL)

An acoustic inversion for bottom geoacoustic properties from an unexpected explosion in the South China Sea (SCS) is presented. Horizontal beamforming is performed to find the source azimuthal direction and the water wave part of the modal dispersion curve is used to find the source range. The localization result shows that the detonation occurred in the direction of east–north east and about 40 km distant from the receiver. To perform the bottom inverse, we use a broadband linear inverse technique employing the modal group velocity, which is inferred from the modal arrival time and source range. In the end, a four-layered bottom model is created, which is consistent with both our explosion data and independent source tow data. The resolution and variance of the estimate are also presented. In conclusion, this bottom model should be widely applicable for modeling broadband sound propagation in the ASIAEX SCS experiment site.

2:00

2pAO5. Integration versus optimization in self-noise geoacoustic inversion. David J. Battle, Peter Gerstoft, William A. Kuperman, William S. Hodgkiss (Marine Physical Lab., Univ. of California–San Diego, La Jolla, CA 92093-0238, davidb@mpl.ucsd.edu), and Martin Siderius (SAIC, La Jolla, CA 92037)

Self-noise geoacoustic inversion involves the estimation of bottom parameters such as sound speeds and densities by analyzing towed-array signals whose origin is the tow platform itself. As well as feeding into more detailed assessments of seabed geology, these parameters enable performance predictions for sonar systems operating in shallow water. In this presentation, the use of the Gibbs sampler to obtain complete probability distributions of seabed parameters is discussed. This contrasts with conventional maximum likelihood inversion, in which only the best fitting model is identified. Advantages of viewing parameter estimation problems from such a probabilistic perspective are discussed in relation to the MAPEX 2000 self-noise data, which have previously been subjected to maximum-likelihood inversion via genetic algorithm optimization.

2:15

2pAO6. Sub-bottom profiling by inverting ambient noise. Chris H. Harrison (SACLANT Undersea Res. Ctr., Viale San Bartolomeo 400, 19138 La Spezia, Italy, harrison@saclantc.nato.int)

In this paper a technique based on spectral factorization for restoring the phase of incoherent bottom sediment reflection coefficient measurements is presented, so that by Fourier transformation one can then obtain the minimum phase impulse response at each grazing angle. The method is developed and discussed in the context of another recently established technique for extracting the seabed's plane wave reflection coefficient from ambient noise data measured on a moored or drifting vertical array (VLA). Limitations of the phase restoration method are discussed, and using modeled data, comparisons are made between the ''true'' impulse response derived from the known complex reflection coefficient and the result of applying spectral factorization to the absolute value of the reflection coefficient. For instance, in both cases one can see clear, matching arrivals from each layer boundary at angles above critical. Finally the method is demonstrated on experimental reflection loss inferred from ambient noise measurements at three moored VLA sites and one VLA drift track in the Mediterranean Sea. Convincing angular variation of the impulse response is shown for the moored sites. Sub-bottom profiles (impulse response vs position) are shown for the drift track demonstrating that one can survey with just a single directional receiver.

2:30

2pAO7. Numerical modeling of air-to-sea transmission of light aircraft noise. Eric M. Giddens and Michael J. Buckingham (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, 8820 Shellback Way, La Jolla, CA 92093-0238, egiddens@ucsd.edu)

Recent experiments at SIO have shown that the acoustic signature of a light aircraft can be detected by sensors in the water column as well as buried in the underlying sediment and a method for extracting the sound speed and attenuation from this Doppler shifted signal has been proposed. To test the accuracy of this geoacoustic inversion technique, a numerical model of the air-water-sediment acoustic propagation, including the effects of a high-speed airborne source, has been developed based on the spectral method. Simulated acoustic data have been generated representing an aircraft flying over a microphone in the atmosphere, a vertical line array in the ocean, and a hydrophone buried 1-m deep in the sediment. The results of the geoacoustic inversion for sound speed and attenuation are compared to the known input parameter values of the model, giving a sense of the relative errors that may be expected when applying the technique to experimental data. [Work supported by ONR.]

2:45

2pAO8. The effects of ignored seabed variability in geo-acoustic inversion. Anna-Liesa S. Lapinski and David M. F. Chapman (Defence R&D Canada Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada)

In recent years, acoustic inversion techniques have been developed to predict ocean environmental properties, source position, etc. However, when applying such techniques to measured acoustic data, assumptions are invariably made with respect to the parametrization of the ocean environment due to unavailable information or to simplify the problem. In this work, the influences of range-dependent variability of the seafloor in the results of inversion are explored. That is, a typical inversion algorithm is applied while making the assumption that all parameters of the ocean-bottom (e.g., compressional speed, water depth, density, attenuation) are constant with range when in fact one parameter varies realistically with range. The inversion algorithm is applied to many unique realizations of a synthetic ocean-bottom; however, for each realization, the nonfluctuating parameters remain constant and the statistics regarding the fluctuating parameter such as the mean parameter value and the standard deviation of the fluctuations about the mean remain the same. It is shown that ignoring the variation in even a single seabed parameter leads to significant and correlated uncertainty in all inverted parameter values.

3:00–3:15 Break

3:15

2pAO9. Are buried river channels sources of geoclutter on the New Jersey Continental Margin? John C. Osler (DRDC Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada, john.osler@drdc-rddc.gc.ca)

Geological features on a continental shelf may be responsible for anomalous acoustic scatter that are identified as (false) targets, or GeoClutter, on active sonar systems. Features on the New Jersey Continental Margin include a drainage system that formed when sea-level was much lower, ran across the shelf, and incised channels approximately 10 meters deep into the surrounding seabed. These channels have since been filled with sediments that are not apparent on bathymetric maps. The potential for these channels to create GeoClutter depends in part on the contrast in geoacoustic properties between the sediments filling the channels and the adjacent flanks. To study this matter, an experiment was conducted to measure the reflection loss from 1 to 10 kHz of channel fill and flank sediments in an area where GeoClutter has been observed and where there is supporting geophysical data. The measurements were made using the WARBLE technique [C. W. Holland and J. C. Osler, J. Acoust. Soc. Am. **107**, 1263–1279 (2000)], adapted for use in rapid environmental assessment using modified sonobuoys. Results from the experiment will be presented and the role of buried channels acting as sources of GeoClutter on the New Jersey Continental Margin will be discussed.

3:30

2pAO10. Propagation and attenuation of sound in a canonical model of shallow water with a thermocline and with a Biot sediment in the bottom. Allan D. Pierce, William C. Carey (Boston Univ., Boston, MA 02215, adp@bu.edu), and James F. Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Objective is benchmark solution, analytic in internal detail, for testing geoacoustic inversion methods. The water column has two isovelocity layers separated by a narrow thermocline. The bottom is a porous (Biot model) sediment, with an elastic matrix coexisting with a fluid filling the pores; coefficients are frequency independent, slowly varying with depth, and analogous to those in works by Stoll, Badiey, and Chotiros. Source and receiver are in the middle region of the water column. Constant frequency problem formulation has field expressed as a Fourier integral over horizontal coordinates, with a z-dependent kernel having discontinuous slope at source depth. Equations for the bottom region yield ratio of derivative to kernel amplitude at bottom interface, this being a complex function of wave number and frequency. Singularities in kernel at discrete complex values of k correspond to the natural modes, imaginary parts of k are attenuation constants. Simplifying approximations, especially those of perturbation theory, are guided by order of magnitude estimates of dimensionless groups, with the suppositions that the orders of magnitude of propagation frequency, water depth, and propagation distance are 100 Hz, 100 m, and from 2 to 100 km, and that Biot model parameter magnitudes are comparable to known measured values.

3:45

2pAO11. Investgation of sediment properties with a rotated coordinates inversion technique. Tracianne Neilsen, Marcia Isakson, and Andrew Worley (Appl. Res. Labs., Univ. of Texas, 10000 Burnet Rd., Austin, TX 78758)

Several models have been developed that describe the interaction of high-frequency sound with ocean sediments. This work presents how estimates of the seabed properties can be obtained with a simulated annealing inversion algorithm that minimizes the difference between measured and modeled reflection loss within the statistical nature of the data and the statistical nature of the parameters. Because there is often a correlation between how the various seabed properties influence the reflection loss of acoustic waves, the efficiency and accuracy of the inversion can be improved by using a rotated set of coordinates to navigate the search space in the inversion. The rotated coordinates are obtained by performing an eigenvalue decomposition of the covariance matrix of the gradients of the cost function. The resulting eigenvectors are the rotated coordinates, show the couplings between the seabed parameters, and are used to vary the parameters in the inversion. The corresponding eigenvalues indicate the relative sensitivities of the cost function to changes in the parameters and provide insights into how the inversion problem can be effectively decoupled [T. B. Neilsen, ''An iterative implementation of rotated coordinates for inverse problems,'' J. Acoust. Soc. Am. **113**, 2574–2586 (2003)]. [Work supported by ONR.]

4:00

2pAO12. A new inversion method (SUB-RIGS) based on range-sensitive frequency-subspacing and reduced iterated grid searches. A. Tolstoy (ATolstoy Sci., 8610 Battailles Court, Annandale, VA 22003)

Propagation for a range-independent, multi-layered test case (Na or Workshop '97) indicates that the highly variable bottom sound-speed profile can be well represented by only a few, e.g., three, linear layers (one layer at the highest frequency, three at the lowest). Thus, at the low frequencies of interest here (less than 500 Hz) transmission loss as a function of range can usually be well predicted using only a few simple sediment layers. This suggested a new inversion method based on range-sensitive frequency-subspacing plus reduced iterated grid spacing (SUB-RIGS). The method is now fully developed and uses the RAMGEO PE propagation model. This paper will discuss the method, pitfalls of the method, and newly successful inversion results for selected test cases (TC1 and CAL from the recent Stennis workshop of May 2001).

4:15

2pAO13. Measurement uncertainty of seabed reflectivity. Charles Holland (Appl. Res. Lab., The Penn State Univ., State College, PA 16804)

The seabed reflection coefficient is a fundamental property of the ocean waveguide. Measurements of the frequency and angular dependence of the reflection coefficient can provide information about the geoacoustic properties of the seabed or can be used as an input to propagation models. The uncertainty of the measurements must be known in order to determine the prediction uncertainties for the acoustic field and/or the geoacoustic properties. An analysis indicates that reflection measurements [Holland and Osler, J. Acoust. Soc. Am. **107**, 1263–1279 (2000)] have a standard deviation from 0.5–1 dB at full angular resolution depending on frequency and experiment geometry. The dominant contribution to the error is source amplitude variability and a new processing approach was developed that reduces the error for frequencies above a few hundred Hz. A further reduction in the uncertainty can be obtained by averaging in angle, for example, a 1 angle averaging leads to a standard deviation of 0.1–0.5 dB. Errors in the angle estimate are a few tenths of a degree from 0–34° grazing angle: the crucial angular range for predicting long-range propagation or for geoacoustic property inversion. [Research sponsored by ONR.]

4:30

2pAO14. Optimally resolving Lambertian surface orientation. Ioannis Bertsatos and Nicholas C. Makris (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

Sonar images of remote surfaces are typically corrupted by signal-dependent noise known as speckle. Relative motion between source, surface, and receiver causes the received field to fluctuate over time with circular complex Gaussian random (CCGR) statistics. In many cases of practical importance, Lambert's law is appropriate to model radiant intensity from the surface. In a previous paper, maximum likelihood estimators (MLE) for Lambertian surface orientation have been derived based on CCGR measurements [N. C. Makris, SACLANT Conference Proceedings Series CP-45, 1997, pp. 339–346]. A Lambertian surface needs to be observed from more than one illumination direction for its orientation to be properly constrained. It is found, however, that MLE performance varies significantly with illumination direction due to the inherently nonlinear

nature of this problem. It is shown that a large number of samples is often required to optimally resolve surface orientation using the optimality criteria of the MLE derived in Naftali and Makris [J. Acoust. Soc. Am. **110**, 1917–1930 (2001)].

4:45

2pAO15. Geocorrection and filtering of 3D bottom images from multi-beam sonar records. Jerzy Demkowicz, Krzysztof Bikonis, Andrzej Stepnowski, and Marek Moszynski (Gdansk Univ. of Technol., Narutowicza 11/12, 80-952 Gdansk, Poland)

For the last decade multibeam sonars have been increasingly used for mapping and visualization of the seafloor to provide the ''physical bases'' for environmental studies. Increasing amount of digital (raster) echo records of high resolution from a multibeam sonar have enhanced the potential of computer modeling of the marine environment to improve our understanding of the bottom processes. However, the 3D bottom images as the result of merging different sonar transects do not comply exact geographical positions and should be corrected. Additionally, the raw sonar records are subject to systematic errors, random noise and outliers. In this paper, Kalman filtering techinque to generating optimal estimates of bottom surface from a noisy raw sonar records is proposed. The experiment on the surface indicates that after applying the Kalman filtering the outliers of raw records can be efficiently removed. Moreover, the two-step Kalman filtering method enables 3D seabed visualization in real time. The paper proposes the geographical corrections applied to the merged mutibeam sonar transects records. The 3D bottom relief before, and after the filtering method are presented.

TUESDAY AFTERNOON, 11 NOVEMBER 2003 TRINITY A ROOM, 1:00 TO 3:30 P.M.

Session 2pBB

Biomedical Ultrasound/Bioresponse to Vibration: HIFU and Scattering

Ibrahim M. Hallaj, Chair

Wolf, Greenfield and Sacks, PC, Federal Reserve Plaza, 600 Atlantic Avenue, Boston, Massachusetts 02210

Contributed Papers

1:00

2pBB1. Synchronization of HIFU therapy system with an arbitrary ultrasound imager. Neil Owen, Michael Bailey, James Hossack, and Lawrence Crum (Ctr. for Industrial and Medical Ultrasound, 1013 NE 40th St., Seattle, WA 98105)

Synchronization for image guided therapy using high intensity focused ultrasound (HIFU) and imaging ultrasound is achieved with a new technique that uses the focused transducer as a receiver that can detect the acoustic pulses created by the imaging probe. Without synchronization, interference from the high intensity source occludes the imager's display unpredictably, degrading the quality of the system. An imaging probe (Sonosite 180) is registered with a HIFU transducer ($d=33$ mm, roc $=55$ mm, $f=3.5$ MHz) such that the scan line bisects the single element focus. When acoustically coupled through a scattering medium, imaging pulses are passively detected with the HIFU transducer and electronically conditioned into a TTL level trigger. A LabVIEW program uses the trigger to create a pulse width modulated signal that controls the timing of HIFU excitation during treatment. Detection takes less than 1% of the time between displayed images when the imager is running at 20 frames per second. HIFU excitations are programmed to occur such that the single element focus is free of interference when viewed with the imager during treatment. With no electrical connections for this new, simple technique, an arbitrary imager can be selected for synchronized image guided therapy. [Work supported by NSBRI.]

1:15

2pBB2. Rapid continuous-wave pressure field calculations for spherically focused radiators. Robert McGough (Dept. of Elec. and Computer Eng., Michigan State Univ., 2120 Eng. Bldg., East Lansing, MI 48824, mcgough@egr.msu.edu)

A new accelerated expression for the continuous-wave pressure field generated by a spherically focused radiator is obtained when the impulse response formulation is transformed and optimized for numerical evaluations. The resulting integral expression converges much more quickly than the impulse response approach, resulting in far fewer function evaluations for the same numerical error. The optimized integral expression is between two and seven times as fast as the impulse response approach, where the increase in speed depends on the peak value of the specified error. In addition, this new result completely eliminates the cone-shaped regions required for impulse response calculations, so the resulting computer code for the accelerated expression is less complicated than the corresponding code for the impulse response. Results also show that the new expression eliminates the numerical artifact that is encountered near the boundary between regions defined for impulse response calculations. All of these features are useful in thermal therapy computer simulations that employ spherically focused transducer geometries.

1:30

2pBB3. Design and evaluation of a 63 element 1.75-dimensional ultrasound phased array for treating benign prostatic hyperplasia. Khaldon Y. Saleh and Nadine B. Smith (Dept. of Bioengineering, 205 Hallowell Bldg., The Penn State Univ., University Park, PA 16802)

Focused ultrasound surgery (FUS) is a clinical method for treating benign prostatic hyperplasia (BPH) in which tissue is noninvasively necrosed by elevating the temperature at the focal point above 60 °C using short sonications. With 1.75-dimensional (1.75-D) arrays, the power and phase to the individual elements can be controlled electronically for focusing and steering. This research describes the design, construction and evaluation of a 1.75-D ultrasound phased array to be used in the treatment of benign prostatic hyperplasia. The array was designed with a steering angle of ± 13.5 deg in the transverse direction, and can move the focus in three parallel planes in the longitudinal direction with a relatively large focus size. A piezoelectric ceramic (PZT-8) was used as the material of the transducer and two matching layers were built for maximum acoustic power transmission to tissue. To verify the capability of the transducer for focusing and steering, exposimetry was performed and the results correlated well with the calculated fields. *In vivo* experiments were performed to verify the capability of the transducer to ablate tissue using short sonications. [Work supported by the Whitaker Foundation and the Department of Defense Congressionally Directed Medical Prostate Cancer Research Program.]

1:45

2pBB4. Optimized hyperthermia treatment of prostate cancer using a novel intravavitary ultrasound array. Osama M. Al-Bataineh, Nadine B. Smith (Dept. of Bioengineering, The Penn State Univ., University Park, PA 16802), Robert M. Keolian, Victor W. Sparrow (The Penn State Univ., University Park, PA 16802), and Lewis E. Harpster (Penn State Milton S. Hershey Medical Ctr., Hershey, PA 17033)

Localized uniformly distributed ultrasound-induced hyperthermia is a useful adjuvant to radiotherapy in the treatment of prostate cancer. A two-dimensional, 20×4 element, transrectal phased-array probe was designed to deliver a uniform and controllable amount of heat directly to the prostate without damaging the rectal wall or surrounding tissue. A three-dimensional prostate model was created using anatomical markers from the Visible Human Project to optimize the array. Sound speed, density, and absorption parameters were mapped to hue, saturation and value of the photographic data to simulate sound propagation through inhomogeneous tissue using the k-space method. To satisfy the requirements of this method from 1.2 to 1.8 MHz, the grid was adjusted to have 5 points per millimeter in each Cartesian direction. A spherical wave pulse was propagated through the model using tapered absorption boundary conditions. The expected temperature rise due to sound was obtained using the bioheat transfer equation. Optimal insonification parameters that uniformly heat the prostate to 43 °C for 40–60 minutes were determined for use in the construction of a clinical hyperthermia array. [Research supported by the Department of Defense Congressionally Directed Medical Prostate Cancer Research Program.]

2:00

2pBB5. Separating thermal coagulation and cavitation effects in HIFU attenuation measurements. Justin Reed, Michael Bailey, Ajay Anand, and Peter Kaczkowski (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Box 355640, Seattle, WA 98105-6698)

HIFU can be used to destroy tumors. The conversion of acoustic energy into heat causes protein coagulation (Lesion) in tissue. Attenuation measurements have been proposed to monitor the progression of thermal therapy. The goal of this work is to study and separate the effects of cavitation and thermal coagulation in attenuation measurements. A HIFU transducer was used to treat Bovine liver. A receiving transducer mounted across from the transmitting HIFU transducer measured attenuation during the treatment. A pressure chamber provided static pressure greater than the pressure amplitude of the HIFU wave, which suppressed cavitation. rf data from a commercial ultrasound scanner was also obtained. A large increase in attenuation was observed with cavitation present, while a subtle increase in attenuation was observed with cavitation suppressed. Attenuation estimated from the RF data showed an increase in attenuation downstream of the location of the lesion with cavitation present, while a subtle increase in attenuation was observed at the location of the lesion with cavitation suppressed. It has been found that attenuation measurements are greatly affected by the presence of cavitation, and the actual effect of thermal coagulation on attenuation is quite small. [Work supported by NIH, NSF, NSBRI.]

2:15

2pBB6. Numerical investigation of dual-frequency HIFU pulsing for lithotripsy. Wayne Kreider, Michael Bailey, and Lawrence Crum (Ctr. for Industrial and Medical Ultrasound, APL, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, wkreider@u.washington.edu)

As an alternative to traditional shock-wave lithotripsy, high-intensity focused ultrasound (HIFU) is currently being investigated for its capability to comminute renal calculi. Because current data indicate that cavitation plays a role in both stone comminution as well as collateral tissue damage, the cavitation effects of HIFU treatment strategies are investigated numerically. In particular, numerical simulations are designed to model the response of bubbles to acoustic excitations generated by a prototype, dual-frequency HIFU transducer for lithotripsy. The prototype transducer is capable of producing both high- (~4-MHz) and low-frequency (~100-kHz) outputs, while the bubble dynamics are modeled by the Gilmore equation for a single spherical bubble subject to diffusion. Numerical simulations are currently ongoing to investigate the effects of the relative phase between high and low-frequency pulses. Initial results demonstrate that the simultaneous application of high and low-frequency pulses can generate maximum pressures several orders of magnitude higher than high-frequency pulses alone.

2:30

2pBB7. The characterization of the lesion growth in time. Marie Nakazawa, Justin A. Reed, Michael R. Bailey, and Yongmin Kim (Dept. of Elec. Eng., Univ. of Washington, 1400 NE Campus Pkwy., Seattle, WA, nakazawa@ns.cradle.titech.ac.jp)

Thermal heating effects of high intensity focused ultrasound (HIFU) on the dynamics of lesion formation were characterized automatically to assess the role of vapor bubbles in distorting the shape. Tissue mimicking phantom was used in experiments by a 4.2 MHz curve-linear transducer with 44 mm diameter and 44 mm radius of curvature. A variety of HIFU intensities were produced by different amplitudes. Images were acquired by a CCD camera and HDI-1000 ultrasound imager, recorded to VHS, and digitized to measure lesion size and shape. Each image was subtracted with noise reduction in order to detect the HIFU on time and to segment the boundaries of the lesions performed by Matlab programming. Area, length, width, and ratio of lesion area proximal to center line over area distal to center line were calculated along HIFU exposure time. Slight increase in HIFU intensity, means hyperecho forms earlier, and lesion shape change. The data supported the hypothesis that lesion dramatically distorts well after hyperecho with only small increase in HIFU intensity. [Work supported by National Space and Biomedical Research Institute.]

2:45

2pBB8. Optimization of angular compounding in scatterer size estimation. Anthony L. Gerig, Quan Chen, and James A. Zagzebski (Dept. of Medical Phys., Univ. of Wisconsin–Madison, 1300 Univ. Ave., Rm. 1530, Madison, WI 53706, algerig@wisc.edu)

Ultrasonic scatterer size estimates generally have large variances due to the inherent noise of the spectral estimates used to calculate size. Compounding partially correlated size estimates associated with the same tissue, but produced with data acquired from different angles of incidence, is an effective way to reduce the variance without making dramatic sacrifices in spatial resolution. This work derives theoretical approximations for the correlation between these size estimates, and between their associated spectral estimates, as functions of data acquisition and processing parameters, where a Gaussian spatial autocorrelation function is assumed to adequately model scatterer shape. Size results exhibit a fair degree of agreement with those of simulation experiments, while spectral results compare favorably with simulation outcomes. Utilization of the theoretical correlation expressions for data acquisition and processing optimization is discussed. Further simplifying approximations, such as the invariance of phase and amplitude terms with rotation angle, are made in order to obtain closed-form solutions to the derived spectral correlation, and permit an analytical optimization analysis. Results indicate that recommended parameter adjustments for performance improvement depend upon whether, for the system under consideration, the primary source of estimate decorrelation with rotation is scatterer phase change or field separation. [Work supported by NIH T32CA09206.]

3:00

2pBB9. Factors affecting scatterer size estimation using a generalized ultrasound attenuation–compensation function to correct for focusing. Timothy A. Bigelow and William D. O'Brien, Jr. (Univ. of Illinois at Urbana-Champaign, 405 N. Mathews Ave., Urbana, IL 61801, bigelow@uiuc.edu)

Over the years many different investigators have attempted to estimate the size of the principle scattering sites in materials by hypothesizing a model for the scattering and then fitting the power spectrum of the backscattered ultrasound signal to the model to obtain an estimate for the scatterer size. Traditionally, these models have assumed that a plane wave is incident on the scattering site limiting the measurement technique to unfocused or weakly focused ultrasound sources. In this investigation, the plane wave assumption was replaced by a model that assumed that the field of a focused source could be modeled as a three-dimensional Gaussian about the focus. Then, strongly focused sources could be used to obtain estimates for the size of the principle scattering sites provided that a generalized attenuation–compensation function was used to correct for the focusing. The improvement provided by the generalized attenuation–compensation function over traditional attenuation–compensation functions that neglected focusing was strongly affected by the type of scatterer being estimated. Also, the wavelength dependence of the axial beamwidth impacted how the improvement provided by the generalized attenuation–compensation function varied with increased focusing (i.e., going from f/1 to f/4). [Work supported by NDSEG Graduate Fellowship and Beckman Institute Graduate Fellowship.]

3:15

2pBB10. Characterization and differentiation of two mammary tumors using parametric imaging with ultrasound. Michael L. Oelze, William D. O'Brien, Jr. (Dept. of Elec. and Computer Eng., Univ. of Illinois, 405 N. Mathews, Urbana, IL 61801, oelze@uiuc.edu), and James F. Zachary (Univ. of Illinois, Urbana, IL 61801)

Two kinds of solid tumors were acquired and scanned *in vivo* ultrasonically. The first tumor series (fibroadenoma) was acquired from tumors that developed spontaneously in rats. The second tumor series was acquired by culturing a carcinoma cell line (4T1-MMT) and injecting the cells into Balb/c mice. The scatterer properties (average scatterer diameter and acoustic concentration) were estimated using a Gaussian form factor from the backscattered ultrasound measured from both kinds of tumors. Parametric images of tumors were constructed utilizing estimated scatterer properties for regions of interest inside the tumors and surrounding normal tissues. The average scatterer diameter and acoustic concentration for the fibroadenomas were estimated at 107 ± 14 micrometers and 15.2 ± 5 dB (mm^{-3}), respectively. The average scatterer diameter and acoustic concentration for the carcinomas was estimated at 30 ± 4.6 micrometers and 10.3 ± 6.9 dB (mm^{-3}), respectively. A comparison with light microscopic evaluations of the fibroadenomas showed cellular structures around 100 micrometers in size, and carcinomas showed cell nuclei with an average size of 12.5 micrometers in diameter (the total cellular size ranging from 50% to 200% larger than the nucleus size). [Work supported by NIH F32 CA96419 to MLO and by the University of Illinois Research Board.]

TUESDAY AFTERNOON, 11 NOVEMBER 2003 SAN ANTONIO ROOM, 1:00 TO 2:30 P.M.

Session 2pMUa

Musical Acoustics and Physical Acoustics: Wind Instruments and Nonlinearity II

Peter L. Hoekje, Chair
Department of Physics and Astronomy, Baldwin-Wallace College, Berea, Ohio 44017

Contributed Papers

1:00

2pMUa1. Modes of vibration of air-driven free reeds in transient and steady state oscillation. Ammon Paquette (Augustana College, Rock Island, IL 61201, ammon-paquette@augustana.edu), Justin Vines (Univ. of Arkansas, Fayetteville, AR 72701), and James P. Cottingham (Coe College, Cedar Rapids, IA 52402)

Most treatments of free reed oscillation approximate the reed vibration as a sinusoidal oscillation of a cantilever beam in the fundamental transverse mode, although some evidence of the presence of the second transverse mode has been reported. [Cottingham *et al.*, J. Acoust. Soc. Am. **105**, 940 (1999)]. Some new measurements of the oscillation of a free reed from an American reed organ mounted on a laboratory wind chest show that the second beam mode is present even at low amplitudes of oscillation, and is often observable in the transient period before the oscillation reaches full amplitude. Some evidence of higher frequency modes has also been observed. In addition to steady state oscillation, reed motion during two types of attack transients has been studied. In one case, with full playing pressure in the wind chest and air flowing through the reed, the reed is restrained in its unblown equilibrium position and suddenly released. In another configuration, the reed is provided with a pallet valve mechanism, and reed oscillation is initiated by a sudden rush of air when the valve is opened. [Work supported by the NSF from REU Grant No. 0139096.]

1:15

2pMUa2. Vibrational modes of the reed in a reed organ pipe. T. M. Huber, B. A. Collins (Dept. of Phys., Gustavus Adolphus College, 800 College Ave., St. Peter, MN 56082, huber@gustavus.edu), M. Pineda (Polytec PI, Inc., Tustin, CA 92780), and C. Hendrickson (Hendrickson Organ Co., St. Peter, MN 56082)

We will describe a series of measurements of the vibrational modes of the reed in a reed organ pipe. These measurements were performed using a Polytec PSV-300 scanning vibrometer, which allows the vibrational deflection shape to be determined at any frequency. In addition to blowing the reed pipe in a standard fashion, a mechanical driver was used to excite the reed. Using both excitation sources, a number of deflection shapes were observed including simple cantilever, torsional, and higher-order shapes corresponding to higher-order cantilever and torsional modes. As expected, the observed frequencies of the mechanically driven modes were not integer multiples of the fundamental, and were consistent with theoretical predictions. The reed pipe was also excited in a standard manner using an organ blower. This raised integer multiples of the fundamental frequency as high as 20 kHz within two decades of the velocity amplitude of the fundamental. Torsional and other deflection shapes were present, however nonlinear interactions in the system caused them to be shifted from their mechanically driven frequencies. In some cases, significant vibration was observed in the tuning wire and the section of reed above the tuning wire that was previously considered to be clamped.

1:30

2pMUa3. An inward striking free reed coupled to a cylindrical pipe. Justin Vines (Univ. of Arkansas, Fayetteville, AR 72701, jvines@uark.edu), Ammon Paquette (Augustana College, Rock Island, IL 61201), and James P. Cottingham (Coe College, Cedar Rapids, IA 52402)

A number of acoustical measurements have made on a reed-pipe combination consisting of a harmonium-type reed from an American reed organ installed at the closed end of a cylindrical pipe. This configuration, which somewhat resembles the configuration of free-reed organ pipes, differs from the reed-pipe combination occurring in the mouth organs of Asia, which use symmetric (outward striking) free reeds and normally operate on both possible directions of airflow. Measurements have been made of the sounding frequency, amplitude of vibration of the reed tongue, and the sound spectrum. Of particular interest is the degree to which the reed frequency can be altered by altering the pipe length, and hence the pipe resonance frequency. In this case the sounding frequency can be pulled considerably below the natural frequency of the reed. These results can be compared with the results of similar measurements on free-reed organ pipes [J. Braasch, C. Ahrens, J. P. Cottingham, and T. D. Rossing, Fortschr. Akust., DAGA (2000)]. In addition, some interesting ''special effects'' have been studied, which can be obtained using unusual pipe lengths and blowing in the ''wrong'' direction. [Work supported by the NSF from REU Grant No. 0139096.]

1:45

2pMUa4. Direct measurement of clarinet air column oscillations. Jesse JonesIV, Chris Rogers (Mech. Eng. Dept., Tufts Univ., 200 College Ave., Medford, MA 02155), and Chris French (The Selmer Co., Elkhart, IN 46514)

The internal oscillation of a clarinet air column has been directly measured through the implementation of hot-wire anemometry. By taking a series of measurements down the centerline of the bore, velocity and pressure modal shapes of individual harmonics are separated, measured, and plotted. Finally, composite averaged power spectra of the internal oscillation are presented and compared to acoustic measurements acquired outside the clarinet. In many cases, the even harmonics of the internal oscillation dominate over the power found in the odd harmonics. This contradicts the classic model of the clarinet as a cylindrical pipe closed at one end and open at the other (where only odd harmonics are produced). Further, the data from the direct velocity measurements also contradict the externally acquired acoustic data, where odd harmonics generally dominate for the lowest 5–9 harmonics. Thus the clarinet, in theory and practice, is generally considered incapable of generating strong even harmonics. In this research, however, it is seen that dominate even harmonics are generated, but the energy for these frequencies is largely trapped inside the clarinet, whereas the energy associated with the odd harmonics is released to the ambient. [This research was conducted with the support of Selmer Musical Instruments.]

2:00

2pMUa5. Airflow patterns in the vicinity of an air-driven free reed. Jesse T. Jensen and James P. Cottingham (Phys. Dept., Coe College, Cedar Rapids, IA 52402, jtjensen@coe.edu)

Free reed instruments are characterized by high-volume airflow rates through the oscillating reeds. Measurements have been made of the average volume flow rate through a single American organ reed as a function of blowing pressure. In addition, measurements of average air velocity have been made at a grid of points close to the vibrating air-driven reed, mounted on the surface of a laboratory wind chest. These measurements supplement earlier investigations, which explored relations among the reed motion, airflow, and acoustic pressure associated with the vibration of free reeds. [M. Busha and J. P. Cottingham, J. Acoust. Soc. Am. **111**, 2376 (2002)]. An attempt has been made to determine whether the left–right asymmetry due to the asymmetric shape (twist) of the reed is reflected in the airflow pattern. A graphing program has been developed so that the airflow pattern can be visualized in two or three dimensions.

2:15

2pMUa6. The acoustic effect of cryogenically treating trumpets. Jesse JonesIV and Chris Rogers (Dept. of Mech. Eng., Tufts Univ., 200 Anderson Hall, 200 College Ave., Medford, MA 02155)

The acoustic effect of cryogenically treating trumpets is investigated. Ten Vincent Bach Stradivarious $B\flat$ trumpets are studied, half of which have been cryogenically treated. The trumpets were played by six players of varying proficiency, with sound samples being recorded directly to disk at a sampling rate of 44.1 kHz. Both the steady-state and initial transient portions of the audio samples are analyzed. When comparing the average power spectra of the treated trumpets to the untreated set, no repeatable, statistically independent differences are observed in the data. Differences observed in player-to-player and trumpet-to-trumpet comparisons overshadow any differences that may have been brought on due to the cryogenic treatment. Qualitatively, players established no clear preference between the treated and untreated trumpets regarding tone and playability, and could not differentiate between the two sets of instruments. All data was collected in a double blind fashion. The treatment itself is a three step process, involving an 8 hour linear cool down period, a 10 hour period of sustained exposure to $-195\,^{\circ}\mathrm{C}$ ($-300\,^{\circ}\mathrm{F}$), and a 20–25 hour period of warming back to room temperature. [Work was completed with the support of Steinway & Sons Pianos and Selmer Musical Instruments.]

TUESDAY AFTERNOON, 11 NOVEMBER 2003 SAN ANTONIO ROOM, 2:45 TO 4:25 P.M.

Session 2pMUb

Musical Acoustics: Where Are They Now? Current Research by Past Student Paper Award Winners

James P. Cottingham, Chair

Physics Department, Coe College, Cedar Rapids, Iowa 52402

Invited Papers

2:45

2pMUb1. Where they are now: An overview. James P. Cottingham (Dept. of Phys., Coe College, Cedar Rapids, IA 52402, jcotting@coe.edu)

Since 1997 there have been fifteen award winners in the Best Student Paper competition in Musical Acoustics. Although some of the award winners are still active in musical acoustics, others are now active in other areas of acoustics or in fields outside acoustics altogether. A brief statistical overview will be presented of the history of the competition and past and current interests of those who have been the award winners. Capsule updates on several award winners who are unable to participate in the session will be presented.

3:10

2pMUb2. Sound radiation and phase mapping of Caribbean steelpans. Andrew Morrison (Dept. of Phys., Northern Illinois Univ., DeKalb, IL 60115)

At the 145th ASA meeting [J. Acoust. Soc. Am. **113**, 2315 (2003)] modal analysis and sound intensity plots of two Caribbean steelpans were presented. Since that meeting the relationship of pan dimension to radiated sound has been examined. Sound radiation was explored using a two-microphone probe to gather sound intensity measurements. Preliminary phase maps of vibrating note areas have also been obtained using holographic interferometry with phase imaging.

3:35

2pMUb3. Future evolution of structure in an accelerating universe. Michael T. Busha (Randall Lab., Univ. of Michigan, 500 E. University Ave., Ann Arbor, MI 48109)

Current cosmological data have confirmed the big bang and dark matter hypothesies, and indicate that our universe contains a substantial component of dark vacuum energy that is driving the cosmos to accelerate. We examine the immediate and long-term consequences of this dark energy on large scale structures formed in the universe. Using spherical model solutions and realistic 3D clustering simulations, we present criteria for test bodies to remain bound to existing structures. We show that collapsed halos become spatially isolated and dynamically relax to a particular kinematic profile. From this state, the asymptotic form of the space-time metric is well specified.

4:00

2pMUb4. Three-dimensional sound signals and their relevance to wave energy quantities and sound interference products. Pantelis Vassilakis[a] (Music, De Paul Univ., 2350 N. Kenmore Ave., JTR 307, Chicago, IL 60614, pantelis@acousticslab.com)

Signals are graphic representations of vibrations/waves and, like every representation, capture only selected attributes of the phenomenon they are meant to represent. The often assumed equivalence between signals and sound waves obscures the fact that two-dimensional signals are not fit to (a) represent wave-energy quantities consistently across frequencies, (b) account for the alternating positive/negative amplitude values of modulated waves with AM-depth$>$100%, and (c) represent the energy content of interference. An alternative sound-signal representation is proposed, based on the complex equation of motion describing a wave. It results in spiral sine signals and twisted-spiral complex signals, similar to complex analytic signals. Spiral sine signals offer a consistent measure of sine-wave energy across frequencies, while twisted spiral complex signals account for the negative amplitudes observed in modulated signals and map the modulation parameters onto the twisting parameters. In terms of interference, 3-D signals illustrate that amplitude fluctuations and the signal envelopes that describe them are not just boundary curves but waves that trace changes in the total instantaneous energy of a signal over time, representing the oscillation between potential and kinetic energies within a wave. Examples of 3-D animations illustrating the proposed signals are presented. [a] Work completed while at the Department of Ethnomusicology, University of California, Los Angeles.

TUESDAY AFTERNOON, 11 NOVEMBER 2003 WEDGWOOD ROOM, 5:00 TO 5:50 P.M.

Session 2pMUc

Musical Acoustics: Performance Session: Concert Featuring Wind Players from the University of Texas

James P. Cottingham, Chair

Physics Department, Coe College, Cedar Rapids, Iowa 52402

Chair's Introduction—5:00

This performance technical session will feature chamber music for winds played by musicians selected from the wind ensembles of the University of Texas School of Music. The program has been organized by Dr. Scott S. Hanna. Dr. Hanna serves as Assistant Director of Bands at the university as well as Associate Director of the Longhorn Band and Music Director of the UT Chamber Winds.

Session 2pNSa

Noise and Engineering Acoustics: Sound Quality for Engineered Products and the Environment

Brigitte Schulte-Fortkamp, Cochair
Institute of Technical Acoustics, Technical University Berlin, Einsteinufer 25, 10587 Berlin, Germany

Klaus Genuit, Cochair
HEAD Acoustics GmbH, Eberstrasse 30a, 52134 Herzogenrath, Germany

Patricia Davies, Cochair
Ray W. Herrick Laboratories, Purdue University, West Lafayette, Indiana 47907-2031

Chair's Introduction—1:00

Invited Papers

1:05

2pNSa1. Speech intelligibility index predictions for young and old listeners in automobile noise: Can the index be improved by incorporating factors other than absolute threshold? Meghan Saweikis, Aimée M. Surprenant (Dept. of Psychol. Sci., Purdue Univ., 703 3rd St., West Lafayette, IN 47907), Patricia Davies, and Don Gallant (Purdue Univ., West Lafayette, IN 47907-2031)

While young and old subjects with comparable audiograms tend to perform comparably on speech recognition tasks in quiet environments, the older subjects have more difficulty than the younger subjects with recognition tasks in degraded listening conditions. This suggests that factors other than an absolute threshold may account for some of the difficulty older listeners have on recognition tasks in noisy environments. Many metrics, including the Speech Intelligibility Index (SII), used to measure speech intelligibility, only consider an absolute threshold when accounting for age related hearing loss. Therefore these metrics tend to overestimate the performance for elderly listeners in noisy environments [Tobias *et al.*, J. Acoust. Soc. Am. **83**, 859–895 (1988)]. The present studies examine the predictive capabilities of the SII in an environment with automobile noise present. This is of interest because people's evaluation of the automobile interior sound is closely linked to their ability to carry on conversations with their fellow passengers. The four studies examine whether, for subjects with age related hearing loss, the accuracy of the SII can be improved by incorporating factors other than an absolute threshold into the model. [Work supported by Ford Motor Company.]

1:25

2pNSa2. Prediction of booming sensation and its difference limen for just noticeable change in frequency. Sung-Hwan Shin and Jeong-Guon Ih (ME3075, NOVIC, Dept. of Mech. Eng., KAIST, Taejon, Korea)

Among many auditory feelings for the car interior noise, the booming sensation is considered the most important nuisance to the passengers. Although there are many origins for the booming noise of vehicles in general, the most important one is the engine boom that consists of tonal components related to fundamental engine rotation and its harmonics including the firing frequency. Because the degree of booming sensation is increased when these tonal components are dominating in car interior noise, it is demanded to extract the aurally relevant tonal components only. Although the pitch extraction model based on the place theory enables to find aurally relevant tonal components, there is a difference between booming sensation and pitch perception according to a frequency change of the tonal component. In this study, a subjective listening test using a tracking method is performed to find the difference limen for just a noticeable change of booming sensation in frequency. By applying the resultant data and also the empirical data by Zwicker, the existing pitch extraction model is modified. This refined model and loudness analysis can be used for predicting the degree of booming sensation. [Work supported by the BK21 project and NRL.]

1:45

2pNSa3. Sound quality evaluation of air conditioning sound rating metric. Kathleen K. Hodgdon, Jonathan A. Peters, Russell C. Burkhardt (Appl. Res. Lab, Penn State Univ., University Park, PA 16802), Anthony A. Atchley, and Ingrid M. Blood (Penn State Univ., University Park, PA 16802)

A product's success can depend on its acoustic signature as much as on the product's performance. The consumer's perception can strongly influence their satisfaction with and confidence in the product. A metric that can rate the content of the spectrum, and predict its consumer preference, is a valuable tool for manufacturers. The current method of assessing acoustic signatures from residential air conditioning units is defined in the Air Conditioning and Refrigeration Institute (ARI 270) 1995 Standard for Sound Rating of Outdoor Unitary Equipment. The ARI 270 metric, and modified versions of that metric, were implemented in software with the flexibility to modify the features applied. Numerous product signatures were analyzed to generate a set of synthesized spectra that targeted spectral configurations that challenged the metric's abilities. A subjective jury evaluation was conducted to establish the consumer preference for those spectra. Statistical correlations were conducted to assess the degree of relationship between the subjective preferences and the various metric calculations. Recommendations were made for modifications to improve the current metric's ability to predict subjective preference. [Research supported by the Air Conditioning and Refrigeration Institute.]

2:05

2pNSa4. Methodology for quantifying the tonal prominence of frequency modulated tones. Kyoung Hoon Lee, Patricia Davies (Ray W. Herrick Labs., Purdue Univ., 140 S. Intramural Dr., West Lafayette, IN 47907-2031), and Aimee Surprenant (Purdue Univ., West Lafayette, IN 47907)

The paper is focused on research conducted to develop a metric to quantify the tonal prominence of frequency modulated tones in noise. Frequency modulated tones are commonly encountered in machinery noise because of RPM variations caused by changing loads and poor control in timing and repeatability of combustion events in engines, particularly diesel engines. When tonal components are noticeable they increase annoyance; thus it is important to quantify both their strength and the resulting contribution to annoyance. In preliminary work on tones with randomly varying frequencies, it was found that it is possible to remove the trackable portion of the frequency variation and use established tonal prominence metrics (Aures' Tonality and Tone-to-Noise Ratio) on the modified signal to predict the perceived tonalness of the sounds (Lee, Hastings, Davies and Surprenant, Internoise 2003). The work presented is an extension of this earlier work: deterministic frequency variations are studied along with different levels of background noise. Modifications to the procedure to determine the strength of the tonal feature are described, as well as the issues that must be addressed before using this approach to analyze more complex machinery sounds. [Work sponsored by Caterpillar Inc.]

2:25

2pNSa5. A loudness model in dichotic conditions. Jeong-Guon Ih and Jeong-Ho Cha (Dept. of Mech. Eng., KAIST, Science Town, Taejon 305-701, Korea, ihih@sorak.kaist.ac.kr)

Existing loudness models are specified only to diotic sounds in spite of the fact that normal human beings hear dichotic sounds. The arithmetic mean of loudness values of both ear signals has been suggested for the approximated value of the resultant perceived loudness. In this study, the dependence of overall loudness perception on the interaural level differences was investigated by the subjective tests. It was found that the larger the interaural level difference, the louder the perception than the mean of calculated loudness values at both ears and the lower the critical band rate or the reference level, the louder the perception than the mean value. A modified loudness model was proposed for the application to dichotic sounds by using the equivalent diotic levels.

2:45–3:00 Break

3:00

2pNSa6. Sound preferences in urban open public spaces. Jian Kang and Wei Yang (School of Architecture, Sheffield Univ., Western Bank, Sheffield S10 2TN, UK, j.kang@sheffield.ac.uk)

This paper studies people's perception of sound, based on an intensive questionnaire survey in fourteen urban open public spaces of five European countries. The questionnaire includes identification of recognized sounds, classification of sound preference, and indication of wanted and unwanted sounds. The results indicate three facets to people's sound preferences. First, people generally prefer natural and culture-related sounds rather than artificial sounds. Vehicle sounds and construction sounds are regarded as the most unpopular, whereas sounds from human activities are normally rated as neutral. Second, cultural background and long-term environmental experience play an important role in people's judgment of sound preference. People from a similar environment may show a similar tendency on their sound preferences, which can be defined as macro-preference. Third, personal differences, such as age and gender, further influence people's sound preference, which can be defined as micro-preference. For example, with increasing age, a higher percentage of people are favorable to, or tolerate, sounds relating to nature, culture or human activities. Male and female exhibit only slight differences. [Work supported by the European Commission.]

3:20

2pNSa7. Environmental noise—a challenge for an acoustical engineer. Klaus Genuit (HEAD acoustics GmbH, Ebertstrae 30a, D-52134 Herzogenrath, Germany, klaus.genuit@head-acoustics.de)

People live in a landscape full of noises which are composed of both natural environmental noises and technically created sounds. Regarding environmental noise, more and more people feel heavily annoyed by noises. Noise is defined as an audible sound which either disturbs the silence or an intentional sound listening or leads to annoyance. Thus, it is clearly defined that the assignment of noise cannot be reduced to simple determining objective parameters such as the A-weighted sound pressure level or the equivalent continuous sound pressure level. The question of whether a sound is judged as noise can only be made after the transformation from the sound event into an auditory event has been accomplished. The evaluation of noise depends on the physical characteristics of the sound event, on the psycho-acoustical features of the human ear, as well as on the psychological aspects of man. For the acoustical design of environmental noise and in order to create a better soundscape the acoustical engineer has to consider these aspects. That means a specific challenge for the sound engineering.

3:40

2pNSa8. (Re-)constructed reality and context—When sound quality becomes soundscape. Brigitte Schulte-Fortkamp (ITA/TU–Berlin, Einsteinufer 25, 10587 Berlin, Germany)

In the interior of a vehicle sound quality is a brandname. Bi-aurale measurement and analysis technology is more or less standard, but there are still no general standards and parameters for sound quality. Evaluation of noises is highly sensitive to context, but testing procedures that can include this realization are exceedingly rare. Soundscapes are specific constellations of a general noise volume in clearly defined ambiences; they even can characterize those ambiences and combine the daily recurrent patterns of sound multifactorial in the process of analysis. The evaluation of noise situations gets interactively modified by their significance for the general residential space. It is contingent upon the respective ponderation of acoustic and nonacoustic modifiers which in turn constitute the soundscape.

Contributed Papers

4:00

2pNSa9. Sound quality descriptors for HVAC equipment from ARI Standards. Stephen J. Lind (Trane, Bldg. 12-1, 3600 Pammel Creek Rd., La Crosse, WI 54601, slind@trane.com)

The Air Conditioning and Refrigeration Institute (ARI) has several standards that provide methods to evaluate the sound quality of heating ventilating and air-conditioning (HVAC) equipment. These include Standard 270 Sound rating of outdoor unitary equipment, Standard 350 Sound rating of non-ducted indoor air-conditioning equipment, and Standard 1140P Procedures for evaluating sound quality of HVAC equipment. The preferred method in these standards is best described in Standard 1140P, which uses one-third octave band sound power levels that are weighted to adjust for the sensitivity to frequency distribution and presence of tones, and are then converted to a single number sound quality indicator. The tone adjustment is based on the projection of a given one-third octave band level relative to the average of the adjacent one-third octave bands. An alternate use of Zwicker method B to determine loudness and loudness level is also provided in ARI Standard 1140P. These standards provide a convenient method by which complex sounds for similar products may be compared.

4:15

2pNSa10. Traffic noise attenuation by scattering, resonance and dispersion. Hasson M. Tavossi (Dept. of Physical and Environ. Sci., Mesa State College, School of Math. & Natural Sci., 1100 North Ave., Grand Junction, CO 81501)

The purpose of this investigation is to analyze various techniques of sound attenuation which can be used to minimize traffic noise from heavy traffic circulation in urban areas, and to improve sound quality in these areas using selective absorption in specific frequency ranges, plus to reduce the degree of annoyance of traffic noise. This study covers uncontrollable noise sources such as trucks, automobiles and trains, in heavy traffic urban areas. Different methods of noise energy reduction are discussed, including sound attenuation by selective absorption, scattering and cavity resonance in the sound barriers. Multiple scattering on sound absorbing road and barriers are considered to increase noise energy dissipation in unwanted frequency ranges. The effects of dispersion and change in the frequency content of the noise with distance are studied. An elimination of the most disturbing frequency components of these uncontrollable noise sources improves sound quality from these sources and increases our overall capabilities in controlling noise pollution in the urban environment.

TUESDAY AFTERNOON, 11 NOVEMBER 2003 CONCHO ROOM, 1:30 TO 3:55 P.M.

Session 2pNSb

Noise, Architectural Acoustics and ASA Committee on Standards: Building Code Noise Compliance

Daniel R. Raichel, Chair
2727 Moore Lane, Fort Collins, Colorado 80526-2192

Chair's Introduction—1:30

Invited Papers

1:35

2pNSb1. Model building codes and acoustical performance: Where are we in 2003? Brandon Tinianov (Johns Manville Tech. Ctr., 10100 W. Ute Ave., Littleton, CO 80127)

The proper acoustical design for multi-family dwellings is an important factor in occupant comfort. Key acoustical design practices are often not mandated by the builder or architect, but by the applicable building codes. In early 2003, the three regional/national building codes agreed to join into a single, unified national building code for residential and commercial construction. The scope and governance of these three codes: the Uniform Building Code (ICBO), the National Building Code (BOCA), the Southern Building Code (SBCCI) are reflected in the International Residential Code (IRC) and the International Building Code (IBC) which was developed by the International Code Council (ICC). With the move to a single code body, those concerned with building acoustical performance welcome the benefit of a single minimum standard. Unfortunately, this new minimum performance requirement does not reflect the state of the science for occupant satisfaction. The acoustical requirements of each of these building codes, the timeline of their development and an overview of the state of the science will be presented. Suggestions for revised performance minimums will also be offered for discussion.

1:55

2pNSb2. Field normalization techniques and practices for determining sound isolation and impact insulation. John J. LoVerde and David W. Dong (Veneklasen Assoc., 1711 Sixteenth St., Santa Monica, CA 90404, jloverde@veneklasen-assoc.com)

The International Building Code (IBC) includes minimum acceptable acoustical performance for floor/ceiling assemblies separating dwelling units. Field tests for determining the Normalized Noise Isolation Class (NNIC) and Field Impact Insulation Class (FIIC) of an assembly require the measurement of the decay rate of the receiving room and the normalization of the sound pressure levels, as set forth in ASTM standards E336 and E1007. The normalization is intended to improve comparability of test results by removing the effects of receiving room absorption. Surprisingly, the normalization procedure is substantially different between these two types of tests. The FIIC normalization is based on the laboratory Impact Insulation Class (IIC) test using a standard room absorption of 10 metric sabins. The NNIC normalization is based on a 0.5-s reverberation time, which is assumed normal in a furnished habitable dwelling space. This difference in normalization procedure for the same receiving room potentially yields different results of NNIC and FIIC as compared with their respective non-normalized values. This will be illustrated with examples of actual field tests. The effectiveness and suitability of the normalization procedures in the field setting are compared and evaluated including the use of the California Building Code modification to FIIC.

2:15

2pNSb3. ANSI S12.60 (2002) testing of modular buildings used as classrooms. Todd A. Busch (Acentech, 1429 E. Thousand Oaks Blvd., Ste. 200, Thousand Oaks, CA 91362)

The acoustics of modular buildings used as classrooms may come under greater scrutiny if schools adopt American National Standards Institute (ANSI) S12.60 (2002) ''Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools'' as required thresholds of performance. To quantify the benefits of series of structural improvements to buildings having an interior volume of less than 10 000 ft^3, both the reverberation time and background noise were measured. Interior spaces were unoccupied as per S12.60, but unfurnished. Reverberation times were 0.4–0.6 s. The background noise level was measured to identify the contributions of exterior sources, interior fluorescent lighting, and AC/Fan systems operating with/without a plenum for the return vent. Targets for the stated room volume are 35 dBA and 55 dBC. For the final configuration, interior noise levels with lights, fan, and cooling active were 43 dBA and 67 dBC; interior noise levels due to exterior sources were 33 dBA and 58 dBC due in part to the surrounding industrial land uses and nearby freeway. For one configuration, with the fan active without AC, pressurized air escaping through a 1/4-in. gap beneath the 36-in.-wide door produced a distinct 35 dBA tone in the 630-Hz third-octave band.

2:35

2pNSb4. Noise codes for residential spaces in Los Angeles. Daniel R. Raichel (Eilar & Assoc., 2727 Moore Ln., Fort Collins, CO 80526, raichel@juno.com)

The State of California is rather unique in using the community noise equivalent level (CNEL) rating to assess the amount of environmental noise transmission into buildings, not just to apply this index to gauge airport noise. When residential housing units in Los Angeles undergo construction or renovation, a certificate of housing compliance must be obtained before tenants or condominium owners are allowed to move in, in accordance with the provisions of Section 91.0318 of the Los Angeles municipal code. As a part of the application procedure for the certificate, a licensed acoustical engineer must prepare an acoustical report that describes the type of construction between dwelling units and the general sound attenuating properties of such construction, including hard numbers for sound attenuation between dwelling units. A field sound transmission class (FSTC) test is mandated, with a minimum required FSTC rating of 45 between the walls of adjacent units. If the FSTC test is not passed, remedial action must be taken so that the proper amount of acoustic insulation is achieved between the dwelling units.

2:55–3:05 Break

3:05

2pNSb5. Evaluation of acoustical parameters of a Brazilian popular housing model. Jos A. C. Ferreira (Acoust. Lab., Dept. of Mech. Eng., UFPR, 81531-990, Curitiba, PR, Brazil), Fabiano B. Diniz, Andressa M. C. Ferreira, and Paulo T. Zannin (Acoust. Lab., UFPR)

This article presents the results obtained from the evaluation of the acoustical insulation parameters determined *in situ* in a popular residence projected to offer an option to combat the housing deficit of the low income Brazilian population. This evaluation has been carried out according to the statements of the standards ISO 140-4 and 140-5, which state about this type of measurement. The results have shown that the surveyed house presents a satisfactory performance if compared to the standard of the Brazilian civil construction, but it is not adequate if compared to the demands of the international standards.

3:25

2pNSb6. Residential reuse: Acoustics and making buildings livable. Jesse J. Ehnert (Arpeggio Acoust. Consulting, LLC, 1947 Aspen Dr. NE, Atlanta, GA 30345, jehnert@arpeggioacoustics.com)

Adapting an existing facility, whether it is industrial, educational, or commercial in nature, to residential uses, such as condominiums and lofts, poses many challenges to the design team. Among other things, the process must include maintaining historic character while addressing current code issues in the context of becoming familiar with and utilizing existing architecture and building systems, which may not always be well-documented or obvious. Of the many unique issues encountered in such adaptive reuse projects, acoustic concerns often present themselves, especially in terms of sound and vibration isolation and noise control within and between residential units. Some of the paramount concerns and pitfalls will be discussed in the context of relevant examples and past experiences.

3:40

2pNSb7. Compressor noise control with a two-dimensional enclosure. Ballard W. George (Environtech Consultants, 1367 Bobolink Circle, Sunnyvale, CA 94087, kingstaco@yahoo.com)

This paper is concerned with the use of a two-dimensional enclosure with masonry block wall to mitigate compressor noise in a northern California location for consistency with city standards. Sensitive receptors consist of a duplex and single residences across the street from the source. Background sound is provided primarily by a nearby highway that is elevated. Noise reduction is accomplished by the barrier effect of the wall along with sound absorption realized from a pegboard wall surface in the vicinity of the compressor, unpainted concrete block surface and acoustic treatment on the doors. Design considerations for the pegboard are discussed with the intention of balancing the overall low- and high-frequency noise reduction from the absorption and from the barrier effect and with the intention of reasonably reducing the number of pegboard holes required. Some considerations from the standpoint of ray acoustics are discussed. A complicating factor for the interior acoustics consists of three full height tanks in a row that serve as sound diffusers. Comments on vibration isolation for equipment and pipes are offered.

TUESDAY AFTERNOON, 11 NOVEMBER 2003 TRINITY B ROOM, 1:30 TO 5:00 P.M.

Session 2pPA

Physical Acoustics: Nonlinear Acoustics and Scattering

Kendall R. Waters, Chair
NIST, Materials Reliability Division, 325 Broadway, Boulder, Colorado 80305

Contributed Papers

1:30

2pPA1. Calculation of the acoustic nonlinearity parameter B/A for four linear alkanes by Lee–Kesler equation of state. Zhiqu Lu (Natl. Ctr. for Physical Acoust., The Univ. of Mississippi, Coliseum Dr., University, MS 38677)

The acoustic nonlinearity parameter B/A has been calculated for four linear alkanes over the temperature range from 273.15 to 473.15 K and the pressure range from 1 atm to 1000 atm by using the Lee–Kesler equation of state. The calculated results are in agreement with the experimental data in the literature. It is found that under low pressure (1 atm and 20 atm), B/A increases linearly with temperature. Under high pressure, the value of B/A increases with temperature first, however, at a slower rate of rise. After a broad maximum, the ratio B/A decreases slowly as the temperature increases. At the fixed temperature, the B/A value decreases with the number of carbon chain of alkanes. It is shown that the B/A decreases monotonically with pressure over the whole temperature and pressure ranges.

1:45

2pPA2. Nonlinear acoustic acceleration waves in porous media flow. Pedro M. Jordan (Naval Res. Lab., Code 7181, Stennis Space Center, MS 39529, pjordan@nrlssc.navy.mil)

Acoustic acceleration waves are defined as jumps in the first derivatives of the velocity, pressure, or density across a propagating singular surface (or wave front). In this talk, the temporal evolution of the amplitude and the propagation speed of such waves are investigated in the context of finite-amplitude acoustic propagation in inelastic Darcy-type porous media. It is shown that there exists a critical value, $\alpha^*(>0)$, of the initial jump amplitude such that the acceleration wave magnitude either goes to zero, as $t\to\infty$, or infinity, in finite time, depending on whether the given initial jump amplitude is greater than or less than $-\alpha^*$. In addition, a connection to traveling wave solutions is noted and the linearized case is examined. Finally, the numerical solution of a one-dimensional, nonlinear IBVP involving sinusoidal signaling in a fluid-saturated porous slab is used to illustrate the finite-time transition from an acceleration wave to a shock wave that occurs when the given initial jump amplitude is less than $-\alpha^*$. [Work supported by ONR/NRL funding.]

2:00

2pPA3. Feasibility of nonlinear acoustic technique with ultrasonic guided waves for nondestructive defect evaluation. Kyung-il Jung and Suk Wang Yoon (Dept. of Phys., SungKyunKwan Univ., Suwon 440-746, Republic of Korea)

Feasibility of nonlinear acoustic modulation technique [K. I. Jung and S. W. Yoon, Proceedings of Internoise 2000, Nice, France, p. 205 (2000)] was discussed for nondestructive defect evaluation. Such nonlinear acoustic technique can be applied to localize a defect in a solid rod and also extended to layer-structured media. In this study, nonlinear acoustic technique with ultrasonic guided waves is introduced. Nonlinear acoustic re-

sponses from a defect are observed with guided waves in a 2-D structure. The defect was oriented in the direction of guided wave propagation. The experimental results show that the nonlinear acoustic technique with ultrasonic guided waves seems feasible for nondestructive evaluation of 2-D structure. [Work supported by BK21 Program in Korea and by the Basic Research Program of KOSEF (R01-2000-000-00014-0).]

2:15

2pPA4. Nonlinear modal methods for crack localization. Alexander Sutin (Stevens Inst. of Technol., Hoboken, NJ 07030), Lev Ostrovsky (Zel Technologies/NOAA ETL, Boulder, CO 80305), and Andrey Lebedev (Inst. of Appl. Phys., Nizhnny Novgorod 603095, Russia)

A nonlinear method for locating defects in solid materials is discussed that is relevant to nonlinear modal tomography based on the signal cross-modulation. The scheme is illustrated by a theoretical model in which a thin plate or bar with a single crack is excited by a strong low-frequency wave and a high-frequency probing wave (ultrasound). A crack is considered as a small contact-type defect which does not perturb the modal structure of sound in linear approximation but creates combinational-frequency components whose amplitudes depend on their closeness to a resonance and crack position. Using different crack models, including the hysteretic ones, the nonlinear part of its volume variations under the given stress and then the combinational wave components in the bar can be determined. Evidently, their amplitude depends strongly on the crack position with respect to the peaks or nodes of the corresponding linear signals which can be used for localization of the crack position. Exciting the sample by sweeping ultrasound frequencies through several resonances (modes) reduces the ambiguity in the localization. Some aspects of inverse problem solution are also discussed, and preliminary experimental results are presented.

2:30

2pPA5. Parametric wave phase conjugation of nonlinear ultrasound waves. Andrew Brysev, Vladislav Mikhalevich, and Vladimir Streltsov (Wave Res. Ctr. of General Phys. Inst., RAS, 38 Vavilov str., 119991 Moscow, Russia, brysev@orc.ru)

Real time acoustic wave phase conjugation (WPC), based on parametric self-consistent physical mechanisms, was realized up to the present time only for the monochromatic waves [A. P. Brysev *et al.*, Phys.-Usp. **41**, 793 (1998)]. Here the possibility of WPC of nonmonochromatic ultrasound waves is considered. For simultaneous WPC of the entire series of spectral components generated by nonlinear propagation of the incident wave we propose the use of phonon-plasmon interaction in piezosemiconductors. WPC of nonlinear acoustic waves can be accomplished by modulation of the electron density provided by a sequence of short laser pulses pumping the sample. If the periodicity of the optical pulses is half the period of the fundamental component of the acoustic wave, such wide-band, excitation leads to self-synchronized parametric conjugation of each spectral component in the incident wave. The conjugation efficiency depends sharply on relations between acoustical frequency content, laser pulse duration, and interband relaxation time. It is shown that under certain conditions the time profile of the conjugate wave may be efficiently controlled by varying the duration of the laser pulses. The time profile of the conjugate wave is investigated for some physical conditions of practical interest.

2:45

2pPA6. Radiation force on a compressible cylinder in a standing wave: Long-wavelength approximation. Wei Wei, David B. Thiessen, and Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

For situations where compressible cylinders are manipulated using acoustic standing waves, it is desirable to have a simple long wavelength approximation of the radiation force analogous to the sphere result published by Yosioka and Kawasima in 1955. That result has also been derived from the Rayleigh scattering approximation of the monopole and dipole scattering contributions. The analogous result for a rigid-movable cylinder (that may have a different density than the surroundings) was given by the first line of Eq. (24) of Wu *et al.* [J. Wu, G. Du, S. S. Work, and D. M. Warshaw, J. Acoust. Soc. Am. **87**, 581–586 (1990)]. The cylinders axis is parallel to the standing wave nodal planes. To generalize that result to a compressible cylinder it is only necessary to replace a monopole term [the final unity term in the curly brackets in the first line of Eq. (24)] by $(1-R)$ where R denotes the ratio of the adiabatic compressibility of the inner fluid to that of the outer fluid. This modification agrees with the form of the monopole scattering contribution for cylinders in the Rayleigh approximation and recovers numerically the low frequency limit for the radiation force-per-length based on the full partial-wave series. [Work supported by NASA.]

3:00–3:15 Break

3:15

2pPA7. Acoustics spectral gaps and transmission in periodic stub tuners. M. S. Kushwaha (Inst. of Phys., Univ. of Puebla, Puebla, Mexico), P. Vasilopoulos, and X. F. Wang (Concordia Univ., Montreal, Canada)

Periodic binary systems can give rise to genuine acoustic stop bands within which sound and vibrations remain forbidden. Extensive band structure and transmission are computed for a periodically stubbed waveguide. In general the waveguide segments and stubs are made of different materials. The acoustic wave in such a system has two independent polarizations: out-of-plane and in-plane modes. The band structure and transmission spectrum are studied for diverse geometries using a simple and efficient version of the transfer-matrix method [Phys. Rev. B **65**, 035107 (2002)]. For the same material between the waveguide and symmetric stubs the width of some gaps can change, upon varying the stub length or width, by more than one order of magnitude. A further modulation can be achieved for different materials between the stubs and the main waveguide or if the stubs are asymmetric. The spectral gaps in the band structure of an infinitely long system correspond to those in the transmission spectrum of the same system but with a *finite* number n of units. For n finite (i) there exist *pseudogaps* that gradually turn into *complete* gaps with increasing n and (ii) the introduction of defects gives rise to states inside the gaps and leads to transmission resonances.

3:30

2pPA8. Acoustic propagation using the direct simulation Monte Carlo method. Amanda L. Danforth and Lyle N. Long (Penn State Univ., 202 Appl. Sci. Bldg., P.O. Box 30, State College, PA 16804, ald227@psu.edu)

In the simulation of fluid dynamics, one can either treat the fluid as a continuum or as discrete particles. Although popular for acoustics, the continuum model is limited to small Knudsen numbers (the ratio of mean free path to a length scale). Particle methods are necessary for, but not limited to, problems with Knudsen numbers greater than 0.1, which can occur in shockwaves, microdevices, or rarefied gases. Some well known particle methods include Molecular Dynamics, Monte Carlo methods, and the Lattice Boltzmann Method. The Direct Simulation Monte Carlo (DSMC) method describes gas flows through direct physical modeling of particle motions and collisions. DSMC can model linear and nonlinear acoustics, as well as the details of viscous dissipation. DSMC results will be shown and compared with continuum theory. A DSMC method has been implemented for 1-dimensional linear and nonlinear acoustics problems on parallel computers using object-oriented C++ and the message

passing interface (MPI). The code has been run on large Beowulf clusters. The results have been compared to solutions to the linear wave equation. The complete paper will include two-dimensional solutions as well as comparisons to nonlinear continuum results. [Work supported by Consortium for Education in Many-Body Applications, Grant No. NSF-DGE-9987589.]

3:45

2pPA9. Modeling of ultrasonic backscatter using ultrasonic radiative transfer theory. Goutam Ghoshal and Joseph A. Turner (Dept. of Eng. Mech., W317.4 Nebraska Hall, Univ. of Nebraska–Lincoln, Lincoln, NE 68588-0526)

The scattering of elastic waves in polycrystalline media is primarily due to the elastic inhomogeneities present between grains. This scattering may be used to extract the microstructural parameters of the material such as grain size and grain texture. In particular, ultrasonic backscatter measurements have been especially useful for extracting microstructural information. From the theoretical perspective, derivations related to ultrasonic radiative transfer equations (URTE) govern the propagation of diffuse intensities and include all multiple scattering effects. In this presentation, a rigorous connection between the URTE theory and the backscatter experiments is discussed. The general URTE formulation is first reduced to the singly scattered problem. The backscattered ultrasonic flux is shown to be the main quantity of interest. The flux is obtained in terms of the relevant scattering parameters. Specific solutions are obtained for a specimen excited by a normally incident longitudinal wave. The backscatter model for a normally incidental longitudinal wave and its mode converted shear wave are discussed. The theoretical results are compared with numerical simulations based on Voronoi polycrystal finite-element calculations. Results are also compared with previous backscattering theories. Relevant applications for materials of common interest are also discussed. [Work supported by DOE.]

4:00

2pPA10. Bragg backscatter and band structure in a sinusoidal waveguide model. Dan Valente (Grad. Prog. in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804, dpv110@psu.edu) and David C. Swanson (Penn State Univ., State College, PA 16804)

The unique characteristics of wave propagation through periodic acoustic structures have been well known for many years. These structures allow only certain bands of frequencies to propagate through the system, resulting in a distinct filter response. The purpose of this work is to investigate the properties of a periodic waveguide whose walls vary sinusoidally along its length. The waveguide is modeled using a transmission line/transfer matrix approach to obtain the power reflection and transmission coefficients at the waveguide input. Typically in the literature, waveguides with distinct periodic scatterers or abrupt area changes have been examined, and a distinct band structure is evident. It is shown that for small amplitude wall variations of the sinusoidal waveguide, the system denies the transmission of only one frequency. This frequency fits the condition for Bragg backscatter, a phenomenon that is analogous to what is observed in the acousto-optic (AO) effect. As the amplitude of the wall variations increase, however, a band structure similar to what is expected for periodic systems begins to arise. [Work supported by FRA.]

4:15

2pPA11. Further results on acoustic scattering from the edge of a wedge. Ronald Hughes, Jan M. Niemiec, Yuri Stoyanov, and Herbert Überall (Naval Surface Weapons Ctr., Carderock Division)

Acoustic monostatic and bistatic scattering from the edge of a solid wedge immersed in water is analyzed further, always assuming a line source parallel to the edge of a wedge. Our previous work on the subject is extended by emphasizing the dependence of the scattering amplitude on the wedge material, which may be uniform or layered, and which can be characterized by surface impedance. Impedance values are obtained for multiple layers by the method of Brekhovskihk. Our numerical analysis is based upon Macdonald functions [A. Erdelyi, *Higher Transcendental Functions* (McGraw–Hill, New York, 1953)] via an integral representation that we were able to render convergent. The scattering amplitude is integral bilinear in the Macdonald functions and dependent upon the wedge impedance, and has been evaluated numerically. [Work supported by NAVSEA.]

4:30

2pPA12. Measurements of the total scattering and absorption cross-sections of the human body. Stephane G. Conti (Southwest Fisheries Sci. Ctr., 8604 La Jolla Shores Dr., La Jolla, CA 92037), Philippe Roux (Scripps Inst. of Oceanogr., Univ. of California–San Diego, La Jolla, CA 92093-0238), David A. Demer (Southwest Fisheries Sci. Ctr., La Jolla, CA 92037), and Julien de Rosny (Laboratoire Ondes et Acoustique, ESPCI, Universit Paris VII, 75005 Paris, France)

Presented here are the first absolute measurements of the total acoustic energy scattered and absorbed by humans. The total scattering and absorption cross-sections were obtained from long-duration acoustic reverberation, for individual humans walking randomly in a room. An extension of the technique is proposed to estimate the absorption cross-section for the human body. Within the audible range, the sound scattering spectra of the human body is similar to that of a hard ellipsoid with the same volume (dimensions proportional to the weight to the one-third power). Moreover, we show that increasing amounts of clothing have little effect on scattering while absorption is greatly increased.

4:45

2pPA13. Sound sources of the screech tone radiated from circular supersonic jet oscillating in the helical mode. Yoshikuni Umeda (Dept. of Aeronautic & Astronautics, Grad. School of Eng., Kyoto Univ., Kyoto 606-8501, Japan) and Ryuji Ishii (Kyoto Univ., Kyoto 606-8501, Japan)

The generation mechanism of the screech tone in the helical oscillation mode is investigated using a series of instantaneous photographs. From the analysis of the photographs, six evanescent sound sources are observed as prominent points along the jet axis. These sound sources move along circular orbits in the planes perpendicular to the jet axis and slightly downstream of the rear edges of each shock cell. The speed of the moving sound sources is supersonic and moving Mach cones generated behind the moving sound sources form the helical-shaped wave fronts of the screech tone. The existence of the moving Mach cones about the jet axis was confirmed by comparing Schlieren photographs and drawings of the envelopes of the moving Mach cones. [Work supported by The Mitsubishi Foundation.]

Session 2pPP

Psychological and Physiological Acoustics: Normal and Pathological Perception

James D. Miller, Chair
Central Institute for the Deaf, 818 South Euclid, St. Louis, Missouri 63110

Chair's Introduction—1:25

Contributed Papers

1:30

2pPP1. Psychoacoustic processing of test signals. Frantisek Kadlec (Dept. of Radioelectronics, Czech Tech. Univ., Technicka 2, 166 27 Prague, Czech Republic, kadlec@fel.cvut.cz)

For the quantitative evaluation of electroacoustic system properties and for psychoacoustic testing it is possible to utilize harmonic signals with fixed frequency, sweeping signals, random signals or their combination. This contribution deals with the design of various test signals with emphasis on audible perception. During the digital generation of signals, some additional undesirable frequency components and noise are produced, which are dependent on signal amplitude and sampling frequency. A mathematical analysis describes the origin of this distortion. By proper selection of signal frequency and amplitude it is possible to minimize those undesirable components. An additional step is to minimize the audible perception of this signal distortion by the application of additional noise (dither). For signals intended for listening tests a dither with triangular or Gaussian probability density function was found to be most effective. Signals modified this way may be further improved by the application of noise shaping, which transposes those undesirable products into frequency regions where they are perceived less, according to psychoacoustic principles. The efficiency of individual processing steps was confirmed both by measurements and by listening tests. [Work supported by the Czech Science Foundation.]

1:45

2pPP2. On temporal integration versus frequency-based discrimination of direct-to-reverberant energy ratio. Erik Larsen, Nandini Iyer, Albert Feng, and Charissa Lansing (Beckman Inst. for Adv. Sci. and Technol., Univ. of Illinois at Urbana-Champaign, 405 N. Mathews Ave., Urbana, IL 61801, elarsen@uiuc.edu)

The direct-to-reverberant (D/R) energy ratio is considered to be an important cue for sound source distance judgments (Zahorik, 2002; Bronkhorst, 2001). The current experiments investigated the processes that might be involved in D/R energy ratio discrimination. Two basic processes may be responsible: temporal integration and frequency domain changes. Temporal integration cues are available only if the on- and offset times of the stimulus are sufficiently rapid relative to the room reverberation time such that distinct direct and reverberant portions of the stimulus are perceived. Therefore, in the first experiment we compared JNDs for signals that either did or did not contain temporal integration cues by varying on- and offset time. We propose that the frequency-based method relies on the standard deviation of the spectral response in a reverberant room, which increases monotonically from 0 dB to 5.57 dB (Schroeder limit) as the D/R energy ratio decreases from approx. 20 dB to 0 dB. Outside this range, the standard deviation asymptotes and thus should not offer any cues for discrimination. Therefore, in the second experiment we compared JNDs for stimuli with D/R energy ratios inside and outside the range of 0–20 dB. The results from both these experiments will provide evidence for the processes of source distance coding based on the D/R energy ratio by the auditory system.

2:00

2pPP3. Detection of high-frequency spectral notches as a function of level. Ana Alves-Pinto and Enrique A. Lopez-Poveda (Instituto de Neurociencias de Castilla y Leon, Universidad de Salamanca, Alfonso X El Sabio s/n, 37007 Salamanca, Spain)

In experiment I, the threshold depth for detecting a spectral notch centered at 8 kHz in an otherwise flat-spectrum noise was measured as a function of the noise spectrum level. Threshold depths increased gradually with spectrum level. This result is explained in terms of the broadening of cochlear filters with level and of the narrow dynamic range of most auditory nerve (AN) fibers. Experiment II investigated the extent that notch detection is affected by stimulus duration. On average, threshold depths were larger for shorter (25-ms) than for longer (220-ms) stimuli, particularly at higher sensation levels. This suggests that the adapted AN response is important for encoding spectral information, despite the fact that spectral features may be more clearly encoded at the onset, where AN fibers have a wider dynamic range. In experiment III, the steady spectral notch was replaced by a dynamic one whose center frequency changed abruptly after 110 ms from 7 to 9 kHz. Depths in this condition improved considerably at high levels. Therefore, at high levels, transient spectral information is detected more easily. This relates to the wide dynamic range of AN fibers for sudden increases in level. [Work supported by FIS PI020343, G03/203, and MCYT.]

2:15

2pPP4. A two level model of the masking property of human ear. Jack Xin (Dept. of Mathematics, Univ. of Texas, Austin, TX 78712, jxin@math.utexas.edu) and Yingyong Qi (Qualcomm, Inc., San Diego, CA 92121)

The absolute hearing threshold and masking are two fundamental phenomena in psychoacoustics. The former is the minimum intensity for the ear to detect sound at a given frequency in quiet. The latter is described by the nonlinearly raised threshold for the ear to detect sound in the vicinity of an existing signal or the masker. A two level model is developed to compute the latter given the former and the masker. The first level is a partial differential equation (PDE) model of the inner ear (cochlea), and the second level is a similarity transform, accounting for the functions of the remaining high level processes of audition. The model has a solid ground on first principles and is adaptive to nonlinearities when compared with existing data-driven empirical models. Modeled masking thresholds of banded noise by tonal signals agree well with existing hearing data. [Work partially supported by ARO Grant No. DAAD 19-00-1-0524 and NSF No. ITR-0219004.]

2:30

2pPP5. Spiral model of pitch. James D. Miller (Dept. of Speech and Hearing, Central Inst. for the Deaf and Washington Univ., 4560 Clayton Ave., St. Louis, MO 63110, jdmiller@artsci.wustl.edu)

A spiral model of pitch interrelates tone chroma, tone height, equal temperament scales, and a cochlear map. Donkin suggested in 1870 that the pitch of tones could be well represented by an equiangular spiral. More recently, the cylindrical helix has been popular for representing tone

chroma and tone height. Here it is shown that tone chroma, tone height, and cochlear position can be conveniently related to tone frequency via a planar spiral. For this ''equal-temperament spiral,'' (ET Spiral) tone chroma is conceived as a circular array with semitones at 30° intervals. The frequency of sound on the cent scale (re 16.351 Hz) is represented by the radius of the spiral defined by $r=(1200/2\pi)\theta_r$, where θ_r is in radians. By these definitions, one revolution represents one octave, 1200 cents, 30° represents a semitone, the radius relates θ to cents in accordance with equal temperament (ET) tuning, and the arclength of the spiral matches the mapping of sound frequency to the basilar membrane. Thus, the ET Spiral gives tone chroma as θ, tone height as the cent scale, and the cochlear map as the arclength. The possible implications and directions for further work are discussed.

2:45

2pPP6. On perception of missing fundamental. Takahide Matsuoka and KeNichi Ito (Utsunomiya Univ., Utsunomiya 321-8585, Japan, matsuoka@cc.utsunomiya-u.ac.jp)

The frequency band of one channel of telephone line is from 300 Hz to 3400 Hz. Although the pitch frequency of speech is not in the frequency band, the frequency can be perceived through telephone. So, the psycho acoustic experiments of missing fundamental f0 were carried out at hearing a compound sound of f1 and f2, where $f1=nf0$, $f2=(n+k)f0$, n and k: natural numbers. It has been confirmed, by a significant difference test, that the missing fundamental can be heard. Then, it was investigated how the phase difference and amplitude difference between f1 and f2 influence the production of f0. The results are the following. The phase difference does not influence on the perception of f0. The amplitude difference in some range does not influence it. Those mechanisms have been made clear by consideration based on the experimental results using cochlea models. Now, the perception of missing fundamental for higher harmonics is being investigated.

3:00

2pPP7. Discrimination of dynamic frequency sounds and a multichannel detector model. Vivek Rajendran, Ashok K. Krishnamurthy (Dept. of Elec. Eng., Ohio State Univ., 2015 Neil Ave., Columbus, OH 43210, rajendrv@ee.eng.ohio-state.edu), and Lawrence L. Feth (Columbus, OH 43210)

A virtual frequency (VF) glide signal refers to the frequency transition perceived in an amplitude-modulated two-tone complex. For example, when the amplitude of one tone increases with duration, while that of the other decreases, the listener hears a rising pitch. This virtual glide is easily distinguished from a linear frequency-modulated (FM) glide because of envelope fluctuations in the virtual glide. They can be rendered more difficult to discriminate by extracting the envelope from the VF glide and imposing it on the FM glide. Discrimination of common envelope VF and FM glides was studied through an adaptive 2Q, 2-AFC task. VF and FM glides (250 ms) were centered on 500-, 1000-, 2000-, and 4000-kHz. Frequency sweeps began at approximately 1/7 octave and were adjusted using the 2 up, 1 Down rule (Levitt, 1971). Center frequency roved over a range of approximately 1/2 octave for each interval. We show that a multichannel detector model (Durlach, 1986) operating on intensity weighted average of instantaneous frequency (IWAIF) values computed by a short-term, multichannel IWAIF model is able to predict listener performance in the common-envelope VF versus FM glide discrimination task. The noise variance required by the model is derived from measured FMDL data.

3:15

2pPP8. Synthetic vowel categorization by hearing-impaired listeners. Michelle R. Molis and Marjorie R. Leek (Army Audiol. and Speech Ctr., Walter Reed Army Medical Ctr., 6900 Georgia Ave. NW, Washington, DC 20307-5001)

Vowel identification performance can be predicted reasonably well for normal-hearing (NH) listeners based on principal components derived from excitation patterns [M. R. Molis, J. Acoust. Soc. Am. **111**, 2433–2434 (2002)]. In this study, vowel categorization was measured in listeners with mild-to-moderate hearing loss. Stimuli were 54 synthesized vowels that varied orthogonally in $F2$ (1081–2120 Hz) and $F3$ (1268–2783 Hz) frequency in equal 0.8-bark steps. Fundamental frequency contour, $F1$ (455 Hz), $F4$ (3250 Hz), $F5$ (3700 Hz), and duration (225 ms) were held constant. Hearing-impaired (HI) listeners categorized the stimuli as the vowels /ɪ/, /ʊ/, or /ɝ/. Estimates of frequency resolution were also obtained, and excitation patterns were constructed for each listener. Categorization performance was more variable for HI listeners relative to NH listeners, both within and between listeners. Many subjects appeared to rely solely on $F2$ frequency and had particular difficulty with the /ʊ/ vs /ɝ/ distinction. Excitation patterns suggested a rather imprecise internal spectral representation of the stimuli. HI response patterns may reflect decreased audibility, spectral smearing of formant structure due to poor frequency resolution, or both. Models of vowel perception based on the impaired excitation patterns will be compared with formant-based models. [Work supported by NIH (DC00626).]

3:30

2pPP9. Dichotic presentation to overcome the effect of increased spectral masking and frequency dependent hearing threshold shifts in persons with bilateral sensorineural impairment. Alice N. Cheeran (BME Group, IIT Bombay, Powai Mumbai-400 076, India) and Prem C. Pandey (IIT Bombay, Powai Mumbai-400 076, India)

A binaural dichotic presentation scheme for reducing the effect of increased spectral masking in persons with bilateral sensorineural loss, using spectral splitting with complementary comb filters based on auditory critical bands, has been earlier reported [Cheeran *et al.*, J. Acoust. Soc. Am. **110**, 2705 (2001)]. The 256-coefficient linear phase FIR filters designed using frequency sampling technique had transition crossovers adjusted within -6 to -4 dB for perceptual balance, and had 78–117 Hz transition, 1 dB passband ripple, and 30 dB stopband attenuation. We evaluated the scheme by conducting listening tests on 5 normal hearing subjects with simulated loss, using a closed set identification of 12 vowel-consonant-vowel syllables. Based on significant improvement, further tests were conducted on 5 hearing-impaired persons with moderate bilateral sensorineural loss. Significant improvement in response time, recognition scores, and transmission of consonantal features, particularly place and duration, was obtained, indicating reduction in the effect of spectral masking. In order to partly compensate for frequency dependent hearing threshold shifts, a pair of filters with frequency response adjusted within a 6 dB range, based on the audiogram for the corresponding ear, was cascaded with the comb filters. These filters resulted in additional improvement, particularly for persons with relatively uniform loss.

3:45

2pPP10. The effect of compression and attention allocation on speech intelligibility. Sangsook Choi and Thomas Carrell (Univ. of Nebraska–Lincoln, 253 Barkley Memorial Ctr., Lincoln, NE 68583, schoi6@bigred.unl.edu)

Research investigating the effects of amplitude compression on speech intelligibility for individuals with sensorineural hearing loss has demonstrated contradictory results [Souza and Turner (1999)]. Because percent-correct measures may not be the best indicator of compression effectiveness, a speech intelligibility and motor coordination task was developed to provide data that may more thoroughly explain the perception of compressed speech signals. In the present study, a pursuit rotor task [Dlhopolsky (2000)] was employed along with word identification task to measure the amount of attention required to perceive compressed and non-compressed words in noise. Monosyllabic words were mixed with speech-shaped noise at a fixed signal-to-noise ratio and compressed using a wide dynamic range compression scheme. Participants with normal hearing identified each word with or without a simultaneous pursuit-rotor task. Also, participants completed the pursuit-rotor task without simultaneous word presentation. It was expected that the performance on the additional motor task would reflect effect of the compression better than simple word-accuracy measures. Results were complex. For example, in some

conditions an irrelevant task actually improved performance on a simultaneous listening task. This suggests there might be an optimal level of attention required for recognition of monosyllabic words.

4:00

2pPP11. A speech compression/expansion method based on subband filtering of the signal envelope. Aslak Bjerkvik, Asbjorn Krokstad, and Barbara Resch[a] (NTNU Acoust. group, Trondheim, Norway)

Compression algorithms are used in hearing aids for hearing-impaired listeners with recruitment, due to these listeners' loss of dynamic range. We propose a method for both compressing and expanding speech signals with the goal to improve speech intelligibility in reverberant conditions where the important rapid variations of the signal are inherently reduced. The method is based on dividing the envelope of the speech signal into two subbands with a division around modulation frequencies of 2 Hz, and compressing the signal based on the lower envelope subband, while expanding the signal based on the higher envelope subband. The sub-band division is accomplished by filtering the envelope of the signal and also by computing separate envelopes for each subband. A secondary goal is to develop low-delay algorithms. Various methods for calculating the envelope of the signal are evaluated, with a focus on the delay. [a] Currently at KTH, Speech processing group, Stockholm, Sweden.

4:15

2pPP12. Adaptive modeling of compression hearing aids: Convergence and tracking issues. Vijay Parsa and Donald Jamieson (Natl. Ctr. for Audiol., Elborn College, Univ. of Western Ontario, London, ON N6G 1H1, Canada, parsa@nca.uwo.ca)

Typical measurements of electroacoustic performance of hearing aids include frequency response, compression ratio, threshold and time constants, equivalent input noise, and total harmonic distortion. These measurements employ artificial test signals and do not relate well to perceptual indices of hearing aid performance. Speech-based electroacoustic measures provide means to quantify the real world performance of hearing aids and have been shown to correlate better with perceptual data. This paper investigates the application of system identification paradigm for deriving the speech-based measures, where the hearing aid is modeled as a linear time-varying system and its response to speech stimuli is predicted using a linear adaptive filter. The performance of three adaptive filtering algorithms, viz. the Least Mean Square (LMS), Normalized LMS, and the Affine Projection Algorithm (APA) was investigated using simulated and real digital hearing aids. In particular, the convergence and tracking behavior of these algorithms in modeling compression hearing aids was thoroughly investigated for a range of compression ratio and threshold parameters, and attack and release time constants. Our results show that the NLMS and APA algorithms are capable of modeling digital hearing aids under a variety of compression conditions, and are suitable for deriving speech-based metrics of hearing aid performance.

4:30

2pPP13. Evaluation of noise reduction techniques for digital hearing aids. Vijay Parsa and Karthikeyan Umapathy (Natl. Ctr. for Audiol., Elborn College, Univ. of Western Ontario, 1201 Western Rd., London, ON N6G 1H1, Canada, parsa@nca.uwo.ca)

Individuals with sensorineural hearing loss have increased difficulty in understanding speech in noisy backgrounds. To combat this issue, there has been a major thrust in recent years toward the development of noise reduction algorithms. The goals of this paper are to quantify the relative benefits of different single-microphone noise reduction algorithms, and to investigate the interaction between the noise reduction and dynamic range compression algorithms. Noise reduction techniques evaluated in this paper include spectral subtraction-based techniques, a wavelet-packet-based technique and a matching pursuit-based technique. All algorithms were tested with HINT signals with SNR levels ranging from -5 to 15 dB, and two different noise types viz. the speech-shaped noise and multi-talker babble. Performance was quantified using the ITU standardized PESQ measure which computes the perceptual similarity between the enhanced signal and the original signal. Initial PESQ results showed that the spectral subtraction-based techniques perform superior to that of the wavelet-packet and matching pursuit-based approaches and that the compression time constants have an impact on the overall performance. Perceptual data collected from hearing impaired listeners on sound quality and noise reduction performance will be presented and their correlation with the objective measurements will be discussed.

4:45

2pPP14. Perception of synthetic speech in quiet and in noise by adults with cochlear implants. Benjamin Munson and Peggy B. Nelson (Dept. of Commun. Disord., Univ. of Minnesota, 115 Shevlin Hall, 165 Pillsbury Dr. SE, Minneapolis, MN 55455)

Recent research has demonstrated that the ability of adults with cochlear implants (CI) to perceive speech in noise is highly variable. There are many potential reasons why this may be so, including poor signal/noise segregation, and poor perception of phonetic features in noise. We examined the CI listeners' perception of four synthetic speech continua in quiet and in noise. In two continua, dynamic spectral information was manipulated (/ra/-/la/ and /wa/-/ja/); in one, static spectral information was manipulated (/i/-/u/); and in one, only temporal information was manipulated (/sei/-/stei/). These were presented to 12 listeners with CIs and 13 normal-hearing listeners in quiet and concurrent with speech-shaped noise, at a $+10$ dB SNR. Noise affected identification (endpoints, boundaries, and slopes) of CI listeners more than NH listeners. Significant group-by-SNR interactions were found for endpoint identification of the /ra/-/la/, /wa/-/ja/, and /sei/-/stei/ continua. CI listeners had significantly shallower identification slopes for the /ra/-/la/ continuum; this was exaggerated in the $+10$ dB SNR condition. In addition, CI listeners showed more /u/ percepts than the NH listeners in the /i/-/u/ continuum at $+10$ dB SNR. Results are discussed with respect to the relative vulnerability of temporal and spectral features to misperception in noisy conditions.

5:00

2pPP15. Cochlear nonlinearity between 500 and 8000 Hz in listeners with impaired hearing. Enrique A. Lopez-Poveda (Instituto de Neurociencias de Castilla y Leon, Universidad de Salamanca, Alfonso X El Sabio s/n, 37007 Salamanca, Spain), Christopher J. Plack, Ray Meddis (Dept. of Psych., Univ. of Essex, Colchester CO4 3SQ, UK), and Jose L. Blanco (Oticon Spain SA, Ctra. Fuencarral 24, Ed. Europa, 28108 Madrid, Spain)

Cochlear nonlinearity was estimated in listeners with impaired hearing using a forward-masking method. For a fixed low-level probe, the masker level required to mask the probe was measured as a function of the masker-probe interval, to produce a temporal masking curve (TMC). TMCs were measured for probe frequencies from 500 to 8000 Hz, and for masker frequencies from 0.5 to 1.6 times the probe frequency. Unlike what happens for normal-hearing (NH) listeners [Lopez-Poveda *et al.*, J. Acoust. Soc. Am. **113**, 951–960 (2003)], TMCs for on-frequency maskers sometimes show a single slope, suggesting linear responses for tones at CF. Sometimes, however, two distinct slopes are visible, suggesting remaining compression. Both patterns are uncorrelated with the absolute threshold and are consistent with selective damage to outer or inner hair cells (IHC), respectively. Remarkably, the slope of the TMCs for very low-frequency maskers is shallower for the impaired ears than for normal ones. This result implies that for NH listeners, the slope of the TMCs for very low-frequency maskers reflects some kind of compression, even at high CFs. It is discussed that this compression is likely to occur at the IHC rather than at the basilar membrane. [Work supported by FIS PI020343, G03/203, and Oticon Spain.]

5:15

2pPP16. Vowel formant discrimination in high-fidelity speech by hearing-impaired listeners. Diane Kewley-Port (Speech and Hearing Sci., Indiana Univ., Bloomington, IN 47405, kewley@indiana.edu), Chang Liu (Univ. at Buffalo, Buffalo, NY 14214), and T. Zachary Burkle (Indiana Univ., Bloomington, IN 47405)

The ability to discriminate differences in vowel formant frequency under a more ordinary listening condition has been reported recently by Kewley-Port and Zheng (1999) for synthetic speech and Liu and Kewley-Port (2003) for high-fidelity speech. Results for normal-hearing (YNH) listeners showed that a longer phonetic context (sentences versus words) degraded formant discrimination while adding a sentence identification task did not. The present study used the same high-fidelity stimuli but employed listeners with moderate hearing impairment (YHI). Experimental factors manipulated included phonetic context (isolated vowels, words and sentences), level [70-dB SPL (partially audible) and 95-dB SPL (fully audible)], and task (discrimination with and without the sentence identification task). Thresholds for F1 and F2 of the vowels /I, ɛ, æ ʌ/ were estimated using adaptive tracking for 71% correct discrimination. The anticipated degrading effects of higher formant frequency and longer phonetic context were obtained. Unexpectedly formant thresholds for high-fidelity vowels at the higher (95-dB SPL) were slightly elevated compared to the 70-dB SPL level, the opposite of results for synthetic vowels. No effect of the added sentence task was seen, similar to the YNH listeners. Details of the differences in vowel processing attributable to moderate hearing impairment will be discussed. [Research supported by NIH-NIDCD.]

TUESDAY AFTERNOON, 11 NOVEMBER 2003 NUECES ROOM, 2:00 TO 4:45 P.M.

Session 2pSA

Structural Acoustics and Vibration, Engineering Acoustics and Physical Acoustics: Measurements of Particle Velocities

Sean F. Wu, Chair
Department of Mechanical Engineering, Wayne State University, 5050 Anthony Wayne Drive, Detroit, Michigan 48202

Invited Papers

2:00

2pSA1. Measurement of acoustic and streaming velocities in fluids using laser Doppler anemometry. Michael W. Thompson and Anthony A. Atchley (Grad. Prog. in Acoust., The Penn State Univ., University Park, PA 16802, mwt126@psu.edu)

The optical technique of laser Doppler anemometry (LDA) can be used to measure both acoustic and streaming velocities in a fluid. In a typical LDA system, two coherent laser beams intersect within the fluid to form a series of interference fringes. Whenever a small tracer particle suspended in the fluid passes through these fringes, light is scattered with an intensity that fluctuates at a frequency proportional to the particle velocity. This intensity is measured using a photodetector, and the resultant signal is processed to yield the fluid velocity. An overview of several signal processing methods will be given, and the inherent strengths and weaknesses of the LDA technique will be discussed. The authors have used LDA with burst spectrum analysis and recursive Fourier averaging to measure standing-wave acoustic velocities in the range of 0.02–15 m/s, and to measure Rayleigh streaming velocities in the range of 0.1–20 cm/s in the presence of standing-wave acoustic velocities that were 2–3 orders of magnitude larger. [Work supported by ONR.]

2:30

2pSA2. Particle image velocimetry measurements in thermoacoustic refrigerators. Philippe Blanc-Benon (Ecole Centrale de Lyon, Ctr. Acoustique, LMFA UMR CNRS 5509, 69134 Ecully Cedex, France)

The knowledge of flow fields in the microchannels and at the edges of the stack plates becomes an increasingly important issue in the design of heat exchangers for thermoacoustic engines. We have developed numerical simulations and conducted experiments in a resonant standing wave thermoacoustic refrigerator model. Recent computational evidence indicates that near the edges of the plates the flow field is dominated by concentrated eddies, whose complex motion significantly affects the performance of the device. Consequently, the effective design and optimization of thermoacoustic refrigerators necessitates a fundamental understanding of these vortical motions, and their dependence on geometric parameters and operating conditions. We present experimental data obtained using Particle Image Velocimetry: velocity profiles across the microchannels, 2D velocity maps including a zoom for the edges of the stack, and vorticity fields calculated with a criterion based on a normalized angular momentum. Results are obtained for two distinct configurations, involving thin and thick stack plates. In the first case, the flow field around the edge of the stack exhibits elongated vorticity layers, while in the latter it is dominated by the shedding and impingement of concentrated vortices. Time-resolved PIV measurements of the velocity and vorticity fields are compared with computational predictions.

3:00

2pSA3. Flow-induced noise on a pressure-acceleration underwater acoustic intensity probe. James A. McConnell (Acoustech Corp., P.O. Box 139, State College, PA 16804) and Gerald C. Lauchle (Grad. Prog. in Acoust., Penn State Univ., 218 Appl. Sci. Bldg., University Park, PA 16802)

The results of an experiment to determine the effect of flow-induced noise on a pressure-acceleration intensity probe are presented. The sensor is a bluff body that conforms to the geometry of a right circular cylinder having a diameter of 10.16 cm and an aspect ratio of unity. A flexural mode hydrophone is used to measure the acoustic pressure and a flexural mode accelerometer is used to measure the acoustic particle acceleration. The hydrophone has a nominal sensitivity of about −185 dB *re*: 1 V/uPa and the accelerometer has an in-water acoustic sensitivity of nearly 1 V/g. The bandwidth of the device covers the 10-Hz to 1-kHz frequency range and the principle axis of sensitivity is coincident with the axis of the cylinder. The sensor was towed in cross-flow at speeds ranging from 0.1 to 0.5 cm/s and data were collected over the 10-Hz to 100-Hz frequency range. Of particular interest is the decomposition of the intensity spectrum into real and reactive parts along with an assessment of any filtering that can be accomplished using the intensity technique to measure far-field sound in the presence of near-field noise. [Work supported by various grants from ONR and NAVAIR.]

3:30–3:45 Break

3:45

2pSA4. Measurements in acoustic boundary layers. G. Huelsz, A. A. Castrejon-Pita, J. R. Castrejon-Pita, F. Lopez-Alquicira, E. Ramos, and R. Tovar (Centro de Investigacion en Energia, UNAM, Mexico)

A summary of the results from experimental measurements of temperature and velocity in the oscillatory boundary layers produced by acoustic waves (acoustic boundary layers) are presented. Temperature measurements were done using the named cold wire anemometer and velocity measurements were done using hot wire and laser Doppler anemometers and particle image velocimetry. Experimental results are compared to theoretical results of a linear theory. The advantages and the limitations of these experimental techniques are discussed, especially for their use in acoustic boundary layers. [Partial support for this work has been provided by DGAPA-UNAM IN104702-2 and CONACYT 32707-U projects.]

Contributed Papers

4:15

2pSA5. A laser-based acoustic velocity sensor. Peter R. Stepanishen (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882-1197) and Lynn Antonelli (Naval Undersea Warfare Ctr., Newport, RI 02841)

A linear systems model is presented for a Laser Doppler Vibrometer (LDV), which is used to detect the acoustically induced normal surface velocity at an air-water interface. Previous acoustic test tank results for signal reception are interpreted using the model. Noise performance is addressed using the model for a commercially available LDV where thermal noise in the fluid is shown to be of lesser importance than the internal noise of the sensor. In addition to the internal sensor noise that can be viewed as an additive noise which is uncorrelated to the signal, it is also noted that sidebands of the signal spectrum can be generated for large displacements of the air-water interface. For the case of harmonic signals the spectral distortion increases as kY increases where k is the acoustic wavenumber and Y is the amplitude of the surface displacement. Results from a series of experiments in-air and in-water using a piezo-ceramic disk with a variable voltage excitation are discussed in light of the model. The in-air experiments use the vibrating face of the disk as the reflecting surface whereas the in-water experiments use the air/water interface in a wave tube excited by the disk.

4:30

2pSA6. Prediction of acoustic radiation via 3D particle velocity measurements. Zhi Ni, Huancai Lu, Sean Wu (Dept. of Mech. Eng., Wayne State Univ., Detroit, MI 48202), and Yang Zhao (Dept. of Elec. & Computer Eng., Wayne State Univ., Detroit, MI 48202)

It has been shown [Wu and Hu, J. Acoust. Soc. Am. **103**, 1763–1774 (1998)] that one can use an explicit integral formulation to predict acoustic radiation from a vibrating object based on a particle velocity distribution over a hypothetical surface enclosing this object. The advantages of this integral formulation are that solutions thus obtained are unique and the efficiency of numerical computations is high. The challenge is how to acquire 3D particle velocities when measurements are taking place in the air, in which seeding particles are hard to control and the signal to noise ratio is low. Here we use laser Doppler Anemometry (LDA) to accomplish this task. Two LDA probes are utilized to measure two components of the particle velocities simultaneously, and the third component is obtained by rotating one probe 90° and measuring again. A traverse system is adopted to sweep the LDA probes over a measurement plane to enhance efficiency. Once particle velocities are specified on all nodes of the enclosing measurement surface, the sound pressure anywhere in an exterior region can be predicted. An interface program for commercial hypermesh software is developed, the Dijkstra algorithm and boundary element method are adopted, and error analyses are presented. [Work supported by NSF.]

Session 2pSCa

Speech Communication: Cross Linguistic Issues (Poster Session)

Katsura Aoyama, Chair
Communication Disorders, Texas Tech University Health Science Center, 3601 4th Street, Lubbock, Texas 79430-6073

Contributed Papers

All posters will be on display from 1:00 p.m. to 4:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:00 p.m. to 2:30 p.m. and contributors of even-numbered papers will be at their posters from 2:30 p.m. to 4:00 p.m.

2pSCa1. Cross-linguistic preference and the phonetics of geminates: Place of articulation and duration of phonologically short and long consonants in Guinaang Bontok. Katsura Aoyama (Dept. of Speech-Lang. & Hearing Sci., Texas Tech Univ., Health Sci. Ctr., 3601 4th St., Stop 6073, Lubbock, TX 79430-6073, katsura.aoyama@ttuhsc.edu)

It is reported that there is a preference for the place-of-articulation in geminates among world languages; alveolar, labial, velar geminate consonants always appear in that order with respect to one another (G. Thurgood, *Papers in Honor of Frederick H. Brengelman*, 1993, pp. 129–137). This study investigated whether the phonetic contrasts in duration are larger in cross-linguistically favored place-of-articulation (e.g., alveolar) than in the less favored place-of-articulation (e.g., velar). The data is from a language called Guinaang Bontok. Four speakers of Guinaang Bontok produced thirty-five words that have either a singleton or a geminate of /p t k m n ŋ/. Eighteen words included a singleton and the other seventeen words included a geminate. The participants produced the target word in isolation first, and then repeated it in a frame sentence twice. The recordings were digitized at 22.05 Hz, and the durations of the consonants were measured. A total of 420 tokens (35 words $\times 3$ repetitions $\times 4$ participants) were analyzed. The results partially supported the hypothesis; the durational contrasts were indeed larger in alveolar consonants (/t n/) than in labial consonants (/p m/) and velar consonants (/ŋ/). However, contrasts in velar stop (/k/) were as large as in the alveolar stop (/t/).

2pSCa2. Contrast enhancement of vowels in Modern Standard German. Amy E. Coren and Cheryl L. Heckmann (Dept. of Psych., Univ. of Texas, 1 University Station A8000, Austin, TX 78712-0187)

In an earlier cross-language study of vowels in and out of utterance focus [Hay *et al.*, J. Acoust. Soc. Am. **111**, 2367 (2002)], it was reported that languages tend to make differential use of three possible means of contrast enhancement: an enlarged vowel space, greater use of vowel inherent spectral change (VISC), and greater systematic variation in duration across vowel categories. In the current study, these same means of vowel contrast enhancement were examined in Modern Standard German. Relative to speakers of American English, German speakers make less use of VISC and greater use of durational variation to enhance vowel distinctions. It appears that phonological properties of a language (e.g., the use of phonemic vowel length differences) are predictive of the particular means used by the language to enhance vowel contrasts in the utterance focus. [Work supported by NIDCD.]

2pSCa3. Perceptual illusion overrides reality: A phoneme detection study of epenthetic vowels in Japanese. Yuki Hirose (The Univ. of Electro-Commun., 1-5-1 Chofugaoka, Chofu 182-8585, Japan, hirose@cs.uec.ac.jp)

Vowel epenthesis is a known phenomenon in Japanese where speakers insert a vowel ([u], by default) inside consonant clusters (e.g., in a VCCV sequence), due to the phonotactic constraint of the language. Dupoux *et al.* (1999) argue that vowel epenthesis occurs at a perceptual level. In their study, Japanese listeners perceived the epenthetic vowel [u] when the interconsonantal vowel [u] in the original stimuli was removed. In the present study, we investigate further whether (1) the perceptual system in Japanese listeners that causes vowel epenthesis is not only constrained by the phonotactic constraint (*VCCV) but also by the quality of the vowel that can be inserted, and (2) whether this constraint overrides existing acoustic information present in the input, e.g., the formant information of some other vowel. In our phoneme-detection experiment, we recorded a set of nonword stimuli (VCVCV) and gradually removed the interconsonantal vowel ([u] or other vowels). Some Japanese listeners perceived [u] not only when the originally present [u] was removed (as found in Dupoux *et al.*), but also when a vowel other than [u] was substituted in the interconsonantal position and that substituted vowel was partially removed, leaving only some remnant information.

2pSCa4. Learning to perceive Mandarin tones: The role of acoustic information. Connie K. So (Dept. of Linguist., Simon Fraser Univ., Burnaby, BC V5A 1S6, Canada, klso@sfu.ca)

Studies show that auditory-training improves non-native listeners' tonal identification; however, persistently perceptual confusion of several tone pairs is still observed. This may imply that learners have not yet mastered/acquired the lexical tones. Their confusion may be reduced further if the essential acoustic information of tones (e.g., duration and pitch contour) could be implemented and emphasized during training. To verify the assumption, the present study examines the impact of employing acoustic information of lexical tones as feedback on non-native listeners' performance during a computer-based perception training of Mandarin tones. Non-native speakers of Mandarin were randomly assigned to one of two groups. Listeners in the control group were merely shown that the answer was right or wrong. In contrast, those in the experimental group received acoustic information by means of both visual and auditory feedback when the response was incorrect (i.e., showing pitch graphs and presenting the audio files for the contrastive tonal pairs). Results indicated a significant improvement in the tonal identifications for listeners who received detailed acoustic information during training. This suggests that

training with acoustic information of lexical tones assists non-native listeners in distinguishing the tone pairs more effectively. [Work supported by SSHRC.]

2pSCa5. Word-final nasals in Romanian. Benjamin Tucker (Univ. of Arizona, 1100 E. University Blvd., Tucson, AZ 85715)

In this study Romanian word-final nasals, e.g., /basm/ ''fairytale,'' are examined acoustically. Acoustic descriptions of Romanian are few, but early research indicated that word-final nasals are devoiced or partially devoiced. In this study an acoustic analysis of cross-linguistically unusual voiceless nasals, as well as information about a language which has not been phonetically well-documented is provided. In the current work the word-final /sm, mn, Vm, Vn/ are examined. The words were placed in various phonological environments (e.g., utterance final, before vowel-initial following word, etc.). Eight Romanian speakers were recorded. A judgment of voicing was made from the spectrogram. The duration of the gap between the /s/ and /m/, the total duration of voiced nasals and the duration of the preceding vowel were measured. The data was also analyzed for the presence of a release noise at the end of the nasal, which indicates that there is an oral closure during the nasal, and hence that the nasal is devoiced rather than deleted. Preliminary data shows that the average duration for /mn/ (0.148 ms) is longer than the duration of the /n/ or /m/ (0.105 ms) further indicating that deletion is not occurring. The preliminary data also shows that 93% of nasals in the /sm/ sequences are devoiced.

2pSCa6. Acoustic comparisons of Japanese and English vowels produced by native speakers of Japanese. Kanae Nishi (Dept. of Speech and Hearing Sci., Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405), Reiko Akahane-Yamada, Rieko Kubo (ATR Intl.-Human Information Sci. Labs., Kyoto 619-0288, Japan), and Winifred Strange (City Univ. of New York–Grad. Ctr., New York, NY 10016)

This study explored acoustic similarities/differences between Japanese (J) and American English (AE) vowels produced by native J speakers and compared production patterns to their perceptual assimilation of AE vowels [Strange *et al.*, J. Phonetics **26**, 311–344 (1998)]. Eight male native J speakers who had served as listeners in Strange *et al.* produced 18 Japanese (J) vowels (5 long-short pairs, 2 double vowels, and 3 long-short palatalized pairs) and 11 American English (AE) vowels in /hVbopena/ disyllables embedded in a carrier sentence. Acoustical parameters included formant frequencies at syllable midpoint ($F1/F2/F3$), formant change from 25% to 75% points in syllable (formant change), and vocalic duration. Results of linear discriminant analyses showed rather poor acoustic differentiation of J vowel categories when $F1/F2/F3$ served as input variables (60% correct classification), which greatly improved when duration and formant change were added. In contrast, correct classification of J speakers' AE vowels using $F1/F2/F3$ was very poor (66%) and did not improve much when duration and dynamic information were added. J speakers used duration to differentiate long/short AE vowel contrasts except for mid-to-low back vowels; these vowels were perceptually assimilated to a single Japanese vowel, and are very difficult for Japanese listeners to identify.

2pSCa7. Discriminability and identification of English vowels by native Japanese speakers in different consonantal contexts. Takeshi Nozawa (Kansai Univ. of Intl. Studies, 1-18 Aoyama Shijimi-cho Miki Hyogo 673-0521, Japan, nozawa@kuins.ac.jp), Elaina M. Frieda (Auburn Univ., Auburn, AL), and Ratree Wayland (Univ. of Florida–Gainesville, Gainesville, FL)

The purpose of the present experiment was to examine the effects of consonantal context on discrimination and identification of English vowels by native Japanese speakers learning English in Japan. A number of studies have assessed the effects of consonantal contexts on the perception of nonnative vowels. For instance, Strange *et al.* (1996, 2001) found that perceptual assimilation of nonnative vowels is affected by consonantal contexts, and Morrison (2002) has shown that Japanese speakers use durational cues to perceive English /i/–/ɪ/. The present study revealed that consonantal context affects discriminability and identification of each English vowel differently. Of all the six vowel contrasts tested, /i/–/ɪ/ was the most likely to be affected by voicing status of the surrounding consonants with it being easier to discriminate in voiceless consonantal contexts. Moreover, /ɪ/ is more likely to be equated with the Japanese short vowel /i/ in a voiceless consonantal context which is in keeping with Morrison (2002). /æ/–/ɑ/, on the other hand, is the most strongly affected by the place of articulation of the preceding consonants. [Work supported by Grant-in-Aid for Scientific Research (C)(1)(1410635).]

2pSCa8. Accent, intelligibility, and comprehensibility in the perception of foreign-accented Lombard speech. Chi-nin Li (Dept. of Linguist., Simon Fraser Univ., Burnaby, BC V5A 1S6, Canada, clia@sfu.ca)

Speech produced in noise (Lombard speech) has been reported to be more intelligible than speech produced in quiet (normal speech). This study examined the perception of non-native Lombard speech in terms of intelligibility, comprehensibility, and degree of foreign accent. Twelve Cantonese speakers and a comparison group of English speakers read simple true and false English statements in quiet and in 70 dB of masking noise. Lombard and normal utterances were mixed with noise at a constant signal-to-noise ratio, and presented along with noise-free stimuli to eight new English listeners who provided transcription scores, comprehensibility ratings, and accent ratings. Analyses showed that, as expected, utterances presented in noise were less well perceived than were noise-free sentences, and that the Cantonese speakers' productions were more accented, but less intelligible and less comprehensible than those of the English speakers. For both groups of speakers, the Lombard sentences were correctly transcribed more often than their normal utterances in noisy conditions. However, the Cantonese-accented Lombard sentences were not rated as easier to understand than was the normal speech in all conditions. The assigned accent ratings were similar throughout all listening conditions. Implications of these findings will be discussed.

2pSCa9. Intonational phonology description of Porteno Spanish. John P. Barjam (Dept. of Linguist., UCLA, 405 Hilgard Ave., Los Angeles, CA 90095-1543, ungaucho@humnet.ucla.edu)

In this study the intonational patterns of Porteno Spanish (PS) are described within an Autosegmental Metrical (AM) (Ladd, 1996) approach. Porteno Spanish is the Spanish spoken in Buenos Aires, Argentina, and because of the influence from Italian and other languages, it differs markedly from varieties of Spanish that have been previously described. Six monolingual PS speakers were recorded saying 99 sentences which include declaratives, interrogatives, imperatives, and focused declarative and interrogative sentences. A ToBI framework, based on the F0 contour, was developed for labeling the sentences intonation. Preliminary results suggest that each phrase contains a somewhat regular pattern: (1) the stressed syllables of content words carry one of five pre-nuclear pitch accents, (2) nuclear pitch accents are more restrictive and tend to vary depending on the sentence type, and (3) boundary tones, both intermediate and intonational, are regular but also vary depending on sentence type.

2pSCa10. Age effects on acquisition of word stress in Spanish–English bilinguals. Susan G. Guion, J. J. Clark (Dept. of Linguist., Univ. of Oregon, Eugene, OR 97403-1290, guion@uoregon.edu), and Tetsuo Harada (Univ. of Oregon)

Based on studies of syntactic and semantic learning, it has been proposed that certain aspects of second language learning may be more adversely affected by delays in language learning than others. Here, this proposal is extended to the phonological domain in which the acquisition of English word stress patterns by early (AOA <6 years) and late (AOA >14 years) Spanish–English bilinguals is investigated. The knowledge of English word stress was investigated by three behavioral tasks. In a production task, participants produced two syllable nonwords in both noun and verb sentence frames. In a perception task, participants indicated a preference for first or last syllable stress on the nonwords. Real words that were phonologically similar to the test items were also collected from each participant. Regression analyses and ANOVAs were conducted to determine the effect of syllable structure, lexical class, and stress pattern of phonologically similar words on the data from the production and perception tasks. Early bilinguals patterned similarly to the native English participants. Late bilinguals showed little evidence of learning prosodically based stress patterns but did show evidence of application of distributional patterns based on lexical class and analogy in stress assignment. [Research supported by NIH.]

2pSCa11. The effect of the dialect on perception of words with different pitch accents in Japanese. Kyoko Okamura (Dept. of Linguist., Indiana Univ., Memorial Hall 322, 1021 E. 3rd St., Bloomington, IN 47405)

Eastern and western dialects of Japanese have been noted to have different pitch accent systems. This study explores how Japanese words with and without accent are perceived by speakers of different dialects. Otake and Cutler (J. Phonet. **27**, 229–253) found little difference in identification between Tokyo and the other variety of speakers. In current study, a speaker of Tokyo Japanese produced words with and without an accent. Using the STRAIGHT resynthesis system, the $F0$ of the accented mora was lowered in 7 steps to the level of the unaccented form, and raised in the unaccented mora to the level of the accented form. Continua were presented to listeners from western and eastern Japan. Listeners exhibit consistent category boundaries according to the difference in F0 between the target mora and the following one. In addition, reaction times are greater near the category boundary, and subjective measures of confidence decrease near the boundary. Speakers from different dialects showed similar identification patterns. Reaction times and confidence measures were different overall, but exhibited the same pattern for the two groups. These results indicate that the accent is a uniform phonetic entity for the two dialects, despite numerous differences in the accentual systems.

2pSCa12. Is there a correlation between Japanese *L*2 learner's perception of English stressed words and acoustic features? Keiko Asano and Toshiko Isei-Jakkola (Yokohama Natl. Univ., 79-1 Tokiwadai Hodogaya-ku Yokohama-city, Kanagawa-prefecture 240-8501, Japan, ll-ed@ynu.ac.jp)

Is there a correlation between Japanese *L*2 learner's perception of English stressed words and acoustic features? [Keiko Asano (Yokohama National University, ll-ed@ynu.ac.jp) and Toshiko Isei-jaakkola (University of Helsinki)]. It is well known that the Japanese have weakness in listening to unstressed words in English, but there are less data on their perception of stressed words. Thus, the listening tests and the acoustic experiments were conducted in terms of (1) relevancy of difficulites depending on part of speech and their English proficiency, (2) the relationship between pitch and intensity of stressed words, and (3) if there is a correlation between their perception and experimental data. In the listening test, an English prose read by an American male speaker was used. The 150 Japanese *L*2 learners were assigned to mark the primary stressed words. The statistical results showed that there was a variance depending on part of speech and more markedly the comparative rating scores of correct words were highly correlated to the learner's English proficiency in any part of speech. In the acoustic experiments, pitch and intensity were measured. It was confirmed that (1) both $F0$ and dB carried the cue to perceive a stressed-word but they were not necessarily correlated, and (2) the relationship between $F0$ and dB might be compared only by relative movement. By further analyzing these acoustic data, prosodic combination of $F0$ and dB might be relevant to the correct ratios of part of speech.

2pSCa13. A note on the acoustic-phonetic characteristics of non-native English vowels produced in noise. Chi-nin Li and Murray J. Munro (Dept. of Linguist., Simon Fraser Univ., Burnaby, BC V5A 1S6, Canada)

The Lombard reflex occurs when people unconsciously raise their vocal levels in the presence of loud background noise. Previous work has established that utterances produced in noisy environments exhibit increases in vowel duration and fundamental frequency ($F0$), and a shift in formant center frequencies for $F1$ and $F2$. Most studies of the Lombard reflex have been conducted with native speakers; research with second-language speakers is much less common. The present study examined the effects of the Lombard reflex on foreign-accented English vowel productions. Seven female Cantonese speakers and a comparison group of English speakers were recorded producing three vowels (/i u a/) in /bVt/ context in quiet and in 70 dB of masking noise. Vowel durations, $F0$, and the first two formants for each of the three vowels were measured. Analyses revealed that vowel durations and $F0$ were greater in the vowels produced in noise than those produced in quiet in most cases. First formants, but not $F2$, were consistently higher in Lombard speech than in normal speech. The findings suggest that non-native English speakers exhibit acoustic-phonetic patterns similar to those of native speakers when producing English vowels in noisy conditions.

2pSCa14. Allophonic variation of Japanese /z/. Yuka Matsugu (Dept. of East Asian Studies, Univ. of Arizona, P.O. Box 210105, Tucson, AZ 85721-0105) and Timothy J. Vance (Univ. of Arizona, Tucson, AZ 85721)

According to impressionistic descriptions, the Japanese phoneme /z/ is pronounced sometimes as a fricative [z] and sometimes as an affricate [dz]. There is no consensus about how the two allophones are distributed, but one claim is that they are in complementary distribution: [dz] word initially or after a nasal and [z] intervocalically. Native speakers of ''standard'' Japanese pronounced words containing /z/ in a carrier sentence. Using spectrograms and waveforms, two (necessarily approximate) measurements were taken for each token of /z/: total duration and closure duration. The results did not show a categorical affricate/fricative distinction. Some word-initial and post-nasal tokens had no measurable closure, while many intervocalic tokens had clear closure. Nonetheless, the average closure duration was significantly longer for word-initial and post-nasal tokens of /z/ than for intervocalic tokens, and a much higher proportion of intervocalic tokens had zero closure. These results partially support the distributional claim above but make it clear that the allophonic variation involves two statistically different ranges of realizations rather than a categorical distinction between two discrete alternatives.

Session 2pSCb

Speech Communication and ASA Committee on Standards: Speech Intelligibility and AAC Devices

Fredericka Bell-Berti, Cochair
Speech Communication Sciences and Theater, St. John's University, 8000 Utopia Parkway, Jamaica, New York 11439

Susan B. Blaeser, Cochair
Acoustical Society of America Standards Secretariat, 35 Pinelawn Road, Suite 114E, Melville, New York 11747

Chair's Introduction—2:00

Invited Papers

2:05

2pSCb1. Advancements in text-to-speech technology and implications for AAC applications. Ann K. Syrdal (AT&T Labs—Res., 180 Park Ave., Rm. D159, Florham Park, NJ 07932-0971)

Intelligibility was the initial focus in text-to-speech (TTS) research, since it is clearly a necessary condition for the application of the technology. Sufficiently high intelligibility (approximating human speech) has been achieved in the last decade by the better formant-based and concatenative TTS systems. This led to commercially available TTS systems for highly motivated users, particularly the blind and vocally impaired. Some unnatural qualities of TTS were exploited by these users, such as very fast speaking rates and altered pitch ranges for flagging relevant information. Recently, the focus in TTS research has turned to improving naturalness, so that synthetic speech sounds more human and less robotic. Unit selection approaches to concatenative synthesis have dramatically improved TTS quality, although at the cost of larger and more complex systems. This advancement in naturalness has made TTS technology more acceptable to the general public. The vocally impaired appreciate a more natural voice with which to represent themselves when communicating with others. Unit selection TTS does not achieve such high speaking rates as the earlier TTS systems, however, which is a disadvantage to some AAC device users. An important new research emphasis is to improve and increase the range of emotional expressiveness of TTS.

2:35

2pSCb2. Assessment of variation between and within speakers. Ruth Huntley Bahr (Dept. of Commun. Sci. Disord., 4202 E. Fowler Ave., PCD 1017, Univ. of South Florida, Tampa, FL 33620)

While few individuals would argue that vocal cues can signal a person's identity, it is difficult to specify exactly which parameter(s) provide the most salient information for speaker identification. Previous literature has suggested that speaking fundamental frequency, long-term spectra, vowel formant frequencies, and speech tempo can provide speaker-specific information. However, investigations focused on automatic speaker identification have provided less than satisfactory results. These findings could be related to how each acoustic parameter is measured or, more probably, to the idea that these acoustic parameters interact in specific ways that may be more obvious in the perceptual realm and may vary across speaking situations. To further complicate matters, individuals may speak more than one language or use multiple dialects. Little is known about the effect of code switching on voice production and identification. The purpose of this presentation is to present some of the relevant literature on voice recognition and factors related to misidentification. The role of intraspeaker variability will be discussed with a special emphasis on bilingualism and bidialectalism. Implications for voice production in augmentative and alternative communication devices will be described.

3:05

2pSCb3. The effects of four variables on the intelligibility of synthesized sentences. Carol Conroy (New York City Board of Education, Technol. Solutions, District 75, 400 First Ave., New York, NY 10010 and Dept. of Speech, Commun. Sci. & Theatre, St. John's Univ., Jamaica, NY 11439), Lawrence J. Raphael (Adelphi Univ., Garden City, NY 11530), and Fredericka Bell-Berti (St. John's Univ., Jamaica, NY 11439)

The experiments reported here examined the effects of four variables on the intelligibilty of synthetic speech: (1) listener age, (2) listener experience, (3) speech rate, and (4) the presence versus absence of interword pauses. The stimuli, eighty IEEE–Harvard Sentences, were generated by a DynaVox augmentative/alternative communication device equipped with a DECtalk synthesizer. The sentences were presented to four groups of 12 listeners each (children (9–11 years), teens (14–16 years), young adults (20–25 years), and adults (38–45 years). In the first experiment the sentences were heard at four rates: 105, 135, 165, and 195 wpm; in the second experiment half of the sentences (presented at two rates: 135 and 165 wpm), contained 250 ms interword pauses. Conditions in both experiments were counterbalanced and no sentence was presented twice. Results indicated a consistent decrease in error rates with increased exposure to the synthesized speech for all age groups. Error rates also varied inversely with listener age. Effects of rate variation were inconsistent across listener groups and between experiments. The presences versus absences of pauses affected listener groups differently: The youngest listeners had higher error rates, and the older listeners lower error rates when interword pauses were included in the stimuli. [Work supported by St. John's University and New York City Board of Education, Technology Solutions, District 75.]

Session 2pSPa

Signal Processing in Acoustics: Time Reversal and Array Processing

Geoffrey S. Edelson, Chair
BAE Systems, P.O. Box 868, Nashua, New Hampshire 03061-0868

Contributed Papers

1:00

2pSPa1. Time-reversal communications in a hostile, reverberative environment. James V. Candy, Alan W. Meyer, Andrew J. Poggio, and Brian L. Guidry (Univ. of California, Lawrence Livermore Natl. Lab., P.O. Box 808, L-156, Livermore, CA 94551)

Time-reversal (T/R) communications is a new application area motivated by the recent advances in T/R theory. T/R receivers offer an interesting solution to the communications problem for a reverberant channel. This paper describes the performance of various realizations of the T/R receiver for an acoustic communications experiment in air. An experiment is developed to evaluate the performance of point-to-point T/R receivers designed to extract a transmitted information sequence propagating in a hostile, highly reverberant environment. It is demonstrated that T/R receivers are capable of extracting the transmitted coded sequence from noisy microphone sensor measurements with reasonable success. The processing required to validate these experimental results is discussed. These results are also compared with those produced by an equivalent linear equalizer or inverse filter, which provides the optimal solution when it incorporates all of the reverberations.

1:15

2pSPa2. Experimental design and processing for time-reversal communications in a highly reverberant environment. Brian L. Guidry, James V. Candy, Andrew J. Poggio, and Alan W. Meyer (Lawrence Livermore Natl. Lab., 7000 East Ave. L-154, Livermore, CA 94550)

A suite of experiments has recently been conducted to validate the utility of using time-reversal (T/R) theory to solve a communications problem for highly reverberant environments. This paper discusses the design and layout of those experiments as well as experimental equipment criteria and selection. Solutions to problems arising from equipment limitations encountered during the experimental design process are examined. The signal processing used to extract information from gathered data is described and it is shown that communications receivers utilizing T/R theory can be used to accurately reproduce messages broadcast through hostile, reverberant communications channels.

1:30

2pSPa3. Deconvolutional beamforming for air and underwater acoustic sensor arrays. Geoffrey S. Edelson, Shelby F. Sullivan, Jr. (BAE Systems, P.O. Box 868, Nashua, NH 03061-0868, geoffrey.s.edelson@baesystems.com), James M. Alsup (AETC, Inc., San Diego, CA 92122), and Harper J. Whitehouse (Linear Measurements, Inc., San Diego, CA 92121)

Autonomous acoustic sensor arrays are often plagued with poor spatial resolution capabilities, especially at the lowest frequencies. Beamforming techniques that are data-adaptive can improve upon the resolution capability of conventional beamformers under certain conditions. An alternate nonadaptive technique that is simply implemented and that achieves sidelobe reduction while retaining or narrowing main lobe beamwidth is presented. The method for weight matrix construction is based on regularized, constrained deconvolution. These deconvolutional beamformer (DBF) beamspace weights, which are precomputed off-line using this nonlinear process, are applied to the data using a linear process. The physical array may be smaller when using the DBF since this beamformer simultaneously reduces the width of the main lobe at low frequencies while reducing the integrated sidelobe level. This is something even ''optimal'' beamformers have difficulty doing since the main lobe generally increases as the sidelobes are reduced. One additional DBF attribute, when compared to most traditional weighting functions, is its compatibility with nonuniformly spaced arrays, including the special case of receive elements that are known to be missing. The automated procedure for determining these beamspace weights is discussed and results are shown for both air and underwater acoustic array data.

1:45

2pSPa4. Time-reversal super-resolution and stability in a weak scattering medium. Christopher D. Jones (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, cjones@apl.washington.edu) and James G. Berryman (Univ. of California, Lawrence Livermore National Lab., Livermore, CA 94551)

The properties of super-resolution and stability in time-reversal focusing of an acoustic wave are investigated using perturbation theory in a finite scattering medium with weak sound-speed fluctuations. Although super-resolution is obscured by weak scattering of the coherent field, the incoherent field still shows time-reversal enhanced focusing and the analytically tractable results may provide physical insight into time-reversal in stronger scattering media. Using first-order perturbation theory, the correlation scales of focused time-reversed field are defined by the scales of the fluctuations in the medium, the scale of effective aperture of the time-reversal array due to scattering and boundary conditions, the bandwidth of the system, and the array geometry. In this weak scattering example analytic expressions for the first and second moments of the time-reversal Green's function operator are found and provide physical insight into the mechanism of time-reversal focusing and the role of bandwidth in super-resolution and stability. Numerical simulation using discrete point scatterers are performed to compare with the analytic results and to investigate the transitional behavior of time-reversal focusing from weak to stronger scattering.

2:00

2pSPa5. Observations of severe storms with infrasound. John M. Noble and Stephen M. Tenney (Army Res. Lab., 2800 Powder Mill Rd., Adelphi, MD 20783, jnoble@arl.army.mil)

We report on data recorded from small aperture (20-m) infrasonic microphone arrays. The array monitors man-made and natural sources of infrasound from 0.1 to 25 Hz. The array runs 24/7 and stores the data into

hourly blocks. Each block of data is passed through a quick-look algorithm to determine if an interesting event occurred. It was noted about a year after operation began that severe thunderstorms would show up as detections and the array would track the passage of the storm system. The data collected during the operation of the array will be presented along with possible ideas on what the array is detecting within the storm. Also the preliminary results of a field study will be presented where a portable infrasound system was placed near severe storm cells in the Midwest during June 2003.

2:15

2pSPa6. A method for making acoustically bright intensity zone. Joung-Woo Choi and Yang-Hann Kim (Ctr. for Noise and Vib. Control, Dept. of Mech. Eng., KAIST, Science Town, Daejeon-shi 305-701, Korea, yanghannkim@kaist.ac.kr)

Acoustic variables on a selected zone can be manipulated by controlling input signals of the fixed sound sources. In the previous work [J.-W. Choi and Y.-H. Kim, J. Acoust. Soc. Am. **111**, 1695 (2002)], acoustic potential energy was controlled to make a bright and dark sound zone in space. Extending this work, this paper addresses a method to manipulate acoustic intensity in the zone of interest. This inevitably has to do with a magnitude as well as direction control, simply because intensity is a vector quantity. To accomplish this objective, it is required to define the acoustic intensity that can represent acoustic intensity distribution in the selected zone. This has been attempted by defining mean intensity projected to the desired direction. This mean intensity is maximized provided that the input power is kept constant. [Work supported by NRL project of KISTEP and the BK21 project initiated by the Ministry of Education and Human Resources Development of Korea.]

2:30

2pSPa7. Target sound source extraction using spectral subtraction with a two-channel microphone array. Takahiro Murakami and Yoshihisa Ishida (Dept. of Electron. and Commun., Meiji Univ., 1-1-1, Higashi-Mita, Tama-Ku, Kawasaki 214-8571, Japan)

A novel method for extracting a target sound source from the measurements of a two-channel microphone array is presented. In this paper, it is assumed that there is a single target sound source located directly in front of the array and a single noise source located in the other direction. The proposed method exploits the power spectrum of the difference signal between the signals observed at two microphones. Since the arrival times of the target sound are equal and those of the noise source are different, the difference signal contains only the noise signal. In the proposed method, the power spectrum of the difference signal is linearly modified on the assumption that the two microphones are located sufficiently close to each other, and then, the target source is extracted by using a spectral subtraction in which the power spectrum of the difference signal is subtracted from that of the observed signal. The proposed method requires neither delay estimation nor direction-of-arrival estimation, which are generally used for modification of the spectrum of the difference signal. Therefore, the proposed method is performed by a simple algebraic calculation.

2:45

2pSPa8. Software-centered implementation of 128 channel huge speaker array with stock PC. Hiroshi Mizoguchi, Yuki Tamai, Koichi Nagashima (Dept. of Mech. Eng., Tokyo Univ. of Sci., 2641 Yamazaki, Noda 278-8510, Japan), Satoshi Kagami, Tachio Takano (Digital Human Res. Ctr., AIST, Japan), and Koichi Nagashima (R-lab., Inc., Japan)

A huge speaker array system of 128 loudspeakers was constructed and experimented. It was implemented as ''software-centered'' style utilizing stock loudspeakers and a PC. No dedicated hardware nor DSP was utilized. Spot forming, instead of beam forming, could be realized by 32 by 4 square layout of the array. Spot means small area of higher sound pressure level. Number of the spot was not limited to one. In the experiment, within 3 m by 3 m area, four spots of different sounds could be simultaneously formed. This spot forming was confirmed by actually measured spatial distribution of sound pressure level. The effect of the spot was also confirmed auditorily. Since the system was software-centered, it was dynamic. By simply changing software parameters, location of the spot can be easily moved even while the system was running. This movability of the spot was intended to be basis for visual steering. To realize the system, a simultaneous 128 channel 14-bit DA converter PCI board was developed. 44.1 kHz sampling rate was achieved by 2.4-GHz Intel Xeon-based PC utilizing the DA board and a real-time OS, named ART-Linux. Approximately 23-μs loop could be realized by software. It was the world's fastest software loop.

3:00

2pSPa9. Development of a handheld bistatic imaging sonar system for underwater search and survey. Alice Chiang, Steven Broadstone, and John Impagliazzo (Teratech Corp., 77–79 Terrace Hall Ave., Burlington, MA 01803)

A high resolution, handheld imaging sonar system is under development by Teratech Corporation for the U.S. Navy. This is a 192 channel, dual frequency bistatic sonar for Navy divers performing search and survey missions for underwater explosives. Our goal is to provide the most compact and energy efficient imaging system for the divers. The system consists of a self-contained handheld unit and a head mounted display integrated into the divers mask. The low power and small volume are a result of the development of Teratechs Charge Domain Processing (CDP) technology. This technology has led to the development of a low power 64-channel beamformer chip. As a result, only three beamformer chips will be required for the 192 channels. Until now, the implementation of small, low power sonar systems containing this many elements and forming enough beams to create an image was considered impossible. Progress in the development of this product will be presented. In-water testing is planned for late summer 2003. Experimental results and test images available will be presented at the conference. [Work sponsored by ONR and OSD Small Business Innovative Research Program, Program manager, Mr. Bruce Johnson, Naval Explosive Ordnance Disposal Technology Division.]

Session 2pSPb

Signal Processing in Acoustics and Speech Communication: Novel and Hybrid Methods (Poster Session)

Ashok Krishnamurthy, Chair
Department of Electrical Engineering, The Ohio State University, Columbus, Ohio 43210

Contributed Papers

All posters will be on display from 1:00 p.m. to 4:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:00 p.m. to 2:30 p.m. and contributors of even-numbered papers will be at their posters from 2:30 p.m. to 4:00 p.m.

2pSPb1. Speech enhancement using parametric spectral subtraction combined with generalized sidelobe canceller. Jaeyoun Cho and Ashok Krishnamurthy (Dept. of Elec. Eng., The Ohio State Univ., 2015 Neil Ave., Columbus, OH 43210, cho.163@osu.edu)

Speech enhancement is an important problem with applications in hearing aid design, speech recognition, speech coding, etc. Parametric spectral subtraction is a common method for speech enhancement when only a single channel of data is available. On the other hand, beamforming methods can be used when multiple channels of spatially separated data are available, such as from a microphone array. In previous work, we have shown that spectral subtraction combined with spatial averaging from multiple microphones leads to improvements in speech SNR and reduction of musical noise compared with either method used alone. In this talk, we extend the previous work to combine parametric spectral subtraction with adaptive beamforming, specifically the generalized sidelobe canceller. The proposed parametric spectral subtraction method determines the parameters adaptively so as to minimize speech distortion. In addition, it is shown that the major drawback of spectral subtraction, so-called musical noise, can be diminished by adaptive beamforming process. We show that the method leads to a reduction of musical noise and results in the enhanced speech having better quality and intelligibility.

2pSPb2. A speech enhancement method using the sliding DFT. Hiroyuki Ono, Takahiro Murakami, and Yoshihisa Ishida (Dept. of Electron. and Commun., Meiji Univ., 1-1-1, Higashi-mita, Tama-ku, Kawasaki 214-8571, Japan)

In this paper, we propose a novel noise reduction of speech signals using sliding DFT (SDFT) which is based on the spectral subtraction method. The spectral subtraction method reduces noise by subtracting the spectrum of noise estimated from an input signal and is a relatively simple and effective technique. We use SDFT to obtain the frequency spectra of speech signals. As is well known, to obtain the successive N-points output, SDFT requires only N times of complex multiplication, although the Fast Fourier Transform (FFT) requires $N\log 2N$ times that. Thus SDFT is computationally simple as compared with FFT. The frame length, N, is determined based on the pitch period. In this paper, the pitch period is estimated by using an ideal low-pass filter and the analytic signal that is obtained from speech signals by the Hilbert converter. Finally, it is shown that the proposed method has good performance for speech enhancement.

2pSPb3. Spoken digits recognition using DP matching combined with a subspace decomposition method. Ken Kusakari, Kurihara Kiyoshi, Takahiro Murakami, and Yoshihisa Ishida (Dept. of Electron. and Commun., Meiji Univ., 1-1-1, Higashi-mita, Tama-ku Kawasaki 214-8571, Japan)

In this paper, we propose a method for spoken digits recognition using DP Matching combined with subspace decomposition that linearly separates into phonetic information from speaker information based on principle component analysis [M. Nishida and Y. Ariki, IEICE Trans. Japan **J85-D-II**, No. 4 (2002)]. This method allows for more robust speech recognition of less standard speech patterns. The use of the spectral envelope by LPC in speech recognition is unable to avoid errors in recognition due to the uncertainty of personalities, the dynamic variation of features, and so on. By using the subspace method, the proposed method eliminates these problems and enables good recognition results of less standard speech patterns. We use DP matching in recognizing, because it allows for more efficient pattern matching by normalizing the length of vowels. Simulation results show that the proposed method, using orthonormal projection to phonetic subspace with less speaker information, is superior to the conventional method using LPC spectra and DP Matching.

2pSPb4. Voice conversion using the radial basis function network and the frequency–domain pitch modification technique. Tomomi Watanabe, Takahiro Murakami, and Yoshihisa Ishida (Dept. of Electron. and Commun., Meiji Univ., 1-1-1, Higashi-mita, Tama-ku, Kawasaki 214-8571, Japan)

This paper presents a novel algorithm for developing a voice conversion system that modifies the speech uttered by a speaker to sound as if produced by another target speaker. And an improved technique for a frequency–domain pitch modification is proposed. The approach is based on modifying spectral envelopes and the pitch of voiced sounds separately. The conversion rules for the spectral envelopes are constructed by the radial basis function (RBF) network, which is one of the well-known artificial neural networks. An excitation signal containing prosodic information is modified to match the average pitch using the proposed technique which proves that the speech quality deteriorates when the pitch is modulated by the Frequency–Domain Pitch-Synchronous Overlap and Add (FD-PSOLA) algorithm. The simulation results show that the proposed method achieves nearly optimal spectral conversion performance.

Moreover, average cepstrum distance to the target speech is reduced by 87%, and the listening tests prove the proposed pitch modification technique maintains higher speech quality than the original FD-PSOLA.

2pSPb5. Musical noise reduction using an adaptive filter. Takeshi Hanada, Takahiro Murakami, Yoshihisa Ishida (Dept. of Electron. and Commun., Meiji Univ., 1-1-1, Higashi-mita, Tama-ku, Kawasaki 214-8571, Japan), and Tetsuya Hoya (Lab. for Adv. Brain Signal Processing BSI-RIKEN 2-1, Hirosawa, Wakoh, Saitama 351-0198, Japan)

This paper presents a method for reducing a particular noise (musical noise). The musical noise is artificially produced by Spectral Subtraction (SS), which is one of the most conventional methods for speech enhancement. The musical noise is the tin-like sound and annoying in human auditory. We know that the duration of the musical noise is considerably short in comparison with that of speech, and that the frequency components of the musical noise are random and isolated. In the ordinary SS-based methods, the musical noise is removed by the post-processing. However, the output of the ordinary post-processing is delayed since the post-processing uses the succeeding frames. In order to improve this problem, we propose a novel method using an adaptive filter. In the proposed system, the observed noisy signal is used as the input signal to the adaptive filter and the output of SS is used as the reference signal. In this paper we exploit the normalized LMS (Least Mean Square) algorithm for the adaptive filter. Simulation results show that the proposed method has improved the intelligibility of the enhanced speech in comparison with the conventional method.

TUESDAY AFTERNOON, 11 NOVEMBER 2003 SABINE ROOM, 1:00 TO 5:00 P.M.

Session 2pUW

Underwater Acoustics, Acoustical Oceanography and Physical Acoustics: The Acoustics of Bubbles in the Marine Boundary Layer

Grant B. Deane, Chair
Marine Physical Laboratory, University of California–San Diego, La Jolla, California 92093-0238

Chair's Introduction—1:00

Invited Papers

1:05

2pUW1. Scattering and attenuation from near-surface bubbles. Peter H. Dahl (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

Scattering from bubbles in the vicinity of the air–sea interface can often be the dominant source of apparent sea surface backscatter, particularly for lower grazing angles and for frequencies of O(10) kHz and above. Such scattering, and by inference, the near-surface bubble concentration, is observed in field data to be highly correlated with a nonlinear function of wind speed, such as a power or exponential law. A hysteresis effect is also observed, wherein for a given wind speed there is a tendency for scattering level to be higher if prior winds had been falling. The situation changes for acquisition geometries far removed from monostatic or bistatic backscattering; for example, in the case of bistatic forward scattering, the scattering level from bubbles rarely reaches the strength of that from the air–sea interface. Here, though, a bubble-mediated attenuation can be observed in field data, which is also wind-speed dependent with hysteresis. However, linking together field measurements of scattering and attenuation associated with near-surface bubbles has long been a vexing problem. This topic is discussed using examples from various field experiments, the most recent of which is ASIAEX East China Sea (2001), and from well-controlled laboratory experiments. [Research supported by ONR.]

1:25

2pUW2. The impact of bubbles on underwater acoustic communications in shallow water environments. James Preisig (Dept. of Appl. Ocean Phys. and Eng., WHOI, Woods Hole, MA 02543)

Two characteristics of the acoustic channel that significantly impact the performance of underwater acoustic communications systems are the total energy throughput of the channel and the delay spread of the channel impulse response. These characteristics impact both coherent (e.g., phase shift keyed) and incoherent (e.g., frequency shift keyed) systems. Experimental data are presented from the surf zone and shallow-water environments demonstrating the impact of bubbles on these characteristics. A novel statistical analysis utilizing a *variance reduction ratio* is used to identify the measurable environmental variables (e.g., significant wave height) that can be used to best predict the impact of the bubbles. [Work supported by ONR Ocean Acoustics.]

1:45

2pUW3. The noise spectral parameters and the energy of breaking waves experimental study. Jaroslaw Tegowski (Inst. of Oceanology, Polish Acad. of Sci., Sopot, Poland, tegowski@iopan.gda.pl)

The process of wave breaking and whitecap creation is one of the most important and least understood phenomena associated with the evolution of the surface gravity waves in the open sea. This process is the main way of energy and momentum transfer between ocean and atmosphere. However, it is very difficult to estimate, under real sea conditions, the frequency of breaking wave events or

the fraction of sea surface covered by whitecaps and the amount of dissipated energy produced by wave breaking. A controlled experiment was carried out in the Ocean Basin Laboratory at MARINTEK, Trondheim (Norway). The simulation of random waves of the prescribed spectra provided a very realistic pattern of the sea surface. The number of breaking waves was estimated using photography method and wave staff recording. Acoustic measurements during the experiments were conducted in order to examine the relationship between the noise spectral parameters and both the whitecap coverage and dissipation energy of breaking waves for different types of waves. A comparison of simultaneous video observations, wave staff records of the surface wave above the hydrophones, with the spectral parameters of acoustical signals made it possible to find physical links between processes.

2:05

2pUW4. Propagation and scattering in bubbly liquids near resonance: A review of recent laboratory measurements. Preston S. Wilson, Eun-Joo Park, William M. Carey, and Ronald A. Roy (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

Acoustic waves propagating through bubbly liquid experience dispersion and attenuation when the excitation frequency is near the resonance frequency region of the bubble distribution. Due to high attenuation in this regime, traditional standing wave and time-of-flight measurement techniques fail and existing theories remain unverified above void fractions of about 10^{-4}. An impedance tube technique was developed in order to investigate this regime by inferring phase speed and attenuation within a bubbly layer from its complex reflection coefficient. Results of experiments using this technique, as well as its limitations, will be reviewed. We found that existing theory described the measured results near the resonance region within the uncertainty of the measured parameters of the bubble population. In a second class of experiments, scattering from laboratory bubble clouds was investigated. Here, the bubbles were much smaller than resonance size and the bubbly liquid was treated as an effective medium, with density given by a mixture law and sound speed given by Wood's equation. Experimental results and the predictions of an effective fluid scattering model will be reviewed for spherical and cylindrical bubbly-liquid targets. [Work supported by U.S. Navy ONR.]

2:25

2pUW5. Laboratory measurements of the 1st and 2nd moments of propagation through bubbles in a flow. Thomas Weber, David Bradley, R. Lee Culver, and Anthony Lyons (Appl. Res. Lab. and the Grad. Prog. in Acoust., P.O. Box 30, State College, PA 16804, tcw141@psu.edu)

The classic theory describing the change in soundspeed and attenuation for waves propagating in bubbly fluids (i.e., the 1st moment of the multiple scattering solution) can be viewed as a statement about the average pressure observed over all possible configurations of bubble position and size. The 2nd moment can also be predicted using the multiple scattering formulation, and both moments can be found independent of any knowledge of the fluid dynamic conditions controlling the mixing of bubbles. Characteristics of the fluid dynamics do appear in the time scales associated with this mixing, however, and this should be reflected in the time scales associated with convergence to the 1st and 2nd moments. This was explored by conducting propagation experiments in a tank filled with bubble-laden water that was subjected to varying flow conditions. Frequency dependent attenuation measurements were made and inverted for the bubble size distribution so that predictions of the 2nd moment could be compared to the observed quantities. High ping rates and long data records were used to observe the flow-dependent time scales associated with the measurements. The results of these experiments will be presented and discussed. [Work supported by ONR Code 321US. First author supported by a National Defense Science and Engineering Graduate Fellowhip.]

2:45

2pUW6. Listening for ambient bubbles in the marine boundary layer. Jeffrey A. Nystuen (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

Bubbles created by breaking waves and raindrop splashes are responsible for much of the high-frequency (500 Hz–50 000 Hz) ambient sound in the ocean. Under high sea state and during heavy rainfall conditions clouds and layers of ambient bubbles in the marine boundary layer absorb sound from bubbles being newly created at the ocean surface. This absorption changes the shape of the ambient sound spectrum below the bubbly surface marine layer, allowing passive detection of the ambient bubbles. This change in the shape of the ambient sound spectrum is especially apparent in the situation of rainfall during high wind conditions, presumably because existing wind-generated turbulent from breaking waves is available to transport rainfall-generated bubbles downward. Estimates of vertically integrated ambient bubble populations can be made from the presumed frequency-dependent attenuation in the ambient sound spectrum. Ambient sound data from many months of deep ocean mooring measurements will be used to document this phenomenon. [Work sponsored by ONR Ocean Acoustics.]

3:05–3:15 Break

Contributed Papers

3:15

2pUW7. Signal propagation and the dispersion relations of bubbly water. Gregory J. Orris, Michael Nicholas, and Dalcio Dacol (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375)

The physics behind the extreme variations of the phase velocity of acoustic wave propagation in a bubbly liquid as a function of frequency have been addressed theoretically. Until recently there has not been a concerted experimental effort to validate this work over the rather large parameter landscape covered by the theories. To this end there there have been several sets of experiments conducted at the Naval Research Laboratory's Salt-Water Tank Facility aimed in part at validating these theories. In order to properly measure phase velocities in a semi-free field environment, one must be able to accurately measure the arrival of a short acoustic pulse. This leads naturally to the question of what one measures as the velocity of sound in a highly dispersive medium where the group velocity

can indeed have infinite, zero, or negative values. We discuss this and its relationship to work in other fields of physics. [Work supported by ONR.]

3:30

2pUW8. Measurements of the attenuation and phase velocities in bubbly fresh and salt water. Michael Nicholas and Gregory J. Orris (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375)

Several types of bubblers have been used to create a dense bubble cloud that fills a 44-m^3 tank, wherein phase velocity and attenuation measurements have been made over a wide range of frequencies. Initial experimental results on the measurement of the phase velocity and attenuation have been previously reported [J. Acoust. Soc. Am. **112**, 2269]. We present in this paper how this work has been extended with respect to the environmental parameter landscape. Paramount among these parameters are the bubble size distribution, void fraction, temperature, and surface tension. In these sets of experiments, void fractions from 0.5% to roughly 4% are investigated over a wide set of temperatures to give a much larger experimental picture of acoustic propagation through near-surface bubble clouds. Current techniques for measuring these quantities will be discussed and results presented within the context of current theories with implications on ocean acoustic experiments. Additionally, initial work using data collected using various concentrations of salt water will also be presented. [Work supported by ONR.]

3:45

2pUW9. Pulse distortion of a signal passing through a bubble cloud. Ralph R. Goodman and Jerald W. Caruthers (Dept. of Marine Sci., The Univ. of Southern Mississippi, Stennis Space Center, MS 39529)

At the 145th Meeting of the Acoustical Society the authors presented a paper on the dispersion and absorption of sound in an idealized bubble field that had a bubble radius size distribution $n(a)$ decreasing as the fourth power [J. Acoust. Soc. Am. **113**, 1278 (2003)]. It was shown that, for this case, exact, relatively simple, analytic solutions exist both for dispersion and absorption. The same model is used here to calculate the effects of dispersion and absorption on pulsed signals as they propagate through a bubbly mixture. The approach is essentially the same as that used by Frank Henyey (unpublished) to analyze the propagation of signals through clouds that were observed in the well-known 1997 Scripps Pier experiment. The effects on signal shapes as a function of range and bubble density and size distribution will be presented.

4:00

2pUW10. Scaling laws for bubble acoustic radiation strength in whitecaps. Grant B. Deane (Scripps Inst. of Oceanogr., Mail Code 0238, La Jolla, CA 92093-0238, grant@mpl.ucsd.edu)

The past decade has seen significant advances in our understanding of open-ocean bubble plumes, including scaling laws for whitecap properties [W. K. Melville and P. Matusov, Nature (London) **417**, 58–63 (2002)] and the physics of bubble creation [Garrett *et al.*, J. Phys. Oceanogr. **30**, 2163–2171 (2000), and G. B. Deane and M. D. Stokes, Nature (London) **418**, 839–844 (2002)]. Despite this progress, we do not yet have a complete understanding of the relationship between noise radiated by whitecaps and the bubble creation processes occurring within the wave crest. Understanding this relationship is important: it would allow bubble-mediated gas transfer rates to be estimated from wind speed, for example, and provide a model for the Knudsen spectrum of underwater ambient noise. An important piece of this puzzle is the dependence of bubble acoustic radiation strength on bubble radius. We will present estimates of the scaling laws for this relationship based on open-ocean measurements of whitecap noise and laboratory measurements of bubble creation rates in plunging breakers. [Work supported by ONR and NSF.]

4:15

2pUW11. An experiment on near-shore sonar performance. Nicolas Le Dantec and Grant B. Deane (Scripps Inst. of Oceanogr., Univ. of California–San Diego, 9500 Gilman Dr., La Jolla, CA 92093)

In order to further investigate the use of sonar systems in the near-shore environment, a simple, short range sonar system has been deployed along with additional devices to characterize the environmental conditions. The goal of the experiment is to determine the effects of volume and boundary conditions on the performance limits of sonar in the near-shore. A 10 kHz source transmitted a series of acoustic pulses with Barker code phase modulation along a 40 m propagation path to spherical targets just outside the surf zone north of Scripps Pier. Both forward and backscatter acoustic signals were recorded. A pressure array mounted on the seafloor measured the shoaling surface waves along the transmission path and the seawater conductivity, temperature and horizontal velocity was measured mid-way between the source and targets. Meteorological data were also collected. The effects of surface gravity waves focusing and multi-path arrivals on target discrimination will be addressed, along with scattering from artificially created bubble clouds. [Work supported by ONR.]

4:30

2pUW12. Bubble acoustics near surfaces. Dalcio K. Dacol and Gregory J. Orris (Acoust. Div., Naval Res. Lab., Washington, DC 20375)

Boundary effects were included in Foldy's effective medium theory of acoustic propagation in a bubbly fluid. This allowed for the discussion of bubble effects on propagation near interfaces and in waveguides such as the ones encountered in ocean acoustics. The implications of this model were examined in a variety of situations from laboratory experiments to acoustic propagation in shallow waters.

4:45

2pUW13. Evaluation of bubble–saturated subsurface sea layer by means of optoacoustics. Sergey V. Egerev (Sci. Council on Acoust., Russian Acad. of Sci., 4 Shvernika St., Moscow 117036, Russia)

Laser sound generation has proved to be a useful tool for probing of various media. The pressure signal outgoing from the point of optoacoustic conversion provides information on the medium properties. The results are presented concerning optoacoustic conversion in the gas-containing bubbly subsurface sea layer. The sea surface is irradiated by a laser pulse. The subsequent acoustic signal, very susceptible to the subsurface gas content and propagating along the vertical axis z is recorded by a submerged hydrophone and interpreted. A model is developed to predict the features of signal, recorded during full–scale experiment conditions, making clear the formation of signal in nonuniform bubble–saturated subsurface layer, characterized by strong dispersion and additional acoustical absorption. The sound velocity increment $\Delta c(z)$ along with the additional attenuation factor $\alpha(z)$ are parameters of a dispersion environment dependent upon the distribution of numerical bubble concentration $n(R,z)$ upon their radii R at a current depth z. Laboratory test results of the phenomena along with the White Sea full-scale experimental results are presented. The equipment issues necessary for such measurements are discussed.

Session 3aAA

Architectural Acoustics, Noise and Speech Communication: Forensic Acoustics I

Jack E. Randorff, Cochair
Randorff and Associates, Inc., 11 West Canyon View Drive, Ransom Canyon, Texas 79366

Peter Ladefoged, Cochair
Linguistics Department, University of California–Los Angeles, Los Angeles, California 90095-1543

Chair's Introduction—9:00

Invited Papers

9:05

3aAA1. Forensic acoustics: An overview of the process. J. T. Weissenburger (Eng. Dynam. Intl., 8420 Delmar Blvd., St. Louis, MO 63124)

There is a potential role for the acoustical expert in litigation. The technical issues may involve aeroacoustics, underwater acoustics, physical effects of sound, environmental acoustics, noise, architectural acoustics, physiological acoustics, speech and hearing, music, psychoacoustics and/or bioacoustics. This brief paper offers an overview of the process of being an expert, the qualifications to be an expert and what is expected of an expert. The six general phases of an expert's involvement—retention, investigation, discovery, deposition, preparation, trial—are addressed. Some antidotal experiences are presented.

9:30

3aAA2. The sounds of a murder. Richard J. Peppin (Scantek, Inc., 7060 Oakland Mills Rd. #L, Columbia, MD 20146, PeppinR@ASME.org)

Often engineers and lawyers cannot communicate, in spite of repeated attempts. The lawyer has an idea and wants the engineer to prove it in front of a jury. As examples: a quiet, or briefly loud source must be shown to cause hearing damage, or a construction project in a backyard must be shown to be nonannoying. Often it is a no brainer, either way. But the testimony must be given! In this paper, I discuss a sad case. A young woman and her baby daughter were murdered. A witness claimed she heard something in the dead of night. If so, it was further evidence of guilt of the accused. If not, it was evidence of the lack of credibility of the witness and helped show innocence. I present the results of a forensic investigation of a very brutal murder based on acoustics of the victims' screams, the structure housing the murder, and the witness. The results of the investigation attempted to help the case.

9:55

3aAA3. Epidemiologic methods in analysis of scientific issues. Linda S. Erdreich (Exponent Health/Epidemiology Practice, 420 Lexington Ave., Ste. 408, New York, NY 10170)

Studies of human populations provide much of the information that is used to evaluate compensation cases for hearing loss, including rates of hearing loss by age, and dose–response relationships. The reference data used to make decisions regarding workman's compensation is based on epidemiologic studies of cohorts of workers exposed to various noise levels. Epidemiology and its methods can be used in other ways in the courtroom; to assess the merits of a complaint, to support Daubert criteria, and to explain scientific issues to the trier of fact, generally a layperson. Using examples other than occupational noise induced hearing loss, these methods will be applied to respond to a complaint that hearing loss followed exposure to a sudden noise, a medication, or an occupational chemical, and thus was caused by said exposure. The standard criteria for assessing the weight of the evidence, and epidemiologic criteria for causality show the limits of such anecdotal data and incorporate quantitative and temporal issues. Reports of clusters of cases are also intuitively convincing to juries. Epidemiologic methods provide a scientific approach to assess whether rates of the outcome are indeed increased, and the extent to which increased rates provide evidence for causality.

10:20

3aAA4. Analysis of audiometric database shows evidence of employee fraud. John Erdreich (Ostergaard Acoust. Assoc., 200 Executive Dr., West Orange, NJ 07052)

Following a lengthy strike, several hundred delivery drivers filed workers compensation claims for occupational hearing loss. We were asked to evaluate the noise exposure of the drivers during their in-plant tasks. In-plant exposures were not predictive of any hearing loss. A comparison of audiometric data for the claimants revealed consistent hearing loss independent of duration of employment or age. These discrepancies between observations and common understanding of dose–response relationships between noise

exposure and hearing loss led to further investigation, ultimately resulting in the dismissal of all claims against the employer who then filed an action against the claimant's attorneys and physician under the Racketeering in Corrupt Organizations Act (RICO). The details of the legal complaint, which reads like a detective novel, can be found at the United States District Court for the Southern District of New York [93 Civ. 7222 (LAP)].

10:45

3aAA5. Legal versus technical evidence of warning signal effectiveness. Sanford Fidell (Fidell Assoc., Inc., 23139 Erwin St., Woodland Hills, CA 91367)

A vast gulf separates technically from legally compelling proof. Attorneys (if not juries) crave deterministic certainty, simplicity, and self-assurance in expression of technical opinions, while scientists are more accustomed to discussing bandwidth-adjusted signal to noise ratios, working hypotheses, confidence intervals, and full disclosure of the limitations of data collection and theoretical interpretation. Given the many constraints imposed by the adversarial nature of civil litigation, presenting a jury with technically meaningful evidence about the effectiveness (or ineffectiveness) of acoustic warnings can be a formidable challenge. Favorable factual information and genuine expertise in interpreting it are helpful, but there are no guarantees that a jury can be persuaded that a technical opinion is correct.

11:10

3aAA6. A method to prove violation of building code sound insulation standards in small condominiums. David Lubman (David Lubman & Assoc., 14301 Middletown Ln., Westminster, CA 92683)

Building code sound insulation standards are health measures intended to protect people from excessive noise. In the USA, minimum values for Field Sound Transmission Loss (FSTC), and Field Impact Insulation Class (FIIC) are widely used to rate noise insulation between adjacent condominiums. Current American Society for Testing and Materials (ASTM) field test standards specify minimum room volumes for each of 16 one-third octave band test frequencies. For field impact insulation class ratings (FIIC), ASTM E 1007 specifies a minimum room volume of 2100 ft^3 for valid measurement in the 100-Hz band. For air-borne sound transmission class ratings (FSTC), ASTM E 336 specifies minimum room volumes of 1400 ft^3 for valid measurement in the 125-Hz band. FIIC and FSTC cannot be determined for bedrooms and kitchens of smaller condominiums which fail the minimum size requirement for the lowest bands. However, upper bounds can be determined merely by excluding invalid lower bands from computations. It is shown that when the upper bound on FIIC or FSTC falls below the code requirement (field values of 45 under the Uniform Building Code and the International Building Code), one may logically conclude that the code requirement is violated under existing standards.

WEDNESDAY MORNING, 12 NOVEMBER 2003 PECOS ROOM, 8:30 A.M. TO 12:00 NOON

Session 3aAO

Acoustical Oceanography: Acoustic Remote Sensing

Aaron M. Thode, Chair

Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238

Contributed Papers

8:30

3aAO1. Integrated acoustics systems for ocean observatories. Bruce M. Howe (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105) and James H. Miller (Univ. of Rhode Island, Narragansett, RI 02882)

Integrated acoustics systems providing navigation and communications and conducting acoustic measurements in support of science applications is, in concept, analogous to the Global Positioning System, but relies on acoustics because the ocean is opaque to electromagnetic waves and transparent to sound. A series of nested systems is envisioned, from small- to regional- to basin-scale. A small number of acoustic sources sending coded, low power signals can service unlimited numbers of inexpensive receivers. Drifting floats with receivers can be tracked accurately while collecting ocean circulation and heat content data, as well as ambient sound data about wind, rain, marine mammals, and seismic activity. The sources can double as transmitters of control data from users to remote instruments; if enabled as receivers, two-way acoustic communications links in large-scale networks can be established. Acoustic-based instrumentation that shares the acoustic bandwidth with, and depends upon, the navigation and communications capabilities completes the concept of integrated acoustic systems.

8:45

3aAO2. Acoustic detectability of squid egg beds. Kenneth G. Foote (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), Roger T. Hanlon (Marine Biological Lab., Woods Hole, MA 02543), Annette E. Henry (California Dept. of Fish and Game, La Jolla, CA 92037), Alfred Hochstaedter (Monterey Peninsula College, Monterey, CA 93940), Rikk Kvitek (California State Univ., Monterey, CA), Deidre Sullivan and Yuko Yogozawa (Monterey Peninsula College, Monterey, CA 93940)

Egg beds of the market squid (*Loligo opalescens*) on the bottom of Monterey Bay seem to have been detected by means of sidescan sonar at 420 kHz. Evidence for this is presented in the form of sidescan sonar

images and egg-bed distribution maps from the same area, as prepared from camera surveys by scuba divers. The general detectability issue is also considered, with specific reference made to preliminary physical measurements performed on two egg capsules. [Work supported by Sea Grant.]

9:00

3aAO3. Fish schooling behavior inferred from differences between backscatter levels in Doppler sonar beams. Len Zedel (Phys. Dept., Memorial Univ., St. John's, NF A1B 3X7, Canada)

The general availability of Doppler profilers on survey ships provides a convenient source of acoustic backscatter data. Aside from calibration issues, caution must be exercised when analyzing this data because each of the diverging beams has a different interaction angle with the scatterers. In particular, for targets with directional scattering characteristics a different backscatter strength will be seen in each beam. The availability of data from multiple beam directions can however provide information on scatterer orientation when groups of such scatterers undertake coherent motion. This effect is demonstrated in observations of Norwegian spring spawning herring (*Clupea harengus*). In data collected while these fish are actively migrating with mean swimming speeds of 20 cm s^{-1}, a difference of 5 dB is seen in volume backscatter strength depending on the direction of fish movement with respect to the acoustic beams. In contrast, in data collected when these fish schools have less well defined movements with a mean swimming speed of less than 10 cm s^{-1}, a difference of less than 1 dB is seen.

9:15

3aAO4. Field trial of a Doppler sonar system for fisheries applications. Cristina D. S. Tollefsen and Len Zedel (Dept. of Phys. and Physical Oceanogr., Memorial Univ. of New Foundland, St. John's, NF A1B 3X7, Canada, cristina@physics.mun.ca)

Various deployments of commercial Doppler current profiling systems have demonstrated that these instruments can detect fish and measure their swimming speeds. However, research into the possible application of Doppler sonar to fisheries problems is limited and has not taken advantage of coherent signal processing schemes. A field trial was undertaken in August 2002 to explore the capabilities of a coherent Doppler sonar when applied to detecting discrete targets. The passage of migrating salmon on the Fraser River in British Columbia provided an ideal test opportunity with fish of well-defined swimming behavior and allowed for comparisons with conventional fisheries acoustics techniques. The instrument tested was a 250-kHz sonar which provided for phase coding of transmit pulses and coherent sampling of successive acoustic returns. The field trial resulted in 11 consecutive days of Doppler sonar data acquired during the peak of the sockeye salmon (*Oncorhynchus nerka*) migration. A total of 7425 individual fish were identified and their swimming speed was measured with an accuracy of between 10 cm s^{-1} and 20 cm s^{-1}, which depended on pulse length, pulse spacing, and target range. By comparison, water velocity measurements made with the same instrument can only achieve a theoretical accuracy of 60 cm s^{-1}.

9:30

3aAO5. Fish schools are the dominant cause of long-range active sonar clutter in the New Jersey Continental Shelf: Quantitative correlations. Deanelle T. Symonds, Purnima Ratilal (MIT, 77 Massachusetts Ave., Cambridge, MA 02139), Redwood W. Nero (Naval Res. Lab., Stennis Space Center, MS 39529), and Nicholas C. Makris (MIT, Cambridge, MA 02139)

Long-range underwater acoustic remote sensing data, acquired using a low- to mid-frequency bistatic sonar system, is compared with data from a downward-directed, fisheries-standard high-frequency fish-finding sonar. These data were simultaneously measured in the New Jersey Continental Shelf environment during the Main Acoustic Clutter Experiment 2003. The long-range sonar is capable of imaging extensive areas, spanning tens of kilometers in range, in near real-time. This makes it possible to continuously observe the spatial and temporal variability of environmental returns over wide areas. In contrast, the fish-finding sonar's single, downward-directed beam typically provides a swath of only 10 m width along the ship track. Dense populations of fish, indicated by prominent returns from the fish-finding sonar, are overlain onto long-range acoustic images. The outputs of the two systems are also directly correlated along the track of the fish-finding sonar. High correlation was found between locations of prominent environmental clutter in the long-range sonar system and locations of densely clustered fish schools measured with the fish-finding sonar. Additionally, regions absent of clutter were also found to be absent of significant fish populations. These correlations were observed repeatedly over a period of roughly 3 weeks.

9:45

3aAO6. Long-range acoustic imaging of the Continental Shelf Environment reveals massive fish schools: 2003 Main Acoustic Clutter Experiment. Nicholas C. Makris, Purnima Ratilal, Yisan Lai, Sunwoong Lee, Deanelle T. Symonds, Lilimar A. Ruhlmann (MIT, 77 Massachusetts Ave., Cambridge, MA 02139), Redwood W. Nero (Naval Res. Lab, Stennis Space Center, MS 39529), John R. Preston (Appl. Res. Lab, Penn State Univ., State College, PA 16804), Edward K. Scheer (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), and Michael T. Sundvik (Naval Undersea Warfare Ctr., Newport, RI 02841)

Results from the three-ship, April–May 2003 Main Acoustics Experiment of the ONR Acoustics Clutter Program are presented. A long-range bistatic sonar system was used to image extensive clutter over wide areas of the New Jersey Continental Shelf in the low- to mid-frequency range in near real-time. A downward-directed high-frequency fish-finding sonar was also simultaneously operated from one of the ships to determine whether fish schools occur in the location of the clutter events. Tracks of the long-range sonar were repeated over days throughout the experiment. Most of the clutter features were observed to evolve both in time and in space moving throughout the survey area and therefore do not consistently correlate with static geologic features. The clutter features were observed to cluster, disperse, and gradually disappear and then re-emerge in the general vicinity at later times. The long-range active sonar data show that the primary source of clutter corresponds to objects moving in the waveguide. This was made certain by our ability to accurately register a number of fixed point targets deployed as controls throughout the experiment. The preliminary finding of this experiment is that the dominant source of clutter in this Continental Shelf environment is marine life, particularly fish schools.

10:00–10:15 Break

10:15

3aAO7. Long-term, large-scale acoustic fluctuations in the Ulleung Basin. Donald R. DelBalzo, LauriAne Winsett, and Erik R. Rike (Neptune Sci., Inc., 40201 Hwy. 190 E., Slidell, LA 70461)

An analysis of long-term, large-scale oceanographic conditions in the Southern Japan/East Sea has identified several interesting phenomena, including a previously undiscovered stable eddy. That analysis was based on temperature profile data observed daily with 23 PIES (pressure inverted echo sounder) instruments spaced at about 40 km throughout the Ulleung Basin during a 2-year period. In the present work, a gridded version of the PIES data was used to study the relation between oceanographic structure and acoustic transmission loss structure from 50 to 500 Hz throughout the area in terms of long-term fluctuations on the order of days to weeks. Significant correlations between oceanographic and acoustic variability were observed with different characteristics during summer and winter. The results indicate predictably stable acoustic paths that correlate with benign ocean conditions and predictably unstable acoustic paths that correlate with dynamic ocean features. The implication is that there are frequencies and sensor geometries that favor long-term acoustic monitoring with controlled uncertainty and that certain stable scenarios preclude the need for daily temperature sampling. [Work sponsored by ONR.]

10:30

3aAO8. Gulf of Mexico oceanography and acoustic-pulse propagation across the DeSoto Canyon. Jerald W. Caruthers, Sergey Vinogradov, Nadya Vinogradova (Dept. of Marine Sci., Univ. of Southern Mississippi, 1020 Balch Blvd., Stennis Space Center, MS 39529), Natalia A. Sidorovskaia (Univ. of Louisiana at Lafayette, Lafayette, LA 70505-4210), George E. Ioup, Juliette W. Ioup (Univ. of New Orleans, New Orleans, LA 70148), and Ilya Udovydchenkov (Univ. of Miami)

To investigate the feasibility of monitoring the movements of the large eddies that dominate circulation in the deep waters of the Gulf of Mexico, acoustic-propagation studies are made using historic and ocean dynamic model data for sound-speed profiles representing the distinctions among the various water masses inside and outside these eddies. The propagation work concentrates on a slice across the DeSoto Canyon in the northeast Gulf from the shelf break off the Mississippi Delta to the Florida shelf—a region in which a deep sound channel exists. The objective for the study of this slice is to determine the potential for monitoring the movement and evolution of eddies as they enter DeSoto Canyon. Two frequency bands of a pulse source (400–600 and 900–1100 Hz) and deep water source placements near the sound-channel axis (900 m) and a vertical array of receivers are used in modeling. Details of the temporal structure in the arrivals show significant variations for the different waters of the order of $0=2E1$ s. Transmission losses in the deep-water channel were of the order of 100 dB over the approximate 200-km propagation path used in the modeling, which gives an estimate of the requirements for the source-receiving system. [Work supported in part by the Littoral Acoustic Demonstration Center sponsored by ONR.]

10:45

3aAO9. The influence of mesoscale eddies on shallow water acoustic propagation. Harry DeFerrari and Donald Olson (RSMAS—Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149)

Acoustic propagation measurements in 150 m depth on the Florida escarpment observe the effects of the passage of a cyclonic eddy. As the stream core of the Florida Current meanders, the eddy is formed and propagates along the shelf edge. The sequence over a roughly a fortnight is as follows: ahead of the eddy, warm surface water and cold bottom water are swept onto the terrace forming a steep thermocline and corresponding strong downward refracting $C(z)$. The gradient produce intense, focused RBR arrivals and the thermocline becomes a duct for internal waves to propagate shoreward. At first, the internal wave energy is minimal and propagation is stable and coherent. As the internal tides attempt to propagate on shelf, the sound speed field and the acoustic signals become increasingly variable. The variability reaches a crescendo as the 200 m long internal tide is blocked from propagating on to the narrower shelf and begins to break and overturn producing small-scale variability. As the eddy passes, nearly iso-thermal conditions are restored along with quiescent internal wave fields and reduced signal variability. Here, the effects are quantized with data from fixed-system acoustic and oceanographic measurements demonstrating that the mesoscale determines acoustic propagation conditions days in advance.

11:00

3aAO10. Angular dependence of fluctuations in acoustic transmissions relative to tidal direction in the SWAT environment. Jacob George and Robert L. Field (NRL Code 7185, Stennis Space Center, MS 39529)

Recently, fluctuations in continuous wave acoustic transmissions through internal tides that are periodic and have a constant amplitude in range have been studied [J. George and R. L. Field, J. Acoust. Soc. Am. **113**, 2333 (2003)]. It was demonstrated through WKB/Pekeris calculations that the fluctuations are significantly greater when the transmission is nearly perpendicular to the tidal direction, and decrease when the transmission becomes more parallel to the tidal direction. Testing the validity of the above results in the realistic ocean provided the motivation for the present work. This was done through simulation using the NRL Coastal Ocean Model (NCOM) fitted to experimental data. The NCOM output provided the sound speed profile as a function of space and time. Results of the simulation will be presented. [Work supported by ONR (PE-62435N), and administered by NRL.]

11:15

3aAO11. A normal-mode formula for the derivative of a waveguide pressure field with respect to an arbitrary three-dimensional sound speed perturbation. Aaron Thode (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238)

Semi-analytic expressions are derived for the first order derivative of a pressure field in a laterally homogeneous depth waveguide, with respect to an arbitrary three-dimensional refractive index perturbation in either the water column or ocean bottom. These expressions for the environmental derivative, derived using an adjoint method, require a three-dimensional spatial correlation between two Greens functions, weighted by an environmental parameter basis function, with the Greens functions expressed in terms of normal modes. When a particular set of orthogonal spatial basis functions is chosen, the three-dimensional spatial integral can be converted into a set of one-dimensional integrations over depth and azimuth. The use of the orthogonal basis permits environmental derivatives to be computed for any arbitrary sound-speed perturbation. To illustrate the formulas, a sensitivity study is presented that explores the impact of three-dimensional plane wave and cylindrical perturbations on the environmental derivative. Under certain circumstances it is found that perturbation components outside the vertical plane connecting the source and receiver have non-negligible effects on the pressure derivative. Potential applications of these formulas include benchmarking three-dimensional propagation codes, computing Cramer-Rao bounds for three-dimensional environmental parameter estimates, and potentially inverting for small three-dimensional refractive index distributions.

11:30

3aAO12. Resonances due to near-bottom soliton packets. Stanley A. Chin-Bing, Alex C. Warn-Varnas, David B. King (Naval Res. Lab., Stennis Space Center, MS 39529-5004), Kevin G. Lamb (Univ. of Waterloo, Waterloo, ON N2L 3G1, Canada), James A. Hawkins (Planning Systems, Inc., Slidell, LA 70458), and James F. Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

At the Acoustical Society of Americas Nashville meeting, 28 April–2 May 2003, we presented oceanographic modeling and acoustic modeling of the ONR winter 1997 Primer4 experiment conducted on the shelfbreak and continental slope south of Cape Cod in the Middle Atlantic Bight. Near-bottom solitary internal wave simulations in this region were generated using a forcing tidal velocity of 0.3 m/s to initiate the primitive equation Lamb model. The simulated internal waves were benchmarked against oceanographic data taken by Lynch. Similar to the acoustic mode coupling mechanism seen in the near-surface solitary wave packets, these near-bottom solitary internal waves also exhibited acoustic mode coupling—in this case the coupling transferred the acoustic energy from the trapped normal modes to the lossy continuous modes. This resulted in a loss in signal intensity, but no accompanying resonances were observed. As an update to that work we will show cases where the resonances are present, and are due to the conversion of the acoustic energy from the trapped normal modes to the lossy continuous modes. [Work supported by ONR/NRL.]

11:45

3aAO13. Time-frequency variability of broadband sound propagation through shallow water internal soliton. Boris Katsnelson, Sergey Pereselkov (Voronezh Univ., 1, Universitetskaya sq, Voronezh 394006, Russia), Mohsen Badiey (Univ. of Delaware, Newark, DE 19716), and James Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Temporal fluctuations of the sound intensity in shallow water are studied using broadband acoustic signal propagation through an internal soliton (IS) wave front. Time-frequency analysis of the received signals at two different vertical arrays is conducted based on the wavelet transform. The

intensity fluctuations as a function of geophysical time are shown for different positions of the IS with respect to the acoustic propagation path. A theoretical model consisting of internal waves coupled with the acoustic normal modes and horizontal rays is presented. The correlation between the signal fluctuations and the internal soliton is analyzed using frequency dependent index of refraction. Model results are compared with the SWARM-95 experimental observation. [Work supported by ONR and RFBR Grant No. 03-05-64568.]

WEDNESDAY MORNING, 12 NOVEMBER 2003 TRINITY A ROOM, 8:00 A.M. TO 12:15 P.M.

Session 3aBB

Biomedical Ultrasound/Bioresponse to Vibration: Multidisciplinary and Non-Conventional Approaches in Biomedical Ultrasound

Stanislav Emelianov, Chair
Department of Biomedical Engineering, University of Texas, Austin, Texas 78712

Chair's Introduction—8:00

Invited Papers

8:05

3aBB1. Optoacoustics for biomedical applications. Takashi Buma, JingYong Ye, Theodore Norris (Ctr. for Ultrafast Optical Sci., 1006 Gerstacker Bldg., Univ. of Michigan, 2200 Bonisteel Blvd., MI 48109-2099), Susanne Milas, Monica Spisar, Kyle Hollmann, Matthew O'Donnell, James Hamilton, Stanislav Emelianov, Lajos Balogh, and James Baker, Jr. (Univ. of Michigan, MI)

We are developing optical techniques to generate and receive ultrasound for various biomedical applications including high frequency 2D arrays, molecular imaging, and microfluidic devices. A 2D synthetic receive array uses a HeNe laser to probe the surface displacements of a thin reflective membrane. Images with near optimal resolution and wide fields of view have been produced at 10 to 50 MHz. A 75 MHz transmitting 2D array element relies on the thermoelastic effect. A 10 ns laser pulse is focused onto a 25 μm thick black polydimethylsiloxane (PDMS) film spin coated on a pure PDMS substrate. Our work in optoacoustic molecular imaging combines ultrafast lasers with high frequency ultrasound. When ultrafast laser pulses are focused into transparent media, laser induced optical breakdown (LIOB) produces acoustic emission and cavitation bubbles. A real-time acoustic technique has been developed to characterize LIOB in dendrimer nanocomposite (DNC) solutions. Lamb waves propagating in thin membranes have found widespread use in microfluidic devices. We use the thermoelastic effect as a noncontact method to generate continuous-wave Lamb waves in gold-coated membranes. We believe these results demonstrate the potential of optoacoustic methods for a broad range of applications in biomedical ultrasonics.

8:30

3aBB2. Ultrasound-aided high-resolution biophotonic imaging. Lihong V. Wang (Texas A&M Univ., 3120 TAMU, 233 Zachry Bldg., College Station, TX 77843-3120)

We develop novel biophotonic imaging for early-cancer detection, a grand challenge in cancer research, using nonionizing electromagnetic and ultrasonic waves. Unlike ionizing x-ray radiation, nonionizing electromagnetic waves such as optical waves are safe for biomedical applications and reveal new contrast mechanisms and functional information. For example, our spectroscopic oblique-incidence reflectometry can detect skin cancers based on functional hemoglobin parameters and cell nuclear size with 95% accuracy. Unfortunately, electromagnetic waves in the nonionizing spectral region do not penetrate biological tissue in straight paths as do x-rays. Consequently, high-resolution tomography based on nonionizing electromagnetic waves alone, as demonstrated by our Mueller optical coherence tomography, is limited to superficial tissue imaging. Ultrasonic imaging, on the contrary, furnishes good imaging resolution but has poor contrast in early-stage tumors and has strong speckle artifacts as well. We developed ultrasound-mediated imaging modalities by combining electromagnetic and ultrasonic waves synergistically. The hybrid modalities yield speckle-free electromagnetic-contrast at ultrasonic resolution in relatively large biological tissue. In ultrasound-modulated (acousto)-optical tomography, a focused ultrasonic wave encodes diffuse laser light in scattering biological tissue. In photo-acoustic (thermo-acoustic) tomography, a low-energy laser (RF) pulse induces ultrasonic waves in biological tissue due to thermoelastic expansion.

8:55

3aBB3. Nonconventional approaches to ultrasonic assessment of skeletal system. Armen Sarvazyan (Artann Labs., 1753 Linvale-Harbourton Rd., Lambertville, NJ 08530)

Conventional bone ultrasonometry is based on the transmission mode measurement of the linear acoustic parameters of bone (speed of sound and broadband attenuation). It is a well-established technique specifically for osteoporosis diagnostics though it is highly limited in its applications. It cannot be used for the assessment of the hip bone which is the most important area in characterization of the skeletal system and can hardly be used in pediatric applications, particularly in neonatology, for the assessment of newborns and premature and low-birth-weight infants skeletal systems, which became especially vital during the last few decades. There are several new ideas on acoustic assessment of the skeletal sites hardly accessible by the conventional bone ultrasonometry and

on principles of the acoustic characterization of bone quality, fracture risk evaluation, and monitoring of therapeutic interventions. New approaches are based on the use of the ultrasound radiation pressure for the remote generation of acoustic waves in bones, on the use of various modes of guided acoustic waves having a propagation speed dependent on both the elasticity modulus and the bone thickness, on the use of geometrical dispersion of sound velocity for bone characterization, and the use of principles of nonlinear acoustic spectroscopy for remote bone testing. [Work supported by NIH.]

Contributed Papers

9:20

3aBB4. A "Fresnel-transducer" for prostate hyperthermia treatment. Robert M. Keolian (Penn State Appl. Res. Lab., P.O. Box 30, State College, PA 16804-0030, bonzo@sabine.acs.psu.edu), Osama M. Al-Bataineh, Nadine B. Smith, Victor W. Sparrow (Penn State, University Park, PA 16802), and Lewis E. Harpster (Penn State Milton S. Hershey Medical Ctr., Hershey, PA 17033)

Simulations and construction methods will be described for a novel "Fresnel-transducer." The transducer is designed for transrectal hyperthermia treatment of prostate cancer as an adjuvant to radiotherapy or chemotherapy. Forty nine 6.3 mm diameter 1.5 MHz PZT elements are arranged in a 3 by 7 cm honeycomb-like pattern. They are individually aimed so that their beams partially converge behind the prostate. The increased beam density away from the transducer compensates for the loss of acoustic intensity due to attenuation. The aiming of the beams is additionally biased toward the periphery of the heated region to compensate for cooling from lateral heat conduction. The elements are divided into three interspersed sets, each driven at a slightly different frequency, to minimize stationary Moire interference bands between the beams. The combined effect is to uniformly raise the prostate temperature to 43 °C without overheating the rectal wall. [Research supported by the Department of Defense Congressionally Directed Medical Prostate Cancer Research Program.]

9:35

3aBB5. Vibration characteristics of implants. Ahmed M. Al-Jumaily (Diagnostics and Control Res. Ctr., Auckland Univ. of Technol., Auckland, New Zealand, ahmed.al-jumaily@aut.ac.nz), Mostafa Fatemi, and James F. Greenleaf (Ultrasound Res. Lab., Mayo Clinic and Foundation, Rochester, MN 55905)

Vibro-acoustography is a technique that uses the radiation force of amplitude-modulated ultrasound to evaluate the dynamic response of an object at the modulation frequency. One potential application of this technique is to assess the reaction of the host tissue to the implant material by evaluating the mechanical properties of the tissue surrounding the implant. To do this, we use vibro-acoustography to measure the frequency response of the implant, and use this information to evaluate the mechanical parameters of the surrounding. Two theoretical models are developed to study the dynamic response of an implant in tissue. The first model is for an implant (1×25-mm steel beam) fully embedded in a viscoelastic medium (tissue). The second model is for a partially embedded implant with one part exposed to the open environment. The exposed end is rigidly connected to a base, and the other end is freely embedded in the viscoelastic medium. At the interface between the two media matching of boundary conditions is achieved to determine the force response. For both models eigenvalues and eigenfunctions are determined and transfer functions are evaluated. The first two natural frequencies compare well with available experimental and finite-element data.

9:50

3aBB6. Sector array transducers for vibro-acoustography. Glauber Silva, Shigao Chen, Randall Kinnick, James Greenleaf, and Mostafa Fatemi (Mayo Clinic and Foundation, Rochester, MN 55905)

Vibro-acoustography is an imaging technique that maps the acoustic response of an object to a localized harmonic radiation force. This force is generated by two interfering continuous-wave ultrasound beams at slightly different frequencies f_1 and f_2. The system point-spread function (PSF) is related to the radiation force on a point-target. Imaging artifacts depend on the PSF sidelobes, which can be reduced by mismatching the sidelobes of each ultrasound beam. Here, we propose a beamforming approach based on an 8-element sector transducer with consecutive elements alternately driven at f_1 and f_2. The transducer generates two ultrasound beams skewed by 22.5° with respect to each other. The system PSF is analytically derived. The theory is validated by experiments using a small steel sphere (radius $=0.2$ mm) as a point-target. A laser vibrometer is used to measure the vibration of the sphere and evaluate the PSF of the system. Theoretically, the PSF sidelobes are under -15.8 dB in eight spots circularly distributed and separated by 22.5°, which agrees with the experimental results. Simulation shows that with 16 elements sidelobes are under -35.2 dB. In conclusion, sector transducers for vibro-acoustography may have lower sidelobes as the number of array elements is increased. [Work supported by Grant Nos. EB00535-01, EB2640, and IMG0100744.]

10:05–10:15 Break

10:15

3aBB7. Radiation force produced by time reversal acoustic focusing system. Armen Sarvazyan and Alexander Sutin (Artann Labs., Inc., 1753 Linvale-Harbourton Rd., Lambertville, NJ 08530)

An ultrasonic induced radiation force is an efficient tool for remote probing of internal anatomical structures and evaluating tissue viscoelastic properties, which are closely related to tissue functional state and abnormalities. Time Reversal Acoustic Focusing System (TRA FS) can provide efficient ultrasound focusing in highly inhomogeneous media. Furthermore, numerous reflections from boundaries, which distort focusing in conventional ultrasound focusing systems and are viewed as a significant technical hurdle, lead to an improvement of the focusing ability of the TRA system. In this work the TRA FS field structure and radiation force in a transcranial phantom were investigated. A simple TRA FS comprising a plane piezoceramic transducer attached to an external resonator such as an aluminum block was acoustically coupled to the tested transcranial phantom. A custom-designed compact electronic unit for TRA FS provided receiving, digitizing, storing, time reversing and transmitting of acoustic signals in a wide frequency range from 0.01 to 10 MHz. The radiation force produced by ultrasonic pulses was investigated as a function of the transmitted ultrasound temporal parameters. The simplest TRA FS provided focusing of 500 kHz ultrasound pulses and the generation of a radiation force with an efficacy hardly achievable using conventional sophisticated phased array transmitters. [Work supported by NIH.]

10:30

3aBB8. Effects of transducer and tissue parameters on motion induced by radiation force during ultrasonic lesion monitoring. Samuel Mikaelian and Frederic L. Lizzi (Riverside Res. Inst., 156 William St., 9th Fl., New York, NY 10038, mikaelian@rrinyc.org)

Tissue motion induced by acoustic radiation force is studied as a means of monitoring the formation of therapeutic lesions with increased stiffness produced by high-intensity focused ultrasound (HIFU). Our analyses and simulations examine the role and interplay of various tissue and system parameters to assist in designing practical systems and interpreting results. The radiation force is generated by a therapeutic transducer excited at levels below lesion-production threshold, while the magnitude and time-course of the induced motion is monitored via a confocal and collinear diagnostic transducer. Lesions are detected by comparing pre- and post-treatment motion patterns. Parameters characterizing properties of the motion-inducing beam, such as its intensity and spatial profile, and tissue parameters signifying its viscoelastic properties, acoustic attenuation, and geometry, are the focus of this investigation. Acoustic attenuation, which can increase significantly during lesion production and directly

affects the radiation-force profile, manifests itself on time scales characteristic of acoustic propagation. In contrast, time scales characteristic of viscoelastic effects are substantially larger. These viscoelastic properties also mediate effects associated with tissue geometry and large-scale spatial beam profile. Effects of diagnostic beam parameters on the sensitivity and accuracy of the method have also been investigated.

10:45

3aBB9. Relationships between cortical bone properties and head wave ultrasonic velocity at 1 MHz: An *in vitro* investigation. Emmanuel Bossy, Maryline Talmant (Lab. Imagerie. Paramétrique, 15 r. Ecole de Médecine, 75006 Paris, France), Françoise Peyrin, Peter Cloetens (ESRF, BP 220, 38043 Grenoble Cedex, France), Leila Akrout, and Pascal Laugier (Lab. Imagerie. Paramétrique, 75006 Paris, France)

The axial transmission technique provides a measurement of the velocity (SOS) of elastic waves propagating along the surface of cortical bone. Currently, SOS is derived from the time of flight of the first arriving signal. Our goal was to investigate the relationship between SOS and several bone properties that independently contribute to bone strength. One-MHz SOS measurements were performed with a proprietary probe on 41 excised human radius with soft tissue removed. Small pieces of cortical bone were removed from the site of ultrasonic testing to be investigated using synchrotron radiation microtomography (resolution 10 μm). Cortical *thickness*, cortical *porosity*, and tissue **mineralization** were derived from the 3-D data set. Mean SOS values were 3940 (3796–4120) m s^{-1}. The reproducibility of the SOS measurements expressed as a variation coefficient (CV) was 0.5%. The relationships between SOS and bone properties were assessed using linear regression and displayed significant correlations : $R^2=0.20$ ($p<10^{-5}$) for cortical thickness, $R^2=0.30$ ($p<10^{-3}$) for cortical porosity, $R^2=0.37$ ($p<10^{-4}$) for tissue mineralization. A multiple regression analysis showed that cortical porosity and tissue mineralization in the periosteal region were the most predictive variables and explained 60% of the total variance of the SOS. These results are in agreement with predictions obtained using 3-D numerical computations based on a finite-difference method.

11:00

3aBB10. Ultrasonic axial transmission on cortical bone: A 3-D simulation study. Emmanuel Bossy, Maryline Talmant, and Pascal Laugier (Lab. Imagerie. Paramétrique, 15 r. Ecole de Médecine, 75006 Paris, France)

In vivo, cortical bone is currently investigated using ultrasonic velocity measurement in the megahertz frequency range. It is virtually impossible to analytically model the extremely complex field resulting from the interaction of an incident wave with cortical bone, taking into account 3-D geometry, boundary conditions, anisotropic and heterogeneous bone properties. Recently developed simulation methods based on finite difference offer a fertile alternative to inextricable analytic formulations. Wave propagation simulation is applied here to the problem of axial transmission along the radius. 3-D computations showed that different types of head wave propagate along the anisotropic cortex depending on the thickness-to-wavelength ratio. Simulated variations of SOS as a function of thickness are in good agreement with *in vivo* [Njeh (1999)] and *in vivo* [Prevrhal (2001)] observations. Cortical porosity leads to a velocity decrease of approximately 50 m s^{-1} (respectively 25 m s^{-1}) per % of porosity increase in the axial (respectively, radial) direction. When the cortical thickness is larger than one compressional wavelength, the signal velocity reflects bone properties over a depth of about half a wavelength. Numerical simulation brings an insightful view into the influence of several bone properties on signal velocity, and can also be of great value in testing inverse calculation procedures.

11:15

3aBB11. Feasibility of bone assessment with leaky Lamb wave in bovine cortical bone. Kang Il Lee and S. W. Yoon (Dept. of Phys., SungKyunKwan Univ., Suwon 440-746, Republic of Korea, swyoon@skku.ac.kr)

Quantitative ultrasound (QUS) technique is now widely used for noninvasive assessment of osteoporosis. The use of leaky Lamb wave is very attractive since it propagates throughout the cortical thickness of long bones, which means that the entire thickness of bones is interrogated. Feasibility of bone assessment with leaky Lamb wave is investigated in a bone phantom and bovine tibia *in vitro* using the axial transmission method commonly used to characterize human long bones. The bone phantom consists of Lucite plates with thicknesses of 1, 3, and 5 mm. The results obtained from the bone phantom show that the peak frequencies and amplitudes of excited Lamb modes are strongly dependent on the frequency-thickness product. In the case of the bovine tibia, the S0 Lamb mode may be more sensitive to any change of elastic properties inside the bone plate due to damage or to cortical thickness changes with aging and osteoporosis. This study suggests that leaky Lamb wave is feasible to ultrasonic bone assessment. [Work supported by BK21 Program in Korea.]

11:30

3aBB12. A new method for measurement of pulse wave velocity in arterial wall. Xiaoming Zhang, Randall R. Kinnick, Mostafa Fatemi, and James F. Greenleaf (Ultrasound Res. Lab., Mayo Clinic, 200 First St. SW, Rochester, MN 55905)

Arterial wall stiffness can be associated with various diseases. The stiffness of an artery can be assessed by measurement of the pulse wave velocity (PWV). PWV is directly related to the Youngs modulus by the well-known Moens–Korteweg equation. Usually, PWV is estimated using the foot-to-foot method. However, the foot of the pressure wave is not very clear due to reflected waves. Also, the pressure wave is normally at a low frequency, hence, the time resolution is low. PWV is an average indicator of artery stiffness between the two measuring points, therefore it is not easy to identify local stiffness. We propose producing a very short pulse wave in the arterial wall using ultrasound radiation force and measuring its propagation speed along the artery by laser. The temporal resolution of this method is in the range of microseconds, which allows PWV to be measured accurately over a few millimeters. Experiments were carried out on a silicone tube in gelatin. PWV was measured by two scanning methods: (1) fixed source and scanning detector, (2) scanning source and fixed detector. Results: PWV was measured at 2 mm/40 μs by both methods. The Doppler technique was also tested which is potentially suitable for clinical applications.

11:45

3aBB13. Adaptive linearized modeling and inversion for 3D tissue characterization. Basak Ulker Karbeyaz, Eric L. Miller (Dept. of Elec. Eng. and Computer Sci., Northeastern Univ., Boston, MA 02115, bulker@ece.neu.edu), Robin O. Cleveland, and Ronald A. Roy (Boston Univ., Boston, MA 02115)

Quantitative information regarding the acoustic velocity, absorption, and density of tissue given broadband ultrasonic (US) scattering data can enhance tissue characterization capabilities. Extracting quantitative information requires the use of a well-calibrated physical model directly within the processing algorithms. The computational size of most relevant 3D problems, with dimensions of tens to thousands of wavelengths on a side, render impractical the use of nonlinear inverse methods. Thus we must rely on linearized models obtained using the well-known first Born or first Rytov approximations. To extend the validity of these models and thereby improve their quantitative accuracy, we introduce an adaptive approach to the linearized inversion of US scattering data. We determine both the perturbations in the three acoustic parameters and the background properties around which the linearizion is performed. To make this adaptive re-linearization tractable for 3D US problems, we demonstrate a novel method for rapidly constructing the Born/Rytov kernels based on a new semi-analytic expression for the impulse response of the transducer. The

approach is validated using simulated broadband US backscatter data as well as measurements in a tissue phantom. [Work supported in part by CenSSIS (NSF Award No. EEC-9986821) and NSF Award No. 0208548.]

12:00

3aBB14. Quantitative object localization and characterization from broadband ultrasonic backscatter. Basak Ulker Karbeyaz, Eric L. Miller (Dept. of Elec. Eng. and Computer Sci., Northeastern Univ., Boston, MA 02115, bulker@ece.neu.edu), Robin O. Cleveland, and Ronald A. Roy (Boston Univ., Boston, MA 02115)

The goal of many image applications in biomedical ultrasound is to ascertain the structure of regions in tissue which are, in a sense, anomalous relative to a background. To obtain quantitatively accurate reconstructions of the location, size, shape, and contrast (in sound speed, attenuation, or density) of the anomaly, one could invert for the acoustic properties on a finely sampled grid. However, this formulation is known to be ill-posed, especially when only limited view data, e.g., backscatter, are available. To avoid the difficulties associated with regularization schemes for overcoming ill-posedness, we introduce an inversion method in which we directly determine from the data the relevant geometric structure of the anomaly (center location, parameters governing shape and orientation) and a low order representation of the spatial structure of the possibly unknown background. By concentrating the information in the limited data to extract only these few unknowns, explicit regularization is not required. The details of this nonlinear inversion method will be provided and its performance and robustness will be demonstrated using both simulated broadband ultrasound backscatter data and monostatic backscatter measurements. [Work supported in part by CenSSIS (NSF Award No. EEC-9986821) and NSF Award No. 0208548.]

WEDNESDAY MORNING, 12 NOVEMBER 2003 CONCHO ROOM, 8:30 TO 11:45 A.M.

Session 3aEA

Engineering Acoustics, Noise and Architectural Acoustics: Honoring the Contributions of Elmer Hixson

Brandon D. Tinianov, Cochair
Johns Manville, 10100 West Ute Avenue, Littleton, Colorado 80127

John T. Post, Cochair
Post Acoustics, 1604 St. Mary's Mountain Road, Altus, Arkansas 72821

Chair's Introduction—8:30

Invited Papers

8:35

3aEA1. The evolution of acoustically absorbent lining systems for anechoic chambers. Jeff Schmitt (JGS Consulting, 2512 Star Grass Circle, Austin, TX 78745)

Materials for the construction of and techniques for the evaluation of acoustically absorbent lining systems for anechoic test chambers have evolved significantly over the past 20 years. This paper will recount the authors experiences in anechoic chamber design, construction and evaluation techniques since his early 1980s thesis work on the subject conducted under the guidance of Dr. Elmer Hixson of the University of Texas.

9:00

3aEA2. Error in the measurement of acoustic impedance by the two-microphone method. John T. Post (Post Acoust. and Post Winery, 1604 St. Mary's Mountain Rd., Altus, AR 72821)

Methods for measuring acoustic impedance have traditionally been slow, physically bulky, and difficult to accurately implement. A technique known as the two-microphone method is presented, which overcomes these previous limitations. The performace of this method is evaluated by an error study which shows the relationship of the measurement error and the measured impedance load and microphone spacing. The two-microphone method is then compared with another method, the reaction-on-the-source method, and the advantages and disadvantages of both are discussed. The two-microphone method is shown to be robust with loads that are dominantly inertial in the low-frequency limit.

9:25

3aEA3. Analogies between the measurement of acoustic impedance via the reaction on the source method and the automatic microwave vector network analyzer technique. James McLean, Robert Sutton (TDK RF Solutions, 1101 Cypress Creek Rd., Cedar Park, TX 78613, jmclean@tdkrf.com), and John Post (Post Acoust., Altus, AR 72821)

One useful method of acoustic impedance measurement involves the measurement of the electrical impedance ''looking into'' the electrical port of a reciprocal electroacoustic transducer. This reaction on the source method greatly facilitates the measurement of acoustic impedance by borrowing highly refined techniques to measure electrical impedance. It is also well suited for *in situ* acoustic impedance measurements. In order to accurately determine acoustic impedance from the measured electrical impedance, the charac-

teristics of the transducer must be accurately known, i.e., the characteristics of the transducer must be ''removed'' completely from the data. The measurement of acoustic impedance via the measurement of the reaction on the source is analogous to modern microwave measurements made with an automatic vector network analyzer. The action of the analyzer is described as de-embedding the desired data (such as acoustic impedance) from the raw data. Such measurements are fundamentally substitution measurements in that the transducer's characteristics are determined by measuring a set of reference standards. The reaction on the source method is extended to take advantage of improvements in microwave measurement techniques which allow calibration via imperfect standard loads. This removes one of the principal weaknesses of the method in that the requirement of high-quality reference standards is relaxed.

9:50

3aEA4. Energy density and intensity measurements in a reverberant sound field. Thomas J. Poterek (Dept. of Elec. and Computer Eng., Univ. of Texas, Austin, TX, tomp787@yahoo.com)

The measurement of reverberant fields using an acoustic energy density sensor has been investigated over the years but has not seen general use. The energy density sensor can also measure acoustic intensity, which is a more commonly used technique in noise control. Measuring three-dimensional acoustic intensity with the energy density sensor shows that the techniques of energy density are complementary to intensity measurement and worthy of broader use Statistical measurements of kinetic and potential energy density were taken using a rotating boom in a reverberation room. Vector intensity was simultaneously measured in three orthogonal directions along with the total intensity. The measurements show that while the vector intensity components were significant, their sum was zero within the limits of the measurement system.

10:15–10:30 Break

10:30

3aEA5. Acoustics, computers and measurements. James J. Truchard (Natl. Instruments, 11500 N. Mopac Expressway, Bldg. C, Austin, TX 78759)

The human ear has created a high standard for the requirements of acoustical measurements. The transient nature of most acoustical signals has limited the success of traditional volt meters. Professor Hixson's pioneering work in electroacoustical measurements at ARL and The University of Texas helped set the stage for modern computer-based measurements. The tremendous performance of modern PCs and extensive libraries of signal processing functions in virtual instrumentation application software has revolutionized the way acoustical measurements are made. Today's analog to digital converters have up to 24 bits of resolution with a dynamic range of over 120 dB and a single PC processor can process 112 channels of FFTs at 4 kHz in real time. Wavelet technology further extends the capabilities for analyzing transients. The tools available for measurements in speech, electroacoustics, noise, and vibration represent some of the most advanced measurement tools available. During the last 50 years, Professor Hixson has helped drive this revolution from simple oscilloscope measurements to the modern high performance computer-based measurements.

10:55

3aEA6. Techniques in audio and acoustic measurement. Thomas D. Kite (Audio Precision, P.O. Box 2209, Beaverton, OR 97075-2209)

Measurement of acoustic devices and spaces is commonly performed with time-delay spectrometry (TDS) or maximum length sequence (MLS) analysis. Both techniques allow an impulse response to be measured with a signal-to-noise ratio (SNR) that can be traded off against the measurement time. However, TDS suffers from long measurement times because of its linear sweep, while MLS suffers from the corruption of the impulse response by distortion. Recently a logarithmic sweep-based method has been devised which offers high SNR, short measurement times, and the ability to separate the linear impulse response from the impulse responses of distortion products. The applicability of these methods to audio and acoustic measurement will be compared.

11:20

3aEA7. Reflections on Elmer Hixson's contributions to architectural acoustics. Paul T. Calamia (Dept. of Computer Sci., Princeton Univ., 35 Olden St., Princeton, NJ 08544)

Although not often thought of in the context of architectural acoustics, many of Elmer Hixson's research interests, for example in areas such as noise control, loudspeaker arrays, and acoustic sensors/measurements, are applicable to the design and analysis of concert halls and other performance spaces. In this presentation the author will discuss recent, current, and future directions of research in architectural acoustics, with emphasis on the contributions and influence of Elmer Hixson. Specific topics will include sound isolation and quiet HVAC design; the use of arrays in sound reinforcement systems; measurement and analysis of room-acoustics parameters; and predictive, three-dimensional computer modeling.

Session 3aMU

Musical Acoustics: Honoring the Contributions of Gabriel Weinreich I: Piano Acoustics

Thomas D. Rossing, Chair
Physics Department, University of Illinois, DeKalb, Illinois 60115

Chair's Introduction—9:00

Invited Papers

9:05

3aMU1. Gabriel Weinreich: The life and style. William M. Hartmann (Phys. and Astron., Michigan State Univ., Lansing, MI 48824)

Gabriel Weinreich (Gabi) was born in Vilna, Poland (now the capitol of Lithuania) one year prior to the founding of the Acoustical Society of America. When the second world war began in central Europe, Gabi's family came, in serial fashion, to New York City—Gabi himself arriving in 1941. Gabi studied physics at Columbia, and received a Ph.D. in 1953 for a thesis on atomic physics directed by the legendary I. I. Rabi. He subsequently worked on fundamental properties of semiconductors, first at Bell Labs, then, starting in 1960, at the University of Michigan. In 1977 he turned his attention to the acoustics of musical instruments, mainly the piano and bowed strings. He studied all phases of the physical elements: string excitation, string vibration, coupling, and radiation. Gabi brought his special style to acoustics—a combination of theory and experiment that imaginatively imports ideas and techniques from one area of physics into another, a willingness to attack traditional problems afresh by returning to first principles, and the ability to present ideas with incisive wit and charm so that information is not only informative but is also entertaining.

9:20

3aMU2. Piano acoustics—A review. Anders Askenfelt (Dept. of Speech, Music and Hearing, Royal Inst. of Technol. (KTH), SE-10044 Stockholm, Sweden)

The design of the piano as we know it today dates back to the second half of the 19th century. The history of studies of the acoustics of the piano begins during the same period. In this talk, known facts and unanswered questions about the acoustics of the piano are reviewed.

9:50

3aMU3. Modeling the dynamics of piano keys. Celine Brenon (Ecole Nationale Superieure d'Arts et Metiers, B.P. 3525, 49 035 Angers Cedex, France, Celine.Brenon@angers.ensam.fr) and Xavier Boutillon (Lab. for Mech. of Solids, CNRS—Ecole Polytechnique, 91128 Palaiseau Cedex, France)

The models of piano keys available in the literature are crude: two degrees of freedom and a very few dynamical or geometrical parameters. Experiments on different piano mechanisms (upright, grand, one type of numerical keyboard) exhibit strong differences in the two successive phases of the key motion which are controlled by the finger. Understanding the controllability of the escapement velocity (typically a few percents for professional pianists), the differences between upright and grand pianos, the rationale for the numerous independent adjustments by technicians, and the feel by the pianist require sophisticated modeling. In addition to the inertia of the six independently moving parts of a grand piano mechanism, a careful modeling of friction at pivots and between the jack and the roll, of damping and nonlinearities in felts, and of internal springs will be presented. Simulations will be confronted to the measurements of the motions of the different parts. Currently, the first phase of the motion and the transition to the second phase are well understood while some progress must still be made in order to describe correctly this short but important phase before the escapement of the hammer. [Work done in part at the Laboratory for Musical Acoustics, Paris.]

10:20

3aMU4. Physical modeling of the piano. N. Giordano and M. Jiang (Dept. of Phys., Purdue Univ., 525 Northwestern Ave., West Lafayette, IN 47907)

Over the past several years, this project has been aimed at constructing a physical model of the piano. The goal is to use Newton's laws to describe the motion of the hammers, strings, soundboard, and surrounding air, and thereby calculate the sound produced by the instrument entirely from first principles. The structure of the model is described, along with experiments that have provided essential tests and guidance to the calculations. The state of the model and, especially, how this work can lead to new insights and understanding into the piano are discussed. In many cases the work and the specific questions addressed along the way have followed paths initially inspired and developed by Gabriel Weinreich. [Work supported by NSF.]

10:50

3aMU5. Comparison between musical acoustics parameters of an upright and a grand piano. Klaus Wogram (PTB, Braunschweig, Germany) and Taro Mori (Tokushima-shi, Japan)

''Why does a grand piano sound so much better than an upright piano?'' This question is frequently discussed. Is it the sound itself, or the inharmonicity, or is it the way of playing the different instruments, or the sound pattern? All these questions have been investigated in the former laboratory of musical acoustics of the PTB by Taro Mori with very interesting results. Results of measurements and subjective hearing tests will be presented, and the method itself will be explained. One of the most interesting findings is that it is nearly impossible to distinguish between an upright and a grand piano in hearing tests by listening only to single tones, sequences of tones or chords. Better results are found by performing complete musical pieces on the instruments under test.

11:20

3aMU6. Relating the radiated piano sound field to the vibrational modes of the soundboard. Uwe J. Hansen (Indiana State Univ., Terre Haute, IN 47809), Ingolf Bork (PTB, Braunschweig, Germany), and Thomas D. Rossing (Northen Illinois Univ., DeKalb, IL 60115)

The sound field near a piano sound board is determined by moving a microphone over a grid of points above and below the sound board as well as in a plane in front of the piano using the experimental techniques of modal analysis with soundboard excitation at a bridge point by a swept sine signal. Since the standard modal analysis signal processing technique relies on tracking phase relations between excitation and response, it is possible to relate the sound field in terms of the vibrating structure which radiated it. Animations of sound board motion and sound field pressure variations are shown for the lowest four modes. It is noted that in all modes the locations of maximal excursion correlate with the sound pressure maxima or minima, respectively.

WEDNESDAY MORNING, 12 NOVEMBER 2003 BOSQUE ROOM, 9:00 A.M. TO 12:00 NOON

3a WED. AM

Session 3aNS

Noise, Engineering Acoustics and Signal Processing in Acoustics: New Technologies in Active Control

Robert D. Corsaro, Cochair
Naval Research Laboratory, Code 7135, Washington, D.C. 20375-5350

Jiri Tichy, Cochair
Graduate Program in Acoustics, Pennsylvania State University, 217B Applied Science Building, University Park, Pennsylvania 16802

Invited Papers

9:00

3aNS1. Overview of the University of Wyoming's hexapod program. John McInroy (Dept. of Elec. Eng., Univ. of Wyoming, Laramie, WY 82071, mcinroy@uwyo.edu)

For the past eight years, the University of Wyoming (UW) has been developing parallel kinematics machines to perform fault tolerant precision pointing and vibration isolation. These machines have been predominantly of the Stewart platform, or hexapod, type although other configurations including redundant fast steering mirrors have been considered. A brief history of the program will be followed by an overview of the experimental hardware. Then, UW's technical contributions will be summarized. These include the following: (1) dynamic modeling of flexure jointed hexapods, (2) decoupled control, (3) fault tolerant pointing and isolation, (4) adaptive cancellation of repetitive vibrations, (5) division of labor among fine and very fine actuators, and (6) geometric design of Stewart platforms and optimization.

9:20

3aNS2. Flight performance of the International Space Station active rack isolation system. Glenn S. Bushnell (The Boeing Co., Seattle, WA 98124), Ian J. Fialho, James L. Allen (The Boeing Co., Houston, TX), and Naveed Quraishi (NASA Johnson Space Ctr., Houston, TX)

Space flight experiment test results of a Space Station Active Rack Isolation System (ARIS) are presented. The purpose of ARIS is to isolate microgravity sensitive science experiments mounted in Space Station racks from structural vibrations present on the large Space Station orbital structure. The ARIS is shown to solve the very difficult and challenging low frequency isolation problem by providing over an order of magnitude reduction in the acceleration at 0.1 Hz. The Station displacement response to crew motion is

discussed along with the control method that ARIS employs to maintain microgravity performance while limiting the motion between the Station and the isolated rack. The dramatic difference between the Station acceleration levels during crew awake and sleep periods are presented. Some microgravity experiments are sensitive to angular acceleration, so both the translational and angular accelerations of the isolated rack are presented. The performance at frequencies up to 300 Hz was measured by exciting the Station structure with a proof-mass shaker and a hammer and these results, and the impacts from payload fans are presented. ARIS has been in operation for two years and three Zeolite Crystal Growth Experiments have been supported.

9:40

3aNS3. Smart panels with SISO or MIMO velocity feedback control systems. Paolo Gardonio and Stephen Elliott (Inst. of Sound and Vib. Res., Univ. of Southampton, Highfield, Southampton SO17 1BJ, UK)

This paper summarizes and contrasts the research work that has been carried out on two types of smart panels with ASAC control systems: first, a SISO velocity feedback control system with a sensor that detects the volumetric component of the vibration of the panel and an actuator that exerts a uniform force over the surface of the panel and, second, a MIMO decentralized feedback control system using a grid of sensors that measure the point velocities of the panel and a grid of actuators that generates point excitations in the locations of the sensors. Both theoretical predictions and experimental results are presented for a reference problem of control of sound radiation by a rectangular panel clamped along the perimeter. The simulations carried out for the two systems have shown that the smart panel with the volume velocity sensor and uniform force actuator can in principle give better control performance than the smart panel with the sixteen decentralized control units. However, the practical implementation of the smart panel with the volume velocity sensor and uniform force actuator is limited by stability problems which are instead a minor issue for the panel with 16 decentralized velocity feedback control units.

10:00

3aNS4. Emerging technologies in ANC at VAL-Virginia Tech. Marty Johnson, Chris Fuller, and Kenji Homma (Vib. and Acoust. Lab., Dept. of Mech. Eng., Virginia Tech, VA 24061-0238)

This presentation will overview the development and application of three new active noise control technologies at VAL-Virginia Tech. Distributed Active Vibration Absorbers (DAVAs) consist of large area spring elements with embedded active elements and covered with a dynamic mass layer. Development of conformal DAVAs and their application to noise control in aircraft and launch vehicle fairings will be overviewed. New work in applying active control techniques to developing virtual acoustic systems and its application to virtual prototyping and virtual reality applications will be discussed. The development a modular, compact active noise control duct which utilizes very small active sources, transducers and a cost effective controller to reducing the total radiated sound from small fans will be presented. The application of the modular ANC duct to a small computer cooling fan will be presented.

10:20

3aNS5. Vibro-acoustic launch protection experiment (VALPE). Benjamin Henderson, Charlotte Gerhart, Steven Lane, Elizabeth Jensen (Air Force Res. Lab., Space Vehicles, AFRL/VSSV, 3550 Aberdeen Ave. SE, Kirtland AFB, NM 87117-5776, benjamin.henderson@kirtland.af.mil), Steve Griffin, and Anthony Lazzaro (Boeing-SVS, 4411 The 25 Way NE, Ste. 350, Albuquerque, NM 87109)

Launch acoustic and vibration loads have the potential to damage sensitive payloads within a payload fairing, often requiring more structural mass to withstand these loads than would otherwise be necessary to survive launch. Experiments demonstrating several vibro-acoustic mitigation technologies developed by AFRL/VS and its contractors flew on the Vibro-Acoustic Launch Protection Experiment 2 (VALPE-2) aboard a Terrier-Improved Orion sounding rocket from Wallops Island Flight Facility in August 2003. Flight data collected in November 2002 from a nearly identical launch (VALPE-1) was used to characterize the fairing environment for comparison. Preparations for the flight experiments are discussed along with the performance of the various experiments in flight. The several experiments include an Adaptive Vibro-Acoustic Device (AVAD) to mitigate acoustic loads, an active/passive hybrid vibration isolation system using voice-coil actuation and a ShockRing passive component, a voice-coil regenerative electronics vibration isolation system to absorb a portion of the vibration energy during launch and use it to power an active isolation system during a staging event, and a ChamberCore composite fairing with implications for passive acoustic performance.

10:40

3aNS6. A detailed experimental comparison between active and passive approaches for the reduction of total far-field radiated sound power from a point-driven simply support plate. Peter C. Herdic, Brian H. Houston, and Robert D. Corsaro (Naval Res. Lab., Code 7136, Washington, DC 20375)

Passive and active control of sound radiation from plates has been a challenging problem for a number of years where numerous surface treatments have been examined. The two most viable active surface treatments are active constrained layer damping (ACLD) and modal restructuring (MR). ACLD minimizes the total radiated power through a combination passive/active damping mechanisms

as well as by applying restoring forces/moments. MR, on the other hand, achieves far-field attenuation by reducing the plates radiation efficiency through a restructuring of its mode shape. These active approaches show considerably more performance when compared to an optimized (Ross–Kerwin–Ungar) passive constraint layer damping design. The performance is evaluated by scanning near-field acoustic holography (NAH) measurements at several mode frequencies associated with strong radiation. The quantities determined via NAH include the power radiated in the half space as well as spatially dense surface motions of the plate. Both ACLD and MR show ~50 dB reductions of total far-field radiated sound power (per unit input power). These experimental results will be discussed with particular emphasis placed on the tradeoffs between the passive and active surface treatments as well as the level of performance that can be obtained through other control approaches.

11:00

3aNS7. On the application of magneto-rheological dampers for reducing floor vibrations. Mehdi Ahmadian and Jeong-Hoi Koo (Dept. of Mech. Eng., Virginia Tech, MC-0238, Blacksburg, VA 24061)

The primary purpose of this study is to experimentally evaluate the dynamics of a magneto-rheological tuned vibration absorber (MR TVA) with several semi-active control schemes. A test rig is built to represent a two degree of freedom primary structure model coupled with an MR TVA, and four semi-active control policies are considered. The four control policies include: velocity-based, on–off groundhook control (on–off VBG); velocity-based, continuous groundhook control (continuous VBG); displacement-based, on–off groundhook control (on–off DBG); and displacement-based, continuous groundhook control (continuous DBG). Using the test apparatus, a series of tests are conducted to investigate the dynamics of the MR TVA with each control policy. The performance of each of the cases is then evaluated along with an equivalent passive TVA, using a transmissibility analysis. The experimental results indicate that the MR TVA with all of the semi-active control policies outperforms the passive TVA in reducing structural vibrations. Furthermore, the displacement-based groundhook control policies perform better in reducing the resonant vibrations of the primary structure than the velocity-based groundhook control schemes.

11:20

3aNS8. Vibration-controlled modules for use in active segmented partitions. Timothy Leishman (Dept. of Phys. and Astron., Brigham Young Univ., N335 Eyring Sci. Ctr., Provo, UT 84602, tim_leishman@byu.edu)

Active segmented partitions (ASPs) have become an important topic in the research of active sound transmission control (ASTC). Many ASTC applications require lightweight partitions, high transmission loss over a broad frequency range, and applicability to a variety of source and receiving space conditions. This paper provides theoretical, numerical, and experimental evaluations of specific ASP module configurations used to satisfy these requirements. The control objective for each is to induce global vibration control of its various transmitting surfaces through direct minimization of the normal vibration of its principal transmitting surface. Normal-incidence transmission loss estimates are based on equivalent circuit analysis and corresponding experimental measurements are made using plane wave tube techniques. The paper explains why characteristic single diaphragm configurations are unable to use this strategy effectively. It also discusses two dual diaphragm configurations that use the control scheme particularly well to produce efficient global control of their transmitting surfaces and achieve very high transmission loss over a broad frequency range. The designs facilitate incorporation into full ASP arrays and satisfy other useful criteria.

11:40

3aNS9. Active control of low-speed turbofan tonal noise. Scott D. Sommerfeldt (Dept. of Phys., Brigham Young Univ., Provo, UT 84602-4673, scott_sommerfeldt@byu.edu) and Paul J. Remington (BBN Technologies, Cambridge, MA 02138)

Active noise control has been proposed as a technique for reducing the tonal noise radiated from turbofan engines. The sound field in the duct of a turbofan engine is characterized by acoustic modes, which exhibit both a radial and a circumferential spatial dependence. The dominant circumferential modes are determined by the relationship between the number of rotor and stator blades. Using these concepts, an active noise control system has been developed to measure and minimize the modes in the duct of a turbofan engine. By using multiple source and sensor locations, it has also been shown that it is possible to control multiple radial modes within the engine duct. Some of the issues associated with the design of the control system will be reviewed, and results obtained using the Active Noise Control Fan (ANCF) at NASA Glenn Research Center will be presented. [Work supported by NASA.]

Session 3aPA

Physical Acoustics: Bubbles and Particles

G. Douglas Meegan, Chair
Applied Research Laboratories, The University of Texas–Austin, Austin, Texas 78713-8029

Contributed Papers

9:00

3aPA1. Dynamics of concerted bubble cluster collapse in shock wave lithotripsy. Yuri A. Pishchalnikov, James A. McAteer, Andrew P. Evan (Dept. of Anatomy and Cell Biol., Indiana Univ., 635 Barnhill Dr., Indianapolis, IN 46202, yura@anatomy.iupui.edu), Oleg A. Sapozhnikov (Moscow State Univ., Moscow 119992, Russia), Robin O. Cleveland (Boston Univ., 110 Cummington St., Boston, MA 02215), Tim Colonius (California Inst. of Technol., Pasadena, CA 91125), Michael R. Bailey, and Lawrence A. Crum (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Seattle, WA)

Cavitation bubble cluster collapse at the surface of artificial kidney stones during shock wave lithotripsy was investigated *in vitro* by means of multiframe high-speed photography, passive cavitation detection (PCD), and pressure waveform measurements using a fiber-optic probe hydrophone (FOPH). It was observed that after the passage of the lithotripter shock pulse the stone was covered by numerous individual bubbles. During their growth phase the bubbles coalesced into bubble clusters, with the biggest cluster at the proximal face of the stone. High-speed camera images suggested that cluster collapse started at the periphery and ended with a violent collapse in a small region in the center of the surface of the stone. Shadowgraphy resolved numerous secondary shock waves emitted during this focused collapse. Shock wave emission during cluster collapse was confirmed by PCD. Measurement with the FOPH showed that these shock waves were typically of short duration (0.2 μs). The majority of the shock waves emanating from cluster collapse were low amplitude but some shock waves registered amplitudes on the order of the incident shock pulse (tens of MPa). [Work supported by NIH DK43881, DK55674.]

9:15

3aPA2. Separating nonlinear propagation and cavitation effects in HIFU. Justin Reed, Michael Bailey, Marie Nakazawa (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Box 355640, Seattle, WA 98105-6698), and Vera Khokhlova (Moscow State Univ., Moscow 119992, Russia)

HIFU waves can effectively destroy tumors or stop internal bleeding. The primary physical mechanism in HIFU is the conversion of acoustic energy into heat, which is often enhanced by nonlinear acoustic propagation and nonlinear scattering from bubbles. The goal of this work is to study and separate the effects of nonlinear propagation and cavitation on HIFU heating of tissue. Transparent polyacrylamide gel was used as a tissue-mimicking phantom. The phantom contained protein that turned opaque at high temperature. The lesions were produced for the same time-averaged intensities, but with variation of acoustic pressure amplitude combined with different duty cycles. A computer program was written to quantify the size and shape of lesions captured on CCD. Substantial increase of the lesion size was observed with high-amplitude waves. In order to separate the two effects, the experiments were repeated under static pressure greater than the pressure amplitude of the sound waves, which suppressed cavitation. It was found that with the same time-averaged power, a shorter, higher amplitude pulse creates a lesion faster and grows the lesion larger, than a longer, lower amplitude pulse. Cavitation was found to play a subtle role in the initial formation of the lesion. [Work supported by NIH, NSF, NSBRI.]

9:30

3aPA3. The effect of surface-active solutes on acoustic cavitation bubble populations. Judy Lee, Sandra Kentish (Dept. of Chemical and Biomolecular Eng., Particulate Fluids Processing Ctr., Univ. of Melbourne, VIC 3010, Australia), Muthupandian Ashokkumar, and Franz Grieser (Particulate Univ. of Melbourne, VIC 3101, Australia)

Ultrasound in liquids can cause bubbles to grow to the resonance size via either the process of rectified diffusion or bubble-bubble coalescence. At the resonance size bubbles collapse and emit light, known as sonoluminescence (SL). In this study the effect of surface-active solutes on the initial growth of the active bubble population in the presence of ultrasound (a train of 4 ms pulses of 515 kHz) was monitored. This was achieved by following the SL intensity emitted by the active bubbles, until an equilibrium population of active bubbles was reached, as indicated by a maximum SL intensity. The number of pulses required to reach the maximum SL intensity was found to increase with increasing solute concentration to a maximum and then decrease. The effect of the same surface-active solutes on the total bubble volume in the presence of ultrasound was also investigated. A sharp decrease in the bubble volume is observed with increasing solute concentration. This decrease is believed to be the result of the inhibition of bubble–bubble coalescence. A correlation between the number of pulses required to reach the maximum SL intensity and the bubble–bubble coalescence data will also be discussed in the presentation.

9:45

3aPA4. Cavitation modeling and diesel engine cylinder liners. Gautam Chandekar and Sally Pardue (Mech. Eng., Tennessee Technolog. Univ., P.O. Box 5014, Cookeville, TN 38505, spardue@tntech.edu)

A common occurrence of cavitation damage is the waterside pitting of a wet sleeve liner in a diesel engine. The automotive industry utilizes an ultrasonic test of 20 kHz according to ASTM standards to quantify the effectiveness of engine coolant additives to prevent damage. However, recent tests indicate a mismatch between the ultrasonic test results and actual engine test runs. The focus of this study is to generate numerical models of bubble dynamics using already published literature. In most of the published papers higher-range frequencies (ultrasonic >15 kHz) are used. It is useful to explore the results of lower excitation frequencies as the vibrating frequencies of a diesel engine liner are between 500–9000 Hz. A Rayleigh–Plesset equation, nonlinear in nature, is used to plot the relation between bubble radius and time. Plots of the numerical solution from MATLAB are compared with plots published in the literature. Results from when the frequency of excitation is changed to the liner wall frequency and the fluid properties are changed to approximate engine conditions will be presented. Future work will examine the energy released by the bubble collapse and its correlation with erosion measured as mass change in a standard test button.

10:00

3aPA5. Low-power, cylindrical, air-coupled acoustic levitation/concentration devices. Gregory Kaduchak, Christopher Kwiatkowski, and Dipen Sinha (Los Alamos Natl. Lab., Electron. and Electrochemical Mater. and Devices Group, Los Alamos, NM 87545)

The detection, identification, and separation of aerosols have applications in a wide variety of fields. Particle size, size distribution, and particle composition are quantities that are important to scientists studying both the environmental impact of industrial smog and the detection of chemical or biological agents in air. One common rule in aerosol detection is that most aerosol detection devices (particle sizers, optical classifiers, etc.) are more sensitive at higher aerosol concentrations. The present research describes a new class of acoustic aerosol concentration devices that alleviate many of the alignment and power consumption requirements of traditional acoustic levitators/concentrators [Kaduchak *et al.*, Rev. Sci. Instrum. **73**, 1332–1336 (2002)]. These devices are inexpensive, low-power, and never require alignment. They are constructed from a single PZT tube. The devices operate on the premise that the lowest-order structural modes of the tube are tuned to match a corresponding resonant mode of the interior air-filled cavity. The coupled mode can be driven efficiently; drops of water in excess of 1 mm diameter can be levitated for approximately 100 mW of input electrical power. Both axisymmetric and dipole vibrational modes will be demonstrated in aerosol concentration and collection experiments.

10:15

3aPA6. Ultrasonic particle concentration in a line driven cylindrical tube. Greg Goddard and Gregory Kaduchak (MST-11, Los Alamos Natl. Lab., P.O. Box 1663 MSD-429, Los Alamos, NM 87545)

Acoustic particle manipulation has many uses in flow cytometry and micro-fluidic array applications. Currently, most ultrasonic particle positioning devices utilize a quasi-one-dimensional geometry to set up the positioning field. A transducer fit with a quarter-wave matching layer locally drives a cavity of width one-half wavelength. Particles within the cavity experience a time-averaged drift force that transports them to a nodal position. Present research investigates an acoustic particle-positioning device where the acoustic excitation is generated by the entire structure, as opposed to a localized transducer. The lowest-order structural modes of a long cylindrical glass tube driven by a piezo-ceramic with a line contact are tuned, via material properties and aspect ratio, to match resonant modes of the fluid-filled cavity. The cylindrical geometry eliminates the need for accurate alignment of a transducer/reflector system, in contrast to the case of planar or confocal fields. Experiments show that the lower energy density in the cavity, brought about through excitation of the whole cylinder, results in reduced cavitation, convection, and thermal gradients. The effect of excitation and material parameters on concentration quality and power consumption is theoretically evaluated, using two-dimensional elastodynamic equations describing the fluid-filled cylindrical shell with a line excitation.

10:30

3aPA7. Two-dimensional acoustic aerosol concentration. G. Douglas Meegan and Yurii A. Ilinskii (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

The detection of very dilute but dangerous biological and/or chemical aerosols can be significantly improved by densely concentrating the aerosols in a small air stream before they enter existing analyzers. Aerosol concentration can be accomplished by passing the sampled air stream through a chamber containing a sound field that forces aerosols to align themselves along a single line while the air continues to flow straight through the device. Several well-known physical mechanisms responsible for aerosol concentration include Stokes drag, Kings force, and nonlinear viscosity. However, these existing models assume one-dimensional acoustic fields and other simplifying approximations that are violated in most realistic situations. In this talk, a reformulation of the three physical mechanisms is presented that removes simplifying assumptions and describes aerosol concentration in a multi-dimensional sound field (cylindrical and Cartesian geometries). The influence of the boundary layer surrounding individual particles is included in the calculations. It was found that selective acoustic aerosol concentration (SAAC) is possible whereby aerosols in a specific particle size range can be concentrated and particles outside that size range can be excluded. Experiments will be described that convey the unique capabilities of SAAC that other methods, such as electrostatic aerosol concentration, cannot achieve.

10:45

3aPA8. Accelerated adsorption by activated carbon powders in the presence of an acoustic field. Timothy W. Hawkins, G. Douglas Meegan, and Chris E. Peterson (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029, hawkins@arlut.utexas.edu)

Vaporous contaminants can be removed from a gas stream by adsorption into powdered activated carbon (PAC). For example, the injection of PAC upstream of particle collectors is the leading approach for the removal of mercury vapor from the exhaust streams of incinerators and coal fired power plants. The removal of dilute vapors, though, can require significant time and significant quantities of PAC. A numerical model will be described that suggests that the adsorption process can be accelerated by applying an intense sound field to the gas after the PAC is injected. It is theorized that a sound field improves diffusion limited adsorption reactions by creating an oscillatory motion of the gas relative to the activated carbon particles. The translation effectively mixes the particles and gas on a microscopic scale and improves the concentration gradient at the surface of the PAC. Results from the numerical model suggest that adsorption rates can be increased by over 50% with the application of an appropriate sound field. Simple laboratory experiments have been performed studying the adsorption of dilute vaporous ethanol in air, and the results support the model. Data has also been collected at a coal fired power plant that demonstrates the improved adsorption of mercury vapor.

11:00

3aPA9. Particle sedimentation measurements using active resonator shifts. Christopher S. Kwiatkowski (Electron. and Electrochemical Mater. and Devices, Los Alamos Natl. Lab., Los Alamos, NM 87545)

A measurement of the properties of fluids with suspended particles is important in fields such as biology, chemistry, geology, and industrial processing. The limitation of being inside closed or otherwise inaccessible containers provides a challenge to typical measurement systems, especially where an invasive technique is either disadvantageous or impossible. These situations occur in opaque containers or within pre-existing pipes. The technique presented describes using an active all-ultrasonic technique, where the time-averaged radiation force on particles in an acoustic standing wave causes them to migrate. The migration of particles has a measurable effect on the resonant frequency of the cavity that depends on the properties of the particles and the host fluid. This technique is used to initially determine the presence of suspended particles in a fluid, to characterize the particles, and provide a measurement of the total volume fraction of suspended particles. The measurement of the total volume fraction as a function of time gives information on the particle sedimentation rate. The technique is noninvasive, uses no moving parts, consumes little electrical power, and is capable of relatively fast measurement times on the order of a few seconds.

11:15

3aPA10. Ultrasonic stabilization of small lifted flames. David B. Thiessen, Wei Wei, and Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

Combustion of a fuel gas (such as methane) flowing at low speeds from a nozzle into air produces a flame commonly described as a ''diffusion'' flame. As the flow rate of fuel is increased, the flame is known to lift off the nozzle. This premixes fuel with air and improves certain aspects of the combustion. Ordinarily, however, at large flow rates the flame becomes unstable and blows-off. Furthermore, decreasing the rate of flow to a previously lifted flame ordinarily causes the flame to drop-back and reattach to the nozzle. We demonstrate, however, that the stable range of a lifted flame can be enhanced by applying appropriately positioned ultrasonic standing waves. In the simplest case, a transverse standing wave close to the nozzle suppresses reattachment and facilitates improved pre-mixing at low flow rates. In the other case, an ultrasonic standing wave applied to the flame increases the flow rate for blow-off. These results are apparently the consequence of ultrasonic radiation forces on the flame directed, respectively, away from or toward the nozzle. We previously demonstrated the deformation of small diffusion flames within an ultrasonic standing wave [D. B. Thiessen *et al.*, J. Acoust. Soc. Am. **112**, 2240 (2002)]. [Work supported in part by NASA.]

11:30

3aPA11. Gas mixing diagnostics using acoustic absorption measurements. Aurelien Cottet, Andrew Meyers, David Scarborough, and Tim Lieuwen (School of Aerosp. Eng., Georgia Inst. of Technol., Atlanta, GA 30332)

This paper demonstrates the use of acoustic absorption measurements to quantify the level of mixing between two gases. This work is motivated by the need for quantification of mixing effectiveness in a number of applications, such as in fuel/air preparation in combustion systems or various process control applications in chemical plants. The basic idea of the technique is to interrogate the medium with acoustic waves in the 20–200 kHz region where vibrational relaxation processes are the dominant factor affecting absorption. Our approach takes advantage of the fact that the relaxation frequency for a particular gas is often a strong function of the other species it is in molecular contact with. Thus, the relaxation frequency of a dual component gas mixture can vary substantially with the level of molecular mixedness of the constituent species. We present experimental data illustrating the technique which show that the level of acoustic absorption varies by a factor of five, depending upon whether the constituent gases are unmixed or perfectly mixed. These values agree favorably with theoretical results generated from classical acoustic absorption equations and empirical values of relaxation frequencies.

WEDNESDAY MORNING, 12 NOVEMBER 2003 FRIO ROOM, 9:00 TO 11:00 A.M.

Session 3aPP

Psychological and Physiological Acoustics: Ecology of Complex Acoustics

Edward L. Goshorn, Chair
Department of Speech, Louisiana Tech University, P.O. Box 3165 Tech Station, Ruston, Louisiana 71272

Chair's Introduction—9:00

Contributed Papers

9:05

3aPP1. Effects of eye-glasses, hair, headgear, and clothing on measured head-related transfer functions Part Ib. Klaus A. J. Riederer (Lab. of Acoust. and Signal Processing, Helsinki Univ. of Technol., Otakaari 5 A, P.O. Box 3000, FIN-02015 TKK, Finland, klaus.riederer@hut.fi)

Extensive head-related transfer function (HRTF) measurements show high HRTF repeatability, consequences of different measurement methods, and conditions covering the whole three-dimensional space [Riederer, J. Audio Eng. Soc. (Abstracts) **46**, 1036 (1998), preprint 4846]. This study concentrates on specific effects on HRTFs carefully re-measured on the same Cortex dummy head applying Sennheiser KE4-211-2 microphones at its silicone putty *blocked ear-canal entrances*, employing 252 sound incidents including seven elevations. The effects of five different wigs (synthetic, natural, thick, thin, long and short hair) with varied hairstyles, four hats (cap, bicycle helmet, mens and womens trilby), clothes (alpaca pullover, bicycling drymax-jacket) and spectacles were investigated under 28 combinations. The influences are highly dependent on direction, frequency, and case. Clothes and eye-glasses affect minimally HRTF; hair has a stronger effect, depending on the actual *hairdo* (typically above 7 kHz). Hats alter intensively HRTFs (typically above 5 kHz), depending on the model. The measurements give deeper insight to the development of idiosyncratic features in binaural localization cues. The second part of the study addresses their *perceptual effects* [Riederer, J. Acoust. Soc. Am., this issue]. [Work supported by Graduate School of Electronics, Telecommunication and Automation; thanks to Finnish Broadcasting Company, Mr. Hellstrom; Mrs. Chen.]

9:20

3aPP2. Modeling HRTF of a thin elastic layered spherical model of head. Tran H. Dat (CIAIR, Nagoya Univ., Japan), Itakura Fumitada (Itakura Lab., Nagoya Univ., Japan), and Igor L. Oboznenko (NTUU, Kiev, Ukraine)

The understanding and simplified modeling of the Head Related Transfer Function (HRTF) is very important for applications in virtual audio reality. However HRTF depends on many factors such as the shape of head and pinnae, torso, refraction by the body, and others. In this work we investigated the influences of a thin-elastic layered structure of a head on the HRTF. The analytical solution is applied to a six-layer model of the head. The elastic properties of layers are taken into account in the modeling. The results are compared with the results of spherical rigid model and experiments. The numerical results showed the new peaks of HRTF caused by elastic scattering and re-reflecting from layers. However this effect is very small and can be neglected at high frequency.

9:35

3aPP3. Effects of congruency on localization of audiovisual three-dimensional speech sounds Part IIa. Klaus A. J. Riederer (Lab. of Acoust. and Signal Processing, Helsinki Univ. of Technol., Otakaari 5 A, P.O. Box 3000, FIN-02015 TKK, Finland, klaus.riederer@hut.fi)

Part two of the current study [J. Acoust. Soc. Am., this issue] investigated localization of virtual audiovisual speech under exactly the same conditions. Perceived directions were signified by pushing keypad-buttons. Inside-the-head localization occurred almost only for the median-plane stimuli, being insignificant of the *stimulus type* (7.62% *congruent*, 9.38% *incongruent*, and 6.54% *auditory-only*) and disregarded from further analyses. The mean of correct answers was 46.81%. Factorial within-subjects ANOVA showed no significance on acoustic stimuli (/ipi/, /iti/) or stimulus type but showed *strong dependence* on *direction* ($p = 0.000\,015$) and its interactions with *acoustic stimuli* ($p = 0.015\,374$) and *stimulus type* ($p = 0.008\,12$). *Reaction times* were highly dependent on *direction* ($p = 0.000\,002$). From the 384 frontal location answers (azimuths 0°, ±40°) 25.52% congruent, 28.39% incongruent, and 28.65% auditory-only were perceived as backward confused, for 0° azimuth only the corresponding values were 28.13%, 28.13%, and 35.94%. Back–front confusions were 13.80%, 9.64%, and 8.85% (azimuths 180°, ±130°), and 18.75%, 14.06%, and 14.06% (azimuth 180°). Seeing the (congruently) talking face biased the localization more to the front, especially for the median-backward sounds. Obviously, vision overcomes weaker monaural localization cues as in the *ventriloquism effect* [Driver, Nature (London) **381**, 66–68 (1996)]. [Work supported by Graduate School of Electronics, Telecommunication and Automation.]

9:50

3aPP4. Horizontal plane localization in children. Edward L. Goshorn, Jennifer D. Goshorn, and E. Mary Cone (Speech Dept., Louisiana Tech Univ., P.O. Box 3165 Tech Station, Ruston, LA 71272)

Recently, an ASHA (1996) task force on auditory processing disorders (APD) recommended the assessment of several psychoacoustic processes for identifying APD. One of the processes listed was localization. To date, little data exist on localization capabilities in children. Therefore, this project investigated the feasibility of assessing localization in children with the purpose of developing an APD screening test. Six children, age 6.5 to 7.5 years, have thus far been evaluated. This age range was chosen because it is an optimum age, for educational purposes, to identify APD. Three stimuli were used: a 500-Hz tone, a 4000-Hz tone, and NU-CHIPS monosyllabic words. Tone duration was 1 s; the average word duration was 469 ms. Right–left localization was evaluated at 45 and 315 deg azimuth and front–rear localization at 0 and 180 deg. Each source was presented 24 times in random order to each subject. Although the data were highly variable, results revealed significantly better localization for right–left (88.2%) than for front–rear (80.5%). Within the front–rear condition, performance was better for rear (84.7%) than for front (76.4%) localization. There was no significant difference between right versus left localizations. Across stimuli performance was best for 500 Hz and poorest for words.

10:05–10:30 Break

10:30

3aPP5. Effects of hair, clothing, and headgear on localization of three-dimensional sounds Part IIb. Klaus A. J. Riederer (Lab. of Acoust. and Signal Processing, Helsinki Univ. of Technol., Otakaari 5 A, P.O. Box 3000, FIN-02015 TKK, Finland, klaus.riederer@hut.fi)

Seven 20–25-year-old normal hearing (≤20 dBHL) native male-undergraduates listened twice to treatments of 85 virtual source locations in a large dark anechoic chamber. The 3-D-stimuli were anew-calculated white noise bursts, amplitude modulated (40-Hz sine), repeated after a pause (total duration $3 \times 275 = 825$ ms), HRTF-convolved and headphone-equalized (Sennheiser HD580). The HRTFs were measured from a Cortex dummy head wearing different *garments*: 1 = alpaca pullover only; 2 = 1 + curly pony-tailed thick-hair + eye-glasses; 3 = 1 + long thin-hair (ear-covering); 4 = 1 + mens trilby; 5 = 2 + bicycle helmet + jacket [Riederer, J. Acoust. Soc. Am., this issue]. Perceived directions were signified by placing a tailored digitizer-stylus over an illuminated ball darkened after the response. Subjects did the experiments during three days, each consisting of a 2-h session of several randomized sets with multiple breaks. *Azimuth and elevation errors* were investigated separately in factorial within-subjects ANOVA showing *strong dependence* $p(\leq 0.004)$ *on all main effects and interactions (garment, elevation, azimuth)*. The grand mean errors were approximately 16°–19°. Confused angles were retained around the ±90°-interaural axis and cos(elev)-weighting was applied to azimuth errors. The total front−back/back−front confusion rate was 18.38% and up−down/down−up 12.21%. The *confusions* (except left−right/right−left, 2.07%) and *reaction times* depended strongly on *azimuth* (main effect) and *garment* (interaction). [Work supported by Graduate School of Electronics, Telecommunication and Automation.]

10:45

3aPP6. Negative temporal summation of the responses to pairs of tone bursts in albino mice inferior colliculus. Nikolay G. Bibikov (N. N. Andreyev Acoust. Inst., Schvernik st, 4, Moscow 117036, Russia, bibikov@akin.ru), Chen Qi Cai, and Tang Jie (Central China Normal Univ., Wuhan 430079, PROC)

The extracellular activities of single units in an inferior colliculus of narcotized albino mice have been studied. As a stimuli pairs of best frequency (BF) tone bursts with different duration have been used and forward masking has been studied. The test tone usually has a 40 ms duration at intensity 5 dB above threshold. The intensity and duration of the masker could be changed. It was shown that the forward masking essentially depends upon the duration of the first burst. In many cases, the negative temporal summation can be seen. The increase in the duration of first burst (or masker) leads to the decrease in the whole response. Moreover, the BF tone burst which did not evoke any spike response could inhibit the response to the second (test) tone in some cases. Therefore in many units the inhibitory threshold was lower than the excitatory threshold even at the best frequency. The local application of bicuculline through a multibarrel-electrode increased the pulse activity considerably. However, the effect of forward masking usually left even after an inhibitory antagonist (bicuculline) application. [Work supported by grants 39970251 from NSFC, T010360056 from the Foreign Expert Bureau of the State Council of China, and 02-04-3900 from RFBR-NSFC.]

Session 3aSA

Structural Acoustics and Vibration, Architectural Acoustics and Noise: Control of Sound Transmission Through Partitions

Timothy W. Leishman, Cochair
Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602-4652

Scott D. Sommerfeldt, Cochair
Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602-4652

Invited Papers

8:15

3aSA1. On using ordinary mobilities to predict structure borne power flow from studs to directly attached gypsum board of a wood stud wall—Effect of simplifying assumptions. Trevor Nightingale (Acoust. Lab., Inst. for Res. in Construction, Natl. Res. Council Canada, Ottawa, ON K1A 0R6, Canada) and Katrin Kohler (Bauphysik, Univ. of Appl. Sci. Stuttgart, Schellingstrasse 24, D-70174 Stuttgart, Germany)

This paper reports on the effect of simplifying assumptions made when applying ordinary point mobilities (from thin plate/beam theory) to predict structure borne power, and hence velocity level difference (VLD), from a stud to gypsum board attached by screws. The systematic study showed for practical purposes the fastening points of a gypsum board sheet were sufficiently far from an edge so that the measured mobilities approached those of a drive point located near the sheet center. The mobility of a stud does not change appreciably when attached to the wall head and sole plates. Above the transition frequency from line to point connection where fastener spacing is greater than one half wavelength VLD is inversely proportional to the number of fasteners. Measurements indicate above 1000 Hz local and volumetric deformation of the stud are important causing a very significant VLD across the stud. Deep beam theory, which includes area of the contact between the stud and gypsum board must be included. Between 125–1000 Hz, where the basic assumptions are reasonably satisfied, there is very good agreement between measured VLD and predicted using thin plate/beam mobilities. Predictions indicate agreement above 1000 Hz is greatly improved using deep beam theory.

8:45

3aSA2. Research in the field of active sound transmission control. Timothy Leishman (Dept. of Phys. and Astron., Brigham Young Univ., N335 Eyring Sci. Ctr., Provo, UT 84602, tim_leishman@byu.edu)

Sound transmission through an unperforated partition may be described as the vibro-acoustic process in which fluid-borne sound waves incident upon the partition from the source side force it to vibrate and consequently radiate transmitted fluid-borne sound waves into the space on the opposite (receiving) side. Active reduction of sound transmission through partitions, or active sound transmission control (ASTC), denotes an active modification to this process. It may be defined as the reduction of transmitted sound through the action of electro-mechano-acoustical devices regulating any or all stages of the vibro-acoustic sound transmission process. In certain analyses, this branch of active noise control has referred to situations in which electro-mechanical actuation is applied only to a partition through which sound is transmitting. However, since ASTC may be achieved by other means, and since much research in the field has utilized other techniques, this limitation is clearly too restrictive. This paper reviews and categorizes many research efforts in the field, and provides a brief history of the efforts based on an extensive bibliography of the subject matter. It also identifies research areas that could be pursued in the future.

9:15

3aSA3. Control of sound transmission into a large composite cylinder using passive, adaptive–passive and active control strategies. Marty Johnson, Simon Esteve, and Chris Fuller (Vib. and Acoust. Labs., Dept. of Mech. Eng., Virginia Tech, VA 24061-0238)

This work is directed at the goal of controlling the high noise levels experienced in launch vehicle payload fairings. The problem frequency range is typically below 300 Hz where passive structural damping and sound absorption treatments perform inadequately. In this paper the work conducted at Virginia Tech over the last 3 years on techniques for reducing the interior noise levels at launch using an eight ft. diameter composite cylinder as a test platform is summarized. First the theoretical development will be presented with an overview of the mechanisms of sound transmission. Then, passive control methods using optimally damped structural distributed vibration absorbers (DVAs) and acoustic Helmholtz resonators will be presented. Both theoretical results and experimental

results using high level acoustic excitation (132 dB) will be described. A set of adaptive Helmholtz resonators were also built and tested and are capable of tracking changes in acoustic resonances automatically. Finally, the results from a fully active structural acoustic control experiment using distributed active vibration absorbers will be presented.

9:45

3aSA4. Transducer placement for robustness to variations in boundary conditions for active structural acoustic control. Joseph D. Sprofera, Robert L. Clark (Duke Univ., Durham, NC 27708), Gary P. Gibbs, and Randolph H. Cabell (NASA Langley Res. Ctr., Hampton, VA 23681)

The study of control strategies aimed at the reduction of turbulent boundary layer noise transmission into the fuselage of an aircraft has been a topic of academic and industrial research for several years. In this work we focus on an approach that will attempt to address a practical application issue: the impact of uncertainty in boundary conditions on the choice of actuator and sensor locations for structural acoustic control. The selection of an optimized set of transducers and the creation of a suitably adjustable test system is used to demonstrate that robustness to bounded variations in boundary conditions of a plate is achievable in active structural acoustic control. This robustness is achieved through the optimization of transducer placement with respect to maximizing control of structural acoustic radiation over the desired range of boundary conditions. This project incorporates energy based modeling of the structure, electromechanical piezo coupling, radiation filter modeling, and control analysis through the use of Hankel Singular Value (HSV) estimates and optimization based upon the genetic algorithm. Testing in a transmission loss facility will be used to validate the selected transducer placements and demonstrate a reduction in radiated sound power. [Work supported by NASA.]

10:15–10:30 Break

10:30

3aSA5. Active control of flow-induced noise transmitted through aircraft sidewalls. Cedric Maury, Paolo Gardonio, and Stephen J. Elliott (Inst. of Sound and Vib. Res., Univ. of Southampton, Highfield, Southampton SO17 1BJ, UK, cm@isvr.soton.ac.uk)

In this paper, the theory and the results are presented of a study concerned with the active control of sound transmitted through aircraft sidewalls when excited by a Turbulent Boundary Layer. The motivation for this study is the control within an aircraft cabin of the flow-induced noise due to wall–pressure fluctuations over the fuselage. First, a theoretical model accounting for the effect of elastic stiffeners is obtained to calculate the boundary layer noise transmitted through tensioned aircraft sidewalls. Parametric studies including comparisons between different models of aircraft fuselage panels valid over different frequency ranges are analyzed. Second, the best possible noise levels reduction that can be achieved are determined using idealized feedback control of the boundary layer noise transmitted through tensioned and stiffened double-panel partitions. It is shown that it is the strategy based on the active suppression of the skin panel structural modes that is the most efficient in terms of the broadband attenuation of the total sound power inwardly radiated. However, it may be more feasible to control the vibration of the trim panel or the pressure in the cavity, and the simulations suggest that the inwardly radiated sound power can still be reduced by a significant amount.

11:00

3aSA6. Optimal control forces for absorption and transmission through composite porous layers. Michael Kidner (Vib. and Acoust. Lab., Dept. of Mech. Eng., Virginia Polytechic Inst. and State Univ., Blacksburg, VA 24060)

A composite infinite panel consisting of a plate, a foam layer, a layer of active forces, and another foam layer could be constructed such that the absorbtion coefficient on the foam side is equal to unity and the TL is infinite. This paper derives the required active force and discusses conflicts between achieving these goals with one set of forces. The practical implications of realizing this optimal force distribution are discussed. The effect of limited power and discrete actuation of the active layer is investigated. The influence of the macroscopic properties of the foam layers on the performance of the system is also considered.

11:15

3aSA7. Experimental and analytical study of active control of energy transmission through double walls using novel piezoelectric actuators. Jianxin Gao and Li Cheng (Dept. of Mech. Eng., The Hong Kong Polytechnic Univ., Hong Kong)

Active noise control has recently been used to increase the sound transmission loss of double wall structures. Vibration energy transmission through a double-plates system is investigated experimentally and analytically in this paper. Novel high performance actuators which use two curved THUNDER actuators as active driving components are developed and mounted between the two plates to control the energy transmission in the double-plates system. The Rayleigh–Ritz method and eigenfunction expansion theorem are used to resolve arbitrary boundary conditions of plates, and the feedforward control strategy is employed in this paper. The time average power transmission between the source, actuators and receiver plate are discussed and utilized as a cost function to obtain optimal control. The optimal control voltages for actuators are obtained by minimizing the cost function. A double-plates system connected by four actuators is set up experimentally in order to verify the models and formulations by comparing with analytical results. The analytical and experimental data show that the new actuator exhibits excellent performance on active control of power transmission.

11:30

3aSA8. Vibrations in buildings due to helicopter operations from rooftop heliports. Jeffrey Zapfe and Eric Ungar (Acentech, Inc., 33 Moulton St., Cambridge, MA 02138-1118)

Most major hospitals are equipped with rooftop heliports. This paper explores the impact of helicopter-induced vibrations inside the host building, particularly as they impact the operation of sensitive equipment. Helicopter-induced forces are characterized for two helicopters: a Eurocopter BK-117C1 and a Robinson R44. Force time histories show that the greatest forces occur just prior to touchdown and just after liftoff. Force spectra show that the principal dynamic forces occur at the blade passage frequency and amount to 2% to 14% of the helicopters weight for the BK-117C1 and R44, respectively. The force data is used to predict vibration in a planned hospital building.

11:45

3aSA9. Impedance tube characterization of a novel low frequency acoustic barrier. B. Kyle Henderson and Scott E. Franke (Air Force Res. Lab., Space Vehicles Directorate, 3500 Aberdeen Ave., Bldg. 472, Kirtland AFB, NM 87117-5776)

The launch acoustic environment impinging on space payloads during ascent can be damaging to sensitive critical spacecraft components. Conventional foam acoustic mitigation for payload fairings tends to be ineffective at low frequencies due to volume and weight constraints. Acoustic barriers may be employed to increase transmission loss at lower frequencies through the use of a heavy septum layer within the foam, but with the concomitant disadvantage of added weight, which reduces payload capacity. A lightweight barrier is presented employing a thin metal foil bonded to a foam layer with spherical masses embedded in the foil in a periodic array. It is theorized that the particles enable the excitation of bending waves in the plane of the foil, causing inter-particle interactions over a frequency range influenced by the foil material properties, thickness, and particle spacing which increase the reflection coefficient of the layer. The transmission loss and absorption spectra are measured using the ASTM C384 standard. The results should allow for optimization of foil-foam performance with respect to geometric and material variables. It is expected that these foil barriers might provide protection over low frequency bandwidths identified by analysis as harmful to a payload.

WEDNESDAY MORNING, 12 NOVEMBER 2003

RIO GRANDE EXHIBIT HALL, 8:30 A.M. TO 12:00 NOON

Session 3aSC

Speech Communication: Speech Production (Poster Session)

Carol Y. Espy-Wilson, Chair

Department of Electrical and Computer Engineering, University of Maryland, College Park, Maryland 20742

Contributed Papers

All posters will be on display from 8:30 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 a.m. to 10:15 a.m. and contributors of even-numbered papers will be at their posters from 10:15 a.m. to 12:00 noon.

3aSC1. Auditory processing deficits among language-learning disordered children and adults. Ratree Wayland (Prog. in Linguist., Univ. of Florida, P.O. Box 115454, Gainesville, FL 32611-5454) and Linda Lombardino (Univ. of Florida, Gainesville, FL 32611)

It has been estimated that approximately 5%–9% of school-aged children in the United States are diagnosed with some kind of learning disorders. Moreover, previous research has established that many of these children exhibited perceptual deficits in response to auditory stimuli, suggesting that an auditory perceptual deficit may underlie their learning disabilities. The goal of this research is to examine the ability to auditorily process speech and nonspeech stimuli among language-learning disabled (LLD) children and adults. The two questions that will be addressed in this study are: (a) Are there subtypes of LLD children/adults based on their auditory processing deficit, and (b) Is there any relationship between types of auditory processing deficits and types of language deficits as measured by a battery of psychoeducational tests.

3aSC2. Patterns of acoustic variance in native and non-native phonemes: The case of Japanese production of /r/ and /l/. Momoko Sato, Andrew J. Lotto, and Randy L. Diehl (Dept. of Psych., Univ. of Texas, Austin, TX 78712, sato@psy.utexas.edu)

Acoustic measures were obtained from productions of word-initial and word-final /r/ and /l/ in different vowel contexts (/i/, /a/, and /u/) by six native English Speakers (ES) and six native Japanese Speakers (JS). JS productions of the Japanese rhotic flap were also obtained in the same vowel contexts. $F1$, $F2$, and $F3$ were measured at vowel midpoint, at utterance onset for word-initial stimuli, and at utterance offset for word-final stimuli. For word-initial /r/-/l/, ES productions were distinguished solely by $F3$, whereas JS productions varied on both $F2$ and $F3$. For the word-final contrast, ES and some JS productions showed structured variation in both $F2$ and $F3$, whereas many JS productions showed no consonant-dependent variation in these measures. $F2$ and $F3$ measures of the Japanese flap (word-initial only) substantially overlapped the /r/-/l/ distributions for ES and JS. In addition, JS productions of /r/ and /l/ and the flap showed significant effects of context vowel with coarticulatory variations in $F2$ and $F3$. In contrast, ES productions showed little context-dependent $F3$ variance. The results suggest that JS difficulties in producing intelligible English liquids may be due to a mismatch between the acoustic variance that is informative for similar phonemes in the two languages.

3aSC3. Toward data-driven modeling of dynamic vocal-tract data. Abhinav Sethy, Shrikanth Narayanan, Sungbok Lee, and Dani Byrd (USC, 3737 Watt Way, PHE336, Los Angeles, CA 90089, sethy@usc.edu)

Speech production modeling relies on various forms of biometric measurements for getting articulatory data. Recent advances in real-time magnetic resonance imaging (rtMRI) capabilities [Narayanan *et al.*, J. Acoust. Soc. Am. **113**, 2258 (2003)] which make it possible to capture vocal tract images at 24-fps promise to illuminate finer details of speech production. This study presents our first effort toward automatically constructing statistical models of the vocal tract from rtMRI data. We present an automated method of extracting the regions of interest from image sequences based on Kalman snakes and optical flow [Cootes *et al.*, IVCV July 1994]. The time resolution of this image tracking system is improved by incorporating information from parallel vocal tract movement data obtained by the electromagnetic articulography (EMA) system [Perkell *et al.*, J. Acoust. Soc. Am. **92**]. EMA data provide kinematic information such as velocity and acceleration at several points along the vocal tract. Data from segmentation and tracking are merged with information from EMA and from acoustic analysis of speech in order to obtain a time series of feature

vectors. Feature space reduction and statistical parametrization schemes such as PCA are then applied to get parametric models for the articulatory system. [Work supported by NIH Grant No. DC 03172.]

3aSC4. Perception of relative location of *F*0 within the speaking range. Douglas N. Honorof and D. H. Whalen (Haskins Labs., 270 Crown St., New Haven, CT 06511, honorof@haskins.yale.edu)

It has been argued that intrinsic fundamental frequency (I*F*0) is an automatic consequence of vowel production [Whalen *et al.*, J. Phon. **27**, 125–142 (1999)], yet speakers do not adjust *F*0 so as to overcome I*F*0. It may be that so adjusting *F*0 would distort information about *F*0 range–information important to the interpretation of *F*0. Therefore, a speech production/perception experiment was designed to determine whether listeners can perceive position within a speaker-specific *F*0 range on the basis of isolated tokens. Ten male and ten female adult native speakers of US English were recorded speaking (not singing) the vowel /a/ on eight different pitches spaced throughout speaker-specific ranges. Recordings were randomized across speakers. Naive listeners made pitch-magnitude estimates of the location of *F*0 relative to each speaker's range. Preliminary results show correlations between estimated and actual location within the range. Adjusting *F*0 to compensate for I*F*0 differences between vowels would seem to obscure voice quality in such a way as to make it difficult for the listener to recover relative *F*0, requiring a greater perceptual adjustment than simply normalizing for I*F*0. [Work supported by NIH Grant No. DC02717.]

3aSC5. Acoustics of the glottal source at closures for fricative and stop consonants. Ariel Salomon (MIT, RLE Speech Commun. Group, 77 Massachusetts Ave. 36-511, Cambridge, MA 02139)

Previous research on acoustic distinctions between manner classes of obstruent consonants has primarily considered differences at the release. This study investigates acoustic distinctions in the glottal waveforms at the closure. It is hypothesized that the higher pressure generated in a complete closure (stops and affricates) results in a more rapid cessation of vibration of the glottis than for a partial closure (fricatives). This can be measured by rate of energy offset at the closure landmark. This work studies measurements of the rate of change of energy at closure in various frequency bands. For lower frequencies, in the range of the first formant or below, it is expected that there is interaction between voicing and manner due to the presence of a ''voicebar'' (continued vibration) during the closure. Preliminary measurements show that both voiced and voiceless stop consonants exhibit more rapid rates of energy offset than fricatives in a frequency band containing the second formant. This measure may be useful in noisy environments where the energy of frication is obscured; the periodic glottal waveform can be more reliably measured in noise, e.g., by autocorrelation-based methods. [Work supported by NIH Grant No. DC02978.]

3aSC6. The specificity of lexical confusability effects on coarticulation. Rebecca Scarborough (Dept. of Linguist., UCLA, 3125 Campbell Hall, Los Angeles, CA 90095, rab@humnet.ucla.edu)

Brown (2001) demonstrated that speakers adjust their production of coarticulation to accommodate potential lexical confusability (the possibility of one word being misheard as another), and that listeners' lexical recognition is facilitated by such adjustments. Speakers produced confusable, ''hard'' words (those with low-usage frequencies and many frequent, phonologically similar neighbors) with significantly more coarticulation than less confusable, ''easy'' ones (with high frequencies and few, low-frequency neighbors). The current study investigates whether these adjustments are general, applying to any potentially confusable words, or whether the adjustments are sublexical, sensitive to the particular neighbors a given word has. A frequency-balanced set of ''easy'' and ''hard'' words containing a vowel+nasal consonant sequence has been found such that half are confusable by the nasal and half are not. Six speakers read these words twice in carrier sentences. The degree of nasal coarticulation during the vowel is measured acoustically. From this it will be seen whether the coarticulatory cues to a segment are more enhanced (i.e., there is more coarticulation) in words confusable by that segment than in words with no neighbors differing by that segment. Preliminary results suggest that coarticulatory adjustments in response to lexical confusability are made at the word level rather than sublexically.

3aSC7. Asymmetric kinematic changes in speaking rate explored with FDA. Sungbok Lee (Dept. of Linguist. and Dept. of Elec. Eng., USC, 3601 Watt Way, Los Angeles, CA 90089-1693, sungbokl@usc.edu), Shrikanth Narayanan, and Dani Byrd (USC, Los Angeles, CA 90089-1693)

We report preliminary results for the effect on tongue movements of varying speaking rate. Speech acoustics and tongue tip and body kinematic data were collected using electromagnetic articulography (EMA) [Perkell *et al.*, J. Acoust. Soc. Am. **92** (1992)]. A subject read each one of 20 sentences three times in a row with a predefined order of combination of three speaking rates (normal, medium, fast). Velocity and acceleration were derived using functional data analysis (FDA) [Ramsay *et al.*, J. Acoust. Soc. Am. **99** (1996)] for smoother estimation of acceleration from the displacement data. Global utterance analyses of phase-space and variability in the displacement-velocity and the velocity-acceleration relations indicate that the fast speech exhibits comparable tongue tip displacements to normal speech but significantly larger accelerations ($p<0.00$). The difference in the order of combination of speaking rates (e.g., normal changing to fast versus fast to normal) shows significant effects on both voicing effort (expressed as rms energy) and the kinematic parameters. Finally, kinematic variability at some points in the utterance is much more constrained than at other points, irrespective of speaking rate. A possible interpretation of this is that particular portions of the utterance require this stability for linguistic or articulatory reasons. [Work supported by NIH.]

3aSC8. Effects of obstruent consonants on the *F*0 contour. Helen M. Hanson (MIT RLE, 36-585, 50 Vassar St., Cambridge, MA 02139)

When a vowel follows an obstruent consonant, the fundamental frequency in the first few tens of milliseconds of the vowel is influenced by the voicing characteristics of the consonant. The goal of the research reported here is to model this influence, with the intention of improving generation of *F*0 contours in rule-based speech synthesis. Data have been recorded from 10 subjects. Stops, fricatives, and the nasal /m/ were paired with the vowels /i,ɑ/ to form CVm syllables. The syllables mVm served as baselines with which to compare the obstruents. The target syllables were embedded in carrier sentences. Intonation was varied so that each target syllable was produced with either a high, low, or no pitch accent. Results vary among subjects, but in general, obstruent effects on *F*0 primarily occur when the syllable carries a high pitch. In that case, *F*0 is increased relative to the baseline following voiceless obstruents, but *F*0 closely follows the baseline following voiced obstruents. After voiceless obstruents, *F*0 may be increased for up to 80 ms following voicing onset. When a syllable carries a low or no pitch accent, *F*0 is increased slightly following all obstruents. [Work supported by NIH Grant No. DC04331.]

3aSC9. An audiovisual database of English speech sounds. Stefan A. Frisch and Dee Adams Nikjeh (Dept. of CSD, Univ. of South Florida, 4202 E. Fowler Ave. PCD1017, Tampa, FL 33620, frisch@chuma1.cas.usf.edu)

A preliminary audiovisual database of English speech sounds has been developed for teaching purposes. This database contains all Standard English speech sounds produced in isolated words in word initial, word medial, and word final position, unless not allowed by English phonotactics. There is one example of each word spoken by a male and a female talker.

The database consists of an audio recording, video of the face from a 45 deg angle off of center, and ultrasound video of the tongue in the midsaggital plane. The files contained in the database are suitable for examination by the Wavesurfer freeware program in audio or video modes [Sjolander and Beskow, KTH Stockholm]. This database is intended as a multimedia reference for students in phonetics or speech science. A demonstration and plans for further development will be presented.

3aSC10. Simulation of VCV syllables with a parametric area function model of the vocal tract. Brad Story (Dept. of Speech and Hearing Sci., Univ. of Arizona, P.O. Box 210071, Tucson, AZ 85721-0071)

A model of the vocal tract area function has been developed that consists of three tiers. The first tier is a vowel substrate defined by a system of spatial eigenmodes determined from MRI and/or x-ray microbeam data and a neutral area function. The input parameters to this first tier are coefficient values that, when multiplied by the appropriate eigenmode and added to the neutral area function, construct a desired vowel. The second tier consists of a Gaussian shaping function defined along the length of the vocal tract that modifies the vowel substrate such that a constriction is formed. Input parameters consist of the location, area, and range of the constriction. Location and area roughly correspond to the standard phonetic notions of place and degree of constriction while range defines the amount of vocal tract length over which the constriction will influence the tract shape. The third tier allows length modifications for articulatory maneuvers such as lip rounding/spreading and larynx lowering/raising. All parameters can be specified either as static or time-varying, which allows for multiple levels of coarticulation to be produced. The model will be discussed and then used to simulate several VCV syllables. [Work supported by NIH R01-DC04789.]

3aSC11. An integrated analysis of speech and gestural characteristics in conversational child–computer interactions. Serdar Yildirim, Simona Montanari, Elaine Andersen, and Shrikanth S. Narayanan (Dept. of Elec. Eng. and Linguist., Univ. of Southern California, 3740 McClintock Ave., EEB413, Los Angeles, CA 90089, yildirim@sipi.usc.edu)

Understanding the fine details of children's speech and gestural characteristics helps, among other things, in creating natural computer interfaces. We analyze the acoustic, lexical/non-lexical and spoken/gestural discourse characteristics of young children's speech using audio-video data gathered using a Wizard of Oz technique from 4 to 6 year old children engaged in resolving a series of age-appropriate cognitive challenges. Fundamental and formant frequencies exhibited greater variations between subjects consistent with previous results on read speech [Lee *et al.*, J. Acoust. Soc. Am. **105**, 1455–1468 (1999)]. Also, our analysis showed that, in a given bandwidth, phonemic information contained in the speech of young child is significantly less than that of older ones and adults. To enable an integrated analysis, a multi-track annotation board was constructed using the ANVIL tool kit [M. Kipp, Eurospeech 1367–1370 (2001)]. Along with speech transcriptions and acoustic analysis, nonlexical and discourse characteristics, and child's gesture (facial expressions, body movements, hand/head movements) were annotated in a synchronized multilayer system. Initial results showed that younger children rely more on gestures to emphasize their verbal assertions. Younger children use non-lexical speech (e.g., um, huh) associated with frustration and pondering/reflecting more frequently than older ones. Younger children also repair more with humans than with computer.

3aSC12. Speech synthesis with pitch modification using harmonic plus noise model. Parveen K. Lehana and Prem C. Pandey (EE Dept., IIT Bombay, Mumbai-400 076, India)

In harmonic plus noise model (HNM) based speech synthesis, the input signal is modeled as two parts: the harmonic part using amplitudes and phases of the harmonics of the fundamental and the noise part using an all-pole filter excited by random white Gaussian noise. This method requires relatively less number of parameters and computations, provides good quality output, and permits pitch and time scaling without explicit estimation of vocal tract parameters. Pitch scaling to synthesize the speech with interpolated original amplitudes and phases at the multiples of the scaled pitch frequency results in an unnatural quality. Our investigation for obtaining natural quality output showed that the frequency scale of the amplitudes and phases of the harmonics of the original signal needed to be modified by a speaker dependent warping function. The function was obtained by studying the relationship between pitch frequency and formant frequencies for the three cardinal vowels naturally occurring with different pitches in a passage with intonation. Listening tests showed that good quality speech was obtained by linear frequency scaling of the amplitude and phase spectra, by the same factor as the pitch-scaling.

3aSC13. Temporal attractors for speech onsets. Robert Port and Eric Oglesbee (Dept. of Linguist., Indiana Univ., Bloomington, IN 47405, port@indiana.edu)

When subjects say a single syllable like *da* in time with a metronome, what is the easiest relationship? Superimposed on the metronome pulse, of course. The second easiest way is probably to locate the syllable halfway between pulses. We tested these hypotheses by having subjects repeat *da* at both phase angles at a range of metronome rates. The vowel onset (or *P*-center) was automatically obtained for each token. In-phase targets were produced close to the metronome onset for rates as fast as 3 per second. Antiphase targets were accurate at slow rates ($\sim$2/s) but tended to slip to inphase timing with faster metronomes. These results resemble the findings of Haken *et al.* [Biol. Cybern. **51**, 347–356 (1985)] for oscillatory finger motions. Results suggest a strong attractor for speech onsets at zero phase and a weaker attractor at phase 0.5 that may disappear as rate is increased.

3aSC14. Acoustical study of the development of stop consonants in children. Annika K. Imbrie (Div. of Health Sci. and Technol., Harvard MIT, Rm. 36-545, 77 Massachusetts Ave., Cambridge, MA 02139, imbrie@mit.edu)

This study focuses on the acoustic patterns of stop consonants and adjacent vowels as they develop in young children (ages 2;6–3;3) over a six month period. The acoustic properties that are being measured for stop consonants include spectra of bursts, frication noise and aspiration noise, and formant movements. Additionally, acoustic landmarks are labeled for measurements of durations of events determined by these landmarks. These acoustic measurements are being interpreted in terms of the supraglottal, laryngeal, and respiratory actions that give rise to them. Preliminary data show that some details of the child's gestures are still far from achieving the adult pattern. The burst of frication noise at the release tends to be shorter than adult values, and often consists of multiple bursts. From the burst spectrum, the place of articulation appears to be normal. Finally, coordination of closure of the glottis and release of the primary articulator is still quite variable, as is apparent from a large standard deviation in VOT. Analysis of longitudinal data on young children will result in better

models of the development of the coordination of articulation, phonation, and respiration for motor speech production. [Work supported by NIH Grants Nos. DC00038 and DC00075.]

3aSC15. Developing a corpus of spoken language variability. Lesley Carmichael, Richard Wright, and Alicia Beckford Wassink (Dept. of Linguist., Univ. of Washington, Seattle, WA 98195-4340)

We are developing a novel, searchable corpus as a research tool for investigating phonetic and phonological phenomena across various speech styles. Five speech styles have been well studied independently in previous work: reduced (casual), careful (hyperarticulated), citation (reading), Lombard effect (speech in noise), and ''motherese'' (child-directed speech). Few studies to date have collected a wide range of styles from a single set of speakers, and fewer yet have provided publicly available corpora. The pilot corpus includes recordings of (1) a set of speakers participating in a variety of tasks designed to elicit the five speech styles, and (2) casual peer conversations and wordlists to illustrate regional vowels. The data include high-quality recordings and time-aligned transcriptions linked to text files that can be queried. Initial measures drawn from the database provide comparison across speech styles along the following acoustic dimensions: MLU (changes in unit duration); relative intra-speaker intensity changes (mean and dynamic range); and intra-speaker pitch values (minimum, maximum, mean, range). The corpus design will allow for a variety of analyses requiring control of demographic and style factors, including hyperarticulation variety, disfluencies, intonation, discourse analysis, and detailed spectral measures.

3aSC16. Finite element analysis of airflow in the vocal tract with lateral channels. Zhaoyan Zhang and Carol Y. Espy-Wilson (Dept. of Elec. and Computer Eng., Univ. of Maryland, College Park, MD 20742)

Lateral channels are airflow paths around the tongue produced by the laterally inward movement of the tongue toward the midsagittal plane during American English /l/ sound production. If contact is made with the palate, a closure is formed in the flow path along the midsagittal line. The closure is normally formed in the anterior part of the oral cavity and is about 1–1.5 cm long. However, it is speculated that the flow may split at a location posterior to the closure, thereby giving a longer length of the lateral channels up to 3–4 cm. Lateral channels of length around 3 cm have been shown to have significant effects on the resulting sound spectrum. To investigate the flow and acoustic field involved, finite element analysis was performed on a simplified model of the vocal tract during lateral sound production. The tongue was modeled as a rectangular constriction with a tapering slope on the upstream side and two flow channels on its two sides. The results show that the rising up of the tongue causes the flow to split into three regions of different flow amplitude and phase: one main region above the tongue surface and two regions around the tongue. This flow splitting occurs at the point where the tongue first begins rising up, well before the actual constriction location. The effective length of the lateral channels is therefore much longer than the length of the lingual constriction.

3aSC17. Pause length and speech rate as durational cues for prosody markers. Jeung-Yoon Choi (Dept. of Elec. and Computer Eng., Univ. of Illinois–Urbana-Champaign, 2005 BI, 405 N. Mathews Ave., Urbana, IL 61801)

Two durational cues for prosody are examined for utterances in the Radio Speech Corpus. The acoustic cues are length of pause and speech rate associated with each syllabic segment. Pauses are defined as intervals where probability of voicing and energy (rms) fall below respective thresholds. Speech rate is defined as the reciprocal of the duration of the syllabic segment, without normalization with respect to speaker or segment identity. Each syllabic segment was assigned prosodic markers that are a combination of degree of accent and boundary, i.e., {unaccented, accented} $\times$ {nonboundary, intermediate boundary, intonational boundary}. Distributions of pause length and speech rate for each type of prosodic marker show that in general, mean pause length increases and speech rate decreases as the strength of the boundary increases from nonboundary to intonational boundary. Accented syllabic segments at intermediate or intonational boundaries showed longer associated pause length and slower speech rate than unaccented segments. Also, for different types of intonational boundary tones, mean pause length decreases in the order of: L-L%, H-H%, H-L%, L-H%. Preliminary classification results using the two durational cues show detection rates for intonational boundaries around 75%, with insertion rates around 25%.

3aSC18. Age-related changes in the anticipatory coarticulation in the speech of young children. Mathew Parson, Amanda Lloyd, Kelly Stoddard, and Shawn L. Nissen (Dept. of Audiol. & Speech-Lang. Pathol., Brigham Young Univ., Provo, UT 84602)

This paper investigates the possible patterns of anticipatory coarticulation in the speech of young children. Speech samples were elicited from three groups of children between 3 and 6 years of age and one comparison group of adults. The utterances were recorded online in a quiet room environment using high quality microphones and direct analog-to-digital conversion to computer disk. Formant frequency measures (F1, F2, and F3) were extracted from a centralized and unstressed vowel (schwa) spoken prior to two different sets of productions. The first set of productions consisted of the target vowel followed by a series of real words containing an initial CV(C) syllable (voiceless obstruent–monophthongal vowel) in a range of phonetic contexts, while the second set consisted of a series of nonword productions with a relatively constrained phonetic context. An analysis of variance was utilized to determine if the formant frequencies varied systematically as a function of age, gender, and phonetic context. Results will also be discussed in association with spectral moment measures extracted from the obstruent segment immediately following the target vowel. [Work supported by research funding from Brigham Young University.]

3aSC19. A preliminary ultrasound study of velar fronting. Sylvie M. Wodzinski and Stefan A. Frisch (Dept. of CSD, Univ. of South Florida, 4202 E. Fowler Ave. PCD1017, Tampa, FL 33620, frisch@chuma1.cas.usf.edu)

The purpose of this study is to (1) evaluate the effectiveness of ultrasound imaging to measure velar consonant closure location, and (2) conduct a thorough study of velar fronting by measuring several productions of velar stops in the context of every English vowel. Word onset velar stops were measured in both words (CV or CVC) and nonwords (VCV) within a carrier phrase. Other coarticulatory influences were minimized by using words with no coda or labial coda consonants (e.g., ''Say a gap again,'' ''Say /oIkoI/ again''). Measurements were made at the point of maximal closure. Closure location was measured using the radial angle from the center of the ultrasound probe to the center of the velar closure. Pilot data for one subject has been analyzed to date. Closure location appears consistent across all central and back vowels. For front vowels, the degree of fronting of the velar appears to be correlated with the frontness of the vowel. Measures of closure location for diphthongs followed the back vowel pattern in the word targets. For nonwords, the closure location was influenced by the preceding diphthong offset quality and the following diphthong onset quality. Theoretical implications for the phonetics/phonology interface will be discussed.

3aSC20. The relative amplitude of vowel formants: Acoustic analysis. Ewa Jacewicz (Dept. of Speech and Hearing Sci., The Ohio State Univ., Columbus, OH 43210, jacewicz.1@osu.edu)

Acoustically, vowels as speech signals transport power, which is expressed in their overall intensity level. This power varies intrinsically with vowel category and with its immediate consonantal context. These seminal findings [House and Fairbanks, J. Acoust. Soc. Am. **22**, 105–113 (1953)] have received little systematic consideration over the past 50 years. The

present work is a partial replication and extension of the House and Fairbanks study to formant levels, providing a set of data on intensity variations for eight selected American English vowels. Parametric investigation includes overall rms amplitude, amplitude of $F0$, and relative amplitudes of formants $F1-F4$. The study examines how the relative levels of particular formants contribute to the overall vowel intensity and how this distribution changes in consonantal contexts. Changes in vowel intensity (overall intensity and that of the first four formants) as a function of the consonantal context (both symmetrical and asymmetrical) are examined for each consonantal feature: place and manner of articulation, and voicing. Additionally, the intrinsic intensity is measured from sustained vowels recorded in isolation. The results are discussed in relation to the contributions of primary (i.e., formant frequency pattern) and secondary cues to vowel identity. [Work supported by NIDCD R03 DC005560−01.]

3aSC21. Orthography and underlying form in incomplete neutralization. Natasha Warner (Univ. of Arizona and Max Planck Inst., Dept. of Linguist., P.O. Box 210028, Tucson, AZ 85721-0028, nwarner@u.arizona.edu), Erin Good (Univ. of Arizona, Tucson, AZ 85721-0028), Allard Jongman, and Joan Sereno (Univ. of Kansas, Lawrence, KS 66045)

Either orthographic distinctions or underlying morphological distinctions might cause incomplete neutralization. Much previous work has found small durational differences between pairs such as German ''Rad'' ''wheel'' and ''Rat'' ''advice,'' which are traditionally said to be neutralized to [rat] in both cases. Incomplete neutralization effects are stronger when the task emphasizes orthography, and when the underlying distinction is orthographically encoded. Previous work by the authors investigated a case where orthography differs, but the underlying form of the two words is arguably the same, e.g., Dutch ''heten'' ''to be called'' and ''heetten'' ''they were called,'' both /he:tən/. Incomplete neutralization was found, indicating a role for orthography rather than an underlying form. However, that case may also be interpreted as having a difference in the underlying form because of the past tense morpheme /-t/ in ''heetten.'' The current work investigates pairs which have the same underlying morphological difference, but no difference in spelling, e.g., ''(ik) laat'' (/la:t/) ''(I) leave'' and ''(hij) laat'' (/la:t-t/) ''(he) leaves.'' These pairs show no significant durational differences, despite an experimental design with quite high power. Thus, incomplete neutralization effects can be caused by an orthographic difference, but cannot be caused by this type of morphological difference.

3aSC22. Vowel production, speech-motor control, and phonological encoding in people who are lesbian, bisexual, or gay, and people who are not. Benjamin Munson and Nancy DeBoe (Dept. of Commun. Disord., Univ. of Minnesota, 115 Shevlin Hall, 164 Pillsbury Dr. SE, Minneapolis, MN 55455)

A recent study (Pierrehumbert, Bent, Munson, and Bailey, submitted) found differences in vowel production between people who are lesbian, bisexual, or gay (LBG) and people who are not. The specific differences (more fronted /u/ and /a/ in the non-LB women; an overall more-contracted vowel space in the non-gay men) were not amenable to an interpretation based on simple group differences in vocal-tract geometry. Rather, they suggested that differences were either due to group differences in some other skill, such as motor control or phonological encoding, or learned. This paper expands on this research by examining vowel production, speech-motor control (measured by diadochokinetic rates), and phonological encoding (measured by error rates in a tongue-twister task) in people who are LBG and people who are not. Analyses focus on whether the findings of Pierrehumbert *et al.* (submitted) are replicable, and whether group differences in vowel production are related to group differences in speech-motor control or phonological encoding. To date, 20 LB women, 20 non-LB women, 7 gay men, and 7 non-gay men have participated. Preliminary analyses suggest that there are no group differences in speech motor control or phonological encoding, suggesting that the earlier findings of Pierrehumbert *et al.* reflected learned behaviors.

3aSC23. Prosodic domain-initial effects on the acoustic structure of vowels. Robert Allen Fox, Ewa Jacewicz (Dept. of Speech and Hearing Sci., The Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210, fox.2@osu.edu), and Joseph Salmons (Univ. of Wisconsin, Madison, Madison, WI 53706)

In the process of language change, vowels tend to shift in ''chains,'' leading to reorganizations of entire vowel systems over time. A long research tradition has described such patterns, but little is understood about what factors motivate such shifts. Drawing data from changes in progress in American English dialects, the broad hypothesis is tested that changes in vowel systems are related to prosodic organization and stress patterns. Changes in vowels under greater prosodic prominence correlate directly with, and likely underlie, historical patterns of shift. This study examines acoustic characteristics of vowels at initial edges of prosodic domains [Fougeron and Keating, J. Acoust. Soc. Am. **101**, 3728–3740 (1997)]. The investigation is restricted to three distinct prosodic levels: utterance (sentence-initial), phonological phrase (strong branch of a foot), and syllable (weak branch of a foot). The predicted changes in vowels /e/ and /ɛ/ in two American English dialects (from Ohio and Wisconsin) are examined along a set of acoustic parameters: duration, formant frequencies (including dynamic changes over time), and fundamental frequency (F0). In addition to traditional methodology which elicits list-like intonation, a design is adapted to examine prosodic patterns in more typical sentence intonations. [Work partially supported by NIDCD R03 DC005560-01.]

3aSC24. The effect of masking noise on acoustic-phonetic contrasts in post-lingually deafened cochlear implant users. Jennell C. Vick, Joseph S. Perkell, Ellen Stockmann, Majid Zandipour (Speech Commun. Group, Res. Lab. of Elec., MIT, Cambridge, MA 02139, jennell@speech.mit.edu), Harlan Lane (Northeastern Univ., Boston, MA), and Mark Tiede (Yale Univ., New Haven, CT)

This study examined the effect on the vowel contrast distance (average inter-vowel distance in the $F1-F2$ plane) of gradually decreasing the signal-to-noise ratio (SNR) in the auditory feedback of a post-lingually deafened cochlear implant (CI) user at 1-month and 1-year following CI processor activation. Masking noise, mixed with normal levels of speech feedback, was presented through the headpiece of a research sound processor to the CI user. As a control, an analogous procedure was used for a normal-hearing speaker where the masking noise and speech feedback were delivered over headphones. The SNR was gradually decreased over seven steps as the speakers produced ten repetitions of two vowel contrasts (æ\[g\] and i\u). Speech SPL and vowel contrast distance were measured at all seven masking noise levels. Data from both subjects showed that SPL gradually increased with decreased SNR, while contrast distance decreased. The effect was greater after 1 year of experience with a CI than at 1 month. The effect in the NH speaker was similar to that noted in the CI user after 1 year of experience. Data from additional subjects will be analyzed and reported. [Work supported by NIH Grant No. R01 DC03007.]

3aSC25. Cheating Heisenberg: Achieving certainty in wideband spectrography. Sean Fulop (Dept. of Linguist. and Dept. of Computer Sci., The Univ. of Chicago, 1010 E. 59th St., Chicago, IL 60637, sfulop@uchicago.edu)

The spectrographic analysis of sound has been with us some 58 years, and one of the key properties of the process is the trade-off in resolution between the time and frequency dimensions in the computed graph. While spectrography has greatly advanced the development of phonetics, the uncertainty principle has always been a source of frustration to phoneticians because so many of the interesting features of speech must be observed by computing Fourier spectra over very short time frames—i.e., using a ''wideband'' spectrogram. Since the uncertainty relation between time and frequency is unbreakable, the only option for improvement is to make a new kind of spectrogram that does not graph time and frequency. An algorithm is described and demonstrated which computes a new kind

of spectrogram in which Fourier transform frequency is replaced by the *channelized instantaneous frequency*, and time is adjusted by the *local group delay*. The theory behind this procedure was clarified in Nelson [J. Acoust. Soc. Am. **110**, 2575–2592 (2001)]. The resulting wideband spectrograms show dramatically improved resolution of speech features, which will be demonstrated with sample figures. It is thus suggested that phoneticians should be more interested in the instantaneous frequency spectrum than in the Fourier transform.

3aSC26. The impact of intended emotion on dynamic changes in talker amplitude. Kimberly Wieberg and Michael D. Hall (Univ. of Nevada–Las Vegas, 4505 Maryland Pkwy., Box 455030, Las Vegas, NV 89154-5030)

Relatively little research has been conducted on auditory affect, and less research has focused on quantifying the nature of potential acoustic cues to emotion. A production study was conducted to address the possibility that one dynamic acoustic variable, talker amplitude, may systematically vary as a function of intended affect. Six talkers (three male, three female) were instructed to produce 200 target words with different emotional valences (e.g., joy versus sadness). Target words were phonetically balanced (PB), and each word was produced within a standard carrier phrase. Temporally specific intensity measurements were recorded for each target word. Descriptively and statistically distinct contours were obtained as a function of intended emotion. Furthermore, talkers differed in the way they produced amplitude contours for each emotion. The perceptual relevance of the observed affect-based differences in amplitude then was evaluated. Average contours for joy and sadness were applied to an emotionally neutral set of (CV) stimuli. Listeners discriminated the resulting contours within a two-alternative forced choice task, and sensitivity was evaluated using signal detection analyses. Implications for the potential role of amplitude in the perception of emotion, as well as for the representation of talker information, in words will be discussed.

3aSC27. Control of closure/constriction duration in lingual consonants. Anders Lofqvist (Haskins Labs., 270 Crown St., New Haven, CT 06511, lofquist@haskins.yale.edu)

This study examines tongue movements in the production of lingual consonants where the duration of the oral closure/constriction is varied for linguistic purposes. Earlier work has shown that the tongue continues to move during the closure/constriction. The magnitude of the movement path during the closure/constriction is influenced by the vowel environment. Since the tongue has to stay in contact with the hard palate to maintain the closure/constriction, one might expect that the movement during the closure will be about the same for short and long consonants. To maintain the contact, a speaker would thus have to make a slower movement for the long consonants. Tongue and jaw movements were recorded in native Japanese speakers using a magnetometer system. Preliminary results for three speakers show that the closure/constriction duration for the long consonants was usually more than twice as long as that for the short consonants. The results also show a slightly longer movement path during the closure/constriction for the long consonants. As expected, the average speed of the tongue movement during the closure/constriction was systematically slower for the long consonants. In addition there was a positive correlation between closure/constriction duration and the path during the closure/constriction. [Work supported by NIH.]

3aSC28. Intonation contour in synchronous speech. Bei Wang and Fred Cummins (Dept. of Computer Sci., Univ. College Dublin, Belfield, Dublin 14, Ireland)

Synchronous Speech (Syn-S), obtained by having pairs of speakers read a prepared text together, has been shown to result in interesting properties in the temporal domain, especially in the reduction of inter-speaker variability in supersegmental timing [F. Cummins, ARLO **3**, 7–11 (2002)]. Here we investigate the effect of synchronization among speakers on the intonation contour, with a view to informing models of intonation. Six pairs of speakers (all females) read a short text (176 words) both synchronously and solo. Results show that (1) the pitch accent height above a declining baseline is reduced in Syn-S, compared with solo speech, while the pitch accent location is consistent across speakers in both conditions; (2) in contrast to previous findings on duration matching, there is an asymmetry between speakers, with one speaker exerting a stronger influence on the observed intonation contour than the other; (3) agreement on the boundaries of intonational phrases is greater in Syn-S and intonation contours are well matched from the first syllable of the phrase and throughout.

3aSC29. Exploration of question intonation in read American English. Ann K. Syrdal and Matthias Jilka (AT&T Labs—Res., 180 Park Ave., Florham Park, NJ 07932-0971)

Several generally accepted intonational features of questions in American English have not been the subject of much empirical study: namely that wh-questions end in L-L% phrasal accents, and that their intonational contours are identical to those of declarative sentences, while yes/no questions end in H-H% phrasal accents. The study addresses the following questions about question intonation: How frequently do yes/no questions end in H-H% phrasal tones, and how often do wh-questions end in L-L% phrasal tones? How similar are the intonational contours and features of declarative sentences and wh-questions with phrase-final falls (L-L%)? How do the sentence pitch ranges of yes/no questions, wh-questions, and declarative sentences compare? Does a speaker's characteristic pitch range affect the character or frequency of occurrence of question phrasal-tones? Speaker and utterance pitch ranges and their relation to prosodic features of pitch accents and phrasal tones were observed in yes/no and in wh-questions, and compared to a sample of simple declarative sentences spoken by the same speakers: 5 female and 3 male American English professional voice talents. The same set of 12 sentences were read by each of the 8 speakers in the same contexts. Theoretical and practical implications of the results will be discussed.

3aSC30. Comparison of human, canine, and ovine laryngeal dimensions. Moon J. Kim (Dept. of Otolaryngol., College of Medicine, Ewha Womans Univ., Seoul, Korea), Eric J. Hunter, and Ingo R. Titze (Natl. Ctr. for Voice and Speech, The Denver Ctr. for Performing Arts, Denver, CO)

Geometric measurement on the laryngeal skeleton is a fundamental step in laryngeal studies, especially in biomechanical modeling. Traditionally, canine larynges have been used as models of the human larynx because of their similarity in size and gross structure, but the search continues for an alternative model because of the diminishing availability of the canine species for research in the U.S. In this study, a revised methodology for defining and labeling laryngeal framework parameters was proposed. Ovine laryngeal cartilages were measured and compared to those of human and canine cartilages previously reported in the literature. The ovine cricoid, thyroid and arytenoids cartilages were significantly different from the human and canine. Also, the lack of a definite border between the true and false vocal folds revealed that the ovine model may not always be suitable for a direct comparison to the human larynx in phonation.

Session 3aSP

Signal Processing in Acoustics, Underwater Acoustics, Speech Communication, Animal Bioacoustics, Noise and Engineering Acoustics: Detection and Classification in Acoustics I

Susan Jarvis, Cochair
Engineering Test and Evaluation Department, NUWC Division Newport, Newport, Rhode Island 02841

David Moretti, Cochair
Engineering Test and Evaluation Department, NUWC Division Newport, Newport, Rhode Island 02841

Invited Papers

8:00

3aSP1. Detection in acoustics. Leon H. Sibul (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804-6531, lhs2@psu.edu)

An overview of classical detectors and their applications to acoustics is presented. Neyman-Pearson and Bayes detectors are discussed. Log likelihood detectors are derived from the Bayes criterion and from information theoretic concepts. It is shown that matched filters, range-Doppler and wavelet domain processors, and estimator-correlators are particular implementations of the log likelihood detector. These processors have wide applications to active and passive sonar. Sequential Bayesian detectors and joint detection and estimation are discussed. Generalization of sequential Bayesian concepts has applications to information fusion from multiple sensors and net-centric processing. Basic properties of the receiver operating characteristics (ROC) are reviewed and applied to different detectors. An important current research topic of detection in uncertain/random media is discussed. Detectors in uncertain/random media can be characterized by mean and variance of the random receiver operating characteristic.

8:45

3aSP2. Analytic conditions for asymptotic optimality of the maximum-likelihood estimator with application to problems in remote sensing. Nicholas C. Makris (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

Analytic conditions necessary for a general multivariate maximum-likelihood estimate (MLE) to become asymptotically optimal are derived from the first principles of estimation theory following Makris and Naftali [J. Acoust. Soc. Am. **110**, 1917–1930 (2001)]. The conditions are given as sample size or signal-to-noise ratio requirements for the MLE to become asymptotically unbiased and attain the minimum variance possible for any unbiased estimate known as the Cramer–Rao lower bound (CRLB). Applications to problems common in remote sensing such as time-delay and Doppler shift estimation, beamforming, source localization in a waveguide, pattern recognition in 2-D intensity images, surface orientation estimation, and Kalman filtering will be presented.

Contributed Papers

9:15

3aSP3. Validity conditions for the linear statistical assumptions of Kalman filtering. Michele Zanolin and Nicholas C Makris (MIT, 77 Masssachusetts Ave., Cambridge, MA 02139)

The fundamental assumption of a Kalman filter is that both the state and evolution equations are linear and that the noise in each of these equations is signal-independent, additive and Gaussian. For many practical problems in Kalman filtering and recursive estimation, however, both the state and evolution equations are actually nonlinear and may involve random variables that are neither additive nor Gaussian. Ad-hoc linearizations with little or no justification are often made to apply a Kalman filter in these cases. Validity conditions for the fundamental linear statistical assumptions of Kalman filtering are derived from the first principles of estimation theory by asymptotic analysis of the likelihood function of the general nonlinear model [E. Naftali and N. Makris, J. Acoust. Soc. Am. **110** (2001)]. An illustrative example involving remote acoustic tracking of a fluctuating target with active sonar is presented.

9:30

3aSP4. Image processing techniques for underwater acoustic image enhancement. Frank Boyle (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713)

Underwater acoustic images are often dominated by noise, which obscures signals of interest. When the spatial character of the noise differs from that of the signal, image processing techniques can be employed to enhance the image. In underwater acoustic time-bearing displays, the signals of interest form continuous tracks, while noise often takes on the form of isolated speckle. In this case, a speckle suppression technique related to

the Hough transform can be used to reduce the noise and enhance the tracks. The presentation will include a discussion of variations of the technique that are under consideration, as well as considerations of optimum values for the input parameters.

9:45

3aSP5. Reducing ambient noise in time series data with the Bartlett spectrogram. Paul Elmore and Jevon Harding (Naval Res. Lab., Marine Geosciences Div., Stennis Space Center, MS 39529)

A signal processing technique for enhancing the signal-to noise ratio of transient acoustic signals is analyzed. This method subtracts the average power spectral density as calculated from the Bartlett spectrogram from the data in the frequency domain, and then forms a new time series from the resultant spectrogram and original phase information. We examine signal-to-noise ratio improvements that occur. Noisy time-series containing transient signals from synthetic data and real marine mammal recordings are used as test data sets. Results show that improvements of around 10 dB can be achieved theoretically. Limitations and pre-application considerations also are discussed.

10:00–10:15 Break

10:15

3aSP6. Multiple time-reversed guide-sources in shallow water. Charles F. Gaumond, David M. Fromm, Joseph F. Lingevitch, Roger C. Gauss, and Richard Menis (Acoust. Div. Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375-5320)

Detection in a monostatic, broadband, active sonar system in shallow water is degraded by propagation-induced spreading. The detection improvement from multiple spatially separated guide sources (GSs) is presented as a method to mitigate this degradation. The improvement of detection by using information in a set of one-way transmissions from a variety of positions is shown using sea data. The experimental area is south of the Hudson Canyon off the coast of New Jersey. The data were taken using five elements of a time-reversing VLA. The five elements were contiguous and at midwater depth. The target and guide source was an echo repeater positioned at various ranges and at middepth. The transmitted signals were 3.0- to 3.5-kHz LFMs. The data are analyzed to show the amount of information present in the collection, a baseline probability of detection (PD) not using the collection of GS signals, the improvement in PD from the use of various sets of GS signals. The dependence of the improvement as a function of range is also shown. [The authors acknowledge support from Dr. Jeffrey Simmen, ONR321OS, and the chief scientist Dr. Charles Holland. Work supported by ONR.]

10:30

3aSP7. Detection in shallow water using broadband—DORT. David M. Fromm, Charles F. Gaumond, Joseph F. Lingevitch, Roger C. Gauss, and Richard Menis (Acoust. Div., Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375-5320)

The decomposition of the time-reversal operator (DORT) [Prada *et al.*, J. Acoust. Soc. Am. **99**, 2067–2076 (1996)] has been extended into a coherent, broadband method. Broadband DORT has also been shown to isolate resolvable scatterers at various depths and ranges in a bistatic, active sonar in shallow water. Results are shown from the application of DORT to sea data taken in an area south of Hudson Canyon off the New Jersey coast during Geoclutter II. The vertical source/receiver array with 56 hydrophones spanning the water column was operated between 3.0 and 3.5 kHz. The elements were divided into four groups, with each group acting as a coherent, broadside source. Two methods were used for exciting the separate channels. One method was the use of subsequent LFMs and the other was the use of simultaneous transmission of four pseudorandom-noise signals. The target was a midwater column echo-repeater. Results are compared with modeling based on *in situ* environmental measurements during the experiment. [The authors acknowledge signal-processing expertise from Dr. Ning Xiang, University of Mississippi, and ENS Alan Meyer, LLNL, support from Dr. Jeff Simmen, ONR, and assistance from Dr. Charles Holland, ARL/PSU. Work supported by ONR.]

10:45

3aSP8. Kalman filtering with neural networks for change detection in simulated sidescan sonar data. Pamela J. McDowell, Marlin L. Gendron, Patrick M. McDowell (Naval Res. Lab., Stennis Space Center, MS 39529), Juliette W. Ioup, and George E. Ioup (Univ. of New Orleans, New Orleans, LA 70148)

Sidescan sonars produce acoustical imagery which is used to detect bottom objects and characterize features of the seafloor. Change detection is a method that can be used to flag new bottom objects which were not detected during previous sidescan sweeps. Improvements in object detection and classification are critically needed to improve change detection methods. Adaptive filtering techniques may be used to identify objects known or previously marked from historical data, and flag new objects detected as changed. In this study Kalman Filter techniques will be used to estimate weights for a supervised feed-forward perceptron neural net classifer. Both noise-free and noisy data are considered. Preliminary results from these techniques using simulations that model sidescan sonar data sets will be presented.

11:00

3aSP9. Study of environmental sound source identification based on hidden Markov model for robust speech recognition. Takanobu Nishiura (Wakayama Univ., 930 Sakaedani Wakayama, 640-8510 Wakayama, Japan) and Satoshi Nakamura (ATR Spoken Lang. Translation Res. Labs., Kyoto, Japan)

Humans communicate with each other through speech by focusing on the target speech among environmental sounds in real acoustic environments. We can easily identify the target sound from other environmental sounds. For hands-free speech recognition, the identification of the target speech from environmental sounds is imperative. This mechanism may also be important for a self-moving robot to sense the acoustic environments and communicate with humans. Therefore, this paper first proposes hidden Markov model (HMM)-based environmental sound source identification. Environmental sounds are modeled by three states of HMMs and evaluated using 92 kinds of environmental sounds. The identification accuracy was 95.4%. This paper also proposes a new HMM composition method that composes speech HMMs and an HMM of categorized environmental sounds for robust environmental sound-added speech recognition. As a result of the evaluation experiments, we confirmed that the proposed HMM composition outperforms the conventional HMM composition with speech HMMs and a noise (environmental sound) HMM trained using noise periods prior to the target speech in a captured signal. [Work supported by Ministry of Public Management, Home Affairs, Posts and Telecommunications of Japan.]

11:15

3aSP10. The application of *m*-sequences to bi-static active sonar. Harry A. DeFerrari (RSMAS—Univ. of Miami, 4900 Rickenbacker Cswy., Miami, FL 33149)

The *m*-sequences are ideal pulse compression signals that combine the energy of CW with the resolution of a pulse. Successful applications include numerous acoustic propagation experiments and the Global Positioning System. Yet, early attempts (circa 1960) to apply *m*-sequences to mono-static active sonar were unsuccessful. Through the years, Birdsall, Metzger and others have developed a body of theory, numerical methods and at-sea demonstrations that establish the feasibility of a novel bi-static approach—one that holds promise in high reverberation shallow water environs. An analysis is presented here. The approach includes (1) continuous transmission of long *m*-sequences; (2) synchronous sampling to form a CON (Complete Ortho-Normal) data set; (3) direct blast removal by HCCO (Hyperspace Cancellation by Coordinate Zeroing); and (4) a

full range waveform Doppler search. Ultra-fast Hadamard Transforms speed up the direct waveform pulse m-sequence pulse compression and the inverse pulse waveform transform and thereby allow timely execution of the intensive computational burden. The result is a demonstrable approach that produces a gain of 30 dB over a simple pulse and 10 dB over other sonar signals. In the end, the approach requires continuous transmission and reception as opposed to ping and listen an awkward concept at first.

11:30

3aSP11. Wavelets and neural networks for change detection in simulated sidescan sonar data. Juliette W. Ioup (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148), Marlin L. Gendron, Pamela J. McDowell, Brian S. Bourgeois (Naval Res. Lab., Stennis Space Center, MS 39529), and George E. Ioup (Univ. of New Orleans, New Orleans, LA 70148)

Recently recorded sidescan sonar data can be compared to historical data from the same area to detect any differences or changes. Such comparisons are complicated by the fact that the position is uncertain and the viewing angle or depth of the sidescan sonar may be different in the two data sets. Neural networks offer a possible technique to be used in the comparison procedure. Inputs to neural networks investigated here include wavelet transform coefficients of the data as well as the two-dimensional sidescan sonar data itself. Several different neural network architectures and several different wavelet bases are investigated. Both noise-free and noisy data are considered. Preliminary results for change detection using simulations that model simple historical and recent sidescan sonar data sets will be presented. [Research supported in part by NRL-ASEE Summer Faculty Research Program.]

11:45

3aSP12. Higher-order harmonic signature analysis for loudspeaker defect detection. Shane Thompson, Anthony Pagliaro, Robert Celmer (Acoust. Prog. & Lab., College of Eng., Univ. of Hartford, 200 Bloomfield Ave., West Hartford, CT 06117), Daniel Foley, and Steve Temme (Listen, Inc., Boston, MA 02118)

Loudspeaker assembly faults, such as a rubbing voice coil, bent frame, loose spider, etc., have traditionally been detected using experienced human listeners at the end of a production line. Previous attempts to develop production measurement systems for on-line testing typically analyze only low-order harmonics for the primary purpose of measuring total harmonic distortion (THD), and thus are not specifically designed to detect defective rub, buzz, and ticking sounds. This paper describes a new method wherein the total energy of high-order harmonics groups, for example, 10th through the 20th or 31st through the 40th, are measured and analyzed. By grouping high-order harmonics and resolving their respective total energies, distinct signatures can be obtained that correlate to the root cause of audible rub and buzz distortions [Temme (2000)]. The paper discusses loudspeakers tested with specific defects, as well as results of a computer-based electroacoustic measurement and analysis system used for detection.

WEDNESDAY MORNING, 12 NOVEMBER 2003 SABINE ROOM, 8:30 TO 11:30 A.M.

Session 3aUW

Underwater Acoustics: Localization and Inversion

David P. Knobles, Chair
Applied Research Laboratories, University of Texas–Austin, P.O. Box 8029, Austin, Texas 78713-8029

Contributed Papers

8:30

3aUW1. Matched field processing in a multisource environment. Eric A. Rittermann, Robert A. Koch, Steven A. Stotts, and D. P. Knobles (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

Matched field processing can be utilized to localize a single source in a quiet, well-known medium. In shallow water, the combination of environmental mismatch and low signal to noise ratio causes significantly reduced resolution. Discussed is a simultaneous matched field processing and geoacoustic inversion methodology using a simulated annealing approach applied to a quiet source in the presence of loud interferers. This is accomplished by preprocessing array element acoustic data with a spatial matrix filter, which acts to reduce the effective number of sources present in the received data. Specifically, the effects of the filter on matched field processing and geoacoustic inversion accuracy as a function of interferer strength is investigated. Acoustic data collected from the Gulf of Mexico is employed to validate the proposed simultaneous localization and inversion approach. [Work supported by ONR.]

8:45

3aUW2. Spatial filtering techniques applied to source localization. Steven A. Stotts, Robert A. Koch, and David P. Knobles (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

Matched-field processing outputs may be degraded by the presence of multiple sources. To isolate single sources, a spatial filtering algorithm has been developed. Element level data are transformed to beam space, filtered in bearing, then transformed back to element level data which are then used in simulated annealing inversions. The efficacy of the approach was demonstrated by comparing filtered and unfiltered ambiguity surfaces of depth, range, and bearing using a cross-correlation cost function in a simulated multisource environment with equally spaced array elements [Stotts *et al.*, J. Acoust. Soc. Am. **113**, 2217 (2003)]. Recently, this method was successfully applied to a real data set where localizations had previously been obtained using beam processing, and ground truth bottom properties have been established. Ambiguity surfaces are exhibited to demonstrate filter performance for real data. Examining the spectrum of singular values associated with the filter shows differences between equally and nonuniformly spaced elements, where a modification is required for frequencies greater than ideal. Energy conservation requirements validate filter output. Filter design details and results for both simulated and real data will be presented. Effectiveness as a function of signal-to-noise ratio, measured relative to an interferer, is also examined.

9:00

3aUW3. Comparisons of multi-source localization and spatial filtering techniques. Adam J. Cook, Steve A. Stotts, and Robert A. Koch (P.O. Box 8029, Austin, TX 78713-8029, acook@arlut.utexas.edu)

Localization of moving sources in a multiple source environment has been investigated with simulated annealing algorithms. Results obtained, both by classical annealing and by fast simulated annealing using a rotated coordinate basis, are presented. The fast annealing method varies from the classical method in that each parameter is varied only once per temperature reduction, resulting in many fewer iteration calls per inversion. Ac-

curacy and efficiency for these methods as a function of signal-to-noise ratio are examined using simulated data containing a weak source in the presence of a loud interferer. A method for filtering element level data prior to inversion is compared to subarray processing for effectiveness in localizing the source of interest. The effect of varying the level of the weak source relative to the level of the interferer, as well as the sources azimuthal separation, is also presented. Comparisons between an equally spaced horizontal line array (HLA) and an HLA with nonuniform sensor spacing are made. Finally, comparisons of the inversion techniques for a real data set from a nonuniformly spaced HLA at a Gulf of Mexico site, where ground truth bottom properties have been established, are discussed. [Work sponsored by ONR.]

9:15

3aUW4. Range-depth localization of underwater explosions by matched-field processing of data from a sparse bottom-laid horizontal array. Dag Tollefsen (Norwegian Defence Res. Establishment (FFI) Boks 115, Horten N3191, Norway)

Range-depth localization of underwater explosions by matched-field processing of data from a sparse bottom laid horizontal array in shallow water (the Barents Sea) has been investigated. Data from shallow (18 m) and deep (91 m) SUS charges detonated at ranges of 1–60 km endfire to a 10-element array were considered. A range-independent environment model consisting of a measured sound speed profile in water, an equivalent flat bathymetry and a two-layer fluid seabed model with geoacoustic parameters estimated by matched-field inversion of data from known sources was used. Replica fields were computed using the fast normal mode model C-SNAP. The broadband incoherent Bartlett processor was applied with data from one or two 6 Hz wide frequency bands centered within 38–94 Hz. Estimates were within ten percent from nominal in range and 15 m in depth for all shots but a few. Depth estimates for shallow sources degraded beyond a transition range; these were thereafter incorrectly estimated to be at the sea surface. Results from processing including data recorded concurrently on a vertical array (by incoherently combining data from two sections of an L-shaped array) will also be presented.

9:30

3aUW5. A bathymetry re-mapping technique for computing replica fields on large towed arrays. Martin Siderius, Christopher O. Tiemann, and Michael B. Porter (Sci. Applications Intl. Corp., 10260 Campus Point Dr., San Diego, CA 92121, tiemannc@saic.com)

Matched-field processing is now a well-known technique for scanning a focal spot through the ocean waveguide to localize a source. The focal spot is actually formed by intersections of surface and bottom reflected beams. It is also well known that with vertical receiver arrays bathymetry errors can cause a mirage effect in which the various reflections focus at a biased position. Typically, this bias is not a serious problem and is simply ignored. A more severe effect occurs when arrays with large horizontal aperture are used in the processing. For instance, with a large horizontal array, phones near the source see a different bathymetry than those distant from the source. In that case, using an incorrect bathymetry causes both a bias and a loss of coherence. The latter can be a very serious problem. We show that this loss of coherence can be corrected with a simple re-mapping of the phone coordinates. Through the re-mapping process we can use a flat-bottom model to rapidly predict the field on a regular grid, and then transform that replica field properly to reflect range-dependent bathymetry slices. The result is a vastly more efficient process for calculating replica fields on large, multidimensional arrays in range-dependent environments.

9:45

3aUW6. Focalization in the presence of internal waves. Michael D. Collins and Ralph N. Baer (Naval Res. Lab., Washington, DC 20375)

Focalization is an approach for localizing an acoustic source in a medium with environmental uncertainties [J. Acoust. Soc. Am. **90**, 1410–1422 (1991)]. With this approach, key environmental parameters are adjusted in an attempt to bring the source into focus and determine its location. This approach has proven to be effective for problems involving deterministic environmental parameters. Due to the nature of the parameter space, focalization should also be effective for problems involving uncertain stochastic parameters. Due to a parameter hierarchy in which source parameters are more important than environmental parameters, it is often possible to localize the source without determining the correct environmental parameters. This paper will investigate focalization in the presence of internal waves, which are often sufficient to cause conventional localization techniques to fail. [Work supported by ONR.]

10:00–10:15 Break

10:15

3aUW7. Localization and environmental inversion in a soft sediment environment using rotated coordinates. Rebecca Newman and Robert Koch (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

SARI is used to simultaneously invert data for environmental and source track parameters. The acoustic data were recorded on a bottom mounted, horizontal line array at a soft sediment, shallow water site in the Gulf of Mexico off Port Aransas, Texas. SARI [T. B. Neilsen, J. Acoust. Soc. Am. **113**, 2217 (2003)] is a simulated annealing program that uses rotated coordinates, which represent the coupling and optimal variation of the annealing parameters, to invert for source track and environmental parameters. While previously tested on simulated hard sediment data, soft sediments give rise to significant coupling between source and geoacoustic parameters. A comparison of the inversion results obtained for the Gulf of Mexico annealing with rotated coordinates and without will be presented. [Work supported by ONR.]

10:30

3aUW8. Inversion for source levels and environmental and source track parameters. Robert A. Koch and D. P. Knobles (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713-8025, koch@arlut.utexas.edu)

A methodology for simultaneous inversion for source levels, source track parameters, and environmental parameters is described. An application to acoustic data recorded on a bottom-mounted, horizontal line array at a soft-sediment, shallow-water site in the Gulf of Mexico off Port Aransas, TX is discussed. This inversion approach may be viewed as a technique for making TL measurements with sources of opportunity by a generalization of the inversion of measured TL as a function of range for geoacoustic parameters for which the source depth is known. The technique involves minimizing a cost function formed from a sum over hydrophones and frequencies of the squared difference between received levels and modeled intensities. The inversion methodology will be validated using known cw tones emitted by a towed source and demonstrated for sources of opportunity using acoustic emissions from the RV LONGHORN, the experimental platform for the experiment. [Work supported by ONR.]

10:45

3aUW9. The stability of the time-reversal process in a fluctuating ocean. G. F. Edelmann, S. Kim, W. S. Hodgkiss, W. A. Kuperman, and H. C. Song (Marine Physical Lab., Scripps Inst. of Oceanogr., 9500 Gilman Dr., San Diego, CA 92093-0238, geoff@mpl.ucsd.edu)

Changes in the thermocline perturb the acoustic horizontal wave numbers and to a less extent the mode shapes. The measured acoustic field is the combination of modes and phase terms dependent on these wave numbers. Without a full water column spanning time-reversal mirror, intermodal destructive and constructive interference patterns produce seemingly unpredictable measured fields. As demonstrated in simulation, the field produced by a time-reversal mirror is robust to internal wave perturbation. [Work supported by ONR.]

11:00

3aUW10. Effects of time-reversing array deformation in an ocean waveguide. Karim G. Sabra (Scripps Inst. of Oceanogr., Univ. of California–San Diego, La Jolla, CA 92093) and David R. Dowling (Univ. of Michigan, Ann Arbor, MI 48109)

Active acoustic time reversal is a technique for transmitting and focusing sounds recorded in complex unknown environments back to their remote point(s) of origin. It can be accomplished with a transducer array—a time-reversing array (TRA) or time-reversal mirror (TRM) that can both send and receive sound. Nearly all prior work on underwater TRA performance has involved stationary arrays. This presentation describes how random array deformation influences TRA retrofocusing in a shallow ocean environment. Both theory and simulation results are presented for frequencies of 250, 500, and 750 Hz. For harmonic signals, randomly drifting array elements degrade TRA performance by approximately 20% when the average horizontal wave number times the root-mean-square horizontal element displacement approaches one half. For randomly rising or sinking elements, TRA performance is similarly degraded when the average vertical wave number times the root-mean-square vertical displacement reaches four or five. [Work supported by ONR.]

11:15

3aUW11. A near field technique for acoustic array element localization. Frank Boyle (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713-8029)

A beamformer's effectiveness is sensitively dependent on accurate knowledge of the array element positions. Current techniques for using acoustic data to localize array elements are generally quite effective with distant acoustic sources, but are of questionable accuracy when an array must be localized with sources located in the extreme near field. A geometric technique for localizing arrays was developed and appears effective for near field sources. The method was tested with impulsive acoustic data collected from a line array deployed in shallow water, ensonified with light bulb implosions from various positions very close to the array. The presentation will include a discussion of the technique and the experimental results.

WEDNESDAY AFTERNOON, 12 NOVEMBER 2003 BRAZOS AUDITORIUM, 1:00 TO 3:10 P.M.

Session 3pAA

Architectural Acoustics, Noise and Speech Communication: Forensic Acoustics II

Jack E. Randorff, Cochair
Randorff and Associates, Inc., 11 West Canyon View Drive, Ransom Canyon, Texas 79366

Peter Ladefoged, Cochair
Linguistics Department, University of California–Los Angeles, Los Angeles, California 90095-1543

Chair's Introduction—1:00

Invited Papers

1:05

3pAA1. Forensic acoustics: An opportunity to educate. Bennett Brooks (Brooks Acoust. Corp., 27 Hartford Turnpike, Vernon, CT 06066, bbrooks@brooksacoustics.com)

Forensic science is narrowly defined by some as the gathering of evidence to be used in a criminal court proceeding. However, the word ''forensic'' broadly pertains to arguments made in any public forum. Those acousticians engaged in expert witness service may work in a variety of settings that address the interests of the general public. These can include quasilegal local administrative public hearings, conducted at school board or planning and zoning meetings, civil legal actions, and rarely, a criminal trial. When presenting complex scientific arguments in a public forum, the reception with which that information is met can strongly depend upon the self-interest of meeting participants, as well as the common skepticism toward all things technical. To successfully gain favor for a particular viewpoint, the target audience (board commissioners, judges, juries) must be sufficiently educated to understand the methods of acquiring valid acoustical data, and the impact of acoustics to the situation in question. The challenge is to present credible information with just enough detail to persuade, but not overwhelm, the decision maker. Illustrative case studies will be discussed.

1:30

3pAA2. The interface between the phonetic scientist and forensic investigations. Harry Hollien (Inst. for Adv. Study of the Commun. Proc., Univ. of Florida, Gainesville, FL 32611)

Many scientists find the interface with criminal/civil investigations and the resulting litigation quite challenging. The great variety found among such cases and their (sometimes) shockingly grim aspects, can be most daunting. Moreover, the aid these scientists can expect when attempting to learn their responsibilities (while good) tends to focus only on the roles they will play in the courts. The reality is that they also must serve in a second domain that of investigator/consultant to law enforcement personnel and attorneys. Since training, structure and guidelines are lacking in this area, the relevant problems must be addressed (and solved) directly by the scientist. To do so, he/she must (1) learn about the parallels between laboratory investigations and forensic investigations and organize links, (2) compensate for the differences between them, (3) organize scientific personnel for collaborations in support of the needs of the practitioners, (4) adapt laboratory procedures for forensic application, (5) establish criteria for these applications and (6) develop and articulate what can and cannot be expected of these procedures. Brief case reviews will be presented to illustrate each of these issues.

1:55

3pAA3. Validity of voice identification. Peter Ladefoged (Phonet. Lab, Linguist., UCLA, Los Angeles, CA 90095-1543, oldfogey@ucla.edu)

Courts often need expert opinion as to whether two recordings were made by the same speaker or not. Typically one voice has been recorded by an unknown speaker on a telephone answering machine and the other somehow obtained from a suspect, When such an opinion has been given, the court is bound to ask how reliable it is. There is no answer to this question. There are too many unknown factors. Were the recording conditions similar? Were both the speakers talking in the same style? Was one speaker trying for a disguise? Was there any background noise? How long was the recording? Does the suspect have family or friends with very similar voices? It would be almost impossible to set up a valid experiment that can take all the variables into account. Given two 15 minute recordings of similar material made under studio conditions an expert could state beyond a reasonable doubt whether they were made by the same speaker or not. Given a 5 second bomb threat on a noisy phone line there is no way that this could be done.

2:20

3pAA4. Could the gunshot be heard? Paul Schomer (Schomer and Assoc., Inc., 2117 Robert Dr., Champaign, IL 61821)

Witnesses, friends and family of a man chased through a field by the police, said they heard gunshots. The man chased says that the police shot at him. The police say that no shots were fired. Could the witnesses have heard the gunshots? The main technical issues are source strength and directivity, weather effects on sound propagation, attenuation by forests, sound transmissions through walls, and signal detectability. This paper lays out the material much as the jury heard it.

2:45

3pAA5. Audibility of locomotive horns inside highway–railroad grade crossings. Jack E. Randorff (Randorff and Assoc., 11 W. Canyon View Dr., Ransom Canyon, TX 79366)

Most states require that motorists yield the right-of-way to locomotives and their consists at highway–railroad grade crossings. One method of alerting motorists to the proximity of a train is to sound an audible warning using an air horn. In litigation, the effectiveness of a locomotive's air horn to alert the driver of a vehicle can be a crucial element that allows a jury to determine liability and award financial penalties. Locomotive air horn signal strength, acoustic propagation over distance, excess attenuation, sound isolation of vehicle shells, ambient sound level masking, signal detection ratios, and driver distraction must be considered, measured, and evaluated before the audibility of the air horn can be surmised from the driver's perspective. The final evaluation must then be presented to a jury through the process of direct and cross examination. Several case examples are discussed relating to this process.

Session 3pED

Education in Acoustics: Take "Fives"

Uwe J. Hansen, Cochair
Department of Physics, Indiana State University, Terre Haute, Indiana 47809

Thomas D. Rossing, Cochair
Physics Department, University of Illinois, DeKalb, Illinois 60115

1:00

Do you have a novel demonstration, a new laboratory experiment, a favorite video, a recorded sound example, or a new idea for teaching acoustics? Share it with your colleagues. A sign-up board will be provided for scheduling presentations. No abstracts are printed. Presenters are encouraged to have handouts for distribution. Multiple presentations are acceptable (not consecutively). Presentations are limited to 5 minutes. Keep them short! Keep them fun!

Session 3pID

Interdisciplinary: Hot Topics in Acoustics

Yves H. Berthelot, Chair
School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

Chair's Introduction—2:00

Invited Papers

2:05

3pID1. Gerrymander: An interactive musical composition for clarinet and computer. Russell Pinkston (School of Music, Univ. of Texas, Austin, TX 78712)

While compositions involving traditional musical instruments and pre-recorded electronic sounds have been around since the 1960's, it is only recently that the technology has advanced sufficiently to allow such sounds to be generated in real-time and under the direct control of a the performing musician. Gerrymander is a composition for solo clarinet and computer in which the computer processes and manipulates the sound of a clarinet in real-time to generate a dynamic musical accompaniment to the live performer. The performer is also able to interact with the computer to control various aspects of the accompaniment, including the timing of passages played by the computer and the overall pacing of the composition. The computer accompaniment is controlled primarily by tracking the fundamental pitch of the live clarinet and using it to follow a pre-recorded musical score. Granular synthesizers, pitch shifting delays, and contrapuntal harmonization algorithms, as well as a polyphonic sampler and all the necessary mixing and control logic were implemented in MAX/MSP on a Macintosh G4 powerbook. The paper will consist of a demo performance of the composition using a pre-recorded clarinet part, followed by a detailed description of the software using a video projector.

2:25

3pID2. Hot topics in noise. Michael R. Stinson (Inst. for Microstructural Sci., Natl. Res. Council, Ottawa, ON K1A 0R6, Canada)

Our world continues to be a noisy place and the challenge to ''increase and diffuse knowledge of noise propagation, passive and active noise control, and the effects of noise'' remains. In the last several years, noise in the classroom has emerged as one of the hotter topics: Considerable progress has been made in the underpinning research, the formulation of recommendations, and the process of educating society on the social and personal impact of inadequate acoustical conditions in classrooms. The establishment of the ANSI S12.60-2002 standard for classroom acoustics was a milestone event. Noise in cities and the understanding of our soundscapes are

subjects of ongoing significance. The development of standards and regulations is a continuing process, with urban community noise regulations, aviation noise, and the preservation of natural quiet in national parks being of current concern. New methods to reduce noise are under development and include passive and active methods of noise control, techniques for modeling the performance of noise barriers, and approaches for designing product sound quality.

2:45

3pID3. Hot topics in acoustical oceanography. Darrell R. Jackson (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

Both active and passive acoustic remote sensing methods are used to measure ocean properties and processes. Observations are made on scales ranging from meters to megameters. Several techniques have recently matured from research to application. Beginning with active acoustics on megameter scales, there has been recent progress in basin-wide thermometry, and a global acoustic network has been suggested. On decimeter scales, active acoustics have been used to study hydrothermal flows occurring at oceanic ridges. The ultimate goal of the hydrothermal work is to measure fluxes of heat and particulates. Passive measurement of ambient sound in the ocean has been used successfully to observe a wide variety of processes at all scales. For passive acoustics, this talk will focus on small-scale determinations of rainfall rate and type, as well as studies of wave breaking.

WEDNESDAY AFTERNOON, 12 NOVEMBER 2003 SAN MARCOS ROOM, 1:00 TO 3:00 P.M.

Session 3pSP

Signal Processing in Acoustics, Underwater Acoustics, Speech Communication, Animal Bioacoustics, Noise and Engineering Acoustics: Detection and Classification in Acoustics II

Juliette W. Ioup, Cochair
Physics Department, University of New Orleans, New Orleans, Louisiana 70148

George E. Ioup, Cochair
Physics Department, University of New Orleans, New Orleans, Louisiana 70148

Invited Papers

1:00

3pSP1. A review of transient detection using higher order correlations. Lisa A. Pflug (Naval Res. Lab., Stennis Space Center, MS 39529, pflug@nrlssc.navy.mil), George E. Ioup, and Juliette W. Ioup (Univ. of New Orleans, New Orleans, LA 70148)

Quieter targets and noisier ocean environments prompted a search for detection improvements in the 1990s. Transient signals were a potential source for detection that had not been exploited. Higher order statistics provided a means to capitalize on one of the unique characteristics of transient signals—nonzero higher order moments. Several years of study led to the conclusion that higher order correlation detectors can in fact perform better than the cross-correlation detector for transient signals. This conclusion is supported by both theory and simulations. Theoretical predictions for unknown and known signals in uncorrelated noise have shown that the gains depend on signal moments, sampling rate, and for the unknown signals, data length as well. Derived upper bounds on the gains in uncorrelated noise show the limitations of these processors. Various mechanisms for further improvement were investigated. One-dimensional prefiltering improved the higher-order correlation detectors due to noise reduction in the higher-dimensional correlation space. Functional inputs of certain narrow-band signals resulted in gains for the higher-order correlation detectors which were not realized with the original signal. Finally, studies indicate that the higher order correlation detectors maintain their advantage over the cross-correlation detector in measured, correlated noise. [Research supported by ONR and NRL.]

1:30

3pSP2. Passive monitoring and localization of marine mammals in open ocean environments using widely spaced bottom mounted hydrophones. Susan Jarvis, David Moretti, Ronald Morrissey, and Nancy DiMarzio (Eng. Test and Evaluation Dept., NUWC Div. Newport, Newport, RI 02841)

The Marine Mammal Monitoring on Navy Ranges (M3R) project has developed a toolset for passive detection and localization of marine mammals using the existing infrastructure of Navy's undersea ranges. The Office of Naval Research funded the M3R project as part of the Navy's effort to determine the effects of acoustic and other emissions on marine mammals and threatened/endangered species. A necessary first step in this effort is the creation of a baseline of behavior, which requires long-term monitoring of marine

mammals. Such monitoring, in turn, requires the ability to detect and localize the animals. This paper will present the passive acoustic monitoring and localization tools developed under M3R. It will also present results of the deployment of the M3R tools at the Atlantic Undersea Test and Evaluation Center (AUTEC), Andros Island, Bahamas from June through November 2003. Finally, it will discuss current work to improve automated species classification.

Contributed Papers

2:00

3pSP3. Prediction of performance for ordinary and higher order correlation detection of prefiltered transients. Marcella E. Dean, George E. Ioup, Juliette W. Ioup (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, mdean@uno.edu), and Lisa A. Pflug (Naval Res. Lab., Stennis Space Center, MS 39529)

It has been shown by simulations that bandpass filtering can give significant detection gains [Pflug *et al.*, J. Acoust. Soc. Am. **95**, 1459–1473 (1994) and M. E. Dean, Ph.D. dissertation, 2003]. Prediction formulas have been derived for known and unknown transient source detection at the minimum detectable level by Pflug *et al.* [J. Acoust. Soc. Am. **98**, 248–260 (1995); **103**, 2469–2477 (1998)]. These results have been extended by the derivation of formulas which predict performance for prefiltered signals. The main advantage of prefiltering is noise cancellation, which increases with the order of the correlation and which is represented in the prediction formulas as increasing powers of the ratio of the bandpass filter width to the foldover frequency. Signal moments are also modified by prefiltering. The prediction formulas are tested by comparison with previously given simulation results, in which signal-to-noise ratio values at the minimum detectable level are found using hypothesis testing and Monte Carlo computer simulations. [Research supported by ONR.]

2:15

3pSP4. A comparison of model-based and hyperbolic localization techniques as applied to marine mammal calls. Christopher O. Tiemann and Michael B. Porter (Sci. Applications Intl. Corp., 10260 Campus Point Dr., San Diego, CA 92121, tiemannc@saic.com)

A common technique for the passive acoustic localization of singing marine mammals is that of hyperbolic fixing. This technique assumes straight-line, constant wave speed acoustic propagation to associate travel time with range, but in some geometries, these assumptions can lead to localization errors. A new localization algorithm based on acoustic propagation models can account for waveguide and multipath effects, and it has successfully been tested against real acoustic data from three different environments (Hawaii, California, and Bahamas) and three different species (humpback, blue, and sperm whales). Accuracy of the model-based approach has been difficult to verify given the absence of concurrent visual and acoustic observations of the same animal. However, the model-based algorithm was recently exercised against a controlled source of known position broadcasting recorded whale sounds, and location estimates were then compared to hyperbolic techniques and true source position. In geometries where direct acoustic paths exist, both model-based and hyperbolic techniques perform equally well. However, in geometries where bathymetric and refractive effects are important, such as at long range, the model-based approach shows improved accuracy.

2:30

3pSP5. Travel time estimation for ocean acoustic thermometry. Matthew Dzieciuch (Scripps Inst. of Oceanogr., Univ. of California, San Diego, CA)

The basic datum of ocean acoustic thermometry experiments is the ray travel time, thus an accurate and precise estimator is of paramount importance. It is well known that a matched filter gives the best estimate if the transmitted signal is known exactly. At basin-scale ranges, the signal becomes a random process due to the limited coherence time and coherent bandwidth of the channel. A matched filter is no longer optimum in this case, but since the reception is still partially coherent, an energy detector is not optimum either. A better estimator is the RAKE correlator, which can account for the signal variability at the cost of increased receiver complexity and a reduction of resolution. In this talk, the performance of the RAKE correlator, the matched filter, and the energy detector versus coherence time, coherent bandwidth, and SNR is explored by simulation of realistic cases. The RAKE is then applied to a 4000-km range, 75-Hz sonar transmission and the performance improvement is remarkable. An added benefit of the RAKE is that there is less clutter, or fewer false alarms in the detection theory jargon, and thus tracking an arrival over multiple transmissions is much easier.

2:45

3pSP6. A frequency domain blind deconvolution algorithm in acoustics. Mark R. Gramann, Josh G. Erling, and Michael J. Roan (Appl. Res. Lab, Penn State Univ., P.O. Box 30, State College, PA 16804, mrg227@psu.edu)

It is common in acoustics to measure a signal that has been corrupted by an unknown filtering function during propagation from an unknown source. Blind deconvolution is a technique for learning and applying the inverse of the unknown channel impulse response in order to recover the original source signal. One approach to accomplishing this task is based on an adaptive nonlinear algorithm using mutual information as a cost function [A. J. Bell and T. J. Sejnowski, Neural Comput. **7**, 1129–1159 (1995)]. A new frequency domain implementation of this algorithm is presented which greatly reduces computational cost. The frequency domain approach allows adaptive learning rates to be applied individually to each frequency bin of the inverse filter. This technique can lead to improved convergence times for filters with a large spread of frequency response magnitudes. Preliminary results suggest that a factor of two reduction in convergence time and a factor of ten reduction in computational cost can be attained. Experimental results for several simple acoustical systems are presented comparing the performance of the pre-existing time domain algorithm and the new frequency domain implementation. [Work supported by Dr. David Drumheller, ONR Code 333, Contract No. N00014-00-G-0058.]

Session 3pUW

Underwater Acoustics: Communications and Time Reversal

William A. Kuperman, Chair
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California–San Diego, La Jolla, California 92093-0238

Contributed Papers

1:00

3pUW1. Simultaneous focal spots obtained at sea using time reversal and reciprocity. Philippe Roux, W. A. Kuperman, W. S. Hodgkiss, Hee Chun Song, Tuncay Akal (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0238), and Mark Stevenson (NATO SACLANT Undersea Res. Ctr., La Spezia 19138, Italy)

Time reversal can be implemented between an array of emitters and an array of receivers by invoking reciprocity in the experimental procedure. The actual process requires connectivity between the two arrays but eliminates the requirement of actually having a probe source. Multiple time reversal focal spots can then be achieved simultaneously in the water column. We verify with data taken in a recent experiment that this streamlined, reciprocity-based procedure yields the same results as the classical method. Of greater importance, it provides a versatile method to study a TRM in a fluctuating medium.

1:15

3pUW2. Spatial encoding of communication sequences with an ocean acoustic time-reversal mirror. Philippe Roux, W. A. Kuperman, W. H. Hodgkiss, Hee Chun Song, Tuncay Akal (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0238), and Mark Stevenson (NATO SACLANT Undersea Res. Ctr., La Spezia 19138, Italy)

With the at-sea demonstration of producing simultaneous multiple focal spots with a TRM, spatial encoding of communication sequences is now possible. We present the results of an ocean acoustic experiment in which we demonstrate that different communication sequences can be simultaneously sent to and decoded at individual receivers on a vertical array. This ''multiple input, multiple output (MIMO)'' process significantly increases the acoustic communications band rate. Various incoherent and coherent communication temporal coding schemes are employed together with this spatial procedure.

1:30

3pUW3. The use of a synthetic aperture endfire array in communications. Bill Higley, Philippe Roux, and William A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., NTV Bldg., 4th Fl., 8820 Shellback Way, La Jolla, CA 92093-0238)

An ultrasonic, synthetic aperture, endfire array has been constructed to study its time reversal properties. The endfire time-reversal focal spot exhibits strong spatial sidelobes not present when focusing with a vertical array, but the temporal sidelobes for an endfire focal spot are as low as those obtained using a vertical array. The minimal hardware configuration of a synthetic endfire time-reversal array using only one transmitter and one receiver makes communications a viable application. We have performed experiments to estimate data rates and corresponding bit error rates that could be achieved with such an inexpensive system utilizing either passive or active time-reversal.

1:45

3pUW4. High-frequency underwater acoustic communications using FH-FSK signaling in a reverberant shallow water environment. Wen-Bin Yang and T. C. Yang (Naval Res. Lab., Code 7120, 4555 Overlook Ave. SW, Washington, DC 20375, wyang@wave.nrl.navy.mil)

This paper describes the experimental results of frequency-hopped frequency-shift-key (FH-FSK) signaling operated at 20 kHz with a 4 kHz bandwidth for underwater acoustic communications in a reverberant shallow water environment. The data were collected during the RDS4 (Rapidly Deployable Systems) experiment in a shallow water (<80 m depth) near Halifax, Canada. The measured impulse response function showed multipaths lasting over a second, which is an order of magnitude longer than the symbol length. Time-varying Doppler shifts of 30–70 Hz were found in the data. The long multipath delay and high Doppler shift are found to have a significant impact on data processing. For example, using conventional processing that detects the symbol energy over the symbol duration, the bit error rates (BER) are of the order 30–40%. Using a longer time window allowing integration of multipath energy and using Doppler estimated from trigger signals, the uncoded BER is reduced to 10–15%. The data are error-free after error decoding using a convolutional code with a rate and constraint length of 9. Consequences for acoustic networking will be discussed. [Work supported by ONR.]

2:00

3pUW5. Source motion and environmental effects on orthogonal frequency division multiplexed signaling. Paul Gendron (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375)

An assessment of multipath spread, coherence time, and source motion induced dilation on orthogonal frequency division multiplexed communications is made. Multipath delay spread of over 0.5 s and frequency selectivity of less than 20 Hz were observed off the Nova Scotia coast. Source motion on an 18-kHz center frequency 4-kHz bandwidth signal induce Doppler wander as large as 2 Hz/s. Implications on communications via tightly spaced orthogonal tones is addressed and bit error rates are reported.

2:15

3pUW6. Experimental results of time reversal and optimal inverse filtering performed in a one dimensional waveguide. Pierre M. Dumuid and Ben S. Cazzolato (Dept. of Mech. Eng., Adelaide Univ. SA 5005, South Australia, pmdumuid@mecheng.adelaide.edu.au)

Time reversal is a common technique for improving the temporal and spatial focusing in the transmission of signals and has been demonstrated in a wide variety of applications. An alternative to time reversal is inverse filtering. Inverse filters are similar to time reversal filters in that they also remove the phase aberations, but they have the added advantage that they also counter the amplitude abberations such that the resulting response through the system has a flat response in the frequency domain. Inverse

filtering also provides a means to develop a filter structure that provides a cross-talk cancellation system for multiple input/multiple output systems. In this presentation, results from experimental observations made within a one dimensional waveguide are compared with the commonly used technique of time reversal with an optimal inverse filtering technique. The aim of such experiments is to examine the performance of the time reversal and optimal inverse filtering techniques within a real waveguide and compare these against theoretical predictions. The experiments performed have been developed with a view to extend the research to underwater acoustic communications.

WEDNESDAY AFTERNOON, 12 NOVEMBER 2003 GRAND BALLROOM, 3:15 TO 5:15 P.M.

Plenary Session and Awards Ceremony

Ilene J. Busch-Vishniac, President
Acoustical Society of America

Presentation of Certificates to New Fellows

George L. Augspurger
D. Murray Campbell
Dezhang Chu
Kenneth A. Cunefare
Grant B. Deane
David R. Dowling
Timothy J. Foulkes
Ronald R. Freiheit
Peter Gerstoft
John J. Guinan, Jr.
John M. Harrison
Steven G. Kargl
Joseph F. Lingevitch
Thomas J. Matula
James H. Miller
Jeffrey A. Nystuen
Michael D. Richardson
Brigitte Schulte-Fortkamp
Julius O. Smith, III
Hee-Chun Song
Evan K. Westwood

Presentation of Awards

Science Writing Award for Journalists to Emily A. Thompson

Silver Medal in Physical Acoustics to Philip L. Marston

Trent-Crede Medal to Sabih I. Hayek

Session 4aAAa

Architectural Acoustics, Noise, Engineering Acoustics, and Musical Acoustics: Sound Quality—When Sound is the Essential Quality

Richard H. Lyon, Cochair
RH Lyon Corp, 691 Concord Avenue, Cambridge, Massachusetts 02138

William M. Hartmann, Cochair
Department of Physics, Michigan State University, 4230 BPS Building, East Lansing, Michigan 48824

Chair's Introduction—8:00

Invited Papers

8:05

4aAAa1. Sound quality (SQ) of concert halls: Physical and subjective attributes. Leo L. Beranek (975 Memorial Dr., Cambridge, MA 02138, beranekleo@ieee.org)

Each new concert hall has the following stated goal: ''Acoustics equal to the best in the world.'' The owner can specify the number of seats, areas of public spaces, lighting intensities, etc. But, the attributes of acoustical quality cannot as yet be specified. Most acoustical consultants seem to feel that a ''seat of the pants'' experience is the only possible specification. But the architect's goal is a monument to himself and he believes the acoustical consultant should achieve the ''best in the world'' goal without visible means. Numbers for specifications are needed. In this paper 40 years of pertinent research are described: What are the critical physical attributes of good acoustics, how do we measure them, and how can they be translated into architectural specifications? Four steps have been involved: (1) interviews of conductors and music critics to determine (a) their acoustical rank orderings of a large number of halls and (b) which acoustical characteristics do they believe are important, viz., reverberance, strength of sound, etc.; (2) a determination of which physical measures correlate with their beliefs plus others that are physiologically important; (3) measurements of those physical quantities in the rank-ordered halls; and (5) the correlation of the measured values with the subjective quality ratings.

8:25

4aAAa2. Assessing the sound quality of a grand piano for different tuning standards. Hugo Fastl (Dept. of Human Machine Commun., Munich Univ. of Technol., Germany, fastl@ei.tum.de)

Grand pianos are tuned to a standard frequency of 440 Hz for a4. Eighty years ago, the tuning standard was lower, around 432 Hz. In Germany, a group of music lovers insists that the sound quality of a grand piano when tuned to 432 Hz is much superior to that of the same instrument when tuned to 440 Hz. Therefore, well-controlled psychoacoustic experiments were performed to check the validity of that argument. Using a Welte–Steinway reproduction grand piano of the Deutsches Museum Muenchen, with the exception of the tuning, all other features of the pieces of music used as stimuli could be kept the same. The advantage of using the Welte–Steinway lies in the fact that reproductions of the music of famous (deceased) artists are available, who at their time performed at a lower tuning standard. Recordings of the music played at 432 vs 440 Hz were made on DAT and presented to the subjects via headphones for sound quality evaluation. Psychophysical procedures like semantic differential or preference scaling by ''random access'' that have proven successful for the assessment of sound quality in the context of car interior sounds were used in the experiments.

8:45

4aAAa3. Sound quality and loudspeakers. Neil A. Shaw and Michael A. Klasco (Menlo Sci. Acoust., Inc., P.O. Box 1610, Topanga, CA 90290, menlo@ieee.org)

The sound quality for consumer, professional, and industrial products has been found to be an important part of the product design. For many products, sound is a byproduct of the normal operation of the product. In the case of loudspeakers, the production of sound is the normal operation of the product—its job is to produce sound. A survey of the current popular means of describing and measuring loudspeakers is presented. The introduction of some new means for measuring and describing raw transducers and systems is reviewed. Areas that are still problematic are also discussed.

9:05

4aAAa4. Basic requirements for realistic and unprejudiced evaluation of musical instruments. Klaus Wogram (PTB, Braunschweig, Germany)

In the past, the laboratory of musical acoustics of the PTB had to develop methods of subjective and objective evaluations of all kinds of musical instruments for a competition which was launched by the German ministry of economics. For this work the following questions had to be answered: (1) What makes a musical instrument play well from physical and subjective viewpoints? (2) How does

the musician play the instrument, and what does he think when playing? (3) What is the influence of room acoustical parameters on sound quality? (4) How strong is the influence of prejudices concerning the brand, color, and type? (5) How can we measure the main acoustical parameters objectively? (6) What is the correlation between objective and subjective results? This presentation will give answers to these questions based on 10 years of experience with such evaluations by the author.

9:25

4aAAa5. Evaluating the sound quality of reproduction systems and performance spaces. David Griesinger (Lexicon, Inc., 23 Bellevue Ave., Cambridge, MA 02140)

Evaluation of sound reproduction systems through rapid A/B tests has led to enormous and rapid progress in system design. The sound quality of performance spaces is almost always evaluated through long term listening, and spaces are compared through remembered characteristics. Progress has not been rapid. This paper will present methods for rapid A/B comparisons of spaces and will demonstrate the results of such comparisons. The results can be very different from remembered impressions.

9:45

4aAAa6. Sound quality of brass-wind musical instruments. Robert W. Pyle, Jr. (11 Holworthy Pl., Cambridge, MA 02138, rpyle@post.harvard.edu)

Confronted with the phrase ''sound quality'' applied to brass-instrument sound, the typical player will respond, ''Oh, you mean *tone*.'' For the player, this one word lumps together all spectrum-dependent features of the sound. The player may deliberately vary tone quality depending on the type of music being performed, the other performers involved, and the performance space. After a review of those features of tone quality common to all brass instruments and the differences between the various members of the family, some more subtle questions will be addressed. How and to what extent can the player control tone color without changing equipment? How does the player choose his or her equipment? How does the instrument maker offer choices to the player? How does the tone as heard in the audience differ from that heard by the player? Time permitting, there will be consideration of the difficulties in establishing a meaningful vocabulary for the discussion of tone quality.

10:05–10:20 Break

Contributed Papers

10:20

4aAAa7. Research opportunities in evaluating sound system performance. Bob Thurmond (G. R. Thurmond and Assoc., 9315 Springdale Rd., Austin, TX 78754)

Sound reinforcement has been both ubiquitous and indispensable in our society for many years. With such a volume of experience, it would seem that the accompanying body of knowledge should be sufficient to produce a successful design for every situation. However, the quality of actual systems installed today still varies from good to terrible. A few obvious aspects of this problem have been examined, and papers published, but no serious attempt to list or organize all contributors to sound performance quality has ever been made. In addition, no evaluative techniques or analytical strategies have been established to set standards. There are several valid reasons for this. The actual performance of a sound system is surprisingly complex, both in terms of the number of factors contributing to the overall performance and in the complexity within each factor. As a result, each system is quite unique and difficult to describe in precise terms. Few rigorous investigations have been applied. The results have been meager thus far and are essentially unknown to most practitioners. Research opportunities abound, and their results could have a significant impact. In this paper the current situation is summarized and topics for future study are suggested.

10:35

4aAAa8. Virtual acoustic prototyping. Marty Johnson (Vib. and Acoust. Labs., Dept. of Mech. Eng., Virginia Tech, VA 24061-0238)

In this paper the re-creation of 3-D sound fields so the full psychoacoustic impact of sound sources can be assessed before the manufacture of a product or environment is examined. Using head related transfer functions (HRTFs) coupled with a head tracked set of headphones the sound field at the left and right ears of a listener can be re-created for a set of sound sources. However, the HRTFs require that sources have a defined location and this is not the typical output from numerical codes which describe the sound field as a set of distributed modes. In this paper a method of creating a set of equivalent sources is described such that the standard set of HRTFs can be applied in real time. A structural-acoustic model of a cylinder driving an enclosed acoustic field will be used as an example. It will be shown that equivalent sources can be used to recreate all of the reverberation of the enclosed space. An efficient singular value decomposition technique allows the large number of sources required to be simulated in real time. An introduction to the requirements necessary for 3-D virtual prototyping using high frequency Statistical Energy Analysis models will be presented. [Work supported by AuSim and NASA.]

10:50

4aAAa9. An overview of suitable measurement and assessment techniques for distributed mode loudspeakers. Nick Hill (NXT, Huntingdon PE29 7HJ, UK) and Peter Mapp (Peter Mapp Assoc., Colchester C03 4JZ, UK, Petermapp@btinternet.com)

The sound field generated by a distributed mode loudspeaker (DML) can exhibit some characteristics that may be unfamiliar to conventional cone loudspeaker designers. In particular, the modal vibration of the diaphragm gives rise to a wide range of directivity patterns dependent on the specific design of the unit. This paper outlines some of the possible behaviors and the measurement techniques that should be employed to capture each case. In particular, the features of a diffuse sound field are discussed. It is shown that this can give rise to a pressure response that may vary significantly over a small angle and frequency scale. Here a multipoint measurement, such as a spatial average or acoustic power, gives a more representative assessment than measuring just the acoustic sound pressure level at a single point. In addition, the fine detail of the response may be characterized by a spatially averaged correlation function, giving a measure of the diffusivity that is complementary to more standard measures of energy distribution.

11:05

4aAAa10. Computer modeling of a large fan-shaped auditorium. Heather Smith and Timothy Leishman (Dept. of Phys., Brigham Young Univ., N283 ESC, Provo, UT 84602, hm73@email.byu.edu)

A research project was recently undertaken to analyze the acoustical characteristics of a 21 000-seat fan-shaped auditorium. Careful geometric modeling of the hall has been a significant part of this study. Because of its size, shape, and other architectural features, computer modeling has presented some interesting challenges. For example, it has been shown experimentally that the concavely oriented rows of (moderately absorptive) seats produce significant scattering that aggregates toward the focal point of the hall. This paper will discuss how the seat scattering and scattering from other bodies have been included in the model. Other challenges in modeling a hall this size will also be discussed.

11:20

4aAAa11. Radiant exchange in partially specular architectural environments. C. Walter BeamerIV and Ralph T. Muehleisen (Dept. of Civil, Environ., and Architectural Eng., Univ. of Colorado, Boulder, CO 80309)

The radiant exchange method, also known as radiosity, was originally developed for thermal radiative heat transfer applications. Later it was used to model architectural lighting systems, and more recently it has been extended to model acoustic systems. While there are subtle differences in these applications, the basic method is based on solving a system of energy balance equations, and it is best applied to spaces with mainly diffuse reflecting surfaces. The obvious drawback to this method is that it is based around the assumption that all surfaces in the system are diffuse reflectors. Because almost all architectural systems have at least some partially specular reflecting surfaces in the system it is important to extend the radiant exchange method to deal with this type of surface reflection. [Work supported by NSF.]

11:35

4aAAa12. Citadel in Teotihuacan. Sergio Beristain (P.O. Box 12-1022, Col. Narvarte, 03001 Mexico, D.F. Mexico, sberista@hotmail.com)

Teotihuacan, the largest archaeological site nearby Mexico City, is also a place where traditions are maintained through some ceremonies on specific dates, by the Sun and Moon pyramids, and history telling by the pyramids in day to day light and sound shows. This enormous site has a large square in the south known as The Citadel (La Ciudadela), a place some 200×300 meters (m), surrounded by 2.2 m high pyramid basements, and two pyramids to the East (one in front of the main one dedicated to the good Quetzalcoatl). Near the center of this large square sits a 2.2 m basement 18×20 m where some special sound events (theatre, dance, music, etc.) are occasionally presented. Sound level measurements have proved that due to the site conditions, the sound level decreases 3–4 dB on the average per doubling distance, which makes it suitable for large audiences with the only problem of some minor echoes in small portions of the audience area.

THURSDAY MORNING, 13 NOVEMBER 2003 WEDGWOOD ROOM, 9:00 TO 10:05 A.M.

Session 4aAAb

Architectural Acoustics and Noise: Distinguished Lecture: Analysis of Community Response to Transportation Noise a Quarter Century After Schultz (1978)

Jack E. Randorff, Chair
Randorff and Associates, Inc., 11 West Canyon View Drive, Ransom Canyon, Texas 79366

Chair's Introduction—9:00

Invited Paper

9:05

4aAAb1. Analysis of community response to transportation noise a quarter century after Schultz (1978). Sanford Fidell (Fidell Assoc., Inc., 23139 Erwin St., Woodland Hills, CA 91367)

Transportation noise is a vexing and intrinsically controversial problem that has plagued societies since the beginnings of urban civilization. A function relating cumulative noise exposure to the prevalence of noise-induced annoyance [T. J. Schultz, ''Synthesis of social surveys on noise annoyance,'' J. Acoust. Soc. Am. **64**, 377–405 (1978)] is the foundation for contemporary analyses of transportation noise effects on communities. The expenditures of billions of dollars in airplane ticket and fuel taxes for the construction of an airport infrastructure and for the mitigation of noise impacts in the United States are governed by policies ostensibly supported by a successor to Schultz's original ''synthesis'' curve. Many have grown so comfortable with the last quarter century's paradigm for transportation noise assessment and regulation, however, that they have lost sight of its underpinnings and limitations. A review of the historical and modern states of the art identifies persistent unresolved problems in the prediction and explanation of community response to transportation noise that are not fully addressed by descriptive dosage-effect analysis.

Session 4aAAc

Architectural Acoustics and Noise: Progress in Measurement of Transportation Noise Since Ted Schultz

Sanford Fidell, Chair
Fidell Associates, Inc., 23139 Erwin Street, Woodland Hills, California 91367

Invited Papers

10:15

4aAAc1. Shoehorning in the jet age. Leo L. Beranek (975 Memorial Dr., Cambridge, MA 02138, beranekleo@ieee.org)

The introduction into commercial service of jet air transports preceded the codification of U.S. federal aviation noise regulatory policy by nearly two decades. The immediate problem that had to be solved to permit jet operations in 1958 at New York airports was to determine that the noisiness of jet airplane would be no greater than that of the largest propeller-driven aircraft then in operation. This paper describes the efforts made for the Port of New York Authority to meet and enforce that mandate. BBN made noise measurements of many propeller aircraft takeoffs in the community off the main runway at Idlewild at distances of 2.5 miles and greater from the start of take-off roll. Each measurement was associated with airplane type, gross weight, and altitude. The Boeing 707 was similarly measured at Boeing's airport. Relative annoyance judgments were made in the laboratory to establish equivalent noisiness of jet and propeller spectra, and ''Perceived Noise Levels'' were determined by a process similar to loudness calculations. Boeing was forced to equip the 707 with multi-tube mufflers and, and jet take-off procedures had to be modified to maintain equivalent perceived noise levels. The substantive findings and politics of these efforts are discussed.

10:40

4aAAc2. Origins and application of the European Union Position paper on dose response relationships between transportation noise and annoyance. Bernard F. Berry (Berry Environment Ltd.—BEL, Shepperton, UK)

Dose–response relationships of the sort pioneered by Schultz figure prominently in the current European noise regulation policy. A position paper developed by the Working Group on Dose-effects (part of the EU Expert Network) reviewed a range of potential health effects, but decided that annoyance and sleep disturbance remain the most prevalent and sensitive effects of transportation noise exposure, and those for which the best data were available. A Position paper providing guidance on the dose–effect relations to be used for the assessment of numbers of people annoyed by noise from transportation sources (rail, road and air) may be found at http://europa.eu.int/comm/environment/noise/home.htm This presentation explains the context in which the Paper was developed, outlines the process by which the dose–response relationships were derived and summarizes the key recommendations. Finally some observations are made as the Position Paper is being applied, and related future developments are discussed. The author acknowledges with deep gratitude the assistance of Dr. Henk Miedema of TNO in preparing this paper.

11:05

4aAAc3. Correcting DNL so it works—better. Paul Schomer (Schomer and Assoc., Inc., 2117 Robert Dr., Champaign, IL 61821)

The day–night average sound level (DNL), first developed by the U.S. Environmental Protection Agency, is commonly used to quantify and assess environmental noise. A keystone to noise assessment is the dose–response relationship. However, the dose–response relationship is not an absolute; there is great scatter to the data on which it is based. In an attempt to reduce the scatter to the DNL data, the EPA suggested the use of normalized DNL. Normalized DNL is the basic DNL value with a number of adjustments added to account for specific characteristics and factors of the sound. This paper reviews and analyze the concepts inherent in normalized DNL and provide an updated set of normalization factors that can reduce the scatter to dose–response relationships. Several of these normalizations are contained in the new ISO 1996-1:2003.

11:30–12:00

Panel Discussion

Session 4aAB

Animal Bioacoustics: Neurobiology of Communications

George D. Pollak, Chair
Section of Neurobiology, University of Texas, Austin, Texas 78705

Invited Papers

8:30

4aAB1. Audiomotor integration for active sensing in the echolocating bat, *Eptesicus fuscus*. Cynthia F. Moss, Shiva R. Sinha, and Kaushik Ghose (Neurosci. and Cognit. Sci. Prog., Inst. for Systems Res., Univ. of Maryland, College Park, MD 20742)

Echolocating bats probe the environment with sonar signals that change as they seek, pursue and intercept insect prey on the wing. Coordinating its sonar vocalizations with flight dynamics in response to changing echo information, the bat exhibits a dazzling display of sensorimotor integration. Our work aims at understanding the mechanisms supporting audiomotor integration for echolocation in the FM-bat, *Eptesicus fuscus*. Behavioral studies measure adaptive responses of free-flying bats engaged in complex spatial tasks. The directional aim of the bat's sonar beam and temporal patterning of cries provide explicit data on the motor commands that feed directly back to the auditory system for spatially-guided behavior. Neural studies focus on the superior colliculus (SC), a midbrain structure implicated in species-specific orienting behaviors. A population of SC neurons shows echo-delay tuning, a response property believed to play a role in target range coding. Microstimulation of the SC elicits head and pinna movements, along with sonar vocalizations. SC recordings from tethered, vocalizing bats reveal bursts of neural activity preceding each sonar cry. Collectively, these results suggest that the bat SC plays a functional role in the auditory information processing and orienting behaviors that operate together in echolocation. [Work supported by NSF, NIMH and Whitehall Foundation.]

9:00

4aAB2. Processing and representation of social communication sounds in the brainstem auditory system of bats. George D. Pollak (Section of Neurobiology, Univ. of Texas, Austin, TX 78712, gpollak@mail.utexas.edu)

While bats are best known for their abilities to orient and capture prey via echolocation, they are also highly social animals who use a rich repertoire of species-specific sounds for social communication. This talk explores how communication signals are progressively transformed and represented in the ascending auditory system. One principal transformation that distinguishes the inferior colliculus from lower nuclei is a change from processing that emphasizes response homogeneity among the neuronal population in each lower nucleus, to one that emphasizes heterogeneity and selectivity in the inferior colliculus. Collicular neurons are selective in that each neuron fails to respond to some, or even all calls, even though those calls have energy that encroaches upon their excitatory response regions, and are heterogeneous since each collicular neuron responds to a different subset of calls. The transformation from homogeneity to heterogeneity may largely be a consequence of the difference in the ways that the various excitatory and inhibitory inputs distribute along frequency contours in lower nuclei compared to the inferior colliculus. One important consequence is that those features endow the population in the inferior colliculus with the ability to respond to any signal with a unique and pronounced spatiotemporal pattern of activity. [Work supported by NIH Grant No. DC 00268.]

Contributed Papers

9:30

4aAB3. Computational model of modulation detection by the bullfrog. Andrea M. Simmons and Kyler Eastman (Depts. of Psych. and Neurosci., Brown Univ., Providence, RI 02912, Andrea_Simmons@brown.edu)

Bullfrog eighth-nerve fibers operate as envelope detectors, showing significant phase-locking to amplitude modulation (AM) rates as high as 800 Hz. A computational model that estimates stimulus period from all-order interval histograms (autocorrelation functions) aligned in time across frequency channels mimics fiber responses. The use of autocorrelation implies a mechanism based on delay lines or neural coincidence detection. The goal of this study was to determine the relevance of such a model in predicting responses of neurons in bullfrogs auditory midbrain (torus semicircularis) to AM stimuli. Modeled output of peripheral fibers was passed through a series of delay lines varying in latency, and then compared with actual midbrain data. There is a great diversity in representation of AM in the auditory midbrain, and this diversity is related to response latency and recording location. Neural responses from the cell sparse zone in the caudal midbrain were well-matched by a delay line mechanism, although the model was poorer in predicting response properties of other midbrain areas. [Work supported by NIH and the Brown University Brain Science Program.]

9:45

4aAB4. Otoacoustic emissions measured in rhesus monkeys (*Macaca mulatta*). Dennis McFadden, Edward G. Pasanen (Dept. of Psych., Univ. of Texas, Austin, TX 78712-0187, mcfadden@psy.utexas.edu), Jessica Raper, and Kim Wallen (Emory Univ., Atlanta, GA 30322-2470)

In humans, otoacoustic emissions (OAEs) are stronger in females than in males and stronger in right ears than in left. The physiological bases for these differences are unknown, but several lines of circumstantial evidence suggest that the sex difference is attributable to androgenizing mechanisms operating during prenatal development. Specifically, it appears that exposure to high levels of androgens during prenatal development diminishes the strength of the cochlear amplifiers and thus the strength of the OAEs. Sex and ear differences in OAEs have not been well studied in species other than humans. Accordingly, click-evoked OAEs and distortion-product OAEs were measured in nine female and nine male rhesus monkeys. For CEOAEs, but less clearly for DPOAEs, females exhibited sig-

nificantly stronger OAEs than males. There was no consistent ear difference for either sex for either type of OAE. In order to better study the early components of the CEOAE waveform, a nonlinear procedure [Molenaar *et al.*, Hearing Res. **143**, 197–207 (2002)] was used to collect CEOAEs along with our standard (linear) procedure. This colony also contains animals of each sex that were treated with androgenic or antiandrogenic agents during prenatal development, and OAEs are also currently being measured on those animals. [Work supported by NIDCD.]

10:00

4aAB5. Otoacoustic emissions measured in spotted hyenas (*Crocuta crocuta*). Dennis McFadden, Edward G. Pasanen (Dept. of Psych., Univ. of Texas, Austin, TX 78712-0187, mcfadden@psy.utexas.edu), Mary L. Weldele, Stephen E. Glickman, and Ned J. Place (Univ. of California, Berkeley, CA 94720)

From birth, female spotted hyenas exhibit highly masculinized bodies and behaviors. Their external genitalia greatly resemble those of males, and they are behaviorally dominant over males. This marked masculinization raised the question of whether the otoacoustic emissions (OAEs) of female spotted hyenas also would be masculinized. Click-evoked OAEs were measured in six female and six male hyenas at two click levels. Also, distortion-product OAEs were measured at four or more primary levels in three frequency regions: 2, 3.5, and 5.0 kHz. Both CEOAEs and DPOAEs were strong in both sexes in spotted hyenas. In humans, both CEOAEs and DPOAEs are stronger in females than males and stronger in right ears than left. Unlike humans, both the CEOAEs and DPOAEs in female spotted hyenas were weaker than those in males, and unlike humans, OAEs were not stronger in right ears. The implication is that the same androgenizing processes that masculinize the body and behavior of female hyenas also masculinize those elements of the cochlea responsible for OAEs. That implication is being tested by measuring the OAEs of other hyenas in the Berkeley colony that were treated with antiandrogenic agents during fetal development. [Work supported by NIDCD.]

10:15–10:30 Break

10:30

4aAB6. Distortion product otoacoustic emissions provide clues to hearing mechanisms in the frog. Pantelis Vassilakis[a] (Dept. of Physiological Sci., Univ. of California, Los Angeles, CA 90095-1606, pantelis@acousticslab.com.) and Peter M. Narins (Univ. of California, Los Angeles, CA 90095-1606)

Cubic distortion product otoacoustic emissions (DPOAEs) were recorded from 10 *Rana pipiens* and 10 *Rana catesbeiana*, 5 males and 5 females each. The I/O curves obtained from the amphibian papilla (AP) of both species are very similar to the respective mammalian curves, indicating that, like in the mammalian cochlea, there may be an amplification process active in the frog AP. The DPOAE level dependence on primary levels is also similar to the mammalian case, suggesting a mechanical structure in the frog inner ear may be functioning analogously to the mammalian basilar membrane. DPOAE audiograms were obtained for primary frequencies spanning the animals hearing range and levels determined by the previous experiments. *R. catesbeiana* produce stronger emissions than *R. pipiens* and, consistent with previously reported sexual dimorphism in the mammalian and anuran auditory systems, females from both species produce stronger emissions than males. Additionally, the $2f_1-f_2$ DPOAE is generated primarily at the DPOAE frequency place, while the $2f_2-f_1$ DPOAE is generated primarily at a frequency place between the primaries. This difference in mammalian and frog DPOAEs may be linked to an anatomical difference that results in the acoustic energy following opposite paths through the mammalian and frog inner ears. [Work supported by NIH Grant No. DC-00222 to Peter M. Narins.]

[a] Currently at De Paul Univ., School of Music, Chicago, IL 60614.

10:45

4aAB7. Do goldfish miss the fundamental? Richard R. Fay (Parmly Hearing Inst., Loyola Univ. Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626, rfay@luc.edu)

The perception of harmonic complexes was studied in goldfish using classical respiratory conditioning and a stimulus generalization paradigm. Groups of animals were initially conditioned to several harmonic complexes with a fundamental frequency (f_0) of 100 Hz. ln some cases the f_0 component was present, and in other cases, the f_0 component was absent. After conditioning, animals were tested for generalization to novel harmonic complexes having different f_0's, some with f_0 present and some with f_0 absent. Generalization gradients always peaked at 100 Hz, indicating that the pitch value of the conditioning complexes was consistent with the f_0, whether or not f_0 was present in the conditioning or test complexes. Thus, goldfish do not miss the fundmental with respect to a pitch-like perceptual dimension. However, generalization gradients tended to have different skirt slopes for the f_0-present and f_0-absent conditioning and test stimuli. This suggests that goldfish distinguish between f_0 present/absent stimuli, probably on the basis of a timbre-like perceptual dimension. These and other results demonstrate that goldfish respond to complex sounds as if they possessed perceptual dimensions similar to pitch and timbre as defined for human and other vertebrate listeners. [Work supported by NIH/NIDCD.]

11:00

4aAB8. Low-frequency vocalizations in the Florida manatee (*Trichechus manatus latirostris*). Katherine Frisch (Florida Marine Res. Inst., 100 8th Ave. SE, St. Petersburg, FL 33701) and Stefan Frisch (Univ. of South Florida, Tampa, FL 33620)

Vocalizations produced by Florida manatees (*Trichechus manatus latirostris*) have been characterized as being of relatively high frequency, with fundamental tones ranging from 2500–5000 Hz. These sounds have been variously described as squeaks, squeals, and chirps. Vocalizations below 500 Hz have not been previously reported. Two captive-born Florida manatees were recorded at Mote Marine Laboratory in Sarasota, Florida. The analysis of these vocalizations provides evidence of a new category of low-frequency sounds produced by manatees. These sounds are often heard in conjunction with higher-frequency vocalizations. The low-frequency vocalizations are relatively brief and of low amplitude. These vocalizations are perceived as a series of impulses rather than a low-frequency periodic tone. Knowledge of these low-frequency vocalizations could be useful to those developing future management strategies. Interest has recently increased in the development of acoustic detection and deterrence devices to reduce the number of manatee watercraft interactions. The design of appropriate devices must take into account the apparent ability of manatees to perceive and produce sounds of both high and low frequency. It is also important to consider the possibility that acoustic deterrence devices may disrupt the potentially communicative frequencies of manatee vocalizations.

11:15

4aAB9. A dolphin lower jaw is a hydroacoustic antenna of the traveling wave. Vyacheslav A. Ryabov (Karadag Natural Reserve, Natl. Acad. of Sci. of Ukraine, Kurortnoe, Feodosia 98188, Crimea, Ukraine)

The purpose of the work is the analysis of a possible function of mental foramens as channels through which the echo passes in the lower jaw fat body and the determination of a role of channels and a skull in formation of the directivity of the dolphin echolocation hearing. Concrete problems were studying of the lower jaw morphology, modeling and calculation of a dolphin, *tursiops truncatus* p., echolocation hearing beam pattern. The outcomes of the work indicate those morphological structures of the lower jaw; the left and right half represents two hydroacoustic receiving antennas of the traveling wave type, TWA farther. The mental foramens of a dolphin lower jaw represent nonequidistant array of waveguide delay lines, and determine the phase and amplitude distribution of each of the antenna's array. The beam pattern of the echolocation hearing was calculated with the usage of the TWA model, and the allowance of flat

sound wave diffraction. The beam pattern shape is naturally determined by the echolocation hearing functionality. It is equally well adapted both for echolocation and for pulses echo detection. A steepness of the bearing characteristic is estimated; it reaches 0.7 dB per degree.

11:30

4aAB10. Florida manatee avoidance technology: A pilot program by the Florida Fish and Wildlife Conservation Commission. Katherine Frisch and Elsa Haubold (Florida Marine Res. Inst., 100 8th Ave. SE, St. Petersburg, FL 33701)

Since 1976, approximately 25% of the annual Florida manatee (*Trichechus manatus latirostris*) mortality has been attributed to collisions with watercraft. In 2001, the Florida Legislature appropriated $200,000 in funds for research projects using technological solutions to directly address the problem of collisions between manatees and watercraft. The Florida Fish & Wildlife Conservation Commission initially funded seven projects for the first two fiscal years. The selected proposals were designed to explore technology that had not previously been applied to the manatee/boat collision problem and included many acoustic concepts related to voice recognition, sonar, and an alerting device to be put on boats to warn manatees. The most promising results to date are from projects employing voice-recognition techniques to identify manatee vocalizations and warn boaters of the manatees' presence. Sonar technology, much like that used in fish finders, is promising but has met with regulatory problems regarding permitting and remains to be tested, as has the manatee-alerting device. The state of Florida found results of the initial years of funding compelling and plans to fund further manatee avoidance technology research in a continued effort to mitigate the problem of manatee/boat collisions.

11:45

4aAB11. Massive gas insufflation without effect on esophageal reflectometry profiles. David T. Raphael, Dimiter Arnaudov, and Maxim Benbassat (Dept. of Anesthesiol., Keck School of Medicine, Univ. of Southern California Med. Ctr., Los Angeles, CA 90033, draphael@usc.edu)

Time–domain acoustic reflectometry generates a ''one-dimensional'' image of the interior of a cavity in the form of an area–distance profile. After patient intubation with a breathing tube, the characteristic reflectometry profile consists of a constant-area segment corresponding to the length of the tube, followed either by a rapid increase in the area beyond the carina (lung) or by a sudden decrease in the area to zero (esophagus). In the cardiac arrest setting, during mistaken placement of the breathing tube into the esophagus, followed by aggressive manual ventilation, is it possible to markedly distend the esophagus, such that the esophageal profile looks like a tracheal profile? With approval of the USC IUCAC Committee, an animal study was conducted with anesthetized, tracheally intubated, and mechanically ventilated dogs. With a separate breathing tube in the esophagus, aggressive esophageal ventilation (comparable to that seen in the cardiopulmonary resuscitation setting) was accomplished with a manual resuscitation bag. A Benson Hood Labs two-microphone reflectometer was used to obtain esophageal profiles with and without the above ventilation. In this pilot study, there was no significant esophageal distention as a result of the above ventilation. [Research supported by the Alfred E. Mann Institute.]

THURSDAY MORNING, 13 NOVEMBER 2003 TRINITY B ROOM, 8:00 TO 11:45 A.M.

Session 4aPA

Physical Acoustics and Biomedical Ultrasound/Bioresponse to Vibration: Special Session on Nonlinear Acoustics in Honor of David Blackstock

Mark F. Hamilton, Cochair
Department of Mechnical Engineering, University of Texas–Austin, Austin, Texas 78712-1063

F. Michael Pestorius, Cochair
International Field Office Europe, Office of Naval Research, PSC 802 Box 39, APO/FPO, NY, AE 09499-0700

Invited Papers

8:00

4aPA1. David Blackstock and nonlinear acoustics at UT Austin. Mark F. Hamilton (Dept. of Mech. Eng., Univ. of Texas, Austin, TX 78712-1063)

Except for a period of 15 years, from when he left Austin (with B.S. and M.S. degrees in physics) for active duty in the USAF until he returned for a ''year's visit'' to UT in 1969, Austin has been home to David Blackstock. In the interim he developed hearing protectors for the Air Force, earned a Ph.D. in applied physics from Harvard, conducted research in nonlinear acoustics for industry, and joined the Electrical Engineering faculty at University of Rochester. Following his return, Austin rapidly became known as a home for nonlinear acoustics. This presentation will review the development of nonlinear acoustics at UT, placed in context of the underwater acoustics activities at ARL and the academic programs under David both at ARL and on campus.

8:15

4aPA2. Soap bubbles, sparks, sonic booms, spikes, and David Blackstock. Allan D. Pierce (Boston Univ., Boston, MA 02215, adp@bu.edu)

In the late 1960s people were puzzled by some anomalous features in the measured waveforms of sonic booms recorded at the ground during flyovers of supersonic aircraft. The waveforms were supposed to look like the letter N, but that was not always the case. Sometimes there were strange upward reaching spikes just behind the leading and trailing shocks. There were a lot of explanations

kicking around, and the present author had the idea that it was somehow caused by imperfect focusing of wavefonts after traveling through regions of higher sound velocity and discussed this with David Blackstock, who made the characteristic remark, ''Let's talk experiment.'' What resulted was an experiment that would have been worthy of Lord Rayleigh. The paper that reported this, titled ''Measurements of the Refraction and Diffraction of a Short *N*-Wave by a Gas-Filled Soap Bubble,'' appeared in J. Acoust. Soc. Am. in March 1971. Besides really nailing down the physical phenomena responsible for the spikes on sonic boom waveforms, it illustrated a wealth of physical concepts and experimental techniques. The present talk discusses the background of the paper, the physics that it used, and the influence it had on subsequent research.

8:30

4aPA3. Weak sparks, *N* waves, and the calibration of microphones at ultrasonic frequencies. Wayne M. Wright (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

A fortuitous situation involving the availability of wide-band condenser microphones, curiosity about acoustic waves generated by sparks, and his own development of nonlinear acoustic theory provided two significant contributions to David Blackstock's continuing research. Theoretical understanding of the nonlinear distortion experienced by intense acoustic impulses in air, such as those generated by relatively weak sparks, led to a technique for calibration of microphones that have essentially flat frequency response up to almost 1 MHz. These sparks and calibrated microphones, in turn, provided the means to carry out several graduate student thesis projects that were supervised by David. Diffraction at edges, apertures, and discs was studied in the time domain. Model studies related to sonic boom phenomena were carried out, as was the propagation of *N* waves in waveguides. The most recent application has been a model study of focusing by an ellipsoidal reflector, such as is typically used in lithotripsy. In each case, the various contributions to the received signals were readily identified in the time-domain presentation.

8:45

4aPA4. Application of nonlinear acoustics in the atmosphere and biomedicine. Robin O. Cleveland (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215, robinc@bu.edu)

David Blackstock has made important contributions to understanding nonlinear processes in underwater acoustics, atmospheric acoustics, and biomedical acoustics. In this talk, two of these areas that David has been involved with are addressed by contrasting the evolution of sonic booms in an inhomogeneous atmosphere with the evolution of shock waves in lithotripsy (breaking of kidney stones by shock waves). The distortion of *N* waves generated by supersonic aircraft is affected by geometrical spreading and changes in the acoustic properties of the medium with altitude. In shock wave lithotripsy, the distortion of triangle waves is affected by geometrical spreading and changes in the acoustic properties due to layering of tissue types. In addition, the turbulent boundary layer near the ground and the inhomogeneous nature of tissue, means that both sonic booms and lithotripter shock waves pass through a random media. Calculations are presented using weak shock to compare the distortion that occurs in the two scenarios. Measurements of saturationlike effects are shown for a clinical shock wave therapy device. Numerical simulations for propagation through random media demonstrate the localized focusing and defocusing that can occur for both sonic booms and lithotripter shock waves. [Work supported in part by NIH and NASA.]

9:00

4aPA5. Statistical nonlinear acoustics: First studies and current state. Oleg V. Rudenko (Faculty of Phys., Moscow State Univ., 119992 Moscow, Russia, rudenko@acs366.phys.msu.ru)

Statistical physics of high-intensity noise waves has, in principle, a long-term history. Nonlinear interactions between random and regular waves studied in the 1930s were devoted to sound absorption in solids caused by energy transfer from coherent phonons (signal) to Debye's thermal quantum (noise). However, exhaustive studies of nonequilibrium nonlinear random wave processes have been carried out only in the 1970s, after experimental and theoretical nonlinear acoustics reached an advanced stage of development. David Blackstock, one of the founders of modern nonlinear acoustics, performed pioneering experiments on broadband noise spectra distortion and signal–noise interactions in tubes and outdoors. These experiments confirmed existing data and formed the fundamental basis for both future studies and applications. Along with temporal statistics, spatial ones were considered by D. T. Blackstock during his studies of nonlinear diffraction caused by orifices and screens with irregular edges. These results are also of great significance for applications. A brief overview of the first studies in statistical nonlinear acoustics is given, as well as of new results obtained in recent years.

9:15

4aPA6. Propagation of finite amplitude sound through turbulence. Bart Lipkens (Mech. Eng. Dept., Western New England College, 1215 Wilbraham Rd., Springfield, MA 01119) and Philippe Blanc-Benon (Ctr. Acoustique, LMFA UMR CNRS 5509, Ecole Centrale de Lyon, 69134 Ecully Cedex, France)

Pestorius and Blackstock [''Propagation of finite amplitude noise,'' Finite-amplitude wave effects in fluids, Proceedings of the 1973 International Symposium on Nonlinear Acoustics, edited by L. J. Bjørnø, 1973] developed a method to simulate the propagation of finite amplitude noise in a tube. The method separates the effects of nonlinear distortion and absorption and diffraction. The nonlinear steepening of an arbitrary waveform is performed in the time domain. The absorption and diffraction effects are implemented in the frequency domain. This method has been used widely for finite amplitude wave propagation. Lipkens and Blanc-Benon [''Propagation of finite-amplitude sound through turbulence: a geometrical acoustics approach,'' C. R. Acad. Sci. Paris, Ser. II b **320**, 477–484 (1995)] used the Pestorius model to simulate the propagation of *N* waves through turbulence. Linear geometrical acoustics is used to trace rays through individual realizations of a turbulent field. A nonlinear transport equation describes the propagation of sound through turbulence. The equation is solved using the Pestorius algorithm. The results indicate that the equivalent nonlinear distortion after propagation through turbulence is always less than that for the homogeneous case. In addition, the effect of a random velocity field is more pronounced that that of a random temperature field.

9:30

4aPA7. Aeroacoustics, moving boundaries, and bursting balloons: Acoustic sources revisited. Christopher L. Morfey (Inst. of Sound and Vib. Res, Univ. of Southampton, Southampton SO17 1BJ, UK, clm@isvr.soton.ac.uk)

The use of equivalent acoustic sources to describe scattering in a nonuniform medium dates back to Rayleigh's theory of sound. The idea of equivalent sources in a uniform medium at rest was later developed by Lighthill into his ''acoustic analogy,'' capable of describing the generation of sound by turbulence and other vortical flows. In the present paper Lighthill's acoustic analogy formulation is generalized to encompass initial-value problems; the initial conditions are represented by impulsive sources and dipoles distributed over the domain, and boundary conditions are represented in the usual manner by surface sources and dipoles. David Blackstock's bursting balloon example, discussed in Chapter 3 of *Fundamentals of Physical Acoustics*, can be solved by this method. However, in situations where the medium is of nonuniform density (for example, a gas with a specified temperature distribution at the initial time), the impulsive source distribution obtained by a direct application of time windowing to the acoustic analogy is nonphysical. The apparent paradox is resolved by introducing the energy conservation equation, and reformulating the acoustic analogy with pressure, rather than density, as the wave variable.

9:45

4aPA8. Diffraction of nonlinear acoustic waves: Matched-area method. Lev A. Ostrovsky (Zel Technologies/NOAA ETL, 325 Broadway, Boulder, CO 80305)

This is a brief overview of a series of investigations undertaken by the author together with his collaborators at the Institute of Applied Physics, Russian Academy of Science. In these works, a simplified theory was developed for the acoustic waves which profile is distorted due to both nonlinearity and diffraction. The theory is based on an often existing possibility to separate the regions of nonlinear geometrical acoustics and linear diffraction and match the corresponding solutions in an intermediate area(s) where neither effect is cumulative. The problems addressed on this way include: (i) broadening of the beam pattern of an intensive acoustic radiation (with V. Fridman); (ii) behavior of a nonlinear acoustic wave near a caustic (with E. Pelinovsky and V. Fridman); (iii) radiation and focusing of nonlinear waves (with A. Sutin). In a number of cases, theoretical results were corroborated by comparison with the experimental data obtained by D. Blackstock and other researchers. This approach is believed to be an efficient tool for treating many realistic situations, and to deserve further development. This work benefited greatly from numerous contacts with Professor Blackstock, starting in the early stages of the investigation.

10:00–10:15 Break

10:15

4aPA9. From bench to bedside: Nonlinear bio-ultrasound. Kevin J. Parker (School of Eng. and Appl. Sci., Univ. of Rochester, Lattimore Hall 309, RC Box 270076, Rochester, NY 14627-0076)

In the early 1960s, David Blackstock developed the weak shock theory which characterized the development and evolution of harmonics in a finite-amplitude propagating plane wave. Later work developed an understanding of beam patterns of the fundamental and higher harmonics from piston radiators. All of these concepts were germane to the field of bio-ultrasound. However, nonlinear effects were largely unexploited in bio-ultrasound until the 1980s, when the role of finite-amplitude mechanisms was recognized as an enhancer of heating and lesion production in high intensity applications. Later, in the 1990s, the role and the advantage of higher harmonics in ultrasonic imaging were developed. In particular, the higher harmonics generated by nonlinear effects, especially in the presence of aberrating tissue, were shown to produce dramatic increases in the resolution and contrast of ultrasound images in cardiology and other applications. Today, every high-end medical ultrasound scanner has the capability of ''harmonic imaging'' for improved image quality in a number of applications. The physics and technology of this remarkable development will be reviewed, along with the important guiding role of David Blackstock and his collaborators, particularly at UT Austin and at U. Rochester.

10:30

4aPA10. Cavitation in shock wave lithotripsy. Michael R. Bailey, Lawrence A. Crum, Oleg A. Sapozhnikov (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Seattle, WA 98105), Andrew P. Evan, James A. McAteer (Indiana Univ. School of Anatomy, Indianapolis, IN), Tim Colonius (California Inst. of Technol., Pasadena, CA), and Robin O. Cleveland (Boston Univ., Boston, MA)

A case is presented for the important role of cavitation in stone comminution and tissue injury in shock wave lithotripsy (SWL). Confocal hydrophones and a coincidence algorithm were used to detect cavitation in kidney parenchyma. Elevated hydrostatic pressure dissolved cavitation nuclei and suppressed cell injury and stone comminution *in vitro*. A low-insertion-loss, thin, mylar film nearly eliminated stone erosion and crack formation only when in direct contact with the stone. This result indicates not only that cavitation is important in both cracking and erosion but also that bubbles act at the surface. Time inversion of the shock wave by use of a pressure-release reflector reduced the calculated pressure at bubble collapse and the measured depth of bubble-induced pits in aluminum. Correspondingly tissue injury *in vivo* was nearly eliminated. Cavitation was localized and intensified by the use of

synchronously triggered, facing lithotripters. This dual pulse lithotripter enhanced comminution at its focus and reduced lysis in surrounding blood samples. The enhancement of comminution was lost when stones were placed in glycerol, which retarded bubble implosion. Thus, cavitation is important in comminution and injury and can be controlled to optimize efficacy and safety. [Work supported by NIH DK43381, DK55674, and FIRCA.]

10:45

4aPA11. Nonlinear and linear wave equations for propagation in media with frequency power law losses. Thomas L. Szabo (Boston Univ., 110 Cummington St., Boston, MA 02215, tlszabo@bu.edu)

The Burgers, KZK, and Westervelt wave equations used for simulating wave propagation in nonlinear media are based on absorption that has a quadratic dependence on frequency. Unfortunately, most lossy media, such as tissue, follow a more general frequency power law. The authors first research involved measurements of loss and dispersion associated with a modification to Blackstock's solution to the linear thermoviscous wave equation [J. Acoust. Soc. Am. **41**, 1312 (1967)]. A second paper by Blackstock [J. Acoust. Soc. Am. **77**, 2050 (1985)] showed the loss term in the Burgers equation for plane waves could be modified for other known instances of loss. The authors' work eventually led to comprehensive time-domain convolutional operators that accounted for both dispersion and general frequency power law absorption [Szabo, J. Acoust. Soc. Am. **96**, 491 (1994)]. Versions of appropriate loss terms were developed to extend the standard three nonlinear wave equations to these more general losses. Extensive experimental data has verified the predicted phase velocity dispersion for different power exponents for the linear case. Other groups are now working on methods suitable for solving wave equations numerically for these types of loss directly in the time domain for both linear and nonlinear media.

11:00

4aPA12. The profile of a weak shock wave in an unconsolidated medium. Konstantin Naugolnykh (Univ. of Colorado/Zeltech, 325 Broadway, Boulder, CO 80305, konstantin.naugolnykh@noaa.gov)

The propagation of a high intensity sound wave is determined mainly by the nonlinear and dissipative effects. If the initially sinusoidal wave is intensive, the steepness of the wave fronts increases, resulting in the occurrence of a discontinuity in each period of the wave. On the other hand, the influence of dissipative processes tends to smooth the wave profile, diminishing the gradients of velocity and temperature. Consequently, during the propagation of the intense wave its profile is forming as a result of the balance of nonlinear and dissipative effects. Many aspects of the high intensity sound wave evolution were described in the set of papers of D. T. Blackstock. The standard nonlinear medium with viscosity and thermal conductivity was considered. However, later, M. J. Buckingham has developed the theory of sound propagation in saturated marine sediments which includes a new dissipation term representing internal losses arising from interparticle contacts. The balance of nonlinearity and dissipation in such a medium has specific features that are considered in the present paper.

11:15

4aPA13. A brief history of the nonlinear acoustics of rocks. James A. TenCate (Earth and Environ. Sci., Los Alamos Natl. Lab, Los Alamos, NM 87545, tencate@lanl.gov)

Much of the early measurements done to study the nonlinearity of rocks in the late 1980s were analogs to many of the classic nonlinear acoustics experiments performed by students working under the watchful eye of David Blackstock. However, it soon became apparent that nonlinear waves in rocks did not behave as expected. Wave propagation measurements in a long waveguide (a sandstone core) did not generate the usual harmonic dependencies. The nonlinear resonance of a long thin bar of sandstone did not look at all like the nonlinear resonance of a tube of air. A host of other experiments produced equally puzzling behavior. In general, waves in the rock experience considerable nonlinear distortion, exhibit peculiar hysteresis, have memory, and have confounded researchers looking for a tidy theory to describe them. Moreover, it has recently been discovered that rocks are but one member of a larger class of materials—most all of which are granular—which all exhibit similar behavior. We describe all these experiments and how the results drove us away from classical nonlinear acoustics to new theoretical descriptions and applications. [Work supported by Office of Basic Energy Sciences, Geosciences.]

11:30

4aPA14. Evaluating prediction methods for the spectral evolution of finite-amplitude jet noise. Kent L. Gee and Victor W. Sparrow (Grad. Prog. in Acoust., The Penn State Univ., 217 Appl. Sci. Bldg., University Park, PA 16802, kentgee@psu.edu)

David Blackstock has made substantial contributions to the understanding of the propagation of broadband, finite-amplitude noise. His work with Pestorius, Theobald, Webster, and Menounou, among others, has given us an understanding of much of the underlying physics. Even so, there are still gaps in our knowledge of the noise propagation from supersonic jet flows. A recent study was undertaken to compare several methods for the far field progression of a finite-amplitude broadband spectrum. This paper will provide initial results for that comparison. For each method tested, a Gaussian spectrum and experimental jet noise data were utilized as inputs. [Work supported by Strategic Environmental Research and Development Program.]

Session 4aSA

Structural Acoustics and Vibration: Vibrations in Mechanical Systems

Jeffrey E. Boisvert, Chair
Naval Undersea Warfare Center Division Newport, Code 2133, Newport, Rhode Island 02841

Contributed Papers

9:00

4aSA1. Predicting folded beam waveguide absorber behavior using full translational and rotational degree of freedom coupling. Carl Pray, Robert Campbell, Stephen Hambric, and Andrew Munro (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804, cpray@psu.edu)

Folded beam waveguide absorbers (WGAs) have been shown to be effective low-frequency damping devices. Early WGA studies were unable to accurately predict this damping behavior. These studies used only translational degrees of freedom (DOFs), which resulted in the underestimation of the WGA damping performance. A recent study [Munro and Hambric, ''Modeling folded beam waveguide absorber behavior using translational and rotational degree of freedom frequency response function coupling,'' Proc. NOISE-CON 2003] used translational and rotational DOF frequency response functions to predict folded beam WGA behavior when attached to a thick rectangular plate, where the plate and WGA rotational DOFs were estimated using the finite-differencing method. Each plate and WGA DOF was coupled independently using frequency domain substructure synthesis (FDSS) [Jetmundsen *et al.*, ''Generalized frequency domain synthesis,'' J. Am. Helicopter Soc. 55–64, Jan (1988)], and the damping contributions due to each DOF were summed to give the total WGA damping prediction. This method gives a much improved damping estimate from previous methods but is inefficient for complex problems. In this study, all the DOFs for the plate and WGA are combined simultaneously using FDSS to predict the WGA damping behavior and plate response with folded beam WGAs attached.

9:15

4aSA2. The temporal and spatial effects of a magnetorheological elastomer in squeeze mode. Anne-Marie Albanese and Kenneth Cunefare (School of Mech. Eng., Georgia Inst. of Technol., 771 Ferst Dr., Atlanta, GA 30332)

The behavior of a magnetorheological (MR) elastomer in an adaptive tuned vibration absorber (ATVA) subjected to transient magnetic fields was examined. ATVAs suppress vibration under broadband, variable, and multiple frequency excitations, and are effective because they can rapidly change one or more parameters to retune in very short periods of time. The device examined here features a MR elastomer whose stiffness and static displacement length are both sensitive to magnetic fields; this means that the static displacement transient behavior directly corresponded to the stiffness transient behavior. A magnetic field applied at step intervals was applied to the ATVA, and the static displacement length change and flux density levels through the ATVA were measured as a function of time. For a flux density level change of 0.12 Tesla, a static displacement length change of about 0.4 mm, which amounted to a 4% change in the overall spring length, occurred in 65 milliseconds. The cause of the long transient response is believed to be due to the magnetic–mechanical coupling.

9:30

4aSA3. Prediction of damped circular cylindrical shell vibration using energy flow analysis. Jacob Klos (Structural Acoust. Branch, M.S. 463, NASA Langley Res. Ctr., Hampton, VA 23681, j.klos@larc.nasa.gov) and Robert J. Bernhard (Purdue Univ., West Lafayette, IN 47907)

Approximate energy prediction tools are gaining popularity in the field of engineering noise control because of the need to quickly predict the high frequency dynamic response of complex systems. One approximate energy prediction tool commonly used is statistical energy analysis. However, many problems encountered in the field of engineering noise control do not satisfy the assumptions of lightly damped, lightly coupled systems. A second method, energy flow analysis, has been proposed for the case of moderate damping and coupling. The application of energy flow analysis to plates and beams is well documented in the literature. However, applications of energy flow analysis to shells of curvature are limited and need to be addressed. Two approaches for modeling radial vibration of damped circular cylindrical shells using energy flow analysis are proposed and verified in this presentation. The formulations of the solutions are discussed. Qualitative and quantitative comparisons of the responses predicted by the energy flow models to analytical predictions are made. From the comparisons, it is concluded that energy flow analysis provides a valid method to model radial vibration of circular cylindrical shells subjected to a radial excitation both above and below the ring frequency of the cylinder.

9:45

4aSA4. Are the energy analysis (EA) and the statistical energy analysis (SEA) compatible? G. Maidanik and K. J. Becker (Carderock Div., Naval Surface Warfare Ctr., 9500 MacArthur Blvd., West Bethesda, MD 20817)

The statistical energy analysis (SEA) is being used commercially to find ways of achieving noise control on vehicles of various types. A tenet of SEA states that the modal energy stored in a force-driven (master) dynamic system may not exceed the modal energy stored in a passive (adjunct) dynamic system to which it is coupled. Some recent computations using an energy analysis (EA) of generic coupled dynamic systems shows that this rule may not be obeyed in (EA). To render (EA) compatible with (SEA) a lower limit on the value of the modal overlap parameter in the adjunct dynamic system must be imposed. The value must exceed a threshold for these two analyses to be compatible. [Work supported by ONR.]

10:00

4aSA5. Vibration of in-vacuo elliptic cylindrical shells. Jeffrey E. Boisvert (Naval Undersea Warfare Ctr. Div. Newport, Newport, RI 02841, boisvertje@npt.nuwc.navy.mil) and Sabih I. Hayek (Penn State Univ., University Park, PA 16802)

The equations of motion for the vibration of elliptic cylindrical shells of constant thickness were derived using a Galerkin approach. The elastic strain energy density used in this derivation has seven independent kinematic variables: three displacements, two thickness-shear, and two thickness-stretch. The resulting seven coupled algebraic equations are symmetric and positive definite. The shell has a constant thickness, h,

finite length, L, and is simply supported at its ends, $(z=0,L)$, where z is the axial coordinate. The elliptic cross-section is defined by the shape parameter, a, and the half-length of the major axis, l. The modal solutions are expanded in a doubly infinite series of comparison functions in terms of circular functions in the angular and axial coordinates. The natural frequencies and the mode shapes were obtained by the Galerkin method. Numerical results were obtained for several h/l and L/l ratios, and various shape parameters, including the limiting case of a simply supported cylindrical shell $(a=100)$. [Work supported by ONR and the Navy/ASEE Summer Faculty Program.]

10:15–10:30 Break

10:30

4aSA6. The effect of bearing properties on the eigenvalues of a rotordynamic system. Jerry H. Ginsberg and Benjamin B. Wagner (Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405, jerry.ginsberg@me.gatech.edu)

Natural frequencies and associated modal damping ratios are important in the diagnosis of rotating machinery problems. This work examines how the modal properties of a rotating shaft/disk system are affected by changes in rotation rate, lubricant viscosity, and bearing clearance. The system under analysis is a uniform, elastic, rotating shaft with a single, rigid disk concentrically mounted to the shaft away from mid-span. The shaft is supported by short-length, plain journal bearings. Standard lubrication theory is used to generate the stiffness and damping matrices for the bearings. The clearance and lubricant viscosity are independently adjustable at each bearing. A Ritz series expansion is used to generate the mass, stiffness, and gyroscopic matrices describing the shaft and disk. The combined action of the bearings and shaft/disk system is represented by stationary and rotational inertia, stiffness, and internal and external damping. A nonsymmetric generalized eigenvalue problem solver is used to calculate system eigenvalues from the system matrices over the speed range of interest. The behavior of the system is quantified in terms of dependence of the real and imaginary parts of system eigenvalues on the rotation rate. The real part of the eigenvalue is proportional to the modal damping ratio, and the imaginary part is the natural frequency. The analysis results include plots of natural frequency and damping ratio versus mode for varying bearing clearance and viscosity.

10:45

4aSA7. Effect of modal overlap factor on the ray tracing analysis of the curved beam structure: A computational study. Cheol-Ho Jeong and Jeong-Guon Ih (NOVIC, Dept. of Mech. Eng., Korea Adv. Inst. of Sci. and Technol., KAIST, Daejeon, Korea)

Ray tracing method (RTM) of the geometrical acoustics area is recently known to be useful for the evaluation of spatial distribution of energy density and power flow of vibrating structures at high frequencies. In this study, the spatial distribution of vibrational energy and power flow of curved beam and its connected structures are of interest. Longitudinal, flexural, and torsional waves are considered using the Euler–Bernoulli beam theory. In the analysis of spatial distribution, the band-averaged RTM result varies smoothly in space without local fluctuations. This means that the RTM result can represent a similar tendency to the traveling wave solution. An accuracy of the analysis depends on the modal overlap factor (MOF), which is the function of the length of the beam, frequency range, and structural loss factor. The effect of changing each parameter is investigated and discussed. Similar to other high frequency methods such as statistical energy analysis and power flow analysis, the results become close to the traveling wave solutions as the modal overlap factor increases. [Work supported by the BK21 project and NRL.]

11:00

4aSA8. Sonic inspection of concrete bridge decks. R. Daniel Costley, Gary Boudreaux (Miltec Missiles and Space Co.), and William Gene Ramsey (DIAL, Mississippi State Univ.)

One technique for determining the integrity of concrete structures, such as bridge decks, involves dragging a chain across it and listening to the audible response. A distinctive, hollow sound is produced when a chain is dragged over a section of concrete containing a delamination. This technique has been automated by recording the sound produced by a dragging chain with a suitable microphone and processing these signals with a minicomputer to distinguish between ''good'' and ''bad'' sections of concrete. The equipment is mounted on a hand-pushed cart with chains attached so that they drag along the surface of the deck. In addition, the microphone is mounted in such a way, using standard noise control techniques, so that external noise is minimized. Traffic noise is filtered electronically. These improvements make the technique operator independent and allow inspections to be made in noisy environments. Another advantage is that this approach produces an objective record of the inspection, available both electronically and in hardcopy. These records can be compared to past and future inspections, allowing the inspectors to monitor the health of the structure. Results from bridge deck inspections will be presented, along with a description of the device and the signal processing techniques.

11:15

4aSA9. Landmine vibration modes. Dimitri Donskoy, Andrei Zagrai, and Alexander Ekimov (Stevens Inst. of Technol., Davidson Lab., Hoboken, NJ 07030, ddonskoy@stevens-tech.edu)

Recent experimental and theoretical studies in the field of seismo-acoustic landmine detection proved high potential of this technique. Investigations have also demonstrated that acoustic detection and discrimination of landmines is a complex problem dependent upon interaction between soil and buried mines and their respective properties. Vibration characteristics of a mines casing play a critical role in this problem. Our recent tests, for the first time, revealed the multi-modal vibrations of mines. It was observed that the resonance frequencies of the soil/mine system depend on the burial depth and soil conditions. The present study provides analysis of this phenomenon both experimentally and though physical modeling. The experimental investigations were focused on the effects of burial depth and soil moisture content on vibration response measured on the soil surface above the buried mine. The modeling efforts were based on the electro-mechanical analogy approach where each vibration mode was represented as an oscillator with its own effective parameters. The validity of this consideration was confirmed through comparison against conventional modal analysis approach. The study demonstrated good agreement between the developed model and experimental results.

11:30

4aSA10. Calculation of vibration reduction design for underground sources. Stanislav A. Kostarev (Lab. of Acoust. and Vib. Tunnel Assoc., 21 Sadovo-Spasskaya Str., Moscow 107217, Russia), Samuil A. Rybak (N. N. Andreev Acoust. Inst., Moscow 117036, Russia), and Sergey A. Makhortykh (Inst. of Mathematical Problems of Biol., RAS, Pushchino Moscow reg. 142290, Russia)

The problems of ecology situation control near intensive underground acoustical sources is considered. Studied vibration absorbers are modeled by multi-component system of connected oscillators with damping. Obtained oscillatory equation system was investigated numerically. Frequency dependencies of absorbers efficiency have been calculated. An influence of the physical-mechanical parameters of the surrounding ground on the value of vibration reduction has been determined. Some variants of principal realization of vibration absorbers for the case of underground railway are discussed. [Work supported by Russian Foundation for Basic Researches Grants Nos. 01-02-16127, 02-02-17143.]

Session 4aSC

Speech Communication: Issues in Similarity and Distinctiveness (Lecture/Poster Session)

Lynne E. Bernstein, Cochair
Department of Communication Neurobiology, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

Patricia A. Keating, Cochair
Department of Linguistics, University of California–Los Angeles, Los Angeles, California 90095-1543

Chair's Introduction—8:00

Invited Papers

8:05

4aSC1. Fundamentals of spoken word recognition. Paul A. Luce and Conor McLennan (Lang. Percept. Lab., Dept. of Psych. and Ctr. for Cognit. Sci., Univ. at Buffalo, Buffalo, NY 14260)

Researchers have made significant progress in identifying the basic principles responsible for the normal-listener's rapid and accurate identification of spoken words. In particular, there is now almost uniform consensus that spoken word recognition involves two fundamental processes: *activation* and *competition*. Most current models of recognition propose that stimulus input (i.e., a spoken word) activates a set of representations of *similar* sounding words in memory that subsequently vie for recognition. Despite the fact that *similarity* is afforded a crucial role in activating and discriminating among lexical competitors, we currently have little precise information regarding *perceived* similarity relations among spoken words. We will discuss the fundamental role of similarity in current models of spoken word recognition, past attempts to capture perceived similarity among spoken words, and ongoing efforts in our laboratory to understand more deeply the precise role of similarity in the perception of spoken stimuli. [Work supported by NIH.]

8:30

4aSC2. Two perspectives on similarity between words. Stefan A. Frisch (Dept. of CSD, Univ. of South Florida, 4202 E. Fowler Ave. PCD1017, Tampa, FL 33620, frisch@chuma1.cas.usf.edu)

This presentation examines the similarity between words from both bottom up (phonetic) and top down (phonological/psycholinguistic) perspectives. From the phonological perspective, the influence of structure on similarity is explored using metalinguistic acceptability judgments for multisyllabic nonwords. Results from an experiment suggest that subjects try to align novel words with known words in order to maximize similarities while minimizing dissimilarities. This finding parallels results from psychology on similarity judgments for visual scenes. From the phonetic perspective, the influence of similar gestures on speech error rates is examined using ultrasound measurement of tongue position. In a pilot experiment, subjects, produced tongue twisters containing words where onset and vowel phonemes had similar gestures (e.g., tip, comb) and where the onset and vowel had dissimilar gestures (e.g., tube, keep). Preliminary results suggest that misarticulations are more frequent in the context of dissimilar gestures (e.g., in the tongue twister tip cape keep tape, error rates are higher for /k/ than /t/). These errors appear to be gestural interactions rather than errors at the phonemic or featural level of phonological spellout. Together, these two experiments indicate that similarity relations between words are found at multiple levels, any which are potentially relevant to the structure of phonological systems.

8:55

4aSC3. Stimulus-based similarity and the recognition of spoken words. Edward T. Auer, Jr. (Dept. of Commun. Neurosci., House Ear Inst., 2100 W. Third St., Los Angeles, CA 90057, eauer@hei.org)

Spoken word recognition has been hypothesized to be achieved via a competitive process amongst perceptually similar lexical candidates in the mental lexicon. In this process, lexical candidates are activated as a function of their perceived similarity to the spoken stimulus. The evidence supporting this hypothesis has largely come from studies of auditory word recognition. In this talk, evidence from our studies of visual spoken word recognition will be reviewed. Visual speech provides the opportunity to highlight the importance of stimulus-driven perceptual similarity because it presents a different pattern of segmental similarity than is afforded by auditory speech degraded by noise. Our results are consistent with stimulus-driven activation followed by competition as general spoken word recognition mechanism. In addition, results will be presented from recent investigations of the direct prediction of perceptual similarity from measurements of spoken stimuli. High levels of correlation have been observed between the predicted and perceptually obtained distances for a large set of spoken consonants. These results support the hypothesis that the perceptual structure of English consonants and vowels is predicted by stimulus structure without the need for an intervening level of abstract linguistic representation. [Research supported by NSF IIS 9996088 and NIH DC04856.]

9:20

4aSC4. Perceptual similarity co-existing with lexical dissimilarity. Andrea Weber (Dept. of Psycholinguist., Saarland Univ., 66041 Saarbruecken, Germany, aweber@coli.uni-sb.de) and Anne Cutler (Max Planck Inst. for Psycholinguist., Wundtlaan 1, 6525 XD Nijmegen, The Netherlands)

The extreme case of perceptual similarity is indiscriminability, as when two second-language phonemes map to a single native category. An example is the English had-head vowel contrast for Dutch listeners; Dutch has just one such central vowel, transcribed [E]. We examine whether the failure to discriminate in phonetic categorization implies indiscriminability in other—e.g., lexical—processing. Eyetracking experiments show that Dutch-native listeners instructed in English to ''click on the panda'' look (significantly more than native listeners) at a pictured pencil, suggesting that pan- activates their lexical representation of pencil. The reverse, however, is not the case: ''click on the pencil'' does not induce looks to a panda, suggesting that pen- does not activate panda in the lexicon. Thus prelexically undiscriminated second-language distinctions can nevertheless be maintained in stored lexical representations. The problem of mapping a resulting unitary input to two distinct categories in lexical representations is solved by allowing input to activate only one second-language category. For Dutch listeners to English, this is English [E], as a result of which no vowels in the signal ever map to words containing [ae]. We suggest that the choice of category is here motivated by a more abstract, phonemic, metric of similarity.

9:45–9:55 Break

9:55

4aSC5. The development of sensitivity to speech sound dissimilarity in humans. Curtis Ponton (Neuroscan, 7850 Paseo del Norte, El Paso, TX 79912)

The understanding of spoken language development has been expanded by examining speech-sound evoked brain activity recorded using an oddball presentation paradigm. When brain activity is recorded using this paradigm, a response known as the mismatch negativity (MMN) is generated. The MMN is regarded as a neurophysiological correlate of short-term auditory memory processes that are necessary for behavioral discrimination. Numerous investigations have demonstrated that neural generators underlying the MMN are sensitive to a wide range of acoustic contrasts, including both nonspeech and speech sounds. Recent studies have shown that for at least some speech sound contrasts, the MMN is present at birth and is insensitive to native/non-native speech contrasts. The appearance of language-specific contrasts around six months of age appears to correspond with the emergence of myelinated connections between the thalamus and auditory cortex. Combined, the neurophysiological and anatomical data suggest that this neural mechanism for detecting dissimilarity may have subcortical or cortical loci, depending on the speech contrast. Particularly for vowels, this dissimilarity processing may be the basis for developing categorization.

10:20

4aSC6. Neural and perceptual discrimination of the spectral and temporal modulations in birdsong and speech. Frederic Theunissen (Dept. of Psych., Univ. of California, 3210 Tolman, Berkeley, CA 94720-1650)

My laboratory is interested in the neural basis of complex sound perception, including vocalizations. Our neural studies have focused on the high-level auditory system of songbirds where neurons respond preferentially to birdsong. We quantified the selectivity of these auditory neurons by recording their responses to degraded versions of song. To do so, we added noise to this natural stimulus, affecting the time-frequency structure of the sound at different scales. We found that at particular time-frequency scales the noise has little effect on the response of the neurons, while at other scales it greatly reduces the neural response. We correlated the tuning of the neurons with the statistical structure in the birdsongs that allows the discrimination of songs produced by different birds. We found a good match between the optimal scale of the neurons sensitivity curve and the scale that best captures the variance in the acoustical structure of an ensemble of songs. We have done a similar analysis with speech, where the neural tuning curve in songbirds is replaced by a human perceptual tuning curve based on speech intelligibility. As in songbirds, the scale for optimal speech perception is also best for capturing the statistical structure in the speech signal.

10:45–11:00
Panel Discussion

Contributed Papers

All posters will be on display from 8:00 a.m. to 12:00 noon. Contributors will be at their posters from 11:00 a.m. to 12:00 noon.

4aSC7. Enhancement of distinctiveness in strident fricatives: An example from Mandarin Chinese. Chao-Yang Lee (School of Hearing, Speech and Lang. Sci., Ohio Univ., Athens, OH 45701) and Kenneth N. Stevens (Res. Lab. of Elec., MIT, Cambridge, MA 02139)

Strident fricatives in Mandarin may be followed by fricative vowels, which are made with the tongue body in essentially the same position as in the corresponding fricative (Ladefoged and Maddieson, 1996). Specifically, the vowel [i] may occur only after the palatalized fricative, the apical vowel after the alveolar fricative, and the retroflex vowel after the palatoalveolar fricative. Given the complementary distribution, the three phonetically distinct vowels are generally considered allophones. Under the view that the implementation of discretely specified distinctive features may be enhanced by gradient, non-contrastive articulatory gestures to maintain perceptual distinctiveness (Keyser and Stevens, 2003), it is proposed that the three vowels are derived from an underlying [i] and are modified to enhance the defining articulatory and acoustic attributes of the fricatives. Acoustic data from six speakers indicate that the defining acoustic attributes for the fricatives, i.e., spectral prominences in discrete frequency ranges, are associated with distinct tongue configurations. The articulatory and acoustic properties of the palatalized fricative is most

compatible with those of [i], whereas the tongue body backing gesture in the alveolar and palatoalveolar fricatives are consistent with lowered $F2$ frequency in the apical and retroflex vowels. [Work supported by NIH.]

4aSC8. The importance of duration in the perception of glides and vowels. Amy E. Coren (Dept. of Psych., Univ. of Texas, SEA 4.212, 108 E. Dean Keeton St., Austin, TX 78712-0198, aecoren@mail.utexas.edu) and Willis J. Warren (Univ. of Texas, Austin, TX 78712-0198)

The segmental duration for the perception of glides versus vowels was investigated. A central Texas Spanish speaker produced a single token of radio with a high vowel duration of 150 ms. The duration of the high vowel, [i], was edited at the midpoint in 10-s intervals to see at what length listeners perceive a complete vowel versus a glide, i.e., radyo versus radio. Ten listeners judged ten repitions of the ten tokens. It was determined that the steady state of the vowel must retain somewhere near 80 ms to be perceived as a full vowel.

4aSC9. Changes in $F2-F1$ as a voicing cue. Willis J. Warren (Dept. of Linguist., Univ. of Texas, Calhoun Hall 501, 1 University Station B5100, Austin, TX 78712-0198, warisill@mail.utexas.edu) and Amy E. Coren (Univ. of Texas, Austin, TX 78712-0198)

The interaction between formant transitions and vowel length was measured with respect to syllable final voicing distinctions. A synthesized *ad* VC token of 360 ms was edited in 5-ms intervals from either side, onset or offset, so that 260 ms were preserved. Ten subjects were asked to make final voicing judgments for the words ''odd'' and ''ought'' ([ad] vs [at]) when hearing the 20 edited tokens. Each token was presented five times, randomly, for a total of 1000 judgements. Results showed an overwhelming number of voiced responses when the entire offset was preserved and symmetrical voiceless results with the deletion of offset. A follow-up experiment utilized a similarly synthesized token of 460 ms. The results when adding 100 ms onto the vowel were insignificantly different than the results acquired for formant transitions, suggesting the latter are a more important cue for syllable final voicing distinctions. These findings contradict previous vowel length conclusions [L. J. Raphael, J. Acoust. Soc. Am. **51**, 1296–1303 (1972)] and further suggest that in addition to $F1$ [V. Summers, J. Acoust. Soc. Am. **84**, 485–492 (1988)], $F2$ transitions are also an important cue to final voicing distinctions in low vowel contexts.

4aSC10. Effects of blocking and presentation on the recognition of word and nonsense syllables in noise. José R. Benkí (Dept. of Linguist., Univ. of Michigan, Ann Arbor, MI 48109-1285, benki@umich.edu)

Listener expectations may have significant effects on spoken word recognition, modulating word similarity effects from the lexicon. This study investigates the effect of blocking by lexical status on the recognition of word and nonsense syllables in noise. 240 phonemically matched word and nonsense CVC syllables [Boothroyd and Nittrouer, J. Acoust. Soc. Am. **84**, 101–108 (1988)] were presented to listeners at different S/N ratios for identification. In the mixed condition, listeners were presented with blocks containing both words and nonwords, while listeners in the blocked condition were presented with the trials in blocks containing either words or nonwords. The targets were presented in isolation with 50 ms of preceding and following noise. Preliminary results indicate no effect of blocking on accuracy for either word or nonsense syllables; results from neighborhood density analyses will be presented. Consistent with previous studies, a j-factor analysis indicates that words are perceived as containing at least 0.5 fewer independent units than nonwords in both conditions. Relative to previous work on syllables presented in a frame sentence [Benkí, J. Acoust. Soc. Am. **113**, 1689–1705 (2003)], initial consonants were perceived significantly less accurately, while vowels and final consonants were perceived at comparable rates.

4aSC11. Influence of semantic similarity on spoken word recognition. Jonna L. Armbruster and Michael S. Vitevitch (Dept. of Psych., 1415 Jayhawk Blvd., Lawrence, KS 66045)

Previous research has shown that the number of phonologically similar items influences the processing of spoken words (e.g., Vitevitch, 2002; Vitevitch and Luce, 1998). The present experiment examines the influence of semantic similarity on the speed and accuracy of spoken word recognition using a lexical decision task. Semantic density refers to the number of words that are semantically associated to a given word (Nelson, McEvoy, and Schreiber, 1998). A word that has relatively many semantically associated words has high semantic density and a word that has relatively few semantically associated words has low semantic density. The results showed that words with high semantic density were responded to more quickly and accurately than words with low semantic density, suggesting that semantic information influences spoken word recognition.

4aSC12. The impact of phonetic dissimilarity on the perception of foreign accented speech. Shawn A. Weil (Sytronics, Inc., Dayton, OH and Ohio State Univ., Columbus, OH)

Non-normative speech (i.e., synthetic speech, pathological speech, foreign accented speech) is more difficult to process for native listeners than is normative speech. Does perceptual dissimilarity affect only intelligibility, or are there other costs to processing? The current series of experiments investigates both the intelligibility and time course of foreign accented speech (FAS) perception. Native English listeners heard single English words spoken by both native English speakers and non-native speakers (Mandarin or Russian). Words were chosen based on the similarity between the phonetic inventories of the respective languages. Three experimental designs were used: a cross-modal matching task, a word repetition (shadowing) task, and two subjective ratings tasks which measured impressions of accentedness and effortfulness. The results replicate previous investigations that have found that FAS significantly lowers word intelligibility. Furthermore, in FAS as well as perceptual effort, in the word repetition task, correct responses are slower to accented words than to nonaccented words. An analysis indicates that both intelligibility and reaction time are, in part, functions of the similarity between the talker's utterance and the listener's representation of the word.

Session 4aSP

Signal Processing in Acoustics, Underwater Acoustics, Speech Communication, Animal Bioacoustics, Noise and Engineering Acoustics: Detection and Classification in Acoustics III

Paul M. Baggenstoss, Chair
Naval Undersea Warfare Center, Newport, Rhode Island 02840

Invited Papers

8:00

4aSP1. Audio signal recognition for speech, music, and environmental sounds. Daniel P. W. Ellis (Dept. of Elec. Eng., Columbia Univ., 500 W. 120th St., New York, NY 10027, dpwe@ee.columbia.edu)

Human listeners are very good at all kinds of sound detection and identification tasks, from understanding heavily accented speech to noticing a ringing phone underneath music playing at full blast. Efforts to duplicate these abilities on computer have been particularly intense in the area of speech recognition, and it is instructive to review which approaches have proved most powerful, and which major problems still remain. The features and models developed for speech have found applications in other audio recognition tasks, including musical signal analysis, and the problems of analyzing the general ''ambient'' audio that might be encountered by an auditorily endowed robot. This talk will briefly review statistical pattern recognition for audio signals, giving examples in several of these domains. Particular emphasis will be given to common aspects and lessons learned.

8:45

4aSP2. Recognition of information-bearing elements in speech. Hynek Hermansky (Institut Dalle Molle d'Intelligence Artificielle Perceptive, Rue du Simplon 4, Case Postale 592, CH-1920 Martigny, Switzerland and Intl. Computer Sci. Inst., Berkeley, CA, hynek.hermansky@idiap.ch)

An acoustic speech signal carries many different kinds of information: the basic linguistic message, many characteristics of the speaker of the message, details of the environment in which the message was produced and transmitted, etc. The human auditory/cognitive system is able to detect, decode, and separate all these information sources. Understanding this ability and emulating it on a machine has been an important but elusive scientific and engineering goal for a long time. This talk critically surveys the situation in the speech recognition field. It puts automatic recognition of speech in perspective with other acoustic signal detection and classification tasks, reviews some historical, contemporary, and evolving techniques for machine recognition of speech, critically compares competing techniques, and gives some examples of applications in speech, speaker, and language recognition and identification. The talk is intended for an audience interested but not directly involved in the processing of speech.

9:30

4aSP3. Acoustic classification: An overview of theory and reality. Stephen Greineder (Naval Undersea Warfare Ctr. Div. Newport, 1176 Howell St., Newport, RI 02841, GreinederSG@Npt.NUWC.Navy.Mil)

Historically, Bayes' decision theory has formed the statistical foundation for the development of acoustic classification techniques for problems that contain large numbers of data classes. The successful realization of this approach depends directly on mitigating two important mismatches that are introduced using this classification model. First, the underlying probabilistic structure in the problem is not typically known and therefore must be estimated as part of a classifier training phase. Second, due to the high dimensionality of time series data, a feature extraction stage is added as a preprocessor to most classifiers. Attempting to extract the sufficient information in the problem while limiting the feature set size to a dimension supported by the training data is the ultimate designer's challenge. This is particularly true in the case of finite data sample sizes. This talk will trace the development evolution of a low false alarm rate multiclass operational algorithm that will provide an overview of the fundamental challenges and resulting mitigation approaches. Examples using real data will be used to demonstrate the points being presented.

10:15–10:30 Break

10:30

4aSP4. The class-specific method for classification. Paul M. Baggenstoss (Naval Undersea Warfare Ctr., Newport, RI 02840)

This talk describes a new probabilistic method for classification called the ''class-specific method'' (CSM). CSM is able to avoid the ''curse of dimensionality'' which plagues most classifiers which attempt to determine the decision boundaries in a high-dimensional feature space. Using CSM, it is possible to build a theoretically optimum classifier without a common feature space. Separate low-dimensional features sets may be defined for each class, while the decision functions are projected back to the common *raw data* space. CSM effectively extends classical classification theory to handle multiple feature spaces. It is completely general, and

requires no simplifying assumption such as Gaussianity or that data lies in linear subspaces. In real-data problems, CSM has shown orders of magnitude reductions in the false-alarm rate. CSM achieves this gain because it is able to make use of partial prior knowledge about the data classes. In contrast, the existing theory can only make use of full knowledge—that is when the parametric forms of the data probability density functions (PDFs) are known.

Contributed Papers

11:00

4aSP5. Automatic computational models of acoustical category features: Talking versus singing. David Gerhard (Dept. of Computer Sci., Univ. of Regina, 3737 Wascana Pkwy., Regina, SK S4S 0A2, Canada, david.gerhard@uregina.ca)

The automatic discrimination between acoustical categories has been an increasingly interesting problem in the fields of computer listening, multimedia databases, and music information retrieval. A system is presented which automatically generates classification models, given a set of destination classes and a set of *a priori* labeled acoustic events. Computational models are created using comparative probability density estimations. For the specific example presented, the destination classes are talking and singing. Individual feature models are evaluated using two measures: The Kologorov–Smirnov distance measures feature separation, and accuracy is measured using absolute and relative metrics. The system automatically segments the event set into a user-defined number (n) of development subsets, and runs a development cycle for each set, generating n separate systems, each of which is evaluated using the above metrics to improve overall system accuracy and to reduce inherent data skew from any one development subset. Multiple features for the same acoustical categories are then compared for underlying feature overlap using cross-correlation. Advantages of automated computational models include improved system development and testing, shortened development cycle, and automation of common system evaluation tasks. Numerical results are presented relating to the talking/singing classification problem.

11:15

4aSP6. Simple algorithm for classification of target in shallow water. Angie Sarkissian (Naval Res. Lab., Washington, DC 20375)

When applying classification algorithms to the scattering response of a target in shallow water, algorithms that attempt to deconvolve the free field target response from the environment, such as modal decomposition or time reversal, typically require multiple receivers. A simple algorithm is applied here that requires a single source and a single receiver. The impulse response of the target is Fourier transformed to the frequency domain; the magnitude of the frequency domain response is transformed back to the time domain; the resultant time domain response is truncated. The algorithm is applied to a hemispherically end-capped cylindrical shell under various shallow water conditions. The response obtained for the target in shallow water is correlated against the target free field response. [Work supported by ONR.]

THURSDAY MORNING, 13 NOVEMBER 2003 SABINE ROOM, 8:00 TO 11:40 A.M.

Session 4aUWa

Underwater Acoustics, Engineering Acoustics and Signal Processing in Acoustics: Gradient Array Acoustics I

Paul C. Hines, Cochair
Defence R&D Canada Atlantic, P.O. Box 1012, Dartmouth, Nova Scotia B2Y 3Z7, Canada

Daniel L. Hutt, Cochair
Defence R&D Canada Atlantic, P.O. Box 1012, Dartmouth, Nova Scotia B2Y 3Z7, Canada

Chair's Introduction—8:00

Invited Papers

8:05

4aUWa1. Superdirective and gradient sensor arrays. Harold M. Merklinger (Defence R&D Canada Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada)

During the late 1960s and the 1970s, underwater acoustic investigators examined superdirective and gradient sensor systems in order to enhance submarine detection capabilities for surface ships and maritime aircraft. Simple gradient processing had already been used in both in-air acoustic systems (cardioid and super-cardioid microphones) as well as radio and radar applications. Superdirective techniques were known [R. L. Pritchard, J. Acoust. Soc. Am. **25**, 879 (1953)] and sometimes exploited in radar systems. It was quickly demonstrated that simple gradient sensors and modest degrees of superdirective array processing were possible, although self-noise and the ability to calibrate hydrophones limited the processing gains achievable. Circular superdirective arrays were used extensively by the Defence Research Establishment Atlantic for noise directionality measurements in the frequency range 4 Hz to about 1 kHz and considered for naval ASW applications until the superiority of oil-filled conventional arrays became apparent. Nevertheless, the significant theoretical and practical development of spatial harmonic beamforming and direction finding was completed. Although much of this work was not considered classified, neither was it widely published. This presentation will review the concepts developed and progress made. Beamforming, noise mitigation and calibration issues are covered.

8:30

4aUWa2. Differential and gradient microphone arrays. Gary W. Elko (Avaya Labs, gwe@avaya.com), James E. West (Johns Hopkins Univ.), and Steve Thompson (Knowles Electron. LLC)

Differential microphone arrays have been in existence for more than 7 decades and are the basis of most commercial directional microphones in use today. These microphones obtain directionality by combining the acoustic pressure and the pressure-difference to form what is termed a first-order differential microphone. Differential microphones are inherently superdirectional since they can obtain broadband directional gains of up to 6.0 dB in an array that is physically much smaller than the acoustic wavelength. Differential arrays constructed by subtracting omnidirectional microphones are inherently more flexible in that the directional response can be easily and continuously varied from omnidirectional to hypercardioid. The simultaneous measurement of the acoustic pressure and particle velocity allows one to estimate the complex acoustic intensity along the axis of a microphone pair. A measure of the complex acoustic intensity vector can be obtained using a minimum of four pressure-sensing microphones. Higher-order differential microphones are also possible by using more microphone elements, but the problems of microphone calibration and signal-to-noise combine to practically realize microphones of differential order greater than third order. We will present some of the history of differential microphone array design and discuss some applications related to hands-free communication, hearing aids, and spatial audio recording.

Contributed Papers

8:55

4aUWa3. Performance of a superdirective line array in nonideal environments. Paul C. Hines (Defence R&D Canada Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada), Victor F. Humphrey (Dept. of Phys., Univ. of Bath, Bath BA2 7AY, UK), and Victor Young (Defence R&D Canada Atlantic, Dartmouth, NS B2Y 3Z7, Canada)

Superdirective line arrays can provide high gains relative to their dimensions, whenever the inter-element spacing is much less than half a wavelength. However, their performance can be degraded by system noise. System noise can result from limitations in electronic components, inter-element mismatch in gain or phase, or from scatter from array components. Using state-of-the-art electronics and digital signal processing one can drastically reduce the errors due to electronic noise as well as those due to gain and phase mismatch. Thus, acoustic scatter from the array components can determine the performance limit. Since theoretical developments typically assume plane waves incident on idealized point receivers, it is not all that surprising that array performance fails to meet theoretical expectations! This is especially true if the array is merely one component within a much larger system. Nonetheless, impressive gains can still be realized from superdirective arrays even with significant departures from the idealized model. To support this observation, we present results obtained using a 0.8 m long, 5th order superdirective receiver that is part of a much larger system used to study scattering from marine sediments.

9:10

4aUWa4. Experimental performance analysis of a superdirective line array. Victor F. Humphrey (Dept. of Phys., Univ. of Bath, Bath BA2 7AY, UK, v.f.humphrey@bath.ac.uk), Paul C. Hines, and Victor Young (Defence R&D Canada Atlantic, Dartmouth, NS B2Y 3Z7, Canada)

Superdirective line arrays can provide a significant array gain from a structure that is relatively small in terms of acoustic wavelengths. However, system imperfections, electronic noise and acoustic scatter from the array structure can degrade their performance. An acoustic calibration of a six-element line array, 0.8 m in length, has been performed over the frequency range 1 to 4 kHz in order to investigate the performance of a real array. The data is used to identify the angular variation of the hydrophone outputs and the phase difference between hydrophone pairs. These angular responses are analyzed in terms of a modal series in order to quantify the variations and help identify the source of perturbations. The effects of imperfections are also investigated by synthesising superdirective arrays of order 1 to 5 and monitoring how the array gain varies for both deterministic signals and ambient acoustic noise. These results are compared with theoretical predictions. Further evidence of the variation in performance is gained by comparing the output of different implementations of lower order arrays, synthesized from subsets of the full array. The results indicate the influences that the array structure may have on the performance of the array.

9:25

4aUWa5. Acoustic intensity in the interaction region of a parametric source. G. C. Lauchle, T. B. Gabrielson, D. J. Van Tol, N. F. Kottke (Appl. Res. Lab and Grad. Prog. in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804), and J. A. McConnell (Acoustech Corp., State College, PA 16804)

The goal of this project was to measure acoustic intensity in the strong interaction region of a parametric source in order to obtain a clear definition of the source-generation region and to separate the local generation (the reactive field) from propagation (the real or active field). The acoustic intensity vector was mapped in the interaction region of a parametric projector at Lake Seneca. The source was driven with primary signals at 22 kHz and 27 kHz. Receiving sensors were located 8.5 meters from the projector. At that range, the secondary at 5 kHz was between 40 and 45 dB below either primary. For the primary levels used, the plane-wave shock inception distance would have been at least 14 meters. Furthermore, the Rayleigh distance for the projector was about 4 meters so the measurements at 8.5 meters were in the strong interaction region but not in saturation. Absorption was negligible over these ranges. The intensity measurements were made at fixed range but varying azimuth angle and varying depth thus developing a two-dimensional cross-section of the secondary beam. Measurements of both the active and reactive intensity vectors will be presented along with a discussion of measurement error. [Work supported by ONR Code 321SS.]

9:40

4aUWa6. Performance of vector sensors in noise. Henry Cox (Lockheed Martin Orincon, 4350 N. Fairfax Dr., Arlington, VA 22203) and Arthur Baggeroer (MIT, Cambridge, MA 02139)

Vector sensors are super gain devices that can provide ''array gain'' against ocean noise with a point sensor. As supergain devices they have increased sensitivity to nonacoustic noise components. This paper reviews and summarizes the processing gain that is achievable in various noise fields. Comparisons are made with an omni-directional sensor and with the correlation of a pair of closely spaced omni-directional sensors. Total processing gain that consists of both spatial and temporal gain is considered so that a proper analysis and interpretation of multiplicative processing can be made. The performance of ''intensity sensors'' (pressure times velocity) that are obtained by multiplying the omnidirectional component with a co-located dipole is also considered. A misinterpretation, that is common in the literature, concerning the performance of intensity sensors is discussed. The adaptive cardioid processing of vector sensors is also reviewed.

9:55–10:10 Break

10:10

4aUWa7. Comparison of the performance of vector and sensors using optimum array processing. Arthur Baggeroer (MIT, Cambridge, MA 02139) and Henry Cox (Lockheed Martin Orincon, Arlington, VA 22203)

Vector sensors in sonar are often used for direction finding. With the appropriate processing three component vector sensors are also quite effective at nulling directional interference as well as estimating the ambient acoustic intensity vector. Here we examine optimum array processing (MVDR based) methods for both these applications. We first present the detection performance of a plane wave source operating in directional noise fields as well as noise field based a spheroidial harmonic expansion for modeling diffuse surface, seabed and ducted environments. Next, we use these optimal methods for estimating the intensity vector. Finally, we compare the performance to an array of pressure sensors using the same number of output channels. This indicates that vector sensors are effective against some types of noise fields but not uniformly better than an array of scalar sensors.

10:25

4aUWa8. Highly directional receivers using various combinations of scalar, vector, and dyadic sensors. James A. McConnell (Acoustech Corp., P.O. Box 139, State College, PA 16804)

The generalized theory of directional sensors is presented in the form of the Taylor series expansion of the acoustic pressure about a point in space. If the expansion is truncated to second order, the analysis of scalar, vector, and dyadic sensors can be made and corresponds to the zeroth-, first-, and second-order gradient of the acoustic pressure. This translates into using a sufficient number of omni-directional hydrophones or a multimode hydrophone in conjunction with the appropriate finite-differencing operations to achieve the desired beam patterns. Using the linearized Euler equation, the formulation can be recast in terms of the zeroth-order gradient of the acoustic pressure along with the zeroth- and first-order gradient of the particle acceleration. In this case, the zeroth-order terms can be measured directly with an omni-directional hydrophone and a neutrally buoyant accelerometer. The gradient of the particle acceleration can be measured indirectly using finite differences or directly by measuring the angular acceleration akin to a Rayleigh disk. Of particular interest is the use of scalar, vector, and dyadic sensors to localize sources of sound using arc-tangent-squared processing and cardioid-squared processing as opposed to conventional arc-tangent and cardioid processing. [Work supported by ONR 321SS.]

10:40

4aUWa9. Practical application of a tri-axial intensity array. Victor W. Young, Paul C. Hines, Daniel L. Hutt, and Victor F. Humphrey (DRDC Atlantic, P.O. Box 1012, Dartmouth, NS, Canada, victor.young@drdcrddc.gc.ca)

Sound intensity is a vector quantity representing the magnitude and direction of propagating energy within an acoustic field. In an underwater environment, a single omni-directional hydrophone can be used to measure instantaneous acoustic pressure and a finite difference approximation applied to the pressure signals from a pair of such hydrophones can be used to calculate particle velocity in a single direction. Because the time average of the product of instantaneous pressure and particle velocity is intensity, a pair of hydrophones is all that is required to measure a single component of the intensity vector. The complete three-dimensional intensity vector can be calculated using three orthogonal pairs of hydrophones. To evaluate this concept a tri-axial array consisting of three orthogonal pairs of omni-directional hydrophones has been developed and tested on both calibrated sources at a laboratory facility and sources of opportunity during sea trails in littoral waters. The use of this array to calculate the intensity vector and thereby localize both near-field and far-field acoustic sources and characterize the directionality of ambient noise fields will be discussed. The impact of signal-to-noise ratio and the effect of self-noise will also be examined.

10:55

4aUWa10. Cardioid processsing of reverberation data observed on the STRATAFORM with the Five Octave Research Array (FORA). John R. Preston (ARL, Penn State Univ., P.O. Box 30, State College, PA 16804)

The author recently participated in ONR's 2003 Geoclutter Experiment to study shallow water bottom reverberation and clutter in the STRATAFORM off New Jersey. The experimental effort was lead by M.I.T. and included researchers from the Naval Research Laboratory, Penn State Univ., and the Naval Underwater Weapons Center. Sources were bistatic coherent pulses from 400 to 3600 Hz. The receiver was the new Five Octave Research Array (FORA). The STRATAFORM is known to have benign surface morphology but contains many buried river channels and other sub-surface horizons. Some highlights of the reverberant returns are discussed that include the correlation of returns with a bottom mounted target and probable fish schools. The main objective of this work is to assess the directional characterization of the observed clutter. The cardioid aperture of FORA should, in theory, yield good estimates of the sources of reverberation above ~1800 Hz. Examples from the rverberation data analysis are presented using a cardioid beamforming algorithm developed by SACLANTCEN. [Work supported by ONR Code 32, Grant N00014-03-1-0113.]

11:10

4aUWa11. Separation of signals from near field noise via pressure and velocity field measurements. Peter R. Stepanishen (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882-1197)

The separation of acoustic plane wave signals from near field noise sources is addressed using both pressure and velocity measurements at the same location in the fluid. A compact broadband noise source at a known location is first assumed to generate the noise field. Using a time domain approach a simple signal processor is presented to separate the unknown signal from the noise. Effects of sensor and quantization noise are included in the analysis. A frequency domain analysis of the noise separation approach is presented along with numerical results to illustrate the sensitivity of the method to field point location, noise spectrum, and sensor and quantization noise. The separation of plane wave signals from spatially distributed near field noise sources is then addressed. A signal processor using both pressure and velocity measurements at a single point is developed to separate a plane wave signal from a spectrally pure evanescent noise field generated at a planar interface close to the sensors. Numerical results are presented for several examples to illustrate the noise separation process for a wide range of signals. In particular, the case of overlapping signal and noise spectrums is addressed.

11:25

4aUWa12. Wideband direction finding via shielded gradient beamspace techniques. Terry J. Brudner and Terry L. Henderson (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713-8029)

Monopulse techniques have been used for over 50 years in the radar community to estimate the direction of arrival (DOA) of incoming echoes. In recent years, a variant of the monopulse technique has been developed, termed the shielded gradient technique, which allows DOA estimation for signals of arbitrary bandwidth. The technique maps the array-output M-vector into a frequency-invariant B-dimensional beamspace. The work

presented here describes the shielded gradient beamspace model in its higher-order form, and develops wideband DOA estimation algorithms analogous to the narrow-band MUSIC, root-MUSIC, and ESPRIT algorithms. The performance of these new algorithms is studied through simulation and application to measured, in-water sonar data. They are also compared via simulation to existing wideband DOA estimation algorithms. [Work supported by the Internal Research and Development Program under Contract No. FEE-800.]

THURSDAY MORNING, 13 NOVEMBER 2003 PECOS ROOM, 8:00 TO 11:45 A.M.

Session 4aUWb

Underwater Acoustics: Modeling of Propagation and Scattering

Nicholas C. Makris, Chair
Department of Ocean Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

Contributed Papers

8:00

4aUWb1. Ocean acoustic wave propagation and ray method correspondence: Internal wave fine structure. Nicholas R. Cerruti, Katherine C. Hegewisch, and Steven Tomsovic (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814, ncerruti@wsu.edu)

Acoustical wave fields can only detect structures in the ocean's sound-speed fluctuations that are on the order of the smallest acoustic wavelength component. On the other hand, geometrical ray-tracing methods are sensitive to infinitely fine structures in the wave modeling of these fluctuations. Hence, a proper model seeking agreement between ray methods and wave propagation must, at a minimum, filter out the fine oscillations of the internal wave field. Starting from an efficient numerical scheme for generating the internal waves introduced by Brown and Colosi [J. Acoust. Soc. Am. **103**, 2232 (1998)], which reproduces the Garrett–Munk spectrum, we introduce a smoothing technique that removes the unphysical portion of the internal wave modeling. The key is to find a smoothing that does not significantly alter the propagated wave field, yet eliminates as much of the ''micro-folding'' phenomena that was discussed by Simmen *et al.* [J. Acoust. Soc. Am. **102**, 239 (1997)]. We give a characterization of the extent of smoothing necessary as a function of acoustic frequency and propagation range, and show how it improves the correspondence. Finally, we note that the smoothing allows detailed ray methods to be implemented further in range. [Work supported by ONR.]

8:15

4aUWb2. Covariance of the forward propagated field through a waveguide containing random inhomogeneities. Purnima Ratilal and Nicholas C. Makris (MIT, Cambridge, MA 02139)

Analytic expressions are derived for the spatial covariance of the field from a point source after forward propagation though a waveguide containing random surface and volume inhomogeneities. It is shown that the depth-averaged second moment and expected power of the forward propagated field can be obtained analytically. The mean forward propagated field has also been obtained analytically in terms of modal attenuation and dispersion coefficients in a waveguide [P. Ratilal and N. C. Makris **112**, 2403 (2002)]. The covariance between two receiver depths of the forward propagated field through the entire random medium can then be determined by invoking the equi-partition of modal energy after significant multiple scattering. It is expressible as a sum of modal covariance terms. Each term depends on (1) the modal extinction cross-section [P. Ratilal and N. C. Makris, **110**, 2924–2945 (2001)] of an expected elemental inhomogeneity of the medium, and (2) the scatter function variance of an elemental inhomogeneity which couples each mode to all the other modes.

8:30

4aUWb3. Energy-conserving and single-scattering parabolic equation solutions for elastic media. Elizabeth T. Kusel, William L. Siegmann (Rensselaer Polytechnic Inst., Troy, NY 12180, kusele@rpi.edu), Michael D. Collins, and Joseph F. Lingevitch (Naval Res. Lab., Washington, DC 20375)

Parabolic equation techniques are efficient for solving nonseparable wave propagation problems. When the properties of the medium vary gradually in range, parabolic equation solutions are also very accurate for many problems. The key to achieving accuracy and efficiency simultaneously is to apply energy-conservation or single-scattering corrections to account properly for range dependence. This approach has proven to be very effective for acoustic media. Some progress has been made on the elastic case [J. Acoust. Soc. Am. **94**, 975–982 (1993); **94**, 1815–1825 (1993)], but this problem has not been fully resolved. In this paper we will discuss some recent progress in the formulation of the elastic parabolic equation [W. Jerzak, J. Acoust. Soc. Am. (submitted)], a single-scattering approach for a vector wave problem [J. Acoust. Soc. Am. **104**, 783–790 (1998)], and how they are being used to improve the accuracy of parabolic equation solutions for problems involving elastic sediments. [Work supported by ONR.]

8:45

4aUWb4. Generalization of the rotated parabolic equation to variable slopes. Donald A. Outing, William L. Siegmann (Rensselaer Polytechnic Inst., Troy, NY 12180, outind@rpi.edu), and Michael D. Collins (Naval Res. Lab., Washington, DC 20375)

The parabolic equation method is very efficient for solving range-dependent propagation problems. This approach is also accurate when range dependence is treated properly. This problem was resolved for fluid media by applying energy-conservation [J. Acoust. Soc. Am. **89**, 1058–1075 (1991)] and single-scattering [J. Acoust. Soc. Am. **91**, 1357–1368 (1992)] corrections. Since these corrections were less successful for problems involving elastic layers, other approaches such as mapping [J. Acoust. Soc. Am. **107**, 1937–1942 (2000)] and rotating [J. Acoust. Soc. Am. **87**, 1035–1037 (1990)] coordinates were investigated. In this presentation, the rotated parabolic equation solution is generalized to problems involving variable slope. The medium is divided into a series of regions with constant slope. When changes in slope are encountered, the field is propagated beyond the change and then used to interpolate and extrapolate onto a computational grid that is rotated relative to the previous grid. This approach is implemented and tested for the fluid problem. It should also be applicable to the elastic problem, but it will be necessary to apply a change

of variables each time the slope changes since the dependent variables are the tangential and normal displacements. [Work supported by ONR.]

9:00

4aUWb5. Analysis of measured broadband acoustic propagation using a parabolic equation approach. Mason Gray, D. P. Knobles, and Robert Koch (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

A broadband parabolic equation (PE) approach is employed to simulate data taken from two Shallow Water Acoustic Measurement Instrument (SWAMI) bottom mounted horizontal line array (HLA) experiments in shallow water environments off the east coast of the U.S. and in the Gulf of Mexico. In both experiments the HLA was deployed along an isobath. Light bulbs were imploded at known depths and ranges in both the range-independent (array end fire) and range-dependent (array broadside) directions. For the east coast experimental data, the PE model is used to infer a seabed geoacoustic description in both the range-dependent and range-independent directions. Also, comparisons of modeled time series were made for the range-independent case with a broadband normal mode model to validate the PE calculations. In the Gulf of Mexico experiment, the sediment geoacoustic profile is well known from previous inversions and geophysical measurements. This known seabed description was used to simulate the range-dependent data. A broadband energy-conserving coupled mode approach is also employed to model the range-dependent propagation. This allows the physical mechanisms associated with range-dependent propagation to be examined in a quantitative manner for this shallow water environment. [Work supported by ONR.]

9:15

4aUWb6. Sound scattering by 3-D, time-dependent internal gravity wave fields in the ocean. Oleg A. Godin (CIRES, Univ. of Colorado and NOAA/Environ. Technol. Lab., Boulder, CO 80305-3328, Oleg.Godin@noaa.gov), Alexander G. Voronovich, and Valery U. Zavorotny (NOAA/Environ. Technol. Lab., Boulder, CO 80305-3328)

In a wide range of sound frequencies, internal gravity waves (IWs) are the major cause of underwater sound scattering and resulting fluctuations in acoustic quantities. In this paper, IW-induced variations in acoustic travel times, grazing and azimuthal arrival angles, and frequency spectra are analyzed and their implications for acoustic oceanography are emphasized. Statistical properties of IW-induced acoustic fluctuations are studied within ray and modal representations of the acoustic field. Ray perturbation theory is developed to extend results previously obtained using eikonal perturbation theory to rays which encounter an arbitrary number of caustics on their way from a source to a receiver. Effects of mode coupling on acoustic travel time variance and bias are quantified in range-dependent and horizontally-inhomogeneous environments. Time-dependent, IW-induced variations of sound speed are shown to result in a detectable wander of frequency of the received signal emitted by a CW sound source at rest. Acoustic frequency variance on refracted rays is proportional to spatial density of IW energy. Feasibility of quantifying internal wave energy density and its variation in time through measurements of spectra of acoustic signals received along resolved ray paths from a highly stable source is discussed. [Work supported by ONR.]

9:30

4aUWb7. Nonlinear modes interaction in an acoustic waveguide. Kaelig Castor, Philippe Roux, W. A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92037), and B. E. McDonald (Naval Res. Lab., Washington, DC 20375)

The nonlinear interaction of lower order modes in an acoustic waveguide is investigated. For finite-amplitude wave propagation, nonlinear effects redistribute the modal amplitudes as a function across frequency and wave number. Therefore, the nonlinear propagation in a waveguide can produce an arrival structure quite different from the classical result. Simulations with the NPE code [McDonald and Kuperman, J. Acoust. Soc. Am. **81**, 1406–1417 (1987)] were used to study this frequency mode coupling in a realistic ocean waveguide.

9:45

4aUWb8. Scattering in a 3-D wedge with a rough bottom boundary. D. P. Knobles (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

Scattering problems in ocean waveguides may be formulated in the context of a set of coupled inhomogeneous integral equations. A particular problem of interest is the physics of low frequency reverberation in an ocean waveguide with a bottom described by a rough interface superimposed on a sloped boundary that defines the water–sediment interface. An idealization of such a waveguide is a 3-D wedge with a rough bottom. The Helmholtz equation for the propagation of sound for such an environment is examined by solving the basic coupled integral equations. Reverberation time series for a bistatic source–receiver geometry are computed by Fourier synthesis of the frequency response of the sum of the direct, refracted, and scattered components of the acoustic field. An objective of the computations is to examine the interplay between the refracted component of the field that results from the wedge geometry and the scattering components that result from the rough bottom. Although this picture of reverberation is incomplete since the physics of volume scattering is ignored, it allows one to examine the basic physics of reverberation in a waveguide that has variations in both range and azimuth. [Work supported by ONR.]

10:00–10:15 Break

10:15

4aUWb9. The coupling of elastic, surface-wave modes by a slow, interfacial inclusion. John G. Harris (Ctr. QEFP, Northwestern Univ., 2137 N. Sheridan Rd., Evanston, IL 60208-3020, j-harris8@northwestern.edu) and Gareth Block (Univ. of Texas–Austin, Austin, TX 78713-8029)

The coupling of in-plane, elastic, surface waves guided by a homogeneous layer on a similar substrate, but perturbed by the presence of a second, long, slow, interfacial layer, of slowly varying thickness, is studied. By projecting the elastic-wave equations onto a basis of local eigenmodes, an infinite system of coupled-mode equations describing the evolution of the amplitudes of each mode is obtained. Within the equations, the coupling is manifested by the presence of coupling coefficients that depend critically on the difference between the wavenumbers of adjacent modes. This system is truncated and solved by noting under what conditions the modes can be satisfactorily described by a WKBJ approximation (propagating without coupling) and under what conditions they are coupled by adjacent wavenumbers being brought into proximity by the slowly changing propagation environment. The criteria for coupling is that the difference in neighboring wavenumbers be of the same order as the slope of the inclusion. The case of the first three modes is worked out in detail: it is shown how initially modes one and two couple followed by the coupling of modes two and three.

10:30

4aUWb10. A sub-mesoscale hydrodynamic/acoustic simulation model for continental shelf-break regions. Steven Finette (Acoust. Div., Naval Res. Lab., Washington, DC 20375), Colin Y. Shen, and Thomas E. Evans (Naval Res. Lab., Washington, DC 20375)

A nonhydrostatic, hydrodynamic model of the sound speed field in a continental shelf-break environment has been developed and implemented. The model is based on a vorticity formulation of the equations of motion for an incompressible fluid with a free ocean surface, and it is capable of

simulating the generation and propagation of internal tides and solibores under tidal forcing. The model has been benchmarked with an exact numerical solution for a soliton. A set of space and time evolving sound speed distributions is integrated with a parabolic equation code to compute time and frequency dependent pressure fields. Two-dimensional examples of broad-band signal gain degradation on vertical arrays in this environment are presented, as well as range-frequency maps that illustrate the structure of the waveguide invariant in a shelf-break environment that is changing in time. Implications for source localization are considered. [Work supported by ONR.]

10:45

4aUWb11. Optimal temperature sampling with SPOTS to improve acoustic predictions. Erik R. Rike, Donald R. DelBalzo (Neptune Sci., Inc., 40201 Hwy. 190 E., Slidell, LA 70461), and Brian C. Samuels (Appl. Hydro-Acoust. Res., Inc., Centerville, VA 20120)

The Modular Ocean Data Assimilation System (MODAS) uses optimal interpolation to assimilate data (e.g., XBTs), and to create temperature nowcasts and associated uncertainties. When XBTs are dropped in a uniform grid (during surveys) or in random patterns and spaced according to resources available their assimilation can lead to nowcast errors in complex, littoral regions, especially when only a few measurements are available. To mitigate, Sensor Placement for Optimal Temperature Sampling (SPOTS) [Rike and DelBalzo, Proc. IEEE Oceans (2003)] was developed to rapidly optimize placement of a few XBTs and to maximize MODAS accuracy. This work involves high-density, *in situ* data assimilation into MODAS to create a ground-truth temperature field from which a ground-truth transmission loss field was computed. Optimal XBT location sets were chosen by SPOTS, based on original MODAS uncertainties, and additional sets were chosen, based on subjective choices by an oceanographer. For each XBT set, a MODAS temperature nowcast and associated transmission losses were computed. This work discusses the relationship between temperature uncertainty, temperature error, and acoustic error for the objective SPOTS approach and the subjective oceanographer approach. The SPOTS approach allowed significantly more accurate acoustic calculations, especially when few XBTS were used. [Work sponsored by NAVAIR.]

11:00

4aUWb12. Modeling mid-to-high frequency acoustic signal fluctuations induced by fetch limited sea surface roughness in shallow water. Robert Heitsenrether and Mohsen Badiey (College of Marine Studies, Univ. of Delaware, Newark, DE 19711, rheits@udel.edu)

Surface waves are among several environmental parameters that significantly influence mid-to-high frequency (1–18 kHz) acoustic wave propagation. In coastal regions, surface waves are fetch limited with reduced spectral level at lower frequencies. In order to assess the detail of an acoustic signal interaction with the sea surface in such regions, a combined approach based on experimental observation and modeling of both surface waves and acoustic waves has been adopted. Data from two broadband shallow water acoustic experiments are presented. These data include simultaneous wind speed and acoustic propagation measurements. The experimental design allowed an examination of received signals corresponding to single surface bounced ray paths. Measured data analysis shows a high correlation between time–angle–intensity fluctuations of received signals and varying sea surface conditions. An empirical fetch-limited ocean wave spectrum has been combined with an acoustic ray-based model to study acoustic wave propagation. Rough sea surface realizations are generated and used as sea surface boundaries with the acoustic model. This combined sea surface/acoustic model predicts the variability of acoustic signal fluctuations as a function of varying sea surface. Modeled time–angle–intensity signal fluctuations compare excellently with field data at lower wind speeds.

11:15

4aUWb13. Narrowband signals propagation in randomly inhomogeneous shallow water waveguide. Boris Katsnelson, Sergey Pereselkov (Voronezh Univ., 1, Universitetskaya sq., Voronezh 394006, Russia), and Valery Petnikov (General Phys. Inst., Moscow 113000, Russia)

In the presented work results of experiment and theoretical modeling are considered for the narrowband (ratio of frequency band to frequency ~ 0.1) sound propagation at the long range (up to 200 km) acoustic track in Barents sea. Modes selection for these conditions can be used for acoustical tomography of large-scale perturbations. Characteristics of the sound signals passing through this area (and in turn feasibility of tomographic methods) are determined by mutually competing mechanisms: waveguide dispersion, providing modes filtering, modes attenuation and modes coupling due to random inhomogeneities, masking separation of modes. Experimental results presented in this work show significant influence of noise (or modes coupling). For interpretation of experimental results (frequency and time dependencies of arriving pulses) background internal waves are taken as a main reason of the modes coupling. It is shown that for the modes separation it is necessary to use vertical array along with the measuring difference in arrival time, and correlation processing of received signals. [Work supported by RFBR, grants Nos. 03-05-64568 and 02-02-16509.]

11:30

4aUWb14. Recent developments in underwater acoustic modeling. Paul C. Etter (Northrop Grumman Corp., Baltimore, MD 21203)

This is the fourth paper in an ongoing series of research reviews presented at eight-year intervals [P. C. Etter and R. S. Flum, J. Acoust. Soc. Am. Suppl. 1 **65**, S42 (1979); P. C. Etter, *ibid.* **82**, S102 (1987); J. Acoust. Soc. Am. **97**, 3312 (1995)]. This review of international developments in underwater acoustic modeling reveals an inventory containing 114 propagation models, 17 noise models, 17 reverberation models and 26 sonar performance models representing a 38% increase over the 1995 inventory. As in previous reviews, older models have been retained to provide an historical perspective that enables researchers to pose well-informed questions regarding the need for future developments. When executed in higher-level simulations, these models generate predictive and diagnostic outputs that are useful to sonar technologists or acoustical oceanographers in the analysis of complex systems operating in the undersea environment. Recent applications of underwater acoustic models in naval operations, offshore industries and oceanographic research will be discussed. This review coincides with the appearance of the third edition of *Underwater Acoustic Modeling and Simulation* (Spon Press, London, 2003), which provides a detailed account of research conducted over the past four decades.

Session 4pAA

Architectural Acoustics and Noise: Acoustic Design of Government Buildings

David E. Marsh, Chair
Pelton Marsh Kinsella, 1420 West Mockingbird Lane, Suite 400, Dallas, Texas 75247-4932

Chair's Introduction—2:00

Invited Papers

2:05

4pAA1. Acoustical considerations for secondary uses of government facilities. Jack B. Evans (JEAcoust., 5806 Mesa Dr., Ste. 380, Austin, TX 78731, Evans@JEAcoust.com)

Government buildings are by their nature, public and multi-functional. Whether in meetings, presentations, documentation processing, work instructions or dispatch, speech communications are critical. Full-time occupancy facilities may require sleep or rest areas adjacent to active spaces. Rooms designed for some other primary use may be used for public assembly, receptions or meetings. In addition, environmental noise impacts to the building or from the building should be considered, especially where adjacent to hospitals, hotels, apartments or other urban sensitive land uses. Acoustical criteria and design parameters for reverberation, background noise and sound isolation should enhance speech intelligibility and privacy. This presentation looks at unusual spaces and unexpected uses of spaces with regard to room acoustics and noise control. Examples of various spaces will be discussed, including an atrium used for reception and assembly, multi-jurisdictional (911) emergency control center, frequent or long-duration use of emergency generators, renovations of historically significant buildings, and the juxtaposition of acoustically incompatible functions. Brief case histories of acoustical requirements, constraints and design solutions will be presented, including acoustical measurements, plan illustrations and photographs. Acoustical criteria for secondary functional uses of spaces will be proposed.

2:30

4pAA2. Acoustical and electronic media design aspects of the Texas Capitol restoration. Kenneth Dickensheets (Dickensheets Design Assoc., 12335 Hymeadow Dr., Ste. 200, Austin, TX 78750, ken@dickensheets.com)

A presentation of the studies involved in preparation for the restoration of the Texas State Capitol building and the implementation of acoustical, noise and vibration control and electro-acoustical recommendations to ensure that legislative bodies' and committees' deliberations would be heard while allowing for faithful historical restoration of the facility. Specific attention to and examples in the Senate Chamber will be presented. Solutions to difficult noise and acoustical issues in the context of historical preservation will be discussed. Historical and present photos will be shown.

2:55

4pAA3. The noise quandary of "green" light. Dorie A. Najolia (Acentech, Inc., 33 Moulton St., Cambridge, MA 02138, dnajolia@acentech.com)

In recent years, the government has striven to build more sustainable and efficient facilities by introducing the LEED (Leadership in Energy and Environmental Design) rating system and establishing acoustical privacy standards. However, there are some LEED design credits that conflict with the acoustical goals. This paper reviews the issues that arise when architects attempt to accommodate daylighting and acoustical privacy in their designs for government buildings. In achieving LEED certification, points are earned by designs that introduce daylight throughout occupied spaces in a building. Creating a ''line of sight'' from exterior glazing to interior spaces generally involves using transom, clerestory, side lite windows, or entire glass walls. At the same time, acoustical goals require construction around these spaces to achieve moderate to high levels of sound isolation. Another conflict is that although sound data is available for glass, very little research is available to demonstrate the effect lightweight aluminum framing has on the entire wall system. Incorporating these standards so that they provide the desired lighting and acoustical effects in government buildings needs to be better understood, since achieving both can add cost and complexity to the building design. Alternatively, compromises can be made, if they are understood and planned.

3:20–3:35 Break

3:35

4pAA4. Acoustical design issues in justice facilities. Howard K. Pelton and Ted N. Carnes (Pelton Marsh Kinsella, 1420 W. Mockingbird Ln., Ste. 400, Dallas, TX, peltonhk@c-b.com)

This paper presents various acoustical design challenges experienced by the authors with regard to courtrooms and jail facilities. The challenges in courtrooms include avoiding the sound focusing tendencies of concave curved surfaces and the general preference by courtroom designers for hard (i.e., sound reflecting) surface material. Jail facilities have attempted in recent years to incorporate sound absorbing surfaces to reduce noise, but this is difficult to achieve with durable materials that also must meet the security requirements of such facilities. Another interesting challenge is mitigating the intrusive noise of industrial grade toilets flushing when they are located in holding tanks adjacent to courtrooms.

4:00

4pAA5. From the bank, through the cafeteria line, and on to council chambers—Do the acoustics right! R. Bob Adams (Hoover & Keith, Inc., 11391 Meadowglen, Ste. D, Houston, TX 77082)

A common characteristic of municipal governments is the re-use of existing facilities. When a building's original use and acoustical design are dissimilar to the new acoustical needs, the project can be a challenge. The City of San Antonio Council Chamber was built as a bank in the 1930's and is considered an architectural landmark. This case study of the Council Chamber discusses the acoustical problems observed with intelligibility, focusing, and general ambiance from the highly ornate ceiling, internal arches, and an overlooking mezzanine. Primarily used for open City Council meetings, the room is also used as a ''studio'' for broadcast of other city related issues. The presentation will include photographs identifying the historical aspects of the architecture as well as portions of the EASE model created for the project.

4:25

4pAA6. A new seamless, smooth, interior, absorptive finishing system. Peter D'Antonio (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774)

Government architecture typically employs classic forms of vaults, domes and other focusing or reflective shapes, usually created with hard materials like concrete and plaster. The use of conventional porous absorption is typically rejected as an acoustical surface material for aesthetic reasons. Hence, many of these new and existing facilities have compromised speech intelligibility and music quality. Acousticians have sought a field-applied, absorptive finishing system that resembles a smooth plaster or painted drywall surface, since the dawn of architectural acoustics. Some success has been achieved using sprayed cellulose or cementitious materials, but surface smoothness has been a challenge. A new approach utilizing a thin microporous layer of mineral particles applied over a mineral wool panel will be described. This material can be applied to almost any shape surface, internally pigmented to match almost any color and renovated. Because of these unique characteristics the new seamless, absorptive, finishing system is being specified for many new and renovated spaces. Application examples will be presented.

4:50

4pAA7. Multipurpose council chambers "in the round" poses acoustical challenges. Edward L. Logsdon (D. L. Adams Assoc., 1701 Boulder St., Denver, CO 80211)

The City of Aurora Council Chambers is used for both municipal and public meetings. The room is configured to provide close-in seating with good sightlines from each of the 300 stadium-style seats. Presentations can be made from the central podium location to either the audience or council dais requiring multiple loudspeaker zoning and control. The cylindrical ceiling, shaped to accommodate video projection and lighting equipment, is acoustically treated to eliminate late reflections. The City Council meetings are broadcast to public TV on a regular basis from this room requiring good room acoustics and sound isolation to reduce echo and achieve acceptably low background noise levels while satisfying the aesthetic palette of the interior designers. A case history will be presented along with photographs showing how specialty wood materials, both absorptive and diffusive, were incorporated along with absorptive plaster and cloth-covered fiberglass panels into the design of the building.

Session 4pABa

Animal Bioacoustics and Signal Processing in Acoustics: Medical Imaging Techniques to Understand Auditory Processing

Darlene R. Ketten, Chair

Biology Department, Marine Sensory Systems Group, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Chair's Introduction—1:30

Invited Papers

1:40

4pABa1. Hair cells in motion: Imaging the organ of Corti. David C. Mountain and K. Domenica Karavitaki (Hearing Res. Ctr., Boston Univ., 44 Cummington St., Boston, MA 02215)

The mammalian cochlea contains two types of sensory cells, inner hair cells (IHCs) and outer hair cells (OHCs). The IHCs provide the vast majority of the synaptic input to the auditory nerve while the OHCs express a unique motor protein, prestin, and appear to participate in an electromechanical feedback loop that amplifies the motion of the organ of Corti (OC). To study this amplification process we have employed stroboscopic video microscopy to quantify the motion of various elements of the OC. Extracellular electrical stimulation was used to excite OHC motility and a computer-controlled high-intensity light-emitting diode (LED) is used to illuminate the organ OC in an excised cochlear preparation. Motion is measured by extracting small regions of interest (ROIs) from the images and cross-correlating the ROIs taken during electrical stimulation with a reference image from the same ROIs taken with no stimulation. The observed motion is quite complex with several vibration modes observed. One of the major findings is that there appears to be oscillatory fluid flow within the tunnel of Corti suggesting that the OHC contractions are pumping fluid longitudinally within the organ. [Work funded by NIDCD.]

2:05

4pABa2. Multimodal imaging of the human temporal bone: A comparison of CT and optical scanning techniques. Arne H. Voie (Spencer Technologies, 701 16th Ave., Seattle, WA 98122, voie@spencertechnologies.com), Bruce Whiting, Margaret Skinner, J. Gail Neely, Kenneth Lee, Tim Holden, and Barry Brunsden (Washington Univ. School of Medicine, St. Louis, MO 63110)

A collaborative effort between Washington University in St. Louis and Spencer Technologies in Seattle, WA has been undertaken to create a multimodal 3D reconstruction of the human cochlea and vestibular system. The goal of this project is to improve the accuracy of *in vivo* CT reconstructions of implanted cochleae, and to expand the knowledge of high-resolution anatomical detail provided by orthogonal-plane optical sectioning (OPFOS). At WUSL, computed tomography (CT) images of the cochlea are used to determine the position of cochlear implant electrodes relative to target auditory neurons. The cochlear implant position is determined using pre- and post-operative CT scans. The CT volumes are cross-registered to align the semicircular canals and internal auditory canal, which have a unique configuration in 3-D space. The head of a human body donor was scanned with a clinical CT device, after which the temporal bones were removed, fixed in formalin and trimmed prior to scanning with a laboratory Micro CT scanner. Following CT, the temporal bones were sent to the OPFOS Imaging Lab at Spencer Technologies for a further analysis. 3-D reconstructions of CT and OPFOS imaging modalities were compared, and results are presented. [Work supported by NIDCD Grants R44-03623-5 and R01-00581-13.]

2:30

4pABa3. Functional brain imaging and bioacoustics in the Bottlenose dolphins, *Tursiops truncatus*. Sam Ridgway, James Finneran, Donald Carder, William Van Bonn, Cynthia Smith (U.S. Navy Marine Mammal Prog., Space and Naval Warfare Systems Ctr., San Diego, 53560 Hull St., San Diego, CA 92152-5001), Dorian Houser (Biomimetica, San Diego, CA), Robert Mattrey, and Carl Hoh (Univ. of California, San Diego, CA 92093)

The dolphin brain is the central processing computer for a complex and effective underwater echolocation and communication system. Until now, it has not been possible to study or diagnose disorders of the dolphin brain employing modern functional imaging methods like those used in human medicine. Our most recent studies employ established methods such as behavioral tasks, physiological observations, and computed tomography (CT) and, for the first time, single photon emission computed tomography (SPECT), and positron emission tomography (PET). Trained dolphins slide out of their enclosure on to a mat and are transported by trainers and veterinarians to the laboratory for injection of a ligand. Following ligand injection, brief experiments include trained vocal responses to acoustic, visual, or tactile stimuli. We have used the ligand technetium (Tc-99m) biscisate (Neurolite) to image circulatory flow by SPECT. Fluro-deoxy-d-glucose (18-F-FDG) has been employed to image brain metabolism with PET. Veterinarians carefully monitored dolphins during and after the procedure. Through these methods, we have demonstrated that functional imaging can be employed safely and productively with dolphins to obtain valuable information on brain structure and function for medical and research purposes. Hemispheric differences and variations in flow and metabolism in different brain areas will be shown.

2:55

4pABa4. Extreme variations in skull density of toadfish, *Opsanus ta*. Peggy Edds-Walton (Neurosci. Inst., Marine Biological Lab., Woods Hole, MA 02543) and Darlene Ketten (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Gross observations of the skull in sexually mature toadfish revealed variations in bone density that appear to be most extreme in the otic capsule. Four male and one female toadfish (17–30.5 cm SL) were scanned to obtain images formatted at 100 μm and 1 mm. Consecutive measurements were made from caudal of the otic capsule to the rostral edge of the otic capsule. Attenuation values were recorded from three sites on the skull (left parietal bone, parietal suture, ventral surface of the otic capsule) and along two otoliths (calcareous, associated with the sensory epithelia of the ear) within the otic capsule. In all five fish, the parietal suture had the highest attenuation. Attenuation values for the parietal bone varied with size, indicating increasing density with growth. Among all five fish, the lowest attenuations were obtained for the ventral wall of the otic capsule, with values similar to those of cartilage. In addition, the minimum values were found ventral to the saccular otolith. Given that the saccule is the primary auditory endorgan in this species [Edds-Walton *et al.*, J. Comp. Neurol. **411**, 212–238 (1999)], the co-occurrence of bone thinning in this area of the skull may have functional significance related to audition.

Contributed Paper

3:20

4pABa5. Structural (CT) and functional imaging (PET/SPECT) for the investigation of dolphin bioacoustics. Dorian S. Houser (Biomimetica, 5750 Amaya Dr., Ste. 24, La Mesa, CA 91942), James J. Finneran (U.S. Navy Marine Mammal Prog., Space and Naval Warfare Systems Ctr., San Diego, CA), Robert Mattrey, Carl Hoh (School of Medicine, Univ. of California, San Diego, CA), and Sam Ridgway (U.S. Navy Marine Mammal Prog., Space and Naval Warfare Systems Ctr., Univ. of California, San Diego, CA)

A combination of imaging modalities was used to address physiological and anatomical questions relevant to dolphin bioacoustics. Three dolphins (*Tursiops truncatus*) were scanned with CT to investigate *in vivo* dolphin cranial anatomy. One dolphin underwent SPECT and PET scanning to investigate blood flow and metabolic activity of the cranial tissues. Air spaces were mostly contiguous and covered the periotic bone and auditory bulla dorsally and medially. Cranial air was compartmentalized by the nasal plug and constriction of the palatopharyngeus muscle. Blood flow, determined from SPECT imaging of ^{99}Tc-bicisate distribution, was greatest in the brain, melon, and posterior fats of the lower jaw. Metabolic activity of tissues, assessed by monitoring the uptake of ^{18}F-deoxyglucose via PET, indicated that melon and jaw fats were metabolically inert compared to the brain. Nasal cavity and sinus air volume that is reduced during diving may be replenished with lung air via the palatopharyngeus and Eustachian tube. Air covering the bulla may protect the ears from outgoing echolocation pulses and contribute to spectral and time of arrival cues. Blood flow to the melon and lower jaw fats may serve to either regulate the temperature of acoustic lipids or act as a site of counter-current heat exchange.

THURSDAY AFTERNOON, 13 NOVEMBER 2003 GUADALUPE ROOM, 3:45 TO 5:05 P.M.

Session 4pABb

Animal Bioacoustics: Topics in Animal Bioacoustics

Darlene R. Ketten, Chair

Biology Department, Marine Sensory Systems Group, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Chair's Introduction—3:45

Contributed Papers

3:50

4pABb1. Pure-tone audiograms and hearing loss in the white whale (*Delphinapterus leucas*). James J. Finneran, Donald A. Carder (U.S. Navy Marine Mammal Prog., Space and Naval Warfare Systems Ctr., San Diego, Code 2351, 53560 Hull St., San Diego, CA 92152-5001), Randall Dear (Sci. Applications Intl. Corp., San Diego, CA 92110), Traci Belting (Point Defiance Zoo and Aquarium, Tacoma, WA 98466), and Sam H. Ridgway (Space and Naval Warfare Systems Ctr., San Diego, CA 92152-5001)

A behavioral response paradigm was used to measure pure-tone audiograms for two white whales (*Delphinapterus leucas*). Tests were conducted over a 20 month period at the Point Defiance Zoo and Aquarium, in Tacoma, Washington. Subjects consisted of two males, aged 8–10 and 9–11 during the course of the study. Subjects were born in an oceanarium and had been housed together for all of their lives. Hearing thresholds were measured using a modified up/down staircase procedure and acoustic response paradigm where subjects were trained to whistle in response to hearing test tones and to remain quiet otherwise. Test frequencies ranged from approximately 2 to 130 kHz. Best sensitivities ranged from 40 to 50 dB *re*: 1 Pa. Both subjects had traditional U-shaped mammalian audiograms; however, one subject exhibited significant high-frequency hearing loss, above approximately 37 kHz. The experimental setup and procedure will be presented and the measured hearing thresholds compared to those previously measured in white whales. The potential role of ototoxic antibiotics in the observed hearing loss will be discussed. [Work supported by ONR Marine Mammal S&T Program and the U.S. Navy CNO(N45).]

4:05

4pABb2. Detection and classification of right whales in the Bay of Fundy using independent component analysis. Michael Linford and Brian La Cour (Appl. Res. Labs., Univ. of Texas, 10000 Burnet Rd., Austin, TX 78758)

A novel method of the detection and classification for marine mammals is presented which uses techniques from independent component analysis to solve the blind source separation problem for right whales in the Bay of Fundy. Using the fundamentally non-Gaussian nature of marine mammal vocalizations and data collected on multiple hydrophones, we are able to separate right whale source spectra, up to an unknown scale, from ambient noise. This technique assumes that the array data is a linear combination of the source signals but does not require specific knowledge of the array geometry. A detector/classifier algorithm is demonstrated which compares the estimated source spectra against known right whale vocalizations.

4:20

4pABb3. Dolphin echolocation strategies studied with the Biosonar Measurement Tool. Dorian S. Houser (Biomimetica, 5750 Amaya Dr., Ste. 24, La Mesa, CA 91942), Steve W. Martin, Michael Phillips, Eric Bauer, and Patrick W. Moore (U.S. Navy Marine Mammal Prog., Space and Naval Warfare Systems Ctr., San Diego, CA)

Two free-swimming dolphins (Tt722 and Tt673) were trained to carry the Biosonar Measurement Tool (BMT) during open water, proud target searches in order to explore echolocation behavior without the constraints of traditional experimental designs. The BMT recorded the angular motion, depth, and velocity of the dolphin as well as echolocation clicks and echoes returning from insonified targets. Mean search time for Tt722 was 24.6 ± 7.3 s and 6.5 ± 3.0 s for Tt673 on target present trials, the former strategy resulting in the lower false alarm rate. The majority of clicks exceeded 195 dB *re*: 1 μPa throughout all trials for both animals but each demonstrated preferences for particular frequency bands of echolocation. Considering all trials, only 3.6% of all clicks produced by Tt722 contained peak frequencies greater than 60 kHz whereas Tt673 produced clicks with peak frequencies above 60 kHz 20.4% of the time. Distinctive frequency bands in the distribution of clicks were notable: bands for Tt673 occurred at 38, 54, and 69 kHz with less defined higher order bands; bands for Tt722 occurred at 25, 35, and 40 kHz. Distinctive frequency bands suggest a preferential use or mechanical constraint on harmonically related click frequencies.

4:35

4pABb4. Frameless processing of bird songs using set membership identification. Rafi Mohammad, Mark M. Wilde, and Dale Joachim (Dept. of Elec. Eng. and Computer Sci., Tulane Univ., 211 Stanley Thomas Hall, New Orleans, LA 70118, joachimd@eecs.tulane.edu)

This paper describes a novel approach to feature extraction from bird songs using a set-membership identification (SMI) algorithm. The low computational complexity of the SMI algorithm allows frameless pointwise feature estimation and real-time processing. Both energy-based endpoint detection and set-to-point classification methods are incorporated in this SMI processing for enhanced labeling performance. The described algorithm serves as a front end to a fully automated bird identification system in which training data collection is automated by scheduled computer generated phone calls to a cellular monitoring station. RASTA processing of feature vectors compensates for the telephone channel effects.

4:50

4pABb5. Identification and assessment of constrictions in a branched tube system. Youhua Du and Ahmed M. Al-Jumaily (Diagnostics and Control Res. Ctr., Auckland Univ. of Technol., Auckland, New Zealand)

This paper focuses on developing an analytical methodology, based on the frequency spectrum of the input acoustic impedance, to identify and assess the location and severity of constrictions in a branched-treelike tube system. The method is applied to two models, one is for the trachea with the first-generation asymmetric bronchi and the other is for the trachea with two-generation bronchi. To develop the methodology, a different size constriction is introduced at various locations along the branches and resonant frequencies are analyzed. Correlation is performed between the resonant frequencies and the location and size of these constrictions. For the two models, charts are generated which can be used to determine the location and size of a constriction just by observing the trend of various resonances. For the first model it is shown that in the frequency range of 1–4000 Hz the third and sixth resonances reflect constriction occurrence in the larger branch and the fourth and seventh reflect constriction in the smaller branch. However, for the second model it is indicated that constriction is reflected: in first branch by the sixth resonance, in the second branch by the seventh resonance, and in the third branch by the ninth resonance.

THURSDAY AFTERNOON, 13 NOVEMBER 2003 TRINITY A ROOM, 2:30 TO 4:35 P.M.

Session 4pBB

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Tissue Harmonic Imaging

Robin O. Cleveland, Chair

Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

Invited Papers

2:30

4pBB1. Nonlinear acoustics in tissue harmonic imaging. Michalakis Averkiou (Philips Medical Systems, P.O. Box 3003, Bothell, WA 98041-3003)

In recent years the interest in nonlinear acoustics has dramatically increased in diagnostic ultrasound. There are two main areas where nonlinear acoustics is used in medical imaging: tissue harmonic imaging (THI) and imaging of ultrasound contrast agents. Although similar approaches are used in both of these areas, they are very different in that THI is based on nonlinear propagation of sound in tissue, whereas contrast imaging is based on the nonlinear scattering from resonant microbubbles. The clinical benefits of THI are reduced phase aberration artifacts and overall clutter, improved border delineation, and increased contrast resolution. The

4p THU. PM

basic principles of nonlinear propagation of sound beams in tissue are discussed here. Theoretical and experimental results are used to demonstrate some of the properties of nonlinear propagation and their relation to imaging. To a large extent the clinical benefits of THI are explained with the nonlinear propagation properties. Imaging considerations and techniques like pulse inversion and power modulation are discussed. These techniques are also used in imaging contrast agents (microbubbles) and an effort is made to separate the various issues in these two imaging areas and explain their differences.

2:55

4pBB2. Physics of tissue harmonic imaging. Thomas L. Szabo (Boston Univ., 110 Cummington St., Boston, MA 02215, tlszabo@bu.edu)

Tissue harmonic imaging is used on almost all modern diagnostic ultrasound equipment. The clinical success of this modality involves a combination of fortuitous factors. The improvement in lateral resolution obtained by transmitting on a fundamental frequency and receiving on a harmonic is well known. Other advantages include a deeper focal range with less absorption than would have been obtained by transmitting and receiving at twice the fundamental frequency. Natural apodization considerably reduces the 3-D volume of interaction with tissue. Simulations indicate that harmonic imaging is more robust in the presence of aberration. Analysis shows that the improvements over what would have been achieved by transmitting at twice the frequency are for shallower focal depths. Reasons for the noticeably enhanced contrast of harmonic imaging are discussed.

Contributed Papers

3:20

4pBB3. Phase conjugation of the second harmonic of a focused ultrasound beam as a method for improving C-scan acoustical imaging in nonlinear inhomogeneous media. Leonid M. Krutyansky, Andrew P. Brysev, Roman V. Klopotov (Wave Res. Ctr., General Phys. Inst., Russian Acad. of Sci., 38 Vavilov Str., Moscow 119991, Russia, krut@orc.ru), Philippe J. Pernod, Vladimir L. Preobrazhensky (Institut d'Electronique, de Microelectronique et de Nanotechnologies, IEMN-DOAE, UMR CNRS 8520, Villeneuve d'Ascq 59651, France), Xiang Yan, and Mark F. Hamilton (Univ. of Texas, Austin, TX 78712-1063)

Acoustical imaging in complex media (e.g., biological tissue) can be affected by phase aberrations introduced in a wave during propagation. Wave phase conjugation (WPC) of ultrasound is known for its ability to compensate for phase distortions due to inhomogeneity of the propagation medium, and it can be used for improvement of acoustical imaging under these conditions. In a nonlinear medium harmonics are generated during propagation of an intense beam of ultrasound, and this principle is used in tissue harmonic imaging. The parametric method of WPC permits phase conjugation of a selected frequency component of the probe beam. In this way the peculiarities of WPC can be combined with advantages of harmonic imaging. Automated WPC-focusing of the conjugated second-harmonic component of a focused nonlinear probe beam is studied experimentally and theoretically for the case of a homogeneous medium, and experimentally for a medium with pseudo-random inhomogeneities. The generated conjugate wave can also be sufficiently intense to generate higher-order harmonics, which display enhanced focusing. Improvement of a C-scan harmonic imaging system operating in an inhomogeneous medium is provided as an example.

3:35

4pBB4. Effect of aberration on the acoustic field in tissue harmonic imaging (THI). Yuan Jing and Robin Cleveland (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215, yuanjing@bu.edu)

A numerical simulation was used to study the impact of an aberrating layer on the generation of the fundamental and second-harmonic (SH) field in a tissue harmonic imaging scenario. The simulation used a three-dimensional time-domain code for solving the KZK equation and accounted for arbitrary spatial variations in all acoustic properties. The aberration effect was modeled by assuming that the tissue consisted of two layers where the interface has a spatial variation C that acted like an effective phase screen. Initial experiments were carried out with sinusoidal-shaped interfaces. The sinusoidal interface produced grating lobes which were at least 6 dB larger for the fundamental signal than the SH. The energy outside of the main lobe was found to increase linearly as the amplitude of the interface variation increased. The location of the grating lobes was affected by the spatial period on the interface variation. The inhomogeneous nature of tissue was modeled with an interface with a random spatial variation. With the random interface the average sidelobe level for the fundamental was -30 dB whereas the SH had an average sidelobe level of -36 dB. [Work supported by the NSF through the Center for Subsurface Sensing and Imaging Systems.]

3:50

4pBB5. New formulation of the elastic energy density for soft tissue. Mark F. Hamilton, Yurii A. Ilinskii, and Evgenia A. Zabolotskaya (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713-8029)

Measurements of the elastic constants in Landau's expansion of the strain energy density reported by Catheline *et al.* for a soft tissue phantom reveal that they differ by five orders of magnitude [J. Acoust. Soc. Am. **112**, 2404 (2002)]. A more appropriate expansion would possess constants of the same order. Landau's third-order expansion of the elastic energy density is extended to fourth order. The corresponding elastic constants are related to those for liquids. The principal contribution is an alternative formulation of the energy density that permits separation of effects due to compressibility and shear deformation. To fourth order its expansion is $\mathcal{E} \simeq \mathcal{E}_0(\rho) + \mu I_2 + \frac{1}{3}AI_3 + DI_2^2$, where ρ is density, I_2 and I_3 are the second- and third-order Lagrangian strain invariants used by Landau, μ is the shear modulus, A is one of Landau's third-order elastic constants, and D is a new fourth-order elastic constant. For processes involving mainly compressibility $\mathcal{E} \simeq \mathcal{E}_0(\rho)$, and for processes involving mainly shear deformation $\mathcal{E} \simeq \mu I_2 + \frac{1}{3}AI_3 + DI_2^2$. Comparison with the parameters in Mooney's potential function for incompressible elastic media shows the three constants to be of the same order. [Work supported by the IR&D Program at ARL:UT.]

4:05

4pBB6. Numerical modeling of Harmonic Imaging and Pulse Inversion fields. Victor F. Humphrey, Tracy M. Duncan (Dept. of Phys., Univ. of Bath, Bath BA2 7AY, UK), and Francis Duck (Royal United Hospital, Bath BA1 3NG, UK)

Tissue Harmonic Imaging (THI) and Pulse Inversion (PI) Harmonic Imaging exploit the harmonics generated as a result of nonlinear propagation through tissue to improve the performance of imaging systems. A 3D finite difference model, that solves the KZK equation in the frequency domain, is used to investigate the finite amplitude fields produced by rectangular transducers driven with short pulses and their inverses, in water and homogeneous tissue. This enables the characteristic of the fields and the effective PI field to be calculated. The suppression of the fundamental field in PI is monitored, and the suppression of side lobes and a reduction in the effective beamwidth for each field are calculated. In addition, the differences between the pulse and inverse pulse spectra resulting from the use of very short pulses are noted, and the differences in the location of the fundamental and second harmonic spectral peaks observed.

4:20

4pBB7. Statistical investigation of beam distortion by tissue inhomogeneity in tissue harmonic imaging. Xiang Yan and Mark F. Hamilton (Dept. of Mech. Eng., Univ. of Texas, Austin, TX 78712-1063)

For many patients, tissue harmonic imaging improves resolution by reducing phase distortion due to acoustic propagation through inhomogeneities in the body wall, reverberation in this layer, and artifacts due to sidelobes. Our investigation focuses on the first of these phenomena. A statistical analysis was conducted to quantify the improvement provided by tissue harmonic imaging in the presence of an inhomogeneous layer in close proximity to the source. The inhomogeneity is modeled as a phase screen of zero thickness located directly in front of the source. The fundamental beam and resulting second-harmonic generation are described in the parabolic approximation. The phase aberration is assumed to have zero mean and to be Gaussian correlated in space. For a source with Gaussian amplitude distribution, a solution requiring a double numerical integration was derived for the mean intensity of the second-harmonic beam profile in the target plane. The solution reveals the separate contributions due to the undistorted second-harmonic and the scattered component due to the inhomogeneity. The statistical solution is validated by comparison with ensemble averages of direct numerical simulations. Experiments were also conducted, the results of which confirm the advantages of tissue harmonic imaging. [Work supported by ONR.]

THURSDAY AFTERNOON, 13 NOVEMBER 2003 BRAZOS AUDITORIUM, 2:00 TO 5:00 P.M.

Session 4pMU

Musical Acoustics: Honoring the Contributions of Gabriel Weinreich II: Violin Acoustics

Uwe J. Hansen, Chair
Department of Physics, Indiana State University, Terre Haute, Indiana 47809

Invited Papers

2:00

4pMU1. The role of vibrato in the perception of violin quality. Colin E. Gough (School of Phys. and Astron., Univ. of Birmingham, Birmingham B15 2TT, UK)

The role of vibrato in the characterization of violin tone will be considered from both a physical acoustics and perceptual viewpoint. Musical examples of individual waveforms from violins of widely different qualities will be used to demonstrate the importance of vibrato and other temporal fluctuations in the recognition of the violin as a specific musical instrument and, by inference, of its perceived tonal quality also. Dynamic physical models will be introduced to describe the waveforms of sounds produced by the bowed string when vibrato is used. The sound produced by an individual instrument will be shown to be critically dependent on both the positions and the Q values of the acoustically important resonances excited, in addition to random noise generated by the bowing action. It will be argued that the temporal fluctuations in the sound of a violin played with vibrato are just as important as the temporal fluctuations associated with the very strong frequency-dependence of the directivity of the violin. Weinreich has emphasized the importance of the latter effect in any realistic reproduction of the sound of a violin.

2:30

4pMU2. Physics of the violin. Erik V. Jansson (Dept. of Speech, Music and Hearing, KTH, Dr. Kristinas vag 31, SE-Stockholm, Sweden)

A method to measure acoustical properties of the violin has been developed. The bridge is excited by an impulse force hammer (see www.speech.kth.se/music/acviguit4). Bridge vibrations are recorded by a small magnet and an electrical coil. Measurements can be made in an ordinary room and give a record of properties built into the violin body. An old good Polish violin, B Dankwart, Vilnius *ca*. 1600 shows typical results with peaks P1, P2 and the BH-hill. Our goal is as suggested by Gabriel Weinreich ''to understand, not to copy,''—Stradivarius. By shifting the soundpost position the peaks P1, P2 and BH can be somewhat monitored. It can be shown that the BH is not confined to the bridge only. The feet distance of the bridge is important but also the top plate. Marcin Groblicz, the great Polish violin maker, was court instrument maker in Krakow *ca*. 1600 to Sigismund, King of Poland and Sweden. I believe Gabriel Weinreich has some connection to Vilnius in Poland but unfortunately not to Sweden. We like Gabi.

3:00

4pMU3. Bow response to the string it is bowing. Robert T. Schumacher (Dept. of Phys., Carnegie Mellon Univ., Pittsburgh, PA 15213)

The force that a bowed string exerts on the bow can be reconstructed from the forces on the string's termination. By using one or more accelerometers attached to the bow, the response of the bow to that reconstructed force can be recorded. Bow responses will be shown that were obtained using a bowing machine. At some harmonic frequencies of the string the signal from the longitudinal standing wave excited on the bow hair allows determination of the velocity of propagation of that wave. The consequence of the variable excitation of various harmonics of the standing wave as the bow stroke progresses is a rapidly changing excitation of the bows various normal modes. Three data channels, requiring at least a 128 kHz sampling rate for E-strings, are required for bow force reconstruction and motion detection of the bow. The validity of substituting the force of the string on the bridge instead of the reconstructed bow force is explored. That allows a reduction of the minimum number of data channels from three to two, a reduction in the necessary sampling rate, and use of a bow stroke by a player instead of a bowing machine.

3:30

4pMU4. Modal analysis of violin bodies viewed as three-dimensional structures. Thomas D. Rossing (Phys. Dept., Northern Illinois Univ., DeKalb, IL 60115), Nils-Erik Molin, and Anna Runnemalm (Luleå Univ. of Technol., Luleå, Sweden)

Modal analyses of violins show several strong modes in the low frequency range. Holographic interferograms suggest that four strong modes can be interpreted as doublets having two and three nodal planes that intersect a cylinder with a roughly elliptical cross section at the bridge [A. Runnemalm, N.-E. Molin, and E. Jansson, J. Acoust. Soc. Am. **107**, 3452–3459 (2000); M. Roberts and T. D. Rossing, Catgut Acoust. Soc. J. **3**, 9–15 (1998)]. This is especially clear when the instrument is viewed simultaneously from three sides using mirrors, and the holographic system is made sensitive to in-plane motion as well. These doublets are not unlike those observed in cylindrical vibrators such as bells, and they remind us that a violin is a 3-dimensional object.

4:00

4pMU5. Unusual motions of a vibrating string. Roger J. Hanson (Dept. of Phys., Univ. of Northern Iowa, Cedar Falls, IA 50614, roger.hanson@cfu.net)

The actual motions of a sinusoidally driven vibrating string can be very complex due to nonlinear effects resulting from varying tension and longitudinal motion not included in simple linear theory. Commonly observed effects are: generation of motion perpendicular to the driving force, sudden jumps in amplitude, hysteresis, and generation of higher harmonics. In addition, these effects are profoundly influenced by wire asymmetries which in a brass harpsichord wire can cause a small splitting of each natural frequency of free vibration into two closely spaced frequencies (relative separation $\sim$0.2% to 2%), each associated with transverse motion along two orthogonal characteristic wire axes. Some unusual resulting patterns of complex motions of a point on the wire are exhibited on videotape. Examples include: sudden changes of harmonic content, generation of subharmonics, and motion which appears nearly chaotic but which has a pattern period of over 10 s. Another unusual phenomenon due to entirely different causes can occur when a violin string is bowed with a higher than normal force resulting in sounds ranging from about a musical third to a twelfth lower than the sound produced when the string is plucked.

4:30

4pMU6. Sound radiation by violins radically modal. George Bissinger (Phys. Dept., East Carolina Univ., Greenville, NC 27858)

Following a radically modal path pioneered by Gabi Weinreich in the early 1980s, normal-mode vibration and radiation analysis was applied to quality-rated violins. Being able to describe the violin in terms of normal-mode behaviors leads directly to a violin equation, allows extraction of a critical frequency for the violin, provides estimates of the fraction of vibrational energy radiated and extends to modal-average trendline modeling of the acoustic output, all of which can be linked to violin quality classifications if desired. Some quality-related results mirror Weinreich's conclusions from inverse radiativity measurements on violins, i.e., there is little difference in the radiativity of good and bad violins below 1 kHz. The first corpus bending modes (the baseball modes), first identified by Gabi and Eric Arnold, are a major contributor to violin sound and highlight one of the important aspects of Gabi's research, the coupling between wood and air. The first near-field acoustic holography results indicate that a significant part of the radiation from these structural modes is from the f-holes of the violin. These normal-mode investigations reflect some of the importance of Gabi's work for modern violin acoustics. [Research supported by NSF.]

THURSDAY AFTERNOON, 13 NOVEMBER 2003 BOSQUE ROOM, 1:30 TO 2:45 P.M.

Session 4pNSa

Noise: Topics in Urban and Community Noise

Bennett M. Brooks, Chair

Brooks Acoustics Corporation, 96 Main Street, Talcottville, Connecticut 06066

Contributed Papers

1:30

4pNSa1. Noise impact caused by electrical energy substations in Curitiba, Brazil. Fabiano B. Diniz and Paulo T. Zannin (Environ. Acoust. Lab., Dept. of Mech. Eng., UFPR, Curitiba, PR, Brazil)

This survey is intended to characterize the noise impact due to electrical energy substations in the city of Curitiba over the population living in their vicinity. This impact has been studied with the aid of a computational tool capable of mapping the acoustical field of substations and their vicinity. Several factors have been considered in this survey: sound power of the transformers; vehicle flow on the surrounding roads; positioning of the firewalls, of the buildings and of the walls; terrain topography. Four substations have been analyzed, and an acoustical map has been traced for each of them. With these maps it was possible to visualize what was the incident noise level on the building facades. The predicted noise levels have been compared to the environmental legislation of the noise emissions in effect in the city.

1:45

4pNSa2. CTA "L" train outdoor noise propagation in Wrigley-ville. Matthew McDuffee (Columbia College Chicago, 1232 S. 12th St., St. Charles, IL 60174, mcduffula@hotmail.com)

It is well known to those who reside in Chicago that the elevated mass transit trains, known as the L, are the largest culprit of outdoor noise pollution in most neighborhoods throughout the city. Many potential residents and buyers will make their choice of where to live in the city simply based upon the building's relationship to the L. For those who live directly

next to the tracks vibration becomes a major issue. The study focused on the environmental noise impact of the CTA red line train traveling through Chicago's north side in the Wrigley-ville neighborhood. Approximately 1.5 miles of elevated track stretch between the Addison station (near Wrigley field) and the bridge that crosses Montrose Avenue. Eight locations were chosen throughout an area covering 17 city blocks surrounding the red line tracks to perform field measurements. The field measurements, taken in Leq, were correlated against a computer model designed using Cadna/A outdoor noise propagation software. The models polygons (representing buildings) were created using ESRI Arc View mapping software to view digitized aerial photographs of the north side of Chicago. The models predictions correlated to within plus or minus 1 dB on 6 out of 8 field measurements.

2:00

4pNSa3. Evaluation and prediction of noise pollution levels in urban areas of Cdiz (Spain). Silvia Rivas, Ricardo Hernandez, and Jose Luis Cueto (Acoust. Vib. Lab. C.A.S.E.M, Campus Ro San Pedro 11510, Puerto Real, Cdiz, Spain, silvia.rivas@uca.es)

In the European policy, one of the most important objectives is to achieve a high level of health and environmental protection. The latest studies have shown that more than 20% of the world population lives under unacceptable noise levels and near 60% of the European population is exposed to worrying noise levels during the day. So, nowadays one of the objectives to be pursued is the protection against noise, one of the main environmental problems in Europe. During the last 10 years different studies have been carried out in urban areas of Cdiz, in order to evaluate the noise pollution level and its management. Those studies exposed how the continuous development of legal device, more capable each time, might moderate the upper emission levels. Instead of it, the share of the medium levels of the population hold up and could be detrimental for its normal development, increasing progressively. It will be useful to develop more studies in this area, as we are doing in El Campo de Gibraltar to study those damaging levels and to predict their tendency, in order to establish and improve new methods of environmental noise protection.

2:15

4pNSa4. Analysis over the new data treatment method developed by the community directive 2002/49/ce about environmental noise management. Jose Luis Cueto, Silvia Rivas, and Ricardo Hernandez (Acoust. and Vib. Lab. C.A.S.E.M., Campus Ro San Pedro s/n 11510, Puerto Real, Cdiz, Spain)

Nowadays a European Directive develops useful procedures in order to evaluate, by means of index, the acoustic pollution affecting urban areas. To apply properly those methods, it is necessary to use a long data sequence. Is usual to find in this kind of series some peculiar measures or events that could change indicatively our acoustic pollution index data. So, we suggest a simple data treatment method as a first step in the way of developing a better process of acoustic index construction. This methodology would permit us to introduce the acoustic climate concept.

2:30

4pNSa5. Evaluation of noise pollution in urban parks of Curitiba, Brazil. Andressa M. Ferreira (Environ. Acoust. Lab., Dept. of Mech. Eng., UFPR, 81531-990, Curitiba, PR, Brazil), Fabiano B. Diniz, and Paulo T. Zannin (Environ. Acoust. Lab., UFPR)

A study about the noise pollution found in six urban parks of Curitiba, Paran, Brazil. The equivalent noise levels (Leq) have been measured in points spread throughout the park, and interviews have been conducted with some park visitors. It has been found out that 17.83% out of the measurement sites have presented Leq levels over 65 dB(A), considered by the Preventive Medicine as the maximum level one can be exposed to without being exposed to health impairment risks, and 52.48% out of the meas sites do not satisfy the municipal law 10 625, which states the noise emission level of 55 dB(A) as the limit value for green areas (AV). The results of the questionnaires applied to the local visitors have showed that 39% out of the interviewed people users to visit the park every 75% of them seek for the realization of a physical activity during the realization of their activities in the parks, 22% out of interviewed people point the noise pollution as the source of annoyance and 29% of them point the local security.

THURSDAY AFTERNOON, 13 NOVEMBER 2003 BOSQUE ROOM, 3:00 TO 3:45 P.M.

Session 4pNSb

Noise: Measurement and Modeling of Noise Related Environments

Brandon D. Tinianov, Chair
Acoustical Laboratory, Johns Manville, 10100 West Ute Avenue, P.O. Box 625005, Littleton, Colorado 80162

Contributed Papers

3:00

4pNSb1. A method to separate contribution from coherent and interfering inputs. Valeri V. Lenchine, Hoon Wee, Jae-Man Joo, and Sang-Kyong Oh (Samsung Electr. DA R&D Ctr., 416, Maetan-3Dong, Paldal-gu, Suwon City, Gyeonggi-do 442-742, Korea)

Acoustical development tasks frequently involve the identification of contribution from different sources. A conventional way to detect principal exciters is to separate signals that are completely incoherent between each other and interfering inputs. However the method lacks physical meaning as the principle contributors are coherent. Moreover a conventional technique involves assumptions about independent inputs and the kind of dependence between interfering signals. It is suggested to extract noninterfering but partly coherent virtual signals from measured inputs that are considered as physical inputs. The procedure is based on the determination of a virtual input spectra matrix. As for many practical tasks measured inputs have approximately the same magnitudes of auto, cross spectra and high mutual coherence detection of principal contributors encounters ambiguity. A minimal principle value deviation is employed as criteria to choosing an appropriate virtual input spectra matrix. An example involving acoustical development of a multi indoor units air conditioner that is equipped by compressor generating pressure pulsation into discharge and suction pipeline is considered. These effects cause excessive sound emission of indoor units. The proposed technique is used for the separation of interfering coherent inputs (discharge and suction pressure). It enabled us to develop effective ways for the air conditioner noise reduction.

3:15

4pNSb2. Using the ideal function concept for machine noise control. Brian Landsberger (Caterpillar, Inc., Tech. Ctr., Bldg. E, P.O. Box 1875, Peoria, IL 61656)

An engineered system reaches its ideal function when all its input energy is transformed efficiently into creating the output energy. This leaves less energy available for creating unwanted noise and vibration. A large hydraulic pump evaluation is used to demonstrate the relation between energy efficiency and unwanted noise. Measurements were taken while conducting speed sweeps at various load conditions. Sound in the hemisphere surrounding the pump, acceleration at various locations on and around the pump, and shaft rotation instantaneous phase measurements were used to determine the magnitude, character, and source of the sound. A particularly loud noise level associated with a certain operating region was determined to result from driven vibration of the pump case. The disturbance appeared to originate from pulsations in the pump flow and pressure. This operating region also had a several percent drop in pump efficiency. In this example, more efficient energy transformation would have the dual benefit of higher efficiency and lower noise.

3:30

4pNSb3. Design features for free-field qualification of a new semi-anechoic room, and qualification performance. Kenneth A. Cunefare, Van Biesel, Mark Holdhusen, and Austin Shoemaker (Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

Precision qualification of a semi-anechoic room requires careful attention to the sound source and traversing method. Prior work, with test sources mounted above the reflecting floor of such a room, has indicated the potential for image source problems in the resulting field. To address such shortcomings, the new Georgia Tech semi-anechoic room was constructed with a recessed enclosure in the center of the floor. This enclosure permits the implementation of test sources coincident with the reflecting plane of the floor. In addition, prior work in an anechoic room has indicated the inadequacy of qualification traverses implemented at large spacings. To address this issue, hard-points were designed and implemented within the room to permit installation of traverse cables extending radially from the in-floor source enclosure out to the walls and corners. These traverse cables are an integral component of a custom continuous traverse system. The design features of the chamber which facilitate chamber qualification will be presented, along with the broadband and pure tone results of the qualification performed on the chamber.

THURSDAY AFTERNOON, 13 NOVEMBER 2003 TRINITY B ROOM, 1:30 TO 5:00 P.M.

Session 4pPA

Physical Acoustics: Outdoor Propagation

D. Keith Wilson, Chair

Engineering Research and Development Center, US Army Cold Regions Research Laboratory, 72 Lyme Road, Hanover, New Hampshire 03755-1290

Contributed Papers

1:30

4pPA1. Exact formula for the sound scattering cross section per unit volume in a turbulent atmosphere. Vladimir E. Ostashev (NOAA/Environ. Technol. Lab., 325 Broadway, Boulder, CO 80305 and Dept. of Phys., New Mexico State Univ., Las Cruces, NM 88003) and George H. Goedecke (New Mexico State Univ., Las Cruces, NM 88003)

The sound scattering cross section per unit volume is one of the most important statistical characteristics of a sound wave propagating in a turbulent atmosphere. In the literature, a formula for the sound scattering cross section is derived from an equation for a sound wave propagating in an atmosphere with temperature and velocity fluctuations. Such an equation is obtained from a complete set of linearized equations of fluid dynamics using some approximations. Thus, the classical formula for the sound scattering cross section is intrinsically approximate. In the present paper, sound scattering in a turbulent atmosphere is studied starting from the complete set of linearized equations of fluid dynamics. This approach results in an exact formula for the sound scattering cross section. The exact formula accounts for sound scattering by pressure fluctuations and by the divergence of velocity fluctuations while the classical formula does not. It follows from the exact formula that, in a dry air, a sound wave is scattered at 90 degrees only by pressure fluctuations. This result can be used as a basis for a new remote sensing technique for measuring pressure fluctuations in the atmosphere. [Work supported by the U.S. Army Research Office Grant No. DAAG19-01-1-0640.]

1:45

4pPA2. 3D finite-difference simulation of acoustic waves in turbulent moving media. Neill Symons, David Aldridge (Sandia Natl. Labs., Albuquerque, NM 87185-0750), D. Keith Wilson (U.S. Army Cold Regions Res. and Eng. Lab., Hanover, NH), David Marlin (U.S. Army Res. Lab., White Sands Missile Range, NM), and Vladimir Ostashev (NOAA Environ. Technol. Lab., Boulder, CO)

A finite-difference algorithm appropriate for modeling acoustic waves in a fully heterogeneous moving 3D media has been developed. The model is characterized by: acoustic velocity, density, and the three components of the background media velocity. The approach solves a set of coupled 1st order velocity-pressure differential equations appropriate for an adiabatic divergence-free background velocity. The equations are staggered in time and space and the algorithm uses second order temporal and fourth order spatial finite-differences. Since approximations are not adopted in the solution of the equations all arrivals are modeled with fidelity providing the spatial and temporal grids are chosen appropriately. The algorithm can include either a pressure or velocity free surface on the bottom boundary and absorbing boundaries on other model flanks. Designed to run on large scale parallel computational platforms, the algorithm has been validated for four machine architectures. Comparisons are presented to an analytic solution for a constant wind model and fast-field program results for a vertically stratified wind model. Data resulting from simulations through a kinematic turbulence wind profile developed with the quasi-wavelet method are also presented. Sandia National Laboratories is operated by Sandia Corporation, a Lockheed Martin Company, for the USDOE under Contract No. 94-AL85000.

2:00

4pPA3. Eddy-size decomposition of the scattering cross section using quasi-wavelets. D. Keith Wilson (U.S. Army Cold Regions Res. and Eng. Lab., 72 Lyme Rd., Hanover, NH 03755), Vladimir E. Ostashev (NOAA/Environ. Technol. Lab., 325 Broadway, Boulder, CO 80305), George H. Goedecke (New Mexico State Univ., Las Cruces, NM 88003), and Harry J. Auvermann (Dallas, TX 75228)

Unlike traditional Fourier spectral representations, quasi-wavelets (QWs) describe the structure of turbulence with spatially localized functions. As a result, QWs are particularly well suited to examining the dependence of a turbulent scattering process on the size and spatial location of the eddies. In this paper, the scattering cross section for QWs of an individual size class is derived. It is shown that previous results for the von Kármán spectrum can be reproduced when the scattered energies from a continuous distribution of QW sizes are combined. A Bragg resonance condition is derived for the eddy size that scatters most strongly for a given acoustic wavenumber and scattering angle. An example application of QWs to scattering over noise barriers shows that, for a typical barrier geometry, most of the scattered energy originates from eddies in the size range of approximately one-half to twice the size of the eddies responsible for maximum scattering.

2:15

4pPA4. Time-domain boundary conditions for outdoor ground surfaces. Sandra L. Collier (U.S. Army Res. Lab., ATTN: AMSRL-CI-EE, 2800 Powder Mill Rd., Adelphi, MD 20783-1197, scollier@arl.army.mil), Vladimir E. Ostashev (NOAA/Environ. Technol. Lab., Boulder, CO 80305-3328), D. Keith Wilson (Eng. Res. and Development Ctr., U.S. Army Cold Regions Res. and Eng. Lab., Hanover, NH 03755-1290), and David H. Marlin (U.S. Army Res. Lab, White Sands Missile Range, NM 88002-5501)

Finite-difference time-domain techniques are promising for detailed dynamic simulations of sound propagation in complex atmospheric environments. Success of such simulations requires the development of new techniques to accurately handle the reflective and absorptive properties of a porous ground. One method of treating the ground boundary condition in the time domain [Salomons *et al.*, Acta Acust. **88**, 483–492 (2002)] is to use modified fluid dynamic equations, where the ground is considered as a porous medium described by its physical properties. However, this approach significantly increases computation time, as the domain must be extended into the ground and a large number of grid points are needed. Standard impedance models for the ground boundary condition are frequency-domain models, which generally are non-causal [Y. H. Berthelot, J. Acoust. Soc. Am. **109**, 1736–1739 (2001)]. The development of a time-domain boundary condition from these models requires removing the singularity from the impedance equation when transforming from the frequency domain to the time domain. Alternatively, as the impedance boundary condition is a flux equation, a time-domain boundary condition can be derived from first principles, using the physical properties of the ground. We report on our development of a time-domain ground boundary condition.

2:30

4pPA5. Discussion of errors in the Green's function parabolic equation for outdoor sound propagation. Jennifer Cooper and David C. Swanson (Grad. Prog. in Acoust., Penn State Univ., 202 ASB, University Park, PA 16803, jlc375@psu.edu)

As is to be expected of any numerical method, use of the Green's function parabolic equation (GFPE) does result in small errors that can propagate and increase. The main causes of errors in using the GFPE for outdoor sound propagation with constant sound speeds appear to be inadequate spatial sampling, approximated starting fields, and errors in the numerical integration used to compute the surface wave. The relative significance of these errors as well as a set of appropriate computational parameters for avoiding or minimizing them will be discussed. With the proper parameter selection (for step sizes, attenuation layer depth, etc.), the GFPE produces reliably accurate results for outdoor sound propagation even over finite impedance grounds and beyond barriers or changes in ground height.

2:45

4pPA6. Modeling of secondary sonic booms: Influence of the variability of the atmosphere. Laurent Dallois, Philippe Blanc-Benon, and Julian Scott (Ctr. Acoustique—Ecole Centrale de Lyon, 69134 Ecully, France)

The shock waves generated by a supersonic aircraft are reflected in the upper part of the atmosphere. Back to the ground, they are indirect sonic booms called secondary sonic booms. The recorded signals of secondary sonic booms show a low amplitude and a low frequency. They sound like rumbling noises due to amplitude bursts. These signals strongly depend on the atmospheric conditions, in particular to the amplitude and to the direction of the wind in the stratopause. The propagation of the secondary sonic boom is studied using atmospheric models up to the thermosphere. By solving temporal ray equations, the secondary carpet position is investigated. An amplitude equation including nonlinearity, absorption, and relaxation by various chemical species is coupled to the ray solver in order to get information on the amplitude and on the frequency of the sonic boom at the ground level. Using this propagation model and the atmospheric model, the seasonal dependencies of the secondary sonic boom are investigated. Multipath arrivals are directly linked to wind field or 3-D inhomogeneities. They have been of special concern as a way to explain the rumble noise as a summation over different ray contributions.

3:00–3:15 Break

3:15

4pPA7. Attenuation of blast sound by a mixed stand of pine and hardwood. Michael J. White, Larry L. Pater, Ryan J. Lee, and George W. Swenson, Jr. (US Army ERDC/CERL, P.O. Box 9005, Champaign, IL 61826)

We performed an experiment to determine the attenuation of impulsive sound by a forest in northeastern Texas in July 2002. In the measurement, microphones were placed along a line that extended at one end into approximately 300 m of mostly 20-cm-diam pine trees, and at the other end into an open field of roughly the same extent. Explosive charges of 0.57- and 2.27-kg Composition C-4 were detonated at four locations along the line, at either edge of the open field and at either edge of the trees, in order to compare sound attenuation rates within the woods to those in the open field. Additional microphones were placed 2 and 4 km away to compare propagation from sources in either wooded or open positions to microphones in either wooded or open positions, with all paths substantially forested. Charge size, height-of-burst, and microphone height were varied in order to excite and probe a variety of acoustic propagation modes within the tree layer. We discuss preliminary results of 1/3-octave band analysis of this data set, consider ways to separately identify the observed power law and exponential decay rates, and speculate on their controlling mechanisms.

3:30

4pPA8. Forest physical sampling strategies for investigating noise mitigation benefits. Patrick J. Guertin and Michael J. White (US Army ERDC/CERL, P.O. Box 9005, Champaign, IL 61826)

Acoustic attenuation rates in tree stands have been seen to vary considerably between measurements in various settings, with little hint as to the controlling mechanisms. Usual descriptors for a forest include tree species, stand age, and basal diameter per land surface area, properties that may not correlate with acoustic attenuation rates. As part of a recent forest attenuation measurement, we developed a forest sampling regimen aimed at indicating and distinguishing between three candidate effects: (1) scattering by resilient, wood material, (2) scattering by more compliant, leaf

material, (3) refraction by canopy microclimate influence. Estimates of the wood material were made by an established point sampling process, in which significant tree trunks subtend a minimum angle at breast height. Canopy microclimate effect was characterized using percent optical closure of the upper canopy, canopy thickness, and ground vegetation height. Digital photographs aimed outward in four cardinal directions from each sample point were processed to segregate percentages of wood material, leaf material, and sky. We are currently exploring the connection between these parameters and measured attenuation rates in a single stand of woods. We propose that these parameters could offer improved correlation with measured attenuation rates between forests.

3:45

4pPA9. Sound propagation in an urban environment. I: Preliminary analysis of measurements. Donald G. Albert (US Army ERDC Cold Regions Res. and Eng. Lab., 72 Lyme Rd., Hanover, NH 03755, dalbert@crrel.usace.army.mil) and Lanbo Liu (US Army ERDC Cold Regions Res. and Eng. Lab., Hanover, NH 03755)

Experimental measurements were conducted in a full-scale artificial village to determine the effect of buildings on sound propagation outdoors. The village consisted of 15 concrete block buildings arranged in a 150 $\times$150 m area. Explosive charges were detonated to produce acoustic pulses that were digitally recorded by sensors scattered throughout the village. The measurements confirm that diffraction acts as a low pass filter on acoustic waveforms, greatly reducing the peak pressure received by a sensor in the shadow zone. The measured data present clear examples of the effect of reflections on propagation down a street canyon and diffraction around building corners. Complex signatures formed by multiple reflections and diffractions were recorded in situations where the acoustic waves passed multiple buildings in propagating from the source to the receiver. [Work funded by the U.S. Army.]

4:00

4pPA10. Sound propagation in an urban environment. II: Preliminary FDTD model. Lanbo Liu (US Army ERDC Cold Regions Res. and Eng. Lab., 72 Lyme Rd., Hanover, NH 03755, Lanbo.liu@erdc.usace.army.mil) and Donald G. Albert (US Army ERDC Cold Regions Res. and Eng. Lab., Hanover, NH 03755)

A 2-dimensional finite difference time domain (FDTD) method was used to simulate a data set recorded by a microphone array in an urban terrain. The FDTD model uses a coupled system of first-order differential equations for excess pressure and particle velocity. The perfectly matched layer method was used to truncate the model domain. This technique is much more efficient in diminishing unwanted reflections from the domain boundaries than any other absorptive boundary technique. The numerical integration was carried out with a staggered grid system. This algorithm has been validated with comparison with analytical solutions for cases with uniform physical parameters and simple geometries. For the predicted sound levels, although they are in general agreement with the measured values at 13 sensor locations, there is a significant discrepancy occurred at one particular sensor location (microphone location 9) behind one building (Building A). The observed sound level is about an order of magnitude greater than the predicted level, suggesting that much sound energy has propagated through this building. In contrast, a general consistency, or comparable level, can be reached between the measured and the modeled sound levels within 2-fold for the remaining 13 stations. [Work funded by the US Army.]

4:15

4pPA11. Wave propagation in viscoelastics with chemical relaxation. Timothy Margulies (Dept. of Mathematics, Community College of Philadelphia, 1700 Spring Garden St., Philadelphia, PA 19130)

Acoustic wave properties of fluids have provided a fruitful qualitative and quantitative analysis of fast chemical reactions. The purpose of this paper is to present an extension of the linear acoustic theory of Verdier and Piau [J. Acoust. Soc. Am. **101**, 4 (1997)] to include simultaneous chemical kinetic relaxation. The energy equation, equation of state with thermodynamic relations are formulated to account for chemical reactions. A general wave equation for a compressible viscoelastic with thermal transfer and chemical progress variable perturbations are explicitly modeled. Longitudinal wave propagation is described by a biquadratic equation for the sound absorption and dispersion. Approximations of small thermal transfer and viscous effects yield linearly additive contributions of viscous, thermal, and reaction contributions. Here, linearly viscous or visco-elastic effects can be examined, in addition to the chemical kinetic effects. Sample calculations illustrating the physio-chemical dynamics will be presented.

4:30

4pPA12. Wave propagation in viscoelastic horns using a fractional calculus rheology model. Timothy Margulies (Dept. of Mathematics, Community College of Philadelphia, 1700 Spring Garden St., Philadelphia, PA 19130)

The complex mechanical behavior of materials are characterized by fluid and solid models with fractional calculus differentials to relate stress and strain fields. Fractional derivatives have been shown to describe the viscoelastic stress from polymer chain theory for molecular solutions [Rouse and Sittel, J. Appl. Phys. **24**, 690 (1953)]. Here the propagation of infinitesimal waves in one dimensional horns with a small cross-sectional area change along the longitudinal axis are examined. In particular, the linear, conical, exponential, and catenoidal shapes are studied. The wave amplitudes versus frequency are solved analytically and predicted with mathematical computation. Fractional rheology data from Bagley [J. Rheol. **27**, 201 (1983); Bagley and Torvik, J. Rheol. **30**, 133 (1986)] are incorporated in the simulations. Classical elastic and fluid ''Webster equations'' are recovered in the appropriate limits. Horns with real materials that employ fractional calculus representations can be modeled to examine design trade-offs for engineering or for scientific application.

4:45

4pPA13. TWA Flight 800, explosion airblast unexplained. Jack W. Reed (JWR, Inc., 1128 Monroe SE, Albuquerque, NM 87108, jwreed@nmia.com)

TWA Flight 800 disintegrated off Long Island, NY, on 16 July 1996. Immediate reports from other flyers described what appeared as attacking missiles. Search for terrorists began quickly, with over 1000 FBI agents to collect debris, interview eyewitnesses, and analyze sightings to give a missile launch point. They found no evidence of criminal attack, and turned investigations over to the NTSB to find some accidental cause. The ''empty'' central fuel tank was determined to be the likely explosion source. On the other hand, early witnesses reported a ''loud'' bang after seeing a great fireball fall from the sky, but at 15 km or greater range, they saw and heard two different events. This acoustic discrepancy has not been adequately investigated. When finally released, FBI reports from more than 200 ''ear-witnesses'' give similar observations. Their loudness reports confirm that at least a ton of TNT equivalent explosion had occurred. NASA acousticians engaged by NTSB, however, through spectral analysis techniques for sonic booms, concluded that a 10-kg TNT explosion, from detonating fuel tank vapors, could be heard on Long Island. Evidence for a much larger yield is presented here, but its form remains a mystery.

Session 4pSA

Structural Acoustics and Vibration: Sound Field Reconstructions

Courtney B. Burroughs, Chair
Applied Research Laboratory, Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804-0030

Contributed Papers

2:00

4pSA1. Application of multisource discrimination to the HELS method. Nassif Rayess (Dept. of Mech. Eng., Univ. of Detroit–Mercy, Detroit, MI 48219), Manmohan Moondra, and Sean Wu (Wayne State Univ., Detroit, MI 48202)

In most engineering applications, the acoustic fields are usually generated by multiple incoherent sources. When such acoustic fields are sampled and taken as input data to the Helmholtz equation least-squares (HELS) formulations directly, the reconstruction might not be accurate. Hence there is a need to discriminate the contributions from individual sources and separate the composite sound field into a set of spatially coherent subfields that are also mutually incoherent. In this paper, we apply the principal component decomposition technique to the HELS method to reconstruct acoustic radiation from multiple incoherent sources in a free field. The key ingredients of this technique include a diagonalization by singular value decomposition (SVD) of the cross-spectral matrices generated by a number of reference microphones. Each diagonal term corresponds to a subfield that results from a so-called virtual sound source. Even though such virtual sources are not always representative of the actual sources, they are nonetheless incoherent and thus can be used as input to HELS. Experimental validations of this technique on reconstructing acoustic pressure and time-averaged normal acoustic intensity on the surfaces of multiple incoherent sources are presented. [Work supported by NSF.]

2:15

4pSA2. Analysis and comparison of three energy density probe designs. Lance L. Locey and Scott D. Sommerfeldt (N283 Eyring Sci. Ctr. (ESC), Provo, UT 84602, Lance@Locey.com)

Previous research has demonstrated the utility of acoustic energy density measurements as a means to gain a greater understanding of acoustic fields. Three energy density probes are under development. The first probe has three orthogonal pairs of microphones embedded on the surface of a sphere. The second design is a similarly sized sphere with four surface mounted microphones, equidistant from the center of the sphere. The four microphones are arranged to correspond with the origin and unit vectors of a Cartesian coordinate system, where the origin and the tips of the three unit vectors are on the surface of the sphere. As a result, all four microphones lie on the surface of the top hemisphere. The third design consists of a similarly sized sphere with four surface microphones arranged in a tetrahedron configuration. The author will discuss some of the errors and limitations associated with the four microphone designs as compared to the six microphone design. [Work supported by NASA.]

2:30

4pSA3. Helmholtz equation least squares method for transient nearfield acoustic holography. Huancai Lu and Sean Wu (Dept. of Mech. Eng., Wayne State Univ., Detroit, MI 48202)

The generalized Helmholtz equation least squares (HELS) formulations for reconstructing transient acoustic radiation from an arbitrary object subject to an arbitrary time-dependent excitation are derived. To facilitate the derivations, the Laplace transform is employed and the vibro-acoustic quantities on the source surface are solved explicitly in terms of the acoustic pressures measured on a conformal surface around the source at close range multiplied by transfer functions in the Laplace domain first. The vibro-acoustic responses in the time domain can then be expressed as convolution integrals of the measured acoustic pressure signals over temporal kernels. Replacing the spherical Hankel functions in the transfer functions with polynomial expressions, we can recast the infinite integrals in the inverse Laplace transform as contour integrals and evaluate the temporal kernels by using residue theorem. Once the temporal kernels are determined, the vibro-acoustic quantities anywhere in the field, including those on the source surface can be reconstructed directly. Numerical examples of reconstruction of transient acoustic radiation from a baffled disk subject to impulsive and arbitrarily time-dependent excitations are demonstrated. [Work supported by NSF.]

2:45

4pSA4. Reconstruction of acoustic radiation from vibrating objects in a half space using hybrid nearfield acoustical holography. Xiang Zhao and Sean F. Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202, xiangzhao@wayne.edu)

Hybrid nearfield acoustical holography (NAH) [Wu, J. Acoust. Soc. Am. **113**, 2252 (2003)] is utilized to reconstruct the vibro-acoustic fields generated by vibrating structures in half space. These hybrid formulations are derived from a modified Helmholtz equation least-squares method, where the field acoustic pressure is expanded in terms of both outgoing and incoming spherical waves, and the Helmholtz integral formulations implemented by the boundary element method (BEM). To overcome the ill-conditioning difficulties in the resultant matrix equations, regularization techniques such as the method of Tikhonov and generalized cross validation are employed. The main advantage of this hybrid approach is that satisfactory reconstruction can be obtained with relatively few acoustic pressure measurements taken over a conformal surface around the source. This is because a large portion of the input data is regenerated by the modified HELS, but not actually measured. Hence, the efficiency of reconstruction is enhanced. If the same number of measurements is used in an inverse BEM code, the reconstructed vibro-acoustic quantities on the source surface may be distorted. Numerical examples of a dilating and an oscillating sphere and a cylinder with spherical endcaps in half space are demonstrated. [Work supported by NSF.]

3:00

4pSA5. On the reconstruction of sound field using the equivalent source modeling. Jeong-Guon Ih and In-Youl Jeon (Dept. of Mech. Eng., Korea Adv. Inst. of Sci. and Technol., Science Town, Taejon 305-701, Korea)

The sound field radiated from a vibrating structure is reconstructed by using the equivalent source modeling. In this method, the equivalent multipole sources which are similar to the spherical harmonics in the Helmholtz equation least-squares (HELS) method are distributed over or inside a vibrating surface. To improve the accuracy and usefulness, the effective independence (EfI) method is used to determine the optimal position of equivalent sources and the regularization scheme is adopted to determine the optimal expansion number of spherical harmonics and to suppress the high order poles. After regenerating field pressures using those optimized equivalent source modeling, the vibrating source is effectively recon-

structed with minimal measurement combining the proposed method with the near-field acoustic holography (NAH) using the inverse boundary element method (BEM). In this study, an application example of a commercial vacuum cleaner using the equivalent source modeling will be demonstrated. [Work supported by the BK21 project and NRL.]

3:15

4pSA6. Reconstruction of the surface normal acoustic intensity from noisy pressure data by regularized Helmholtz equation least squares method. Tatiana Semenova and Sean F. Wu (Dept. of Mech. Eng., Wayne State Univ., Detroit, MI 48202)

Reconstructing the normal acoustic intensity on a source surface from the field pressure measurements is critical in identifying the noise sources and their transmission paths. In this paper, we consider the reconstruction of time-averaged normal acoustic intensities from noisy pressure data using the Helmholtz equation least squares (HELS) method with regularization by truncated singular value decomposition or the method of Tikhonov. The regularization parameter is calculated by generalized cross validation, L-curve, and quasioptimality criterion and the best solution is chosen. Numerical results show that the normal velocity reconstruction is more sensitive to noise in the input data than the pressure reconstruction, and depends critically on the choice of the regularization parameter. It is found that the HELS-based Dirichlet-to-Neumann (DtN) map has a similar accuracy as compared to the Green's formula-based DtN map with Tikhonov regularization for 1% noise or larger in the input data, provided that HELS converges sufficiently fast. Moreover, determining the surface normal velocity from the reconstructed surface pressure is more stable than reconstructing it directly from the field pressure data. It is concluded that a regularized HELS, when it converges fast enough, can be an easy-to-use and cost-effective tool for intensity reconstruction [Work supported by NSF.]

THURSDAY AFTERNOON, 13 NOVEMBER 2003 SAN ANTONIO ROOM, 1:00 TO 5:10 P.M.

Session 4pSC

Speech Communication: Statistical Patterns in Speech

Andrew J. Lotto, Cochair
Department of Psychology, University of Texas–Austin, 1 University Station A8000, Austin, Texas 78712-0187

Lori L. Holt, Cochair
Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213

Randy L. Diehl, Cochair
Department of Psychology, University of Texas–Austin, 1 University Station A8000, Austin, Texas 78712-0187

Chair's Introduction—1:00

Invited Papers

1:05

4pSC1. What are the statistics in statistical learning? Lori L. Holt (Dept. of Psych., Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213, lholt@andrew.cmu.edu) and Andrew J. Lotto (Boys Town Natl. Res. Hospital, Omaha, NE 68131)

The idea that speech perception is shaped by the statistical structure of the input is gaining wide enthusiasm and growing empirical support. Nonetheless, statistics and statistical learning are broad terms with many possible interpretations and, perhaps, many potential underlying mechanisms. In order to define the role of statistics in speech perception mechanistically, we will need to more precisely define the statistics of statistical learning and examine similarities and differences across subgroups. In this talk, we examine learning of four types of information: (1) acoustic variance that is defining for contrastive categories, (2) the correlation between acoustic attributes or linguistic features, (3) the probability or frequency of events or a series of events, (4) the shape of input distributions. We present representative data from online speech perception and speech development and discuss inter-relationships among the subgroups. [Work supported by NSF, NIH and the James S. McDonnell Foundation.]

1:30

4pSC2. Bayesian approaches to perception. Randy L. Diehl and Sarah C. Sullivan (Dept. of Psych. and Ctr. for Perceptual Systems, Univ. of Texas, 1 University Station A8000, Austin, TX 78712-0187)

The concepts of Bayesian statistical decision theory have recently transformed research in perception by providing a rigorous mathematical framework for representing the statistical properties of the environment, for describing the tasks that perceivers perform, and for deriving computational theories of optimal performance on those tasks. Such computational theories (ideal observers) offer an appropriate benchmark for evaluating human performance, and, moreover, they can be modified to yield an excellent starting point for

developing testable models. Unlike theories that focus on the role of invariant information for perceived events, the Bayesian approach treats information as probabilistic. The approach is illustrated for cases in which listeners learn to categorize artificial stimulus sets whose statistical properties (prior probabilities and stimulus likelihoods) are controlled by the experimenter. Prospects for extending the Bayesian framework to the analysis of speech perception are discussed in light of the recent progress in developing ideal observers for natural environmental stimuli. A necessary but daunting task will be the measurement of probability distributions that characterize natural speech categories. [Work supported by NIDCD.]

1:55

4pSC3. Explicit pattern recognition models for speech perception. Terrance M. Nearey (Linguist., Univ. of Alberta, Edmonton, AB T6G 0A2, Canada)

Optimal statistical classification of arbitrary input signals can be obtained, in principle, via a Bayesian classifier, given (perfect) knowledge of the distributions of signal properties for the set of target categories. At least for certain constrained problems, such as the perception of isolated vowels, simple (imperfect) statistical pattern recognition techniques can accurately predict human listeners' performance. This paper sketches several relatively successful case studies of the application of static pattern recognition techniques to speech perception. (Static techniques require inputs of a fixed length, e.g., $F1$ and $F2$ for isolated vowels.) Real speech clearly requires dynamic pattern recognition, allowing inputs of arbitrary length. Certain such methods, such as dynamic programming and hidden Markov models, have been widely exploited in automatic speech recognition. The present paper will describe initial attempts to apply variants of such methods to the data from a perception experiment [T. Nearey and R. Smits, J. Acoust. Soc. Am. **111** (2002)] involving the perception of three (VCV) or four (VCCV) segment strings. Practical and conceptual problems in the application of such techniques to human perception will be discussed. [Work supported by SSHRC.]

2:20

4pSC4. Early language acquisition: Statistical learning and social learning. Patricia K. Kuhl (Ctr. for Mind, Brain & Learning, Univ. of Washington, Seattle, WA 98195)

Infants are sensitive to the statistical patterns in language input, and exposure to them alters phonetic perception. Our recent data indicate that first-time exposure to a foreign language at 9 months of age results in learning after only 5 h, suggesting a process that is fairly automatic, given natural language input. At the same time, it appears that early phonetic learning from natural language may be constrained by the need for social interaction. Our work demonstrates that infants learn phonetically when exposed to a live, but not a pre-recorded, speaker. This talk will focus on statistical learning in a social context and develop the thesis that this combination provides an ideal situation for the acquisition of a natural language.

2:45–3:00 Break

3:00

4pSC5. Distributional and statistical bases of allophonic groupings. James L. Morgan, Katherine S. White, Cecilia Kirk (Cognit. & Linguist. Sci., Brown Univ., Providence, RI 02912, James_Morgan@Brown.Edu), Sharon Peperkamp, and Emmanuel Dupoux (Ecole Normale Superieure, 75005 Paris, France)

Segments of a given language do not all have equal status: some are phonetic variants of others. These variants occur only in certain contexts. In addition to occurring in more restricted contexts, however, variant allophones typically occur with lesser frequency. Which (if either) of these asymmetries is important for infants learning of allophonic groupings? Infants were exposed to artificial languages with allophones. Either stops or fricatives were intervocalicly voiced; these voiced segments did not otherwise occur. Infants were then tested on preference for passages like ''rot pazo na bazo rot pazo na bazo'' or ''na zine rot sine na zine rot sine . . . '' All items were grammatical for both groups. For the stop allophone group, /pazo/ and /bazo/ should be one word, whereas /zine/ and /sine/ should be two words; *vice versa* for the fricative allophone group. In previous work using a similar technique, infants listened longer to passages containing what was perceived as two variants. Across a series of experiments, the number and frequency of contexts in which allophonic variants occurred were manipulated. Cross-experiment analyses reveal the relative contributions of distributional and statistical asymmetries. We interpret and discuss our findings in light of current statistical learning approaches to language acquisition.

3:25

4pSC6. Using pronunciation data to study word recognition. Mark A. Pitt (Dept. of Psych., Ohio State Univ., Columbus, OH) and Keith Johnson (Ohio State Univ., Columbus, OH)

Many of the mysteries of spoken word recognition have evolved out of the observation that pronunciation is highly variable yet perception is amazingly stable. Extreme forms of variation can even result in different phonetic percepts, such as consonant assimilation (e.g., ''green ball:'' ''greem ball'') and deletion (e.g., ''and:'' ''an''), yet listeners still perceive the intended word. The results from two lines of work will be presented in which we are studying the regularity of these production phenomena to evaluate models

of how phonological variants are recognized. In one project we examined variation itself. Phonological and acoustic analyses of phonological variation in the Buckeye corpus of conversational speech were carried out in which we asked questions such as the following: How predictable and consistent is regressive assimilation? and How acoustically similar is the assimilated segment to an intended production of that same segment? In a related project, we examined the listener's sensitivity to the variation found in conversational speech. Results reveal the complexities of these production phenomena, the challenges models must overcome to account for how variants are recognized, and listeners' sensitivity to stochastic properties of pronunication variation. [Work supported by NIDCD.]

3:50

4pSC7. Phonotactic probability and neighborhood density effects in the perception and production of speech in adults. Michael S. Vitevitch, Jonna Armbruster, and Julia Fitzer (Dept. of Psych., 1415 Jayhawk Blvd., Univ. of Kansas, Lawrence, KS 66045, mvitevit@ku.edu)

Phonotactic probability refers to the frequency with which phonological segments and sequences of segments occur in a language. Neighborhood density refers to the number of words that sound similar to a given word. Awareness to phonotactic probability and neighborhood density occurs early in life and seems to influence lexical development. Recent research in our lab suggests that these two variables also influence the on-line processing of spoken words in adults. In the case of spoken word recognition, when sub-lexical representations dominate processing, high probability stimuli are responded to more quickly than low probability stimuli. However, when lexical representations dominate processing, stimuli with sparse neighborhoods are responded to more quickly than stimuli with dense neighborhoods. Similarly, in speech production, pictures of high probability words are named more quickly than pictures of low probability words. However, in contrast to speech perception, pictures of words with dense neighborhoods are produced more quickly than pictures of words with sparse neighborhoods, suggesting that statistical regularities in the language may be used differently depending on the demands imposed by a given process or task.

4:15

4pSC8. Statistical learning and the challenge of syntax: Beyond finite state automata. Jeff Elman (Dept. of Cognit. Sci., Univ. of California–San Diego, La Jolla, CA 92093)

Over the past decade, it has been clear that even very young infants are sensitive to the statistical structure of language input presented to them, and use the distributional regularities to induce simple grammars. But can such statistically-driven learning also explain the acquisition of more complex grammar, particularly when the grammar includes recursion? Recent claims (e.g., Hauser, Chomsky, and Fitch, 2002) have suggested that the answer is no, and that at least recursion must be an innate capacity of the human language acquisition device. In this talk evidence will be presented that indicates that, in fact, statistically-driven learning (embodied in recurrent neural networks) can indeed enable the learning of complex grammatical patterns, including those that involve recursion. When the results are generalized to idealized machines, it is found that the networks are at least equivalent to Push Down Automata. Perhaps more interestingly, with limited and finite resources (such as are presumed to exist in the human brain) these systems demonstrate patterns of performance that resemble those in humans.

Contributed Papers

4:40

4pSC9. The effects of distributional information on the formation of nonspeech sound categories by infants. Jessica F. Hay (Dept. of Psych., Univ. of Texas, Austin, TX 78712, hay@psy.utexas.edu)

In order to learn native sound categories, infants may be attending to the distributional information available in their native language environment. The present study examined eight-month-old infants ability to form categories with small amounts of exposure to distributional information. Infants were exposed to a bimodal distribution of nonspeech sounds (narrow-band noise bursts that vary in center frequency). Their ability to discriminate within (same mode) and between category (different mode) differences was then assessed through a habituation-dishabituation procedure. Data were interpreted in terms of dishabituation to change stimuli. The change stimuli were equal in physical step size and either came from within the same mode of the bimodal distribution or from the other mode. The findings may have implications regarding auditory category formation in general and could lend support to the hypothesis that infants use distributional information available in their ambient language environment to learn their native speech categories. [Work supported by NIDCD.]

4:55

4pSC10. Statistical modeling of infant-directed versus adult-directed speech: Insights from speech recognition. Katrin Kirchhoff and Steven Schimmel (Dept. of Elec. Eng., Univ. of Washington, Seattle, WA 98115)

Studies on infant speech perception have shown that infant-directed speech (*motherese*) exhibits exaggerated acoustic properties, which are assumed to guide infants in the acquisition of phonemic categories. Training an automatic speech recognizer on such data might similarly lead to improved performance since classes can be expected to be more clearly separated in the training material. This claim was tested by training automatic speech recognizers on adult-directed (AD) versus infant-directed (ID) speech and testing them under identical versus mismatched conditions. 32 mother–infant conversations and 32 mother–adult conversations were used as training and test data. Both sets of conversations included a set of cue words containing unreduced vowels (e.g., sheep, boot, top, etc.), which mothers were encouraged to use repeatedly. Experiments on continuous speech recognition of the entire data set showed that recognizers

trained on infant-directed speech did perform significantly better than those trained on adult-directed speech. However, isolated word recognition experiments focusing on the above-mentioned cue words showed that the drop in performance of the ID-trained speech recognizer on AD test speech was significantly smaller than vice versa, suggesting that speech with over-emphasized phonetic contrasts may indeed constitute better training material for speech recognition. [Work supported by CMBL, University of Washington.]

THURSDAY AFTERNOON, 13 NOVEMBER 2003 SAN MARCOS ROOM, 2:15 TO 5:00 P.M.

Session 4pSP

Signal Processing in Acoustics, Underwater Acoustics, Speech Communication, Animal Bioacoustics, Noise and Engineering Acoustics: Detection and Classification in Acoustics IV

Patrick J. Loughlin, Chair
Department of Electrical Engineering, University of Pittsburgh, 348 Benedum Hall, Pittsburgh, Pennsylvania 15261

Invited Papers

2:15

4pSP1. Detection and classification of underwater targets by echolocating dolphins. Whitlow Au (Marine Mammal Res. Prog., Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734)

Many experiments have been performed with echolocating dolphins to determine their target detection and discrimination capabilities. Target detection experiments have been performed in a naturally noisy environment, with masking noise and with both phantom echoes and masking noise, and in reverberation. The echo energy to rms noise spectral density for the Atlantic bottlenose dolphin (*Tursiops truncatus*) at the 75% correct response threshold is approximately 7.5 dB whereas for the beluga whale (*Delphinapterus leucas*) the threshold is approximately 1 dB. The dolphin's detection threshold in reverberation is approximately 2.5 dB vs 2 dB for the beluga. The difference in performance between species can probably be ascribed to differences in how both species perceived the task. The bottlenose dolphin may be performing a combination detection/discrimination task whereas the beluga may be performing a simple detection task. Echolocating dolphins also have the capability to make fine discriminate of target properties such as wall thickness difference of water-filled cylinders and material differences in metallic plates. The high resolution property of the animal's echolocation signals and the high dynamic range of its auditory system are important factors in their outstanding discrimination capabilities.

3:00

4pSP2. Speech enhancement for better hearing aid design using chirped wavelets. Adele Doser (Sandia Natl. Labs., P.O. Box 5800, M.S. 1188, Albuquerque, NM 87185-1188)

Wavelet techniques are utilized in the enhancement of speech signals for hearing-impaired persons. The interest is in schemes that isolate and enhance short duration, broadband speech elements (such as unvoiced, fricative consonants) while still capturing narrowband speech components (such as vowels). Discerning unvoiced consonants presents an obstacle for hearing-impaired persons (i.e., ''bark'' and ''park'' may sound identical). A scheme is proposed, using chirped wavelets, to seize the characteristics of broadband speech elements. The method is tested on speech data, in noise-free and noisy conditions, and compared with sinusoidal-based wavelets. Results demonstrate that chirped wavelets are useful in the analysis and enhancement of speech for the hearing impaired. It is hoped these techniques might be incorporated in future hearing aid design.

3:30

4pSP3. Local characteristics of dispersive pulse propagation. Patrick J. Loughlin (Dept. of Elec. Eng., Univ. of Pittsburgh, 348 Benedum Hall, Pittsburgh, PA 15214)

As a wave propagates, its shape and other physical characteristics may change, depending on the initial wave and the propagation environment. In this talk, the effects of structural dispersion on the spatial and temporal spreading of a propagating pulse wave, and on the spectral characteristics of the pulse over time, are considered. These effects are quantified by obtaining the local spatial, temporal, and spectral moments of the pulse. General results for any dispersion relation, as well as the particular case of a two-plate waveguide, are presented. [Work supported by ONR.]

4:00–4:15 Break

4:15

4pSP4. Material discrimination using bispectral signatures. Paul A. Nyffenegger and Melvin J. Hinich (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

A method is presented for material discrimination and characterization using bispectral signatures acquired from an object actively probed with acoustic pulses. Although bispectral techniques have proven useful in a diverse array of fields including passive acoustic ranging, bispectral processing in active acoustic applications has not been widely explored. The mechanisms responsible for the bispectral signatures revealed using active acoustics have not been well studied and little is known about the relative contributions to the bispectrum originating in the physical properties of the target material itself rather than from target structural acoustics and the propagation media. In a pilot experiment, we determine bispectral signatures for three targets of differing composition but similar dimensions using a submerged ultrasonic apparatus. The experiment is designed to isolate effects due to target properties from those attributable to propagation path, source, or receiver. The source wavelet is a broad-spectrum linear frequency modulated pulse. The normalized bispectrum is calculated using conventional nonparametic methods, and is averaged across many frames. Results indicate that at ultrasonic frequencies this technique provides signatures with the potential of discriminating between classes of materials such as plastic, metal, and rock. [Work supported by Applied Research Laboratories.]

4:30

4pSP5. Spatio-temporal gradient signal processing for reconstructing scattered wave fields. Kenbu Teramoto and Kohsuke Tsuruta (Dept. of Mech. Eng., SAGA Univ., Saga-shi 8408502, Japan, tera@me.saga-u.ac.jp)

This paper presents a subspace signal processing method which detects the scattered sound waves based on the spatio-temporal gradient analysis over the Lamb wave field. The proposed method has an ability to classify the surface acoustic wave field through the rank of the covariance matrix defined over the four-dimensional vector space which is spanned by the following components: a vertical displacement, its vertical velocity, and a pair of shear strains of the plate surface. The covariance matrix provides the information about cracks. When a unique sound source exists on the surface without any cracks or scatterers, the rank of the covariance matrix is reduced to two. Increasing the rank up to three, however, one or more scatterers exist on the surface. Therefore, by the singular value decomposition (SVD) of the covariance matrix, optimal, orthonormal bases are obtained. In formulating the scattered wave field by the unknown object, its best reconstruction, in the LMS sense, can be obtained by the sum of orthogonal vector spaces. The computational process in a wave field near the cracks is discussed and their physical meanings are investigated through DTD-simulations and acoustic experiments.

4:45

4pSP6. Constrained maximum consistency multi-path mitigation. George B. Smith (Naval Res. Lab., Code 7185, Stennis Space Center, MS 39529-5004)

Blind deconvolution algorithms can be useful as pre-processors for signal classification algorithms in shallow water. These algorithms remove the distortion of the signal caused by multipath propagation when no knowledge of the environment is available. A framework in which filters that produce signal estimates from each data channel that are as consistent with each other as possible in a least-squares sense has been presented [Smith, J. Acoust. Soc. Am. **107** (2000)]. This framework provides a solution to the blind deconvolution problem. One implementation of this framework yields the cross-relation on which EVAM [Gurelli and Nikias, IEEE Trans. Signal Process. **43** (1995)] and Rietsch [Rietsch, Geophysics **62**(6) (1997)] processing are based. In this presentation, partially blind implementations that have good noise stability properties are compared using Classification Operating Characteristics (CLOC) analysis. [Work supported by ONR under Program Element 62747N and NRL, Stennis Space Center, MS.]

THURSDAY AFTERNOON, 13 NOVEMBER 2003 PECOS ROOM, 2:15 TO 5:00 P.M.

Session 4pUWa

Underwater Acoustics: Seabed Acoustics

Marcia J. Isakson, Chair
Applied Research Laboratory, University of Texas–Austin, 10000 Burnet Road, Austin, Texas 78758

Contributed Papers

2:15

4pUW1. Volume scattering from random porous media. Keiichi Ohkawa (Appl. Marine Phys. Div. RSMAS, Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149)

The new volume scattering model, which enables us to treat the scattering of Biot's slow compressional wave from random porous media, is derived applying the Born approximation to Biot's equations of motion. Within the framework of the Biot theory it is assumed that the fluid-saturated unconsolidated sediment has low values of frame bulk and shear moduli relative to the other moduli of the medium and the shear wave is negligible. This enables us to treat the Biot theory easier. The equations of motion in inhomogeneous media are then simplified and coupled Helmholtz equations for compressional waves are derived applying the mode's decoupling method to the simplified equations of motion. The Born approximation is applied to the coupled Helmholtz equations in random media. The derived equations can be treated as extended forms of fluid bottom models. In this model, the 3-D power spectrum is inferred from porosity and permeability fluctuations instead of velocity and density fluctuations generally used. The permeability fluctuation can be estimated from grain-size distribution. In the limit of low frequency or high porosity, the results coincide with the fluid bottom model [T. Yamamoto, J. Acoust. Soc. Am. **99**, 866–879 (1996)]. The differences of this new model and others will be discussed.

2:30

4pUWa2. Field measurements of attenuation and permeability of shallow water sediments. Tokuo Yamamoto (Geoacoustics Lab., AMP, RSMAS, Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149) and Junichi Sakakibara (JFE Civil Corp.)

The velocity and attenuation of sands and clays in shallow oceans were measured by cross-hole tomography experiments using the pseudo-random binary sequence (PRBS) as source signals with the PRBS frequency of 1 and 3 kHz. Using measured velocity images, 2D velocity and density fluctuation spectra were calculated. The acoustic attenuation due to scattering by velocity and density fluctuations in the sediment volume was calculated using the measured velocity and density fluctuation spectra. The attenuation due to scattering occupies 30 to 98% of the total attenuation measured during the field experiments. The intrinsic attenuations of sands and clays were 0.3 to 1.5 dB/m/kHz and 0.01 to 0.02 dB/m/kHz, respectively. The permeability of the sands and the clays inverted by the Biot theory is 2.0 to 14.0 darcy and 0.1 to 120 md, respectively. The acoustically inverted permeability values agreed very well with the direct measurements of permeability of the sediments. The experimental results indicate that the acoustic attenuation measured in the fields is mainly due to volume scattering and that the intrinsic attenuation is due to the Biot mechanism of pore-fluid within the sediments. [Work supported by ONR Code 321OA and Kansai Power and Light Co.]

2:45

4pUWa3. Reflection and transmission coefficients in terms of pressure for Biot's compressional waves and the importance of the shear modulus for Biot's slow wave. Keiichi Ohkawa (Appl. Marine Phys. Div. RSMAS, Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149)

The method to determine reflection and transmission coefficients in terms of pressure at the fluid–sediment interface is presented. In general, reflection and transmission coefficients derived from the Biot theory have been computed using displacement potentials. Within the framework of the Biot theory it is assumed that the fluid-saturated unconsolidated sediment has low values of frame bulk and shear moduli relative to the other moduli of the medium and the shear wave is negligible. This enables to treat the Biot theory easier. By introducing effective densities, reflection and transmission coefficients can be treated as if the medium is a fluid. This approach using the models decoupling method is fairly different from the effective density fluid model [K. Williams, J. Acoust. Soc. Am. **110**, 2276–2281 (2001)]. It is shown that pressure coefficients give the same results as computed using displacement potentials. This method gives the advantage of directly computing the pressure for problems of reflection, transmission, and scattering. However, boundary conditions should be carefully treated. The pressure coefficient for Biot's slow wave cannot be accurately calculated when the equilibrium of solid pressure at the interface is constrained instead of the normal traction. This disagreement demonstrates that the shear modulus plays an important role in the generation of the slow wave.

3:00

4pUWa4. Spherical wave effects in sediment reflection coefficients. H. John CaminIII and Marcia J. Isakson (Appl. Res. Labs., Univ. of Texas, 10000 Burnet Rd., Austin, TX 78758, misakson@arlut.utexas.edu)

Several research institutions are investigating models to properly describe high frequency, sandy-sediment, acoustic properties because of their importance in shallow water operations. In this talk we will focus on the error contribution of spherical wave effects on each of four common models: the visco-elastic model, the Effective Density Fluid Model (EDFM), Buckinghams micro-sliding model and the Biot/Stoll poro-elastic model. This will be accomplished by comparing the known models of reflection coefficients to laboratory measurements. Theoretical data will be calculated for each of the models using OASES, which utilizes the depth-dependent Green's function and a direct global matrix approach to calculate the full field. Parameters for the visco-elastic model have been measured and verified at the test location. Parameters for the EDFM, Buckingham and Biot/Stoll models were determined from measured values and previous inversions of plane wave reflection data taken at the same location. Air/water interface data will be measured and computed for each model as a test case. Linearly Frequency Modulated (LFM) chirps over the range of 30–160 kHz will be used to establish ranges in frequency. [Work supported by ONR, Ocean Acoustics.]

3:15

4pUWa5. A comparison of high-frequency plane-wave reflection coefficient data taken from a smooth sand/water interface with current models. Andrew Worley and Marcia Isakson (Appl. Res. Labs., Univ. of Texas, 10000 Burnet Rd., Austin, TX 78758-4423, adworley@arlut.utexas.edu)

Dispersion and attenuation data have shown that a fluid model or visco-elastic model is not adequate to describe acoustic interactions in littoral environments. This study will compare laboratory reflection data to five models: the visco-elastic model, the Biot–Stoll poro-elastic model, the effective density fluid model (EDFM), the Buckingham micro-sliding model, and the Biot–Stoll squirt flow + shear model (BISQS). The BISQS model allows for frequency dependence in the frame bulk modulus and the frame shear modulus by considering the squirt flow effects and grain shearing effects at the sand grain interfaces. This model has been shown to produce good agreement with *in situ* dispersion and attenuation data; however, it has not yet been compared to reflection data. A simulated annealing inversion algorithm with rotated coordinates will be used. Special emphasis has been placed on quantifying the effects of the statistical nature of the data on the inversion outcome. [Work supported by ONR, Undersea Signal Processing, John Teague.]

3:30–3:45 Break

3:45

4pUWa6. Experimental study of the poroelastic behavior of a sandy sediment. Michael W. Yargus (Naval Surface Warfare Ctr., Carderock Div., Detachment Bremerton, 530 Farragut Ave., Bremerton, WA 98314-5215) and Darrell R. Jackson (Univ. of Washington, Seattle, WA 98105)

Reflection measurements were conducted on water-saturated sand over the frequency range 200–300 kHz. Three reflection boundaries were used: water-sand, pressure-release-sand, and acrylic-sand. A reflection ratio technique was used to remove unknown calibration and geometric factors, permitting measurement of parameters relating to the Biot fast wave and bounding of parameters of the slow wave. The interpretation of these results was facilitated by use of a mechanical model. Measured (or bounded) parameters include acoustic impedances, effective densities, wave speeds (phase velocities), effective pressures, fluid-frame displacement ratios, pressure reflection coefficients, and material moduli. The acoustic impedance divided by the phase velocity provides the ''effective density'' [K. L. Williams, J. Acoust. Soc. Am. **110**, 2956–2963 (2001)] for the fast wave. As expected, the effective density was less than the total density of the sediment (effective density = 89% of total). The fluid-frame displacement ratio was 2.2 for the fast wave. These results provide strong evidence for the importance of Biot effects in water-saturated sand.

4:00

4pUWa7. Linear coupled electrokinetic-acoustic behavior of consolidated and unconsolidated granular materials. Gareth Block, John G. Harris, and Nicholas Chotiros (Appl. Res. Labs., Univ. of Texas, 10000 Burnet Rd., Austin, TX 78758, gblock@arlut.utexas.edu)

Two electrokinetic (EK) techniques are being developed to study wave propagation in consolidated and unconsolidated granular materials. In contrast to the more common acoustical techniques, EK transduction utilizes the electrokinetic coupling between sediment grain surface chemistry and pore fluid motion. When Biot theory is applicable, coupled electrokinetic-Biot theory predicts that a macroscopic electric potential is generated by fluid motion in electrolyte-saturated porous media. (Our previous research on the reciprocal case—of creating a pressure wave by applying a high voltage impulse to electrolyte-saturated sediments—has been presented at recent ASA meetings.) To assess the validity of Biot theory as a model of

ocean sediments, various EK measurements were taken in the laboratory. Preliminary data using silica sand, as well as unconsolidated and consolidated (sintered) glass microspheres, will be presented. We will also describe the coupled electrokinetic-Biot theory, the underlying volume-averaged poroelastic model upon which it is based, and related issues in homogenization theory. [Work supported by ONR, Ocean Acoustics.]

4:15

4pUWa8. The effect of marine sediments on the underwater penetration of a sonic boom. Oleg A. Godin (CIRES, Univ. of Colorado and NOAA/Environ. Technol. Lab., Boulder, CO 80305-3328, Oleg.Godin@noaa.gov) and David M. F. Chapman (Defence Res. and Development Canada Atlantic, Dartmouth, NS B2Y 3Z7, Canada)

Understanding and quantifying sonic boom penetration into the ocean is of substantial practical interest because of the potential impact of supersonic aircraft on marine mammals. Observations by Desharnais and Chapman [F. Desharnais and D. M. F. Chapman, J. Acoust. Soc. Am. **111**, 544–553 (2002)] indicate, in qualitative agreement with numerical simulations, that sonic boom penetration in coastal regions may be significantly enhanced compared to the deep water case by the resonant excitation of seismo-acoustic waves in marine sediments. In this paper, an analytic theory originally developed to model interface waves in soft sediments with power-law shear speed profiles [O. A. Godin and D. M. F. Chapman, J. Acoust. Soc. Am. **110**, 1890–1907 (2001)] is extended to simulate the shallow-water sound field induced by a shock wave incident from the air side. The theory explains salient features of the observations, including the nonmonotonic depth dependence of the sound exposure spectral density (SESD). Our analysis demonstrates that SESD in shallow water is sensitive to the geoacoustic parameters of the seabed, particularly the shear modulus profile. It may be possible to extract the shear wave attenuation coefficient in the sediment at low frequencies by analyzing sonic boom records from a vertical line array of hydrophones.

4:30

4pUWa9. Estimation of seabed reflection properties from direct blast pulse shape. Mark K. Prior and Christopher H. Harrison (SACLANTCEN, Viale San Bartolomeo 400, 19038 La Spezia, Italy, prior@saclantc.nato.int)

A method is described by which the reflection properties of the seabed can be estimated from the fall-off with time of intensity received on a single hydrophone, tens of water depths away from a broadband source in shallow water. The method is simple and directly applicable to most active sonars. The theoretical basis of the method is described and results presented using experimental data measured in a shallow water region of the Mediterranean Sea. Seabed properties are deduced directly by the method in the form of the rate of change of reflection loss with respect to an angle close to grazing. An extension to the method is proposed whereby seabed sound speed, density and attenuation can be estimated. Good agreement is shown between the results of the method and the results of matched field inversions carried out in the same area using a large-aperture vertical line array. A similar method using target echoes, instead of direct arrivals, is also demonstrated. The features of the method that make it attractive as a candidate for through the sensor probing of the environment are discussed.

4:45

4pUWa10. Geoacoustic inversion of a shallow fresh-water environment. Steven A. Stotts, David P. Knobles, Robert A. Koch, James N. Piper (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029), and Jason A. Keller (Univ. of Texas, Austin, TX 78713-8029)

A recent experiment was conducted at The University of Texas/Applied Research Laboratories test station located at Lake Travis, Austin, TX. Implosive (light bulb), explosive (firecracker), and tonal sources were recorded on a dual receiver system located on the bottom next to a range-independent underwater river channel. Inversion results of the broadband time series obtained over ranges less than 1.5 km were used to predict measured transmission loss at several tonal frequencies in the band from 250–1000 Hz. The average water depth was approximately 38 m along the channel during the experiment. Sound speed profiles were calculated from recorded temperature readings measured as a function of depth. Implosive source spectrums were measured and used to evaluate a model/data correlation cost function in a simulated annealing algorithm. Comparisons of inversion results using both a normal mode and a ray-based plane wave reflection coefficient forward model [Stotts *et al.*, J. Acoust. Soc. Am. (submitted)] are discussed. Predicted transmission loss based on the inversion results are compared to the measured transmission loss. Differences between fluid and elastic layer bottom models will also be presented.

THURSDAY AFTERNOON, 13 NOVEMBER 2003 SABINE ROOM, 2:30 TO 4:10 P.M.

Session 4pUWb

Underwater Acoustics, Engineering Acoustics and Signal Processing in Acoustics: Gradient Array Acoustics II

Paul C. Hines, Cochair
Defence R&D Canada Atlantic, P.O. Box 1012, Dartmouth, Nova Scotia B2Y 3Z7, Canada

Daniel L. Hutt, Cochair
Defence R&D Canada Atlantic, P.O. Box 1012, Dartmouth, Nova Scotia B2Y 3Z7, Canada

Chair's Introduction—2:30

Invited Papers

2:35

4pUWb1. Overview of vector sensors for undersea applications. James F. McEachern (Office of Naval Res., 800 N. Quincy St., Arlington, VA 22217-5660, mceachj@onr.navy.mil)

An overview of vector sensors for sonar applications is presented. The most prolific use of vector or pressure gradient sensors has been in directional sonobuoys to accomplish effective directional measurements with a sensor that is much smaller than the signal wavelength. Common characteristics, implementation issues, and self noise sources of directional hydrophones are reviewed. Arrays

of vector sensors began to emerge in the 1980s and have the attractive property of being able to provide substantially higher directivity for a smaller equivalent aperture than that required for a scalar (omnidirectional) sensor array. Dyadic sensors bring additional directivity but development is required to realize them in a compact configuration. Future applications of vector sensors are discussed. The limited amount of information relative to the ocean acoustic vector fields is noted as a corollary technological requirement for the widespread adoption of vector sensors in sonar systems.

3:00

4pUWb2. Multimicrophone-technology in hearing systems. Benefits and trends. Torsten Niederdraenk (Siemens Audiologische Technik GmbH, D-91058 Erlangen, Germany, torsten.niederdraenk@siemens.com)

Hearing impaired people often suffer from reduced communication skills in real-life situations. In order to increase the signal-to-noise ratio, directional microphones are applied in hearing instruments. Repelling noise from backward directions and focusing on sounds coming to the face of the listener, they provide the hearing impaired with better communication abilities. This contribution reports on higher-order directional microphones in hearing instruments. While the hearing aid user demands an improvement of the signal-to-noise ratio, the directional microphones of hearing instruments should be very small or rather invisible. Some higher-order directional microphone solutions that are positioned on top of a spectacle frame or body-worn devices are already known on the market. Of course, these arrangements are not very practicable in use and cosmetically unappealing. The only relevant approach for the customer seems to be the integration of higher-order directional microphones in the hearing instrument. In terms of the distorted sound field around the head and small microphone distances a special approach has been implemented, that adaptively combines directive microphones of different order.

Contributed Papers

3:25

4pUWb3. Sound source localization with a gradient array using a coherence test. Satish Mohan, Michael E. Lockwood, Douglas L. Jones (Beckman Inst., Univ. of Illinois at Urbana-Champaign, 405 N. Matthews Ave., Urbana, IL 61801, smohan@uiuc.edu), Quang Su, and Ronald N. Miles (State Univ. of New York, Binghamton, NY 13902-6000)

The localization of sound sources is difficult when the number of sources is greater than the number of sensors. An algorithm is described that can localize in such acoustic scenes by utilizing a statistical coherence test to identify desirable time–frequency bins where the MUSIC and minimum-variance spectral estimators are applied. Localization results are then obtained by integrating across the selected time–frequency bins. The algorithm was used to localize signals in azimuth and elevation from a co-located microphone array consisting of one omnidirectional microphone and three gradient microphones. With 2.5 seconds of five speech sources, the azimuth and elevation estimates produced by the algorithm, over a hundred realizations of additive white Gaussian noise, had biases under four degrees and standard deviations less than two degrees. The algorithm was also used to compare the localization performance, in azimuth, of an array of two co-located gradient microphones with an array of two 15-cm-spaced omnidirectional microphones in the free-field.

3:40

4pUWb4. Distortion of interfering speech in the aggregate beamformer. David I. Havelock (Natl. Res. Council, 1200 Montreal Rd., Ottawa, ON K1A 0R6, Canada)

The aggregate beamformer is an alternative to conventional beamformers. It samples an array of sensors randomly, reducing the undesired off-beam signals to noise. Important advantages of this technique over conventional beamforming are the reduced front-end hardware requirements, such as anti-alias filters, the ability to operate at a lower total sampling rate, particularly for arrays with many elements, and improved beamforming time-delay resolution without the need for interpolation. The aggregate beamformer output contains residual noise that is proportional to the level of interfering off-beam signals. On-beam signals do not cause residual noise. The residual noise level is controlled by adjusting the total sampling rate. In speech applications, the residual noise is perceived as a distortion of the off-beam signal. This distortion may help to discriminate desired (on-beam) and undesired (off-beam) speech. The principles of operation of the aggregate beamformer are described and a demonstration of the residual noise in a speech pick-up application is presented.

3:55

4pUWb5. Beamforming with collocated microphone arrays. Michael E. Lockwood, Douglas L. Jones (Beckman Inst., Univ. of Illinois at Urbana-Champaign, 405 N. Matthews Ave., Urbana, IL 61801, melockwo@uiuc.edu), Quang Su, and Ronald N. Miles (State Univ. of New York, Binghamton, NY 13902-6000)

A collocated microphone array, including three gradient microphones with different orientations and one omnidirectional microphone, was used to acquire data in a sound-treated room and in an outdoor environment. This arrangement of gradient microphones represents an acoustic vector sensor used in air. Beamforming techniques traditionally associated with much larger uniformly spaced arrays of omnidirectional sensors are extended to this compact array ($1\ cm^3$) with encouraging results. A frequency–domain minimum-variance beamformer was developed to work with this array. After a calibration of the array, the recovery of sources from any direction is achieved with high fidelity, even in the presence of multiple interferers. SNR gains of 5–12 dB with up to four speech sources were obtained with both indoor and outdoor recordings. This algorithm has been developed for new MEMS-type microphones that further reduce the size of the sensor array.

Session 5aBBa

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Topical Meeting on Shock Waves in Medicine I: Shock Wave Physics

Michael R. Bailey, Cochair
Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, University of Washington, 1013 Northeast 40th Street, Seattle, Washington 98105

Pei Zhong, Cochair
Department of Mechnical Engineering and Materials Science, Duke University, Box 90300, 3000 Science Drive, Durham, North Carolina 27708-0300

Chair's Introduction—8:00

Invited Papers

8:05

5aBBa1. Overview of shock waves in medicine. Robin O. Cleveland (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215, robinc@bu.edu)

A brief overview of three applications of shock waves is presented. Shock wave lithotripsy (SWL) has been in clinical use for more than 20 years. In the United States it is used to treat more than 80% of kidney stone cases and has wide acceptance with patients because it is a noninvasive procedure. Despite SWLs enormous success there is no agreement on how shock waves comminute stones. There is also a general acceptance that shock waves lead to trauma to the soft tissue of the kidney. Yet there has been little forward progress in developing lithotripters which provide comminution with less side-effects, indeed the original machine is still considered the gold standard. The last decade has seen the advent of new shock wave devices for treating principally musculoskeletal indications, such as plantar fasciitis, tennis elbow, and bone fractures that do not heal. This is referred to as shock wave therapy (SWT). The mechanisms by which SWT works are even less well understood than SWL and the consequences of bioeffects have also not been studied in detail. Shock waves have also been shown to be effective at enhancing drug delivery into cells and assisting with gene transfection. [Work partially supported by NIH.]

8:20

5aBBa2. Cloud cavitation effects in shockwave lithotripsy. Tim Colonius and Michel Tanguay (Div. of Eng. and Appl. Sci., California Inst. of Technol., Pasadena, CA 91125)

Cavitation has already been identified as an important damage mechanism in the comminution of kidney stones in shockwave lithotripsy (SWL). However, the precise conditions that maximize the damage caused by the collapsing bubbles are still unknown. Numerical simulations are used to investigate shock propagation and the consequent growth and collapse of a bubble cloud in the focal region of a lithotripter. In the simulations, a continuum two-phase flow model for the ensemble-averaged macroscale is coupled to a Gilmore model for individual spherical bubble dynamics at the microscale. The simulations show agreement with experimental pressure measurements and high-speed photography of the bubble cloud. At void fractions commensurate with experiments, it is found that the collective collapse of the bubble cloud provides a significant increase to the energy available for comminution (beyond what a single bubble would produce). Relatively small increases in the pressure at the center of the cloud in advance of collapse (two orders of magnitude smaller than the initial shock) can more than double the energy of the collapsing bubble.

8:35

5aBBa3. Lithotripter shock interaction with a bubble trapped in an elastic tube. Andrew J. Szeri (Dept. of Mech. Eng., Univ. of California–Berkeley, Berkeley, CA 94720-1740), Pei Zhong (Duke Univ., Durham, NC 27708), and John R. Blake (Univ. of Birmingham, Edgbaston, Birmingham B15 2TT, UK)

In shock wave lithotripsy, it is hypothesized that damage to blood vessels occurs through shock interaction with a bubble trapped within the vessel. This situation is investigated through a computer simulation of a three phase model: liquid, gas bubble, and nonlinear elastic vessel wall. Coupling between the three phases occurs through the pressure field evaluated at interfaces between the different media. It is found that bubble expansion and subsequent violent collapse does indeed cause considerable straining in the vessel wall, and the formation of strong jets that may lead to damage. [Research supported by the U.S. NSF Program in Biomedical Engineering.]

8:50

5aBBa4. Physical properties and behavior of shock waves. David T. Blackstock (Dept. of Mech. Eng. and Appl. Res. Labs., Univ. of Texas, Austin, TX 78712-1063)

The distinguishing feature of a shock wave is its practically instantaneous rise in pressure. Its laws of propagation are different from those for ordinary, continuous-waveform sound waves, even when the sound wave amplitude is finite. The Rankine–Hugoniot shock relations connect the properties on either side of the shock, e.g., pressure, density, and particle velocity, with the shock velocity or Mach number. These relations are valid even for unsteady shocks, e.g., growing or decaying shocks. Weak shock theory provides a means of predicting the propagation of most steady and unsteady shocks that are encountered in acoustics, including those used in medicine. For example, in lithotripsy the focal pressure amplitude rarely exceeds 100 MPa; for such a shock the shock velocity is only $\sim 8\%$ greater than the sound speed in water. Shock rise time τ_{rise} is an interesting property of shocks but often difficult to measure. For example, if the lithotripsy shock cited above was steady and in water, its rise time would be a few tens of picoseconds. When the shock is unsteady, τ_{rise} is much harder to predict.

9:05

5aBBa5. Numerical modeling of nonlinear acoustic waves containing shocks. Vera Khokhlova (Dept. of Acoust., Phys. Faculty, Moscow State Univ., Moscow 119992, Russia)

Theoretical models and corresponding numerical algorithms for the description of propagation of nonlinear acoustic waves containing shocks are presented. Plane waves and high intensity acoustic beams are considered. Various time and frequency domain numerical schemes that include numerical or artificial smoothening of the shocks and their specific features and artifacts are discussed. A modified spectral approach for the description of nonlinear waves containing discontinuous shocks is described. The method utilizes an analytical high frequency asymptote of the abrupt shock, which is used in the numerical algorithm to model sawtooth-like waves with a relatively few number of harmonics. Further generalizations of the method are discussed, in particular, to account for the discontinuity in the first derivative at the shock front, typical for the shock pulses; and for the finite thickness of the shock. The results of modeling of nonlinear waves containing shocks are presented: the effect of saturation in focused acoustic beams, double shock formation in a nearfield of a piston source, saturation of nonlinear waves in active media, the effect of various frequency dependent absorption on nonlinear interactions, and stationary and quasi-stationary solutions of the Burgers-type equations with the high order derivative absorption term. [Work supported by ONRIFO, CRDF, and RFBR.]

9:20

5aBBa6. Regulatory standards and calibration procedures for shock wave devices. Mark E. Schafer (Sonic Tech, Inc., 23 Brookline Court, Ambler, PA 19002-1904)

In order to bring any shock wave device into commercial use, i.e., clinical practice, it must receive regulatory approval from either the U.S Food and Drug Administration (FDA) or the appropriate national agency. A key part of this process involves the complete temporal and spatial description of the shock wave field. This device characterization presents a number of formidable measurement challenges, principally due to the destructive effects of shock waves on the measurement sensor, and shock wave variability (especially for electrohydraulic systems). This presentation reviews the measurement and regulatory approaches used for characterizing shock wave devices, including FDA and international measurement standards. The current approach is a compromise between the desire for a complete characterization of all possible parameters, and the realities of making the measurements. The complete measurement process will be described, including equipment, procedures and pitfalls. Polyvinylidene Fluoride (PVDF) membrane hydrophones have been the key enabling technology, providing sufficient temporal bandwidth and minimal effective sensor area, all at reasonable cost. Other types of sensors, both good and bad, have been used for these measurements. The talk will also present case studies of measurements of several lithotripters measured over the last 15 years.

9:35–9:50

Panel Discussion

Session 5aBBb

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Topical Meeting on Shock Waves in Medicine II: Shock Wave Lithotripsy

Pei Zhong, Cochair
Department of Mechanical Engineering and Materials Science, Duke University, Box 90300, 3000 Science Drive, Durham, North Carolina 27708-0300

James A. McAteer, Cochair
Anatomy MS-5035, Indiana University School of Medicine, 635 Barnhill Drive, Indianapolis, Indiana 46202-5120

Invited Papers

10:10

5aBBb1. Current state of clinical lithotripsy. James E. Lingeman (Methodist Hospital Inst. for Kidney Stone Disease, 1801 N. Senate Blvd., Ste. 220, Indianapolis, IN 46202)

Extracorporeal shock wave lithotripsy (ESWL) has revolutionized the treatment of urolithiasis. Because of the clinical success of the original lithotriptor, the Dornier HM3, numerous manufacturers introduced different approaches to lithotripsy based on empiricism rather than an understanding of the fundamental mechanisms of this new technology. Our understanding of shock wave physics and the physiologic effects of shock waves has progressed greatly over the last decade resulting in insights that hopefully will be reflected favorably in future lithotriptor designs and lithotripsy techniques.

10:25

5aBBb2. Shock wave lithotripsy (SWL) induces significant structural and functional changes in the kidney. Andrew P. Evan, Lynn R. Willis (School of Medicine, Indiana Univ., 635 Barnhill Dr., Indianapolis, IN 46223), and James E. Lingeman (Methodist Res. Inst., Indianapolis, IN 46202)

The foundation for understanding SWL-injury has been well-controlled renal structural and functional studies in pigs, a model that closely mimics the human kidney. A clinical dose (2000 shocks at 24 kV) of SWL administered by the Dornier HM3 induces a predictable, unique vascular injury at F2 that is associated with transient renal vasoconstriction, seen as a reduction in renal plasma flow, in both treated and untreated kidneys. Unilateral renal denervation studies links the fall in blood flow in untreated kidneys to autonomic nerve activity in the treated kidney. SWL-induced trauma is associated with an acute inflammatory process, termed Lithotripsy Nephritis and tubular damage at the site of damage that leads to a focal region of scar. Lesion size increases with shock number and kV level. In addition, risk factors like kidney size and pre-existing renal disease (e.g., pyelonephritis), can exaggerate the predicted level of renal impairment. Our new protection data show that lesion size can be greatly reduced by a pretreatment session with low kV and shock number. The mechanisms of soft tissue injury probably involves shear stress followed by acoustic cavitation. Because of the perceived enhanced level of bioeffects from 3rd generation lithotripters, these observations are more relevant than ever.

10:40

5aBBb3. Generation of shear waves as an effective mechanism of dynamic load of the lithotripter shock wave on the kidney stone. Oleg A. Sapozhnikov (Dept. of Acoust., Faculty of Phys., M. V. Lomonosov Moscow State Univ., Leninskie Gory, Moscow 119992, Russia), Robin O. Cleveland (Boston Univ., Boston, MA 02215), Michael R. Bailey, and Lawrence A. Crum (Univ. of Washington, Seattle, WA 98105)

A number of stone comminution mechanisms have been studied in lithotripsy. Except cavitation erosion, these mechanisms (namely, spallation, dynamic fatigue, shear, and circumferential compression) are associated with stresses generated in the stone by the shock wave. The mechanical load on the stone depends on the waveform and stone structure, size, and shape. We modeled the propagation of lithotripter shock waves through a cylindrical stone with a finite differences simulation based on Lame's equation. The stone parameters were similar to those of natural kidney stones. A new mechanism of tensile stress generation is predicted that may be 5–10 times more efficient than spalling. Shear elasticity of the stone gave rise to the peak tensile strain in the bulk of the stone; this strain occurs near the stone axis due to coherent arrival of shear waves from the front edges of the stone. The position of the region of maximum strain and direction of corresponding tensile forces is similar to those predicted by the spalling mechanism. The modeling also showed that circumferential compression is not activated by the dynamic load produced by a short shock wave typical for lithotripsy. [Work supported by NIH PO1 DK43881, RO1 DK55674 and FOGARTY, CDRF, ONRIFO, and NSBRI.]

10:55

5aBBb4. Innovations in shock wave lithotripsy technology. Pei Zhong, Yufeng Zhou, Songlin Zhu, Franklin Cocks (Dept. of Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708), and Glenn Preminger (Duke Univ. Medical Ctr., Durham, NC 27710)

Since its introduction in early 1980s, shock wave lithotripsy (SWL) has been used widely in clinic for the treatment of kidney and upper urinary stones. Although a variety of methods have been developed for shock wave generation, coupling, and focusing, the core of SWL technology has not changed significantly. In this talk, we will present a summary of our research efforts, aiming to provide innovations in SWL technology. Our strategy is to first better understand the mechanisms by which stone comminution and tissue injury are produced in SWL using various experimental and theoretical techniques. Based on this knowledge, we then developed novel techniques that can optimize the effect of cavitation in SWL via modification of the waveform profile, pressure distribution, and pulse sequence of lithotripter-generated shock waves. These new techniques were upgraded on a Dornier HM-3 lithotripter, the gold standard in SWL. Both *in vitro* phantom and *in vivo* animal experiments were carried out which demonstrated that the performance and safety of the upgraded HM-3 lithotripter is superior to the original HM-3 lithotripter. Finally, strategies to improve stone comminution efficiency while reducing tissue injury in SWL will be presented. [Work supported by NIH DK52985 and DK58266.]

11:10

5aBBb5. Development of lithotripter technology. Wolfgang F. W. Eisenmenger (1. Physikalisches Institut, Univ. of Stuttgart, Pfaffenwaldring 57, D-70550 Stuttgart, Germany)

''Squeezing'' of the stone or cirumferential pressure of the wave propagating at the outside of the stone in the liquid or tissue results in fragmentation in planes perpendicular or parallel to the wave propagation direction. The corresponding pressure zone propagating with the sound velocity in the liquid which is below the sound velocity in the stone, causes an evanescent pressure zone in the stone resulting in tensile stress in planes parallel and perpendicular to the wave plane. A quantitative model predicting the ratio of pulses needed to fragment the stone to 2 mm particle size in relation to the number of pressure pulses needed for the first fragmentation is well in accord with experiments, supporting the ''squeezing mechanism with binary fragmentation.'' On the basis of these results it now appears possible to optimize the pressure pulse parameters measured using the Fiber Optic Probe Hydrophone (FOPH). With correspondingly optimized self-focusing electromagnetic shock wave generator systems a clinical study of the concept ''wide focus and low pressure'' ESWL was performed in a scientific cooperation between the 1. Physical Institute of the University of Stuttgart and the Xixin Medical Instruments Co., Ltd. in Suzhou, China. Literature: W. Eisenmenger, ''The mechanisms of stone fragmentation in ESWL,'' Ultrasound Med. Biol. **27**, 683–693 (2001); W. Eisenmenger *et al.*, ''The first clinical results of 'wide focus and low pressure' ESWL,'' Ultrasound Med. Biol. **28**, 769–774 (2002).

11:25

5aBBb6. Mechanisms of action in dual-pulse lithotripsy. Dahlia L. Sokolov (ROB/NCI/NIH, 9000 Rockville Pike, Bldg. 10, Rm. B3-B69, Bethesda, MD 20892), Michael R. Bailey, and Lawrence A. Crum (CIMU/APL, Univ. of Washington, Seattle, WA 98105)

Dual-pulse lithotripsy has emerged as a promising technique for a safer and more effective shock wave lithotripsy treatment. A single source and confocal sources of various design have been used to produce either synchronous or asynchronous dual pulses. It has been shown that dual pulses enhance stone comminution and reduce tissue injury *in vivo* and *in vitro*. Pressure doubling at the focus of synchronous confocal sources contributes to increased fragmentation by a combination of direct shock-wave effects and increased cavitation, while decreased cell damage near the focus is largely a result of cavitation mitigation by the second, delayed shock wave. It follows that weaker pulses may be used to achieve complete stone fragmentation while further protecting surrounding tissue. Twin sources may also be fired asynchronously to enhance or mitigate cavitation as well as to minimize tissue damage by dividing the impact over two tissue paths. Numerical calculations using the Gilmore model and primary radiation force model as well as high speed camera images show little evidence that bubble translation contributes to the observed bioeffects. These results and the history and mechanisms of dual-pulse lithotripsy will be reported. [Work supported by NIH PO1 DK43881 and RO1 DK55674.]

11:40–12:00

Panel Discussion

Session 5aPA

Physical Acoustics: Acoustic Landmine Detection and Seismic Coupling

Preston S. Wilson, Chair
Mechanical Engineering Department, University of Texas at Austin, Austin, Texas 78712-0292

Contributed Papers

9:00

5aPA1. Seismic sonar sources for buried mine detection. Steven R. Baker, Thomas G. Muir, Frederick E. Gaghan, Jr., Sean M. Fitzpatrick, Kraig E. Sheetz, and Jeremie Guy (Naval Postgrad. School, Code Ph, Monterey, CA 93943)

Prior research on seismo-acoustic sonar for detection of buried targets [J. Acoust. Soc. Am. **103**, 2333–2343 (1998)] has continued with examination of various means for exciting interface waves (Rayleigh or Scholte) used to reflect from targets. Several seismic sources were examined for sand beach applications, including vibrating shakers, shaker devices configured to preferentially excite interface waves, linear force actuators, and arrays of shaker sources to create directional interface wave beams. Burial of some plate-like or rod-like portion of the vibrating devices was found to ensure good coupling to the beach. The preferential interface excitation device employed two degrees of freedom to mimic the two components of elliptically polarized interface waves, and was successfully demonstrated. However, it was found that at long ranges, the medium itself created two component interface waves from vibrating source radiations operating with one degree of freedom in the vertical plane. Linear force actuators were functional in this mode. An array of seven vertical shakers was utilized to create interface waves at ranges of 5 m, in the form of directional beams, some 8 deg wide at the half-power points, at frequencies around 100 Hz. Application of these devices for target detection is discussed in the companion paper. [Work sponsored by ONR.]

9:15

5aPA2. Detection of buried mines with seismic sonar. Thomas G. Muir, Steven R. Baker, Frederick E. Gaghan, Jr., Sean M. Fitzpatrick, Patrick W. Hall, Kraig E. Sheetz, and Jeremie Guy (Naval Postgrad. School, Code Ph, Monterey, CA 93943)

Prior research on seismo-acoustic sonar for detection of buried targets [J. Acoust. Soc. Am. **103**, 2333–2343 (1998)] has continued with examination of the target strengths of buried test targets as well as targets of interest, and has also examined detection and confirmatory classification of these, all using arrays of seismic sources and receivers as well as signal processing techniques to enhance target recognition. The target strengths of two test targets (one a steel gas bottle, the other an aluminum powder keg), buried in a sand beach, were examined as a function of internal mass load, to evaluate theory developed for seismic sonar target strength [J. Acoust. Soc. Am. **103**, 2344–2353 (1998)]. The detection of buried naval and military targets of interest was achieved with an array of 7 shaker sources and 5, three-axis seismometers, at a range of 5 m. Vector polarization filtering was the main signal processing technique for detection. It capitalizes on the fact that the vertical and horizontal components in Rayleigh wave echoes are 90 deg out of phase, enabling complex variable processing to obtain the imaginary component of the signal power versus time, which is unique to Rayleigh waves. Gabor matrix processing of this signal component was the main technique used to determine whether the target was man-made or just a natural target in the environment. [Work sponsored by ONR.]

9:30

5aPA3. Nonlinear acoustic techniques for landmine detection: Tuning curves and two-tone tests. Murray S. Korman (Dept. of Phys., U. S. Naval Acad., Annapolis, MD 21402), James M. Sabatier (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677, sabatier@olemiss.edu), and Douglas J. Fenneman (U.S. Army Night Vision and Electron. Sensors Directorate, Fort Belvoir, VA 22060)

Measurements of the top surface vibration of a buried (inert) VS 2.2 anti-tank plastic landmine reveal significant resonances in the frequency range between 80 and 650 Hz. Resonances from measurements of the normal component of the acoustically induced soil surface particle velocity (due to sufficient acoustic-to-seismic coupling) have been used in detection schemes. Since the interface between the top plate and the soil responds to pressure fluctuations nonlinearly, characteristics of landmines, the soil, and the interface are rich in nonlinear physics and allow for new methods of buried landmine detection not previously exploited. Tuning curve experiments (revealing ''softening'' and a back-bone curve linear in particle velocity amplitude versus frequency) help characterize the nonlinear resonant behavior of the soil-landmine oscillator. When two primary waves f_1 and f_2 drive the soil over the mine, a rich spectrum of nonlinearly generated tones is measured with a geophone on the surface over the buried landmine in agreement with Donskoy [Proc. SPIE **3392** (1998); **3710**, 239–246 (1999)]. In profiling, particular nonlinear tonals can improve the contrast ratio compared to using either primary tone in the spectrum. [Work supported by the United States Army Communications-Electronics Command Night Vision and Electronic Sensors Directorate.]

9:45

5aPA4. Acoustic landmine detection using binary maximum-length sequences. Ning Xiang and James M. Sabatier (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

Acoustic landmine detection using a LDV has demonstrated success in recent field tests [Sabatier and Xiang, IEEE Trans. Geosci. Remote Sens. **39**, 1146–1154 (2001); Xiang and Sabatier, J. Acoust. Soc. Am. **113**, 1333–1341 (2003)]. However, low detector signals and speckle noise are still challenging problems in the LDV-based acoustic-to-seismic detection of buried landmines. Scanning insonified patches over buried targets with the spatial resolution required in minefield applications demands processing of a large quantity of detection data. This work proposes the use of powerful signals-binary maximum-length sequences, as the acoustic excitation. By exploiting their excellent number-theory properties, high signal-to-noise ratios for the scanning results can be achieved. Furthermore, the fast maximum-length sequence transform is developed to allow the landmine detection algorithms to extract the impulse response of the acoustic-to-seismic soil / mine-soil system with efficient frequency-demodulation of the laser Doppler vibrometer signals. This presentation discusses some relevant issues associated with the detection system design and extensive field tests. [Work supported by U.S. Army CERDEC Night Vision and Electronic Sensors Directorate.]

10:00

5aPA5. A multi-carrier technique for ultrasonic Doppler vibrometry. Andi Petculescu and James Sabatier (Natl. Ctr. for Physical Acoust., 1 Coliseum Dr., University, MS 38677)

A novel and simple technique for a signal-to-noise-ratio augmenting in ultrasonic Doppler vibrometry is presented. The approach involves sending two probe frequencies separated by twice the modulation frequency. By adjusting the mutual phase of the carriers, the overlapped sideband can be twice as large as the nonoverlapped ones. Moreover, the phase-demodulated signal is also larger than is obtained in the case of one-carrier operation. Parallel results of electronic phase modulation and ultrasonic Doppler vibrometry (with commensurate modulation strengths) are presented. A practical implementation of the method to probe vibrations of a rough surface is then described. [Work supported by ONR.]

10:15

5aPA6. The importance of angle of incidence on acoustic to seismic coupling. Wheeler B. Howard, Craig J. Hickey, and James M. Sabatier (Natl. Ctr. for Physical Acoust., 1 Coliseum Dr., University, MS 38677)

Acoustic to seismic coupling has been successfully used in the detection of buried mines and more recently in the investigation of sub-surface soil structure. Due to the downward refracting nature of the acoustic wave at the air–earth interface, the mechanism by which airborne waves couple to the earth is thought to be insensitive to changing angles of incidence. In experimental setups, the airborne sound sources have been positioned at angles ranging from normal to grazing with respect to the ground surface. A more controlled study of the importance of angle of incidence has been performed. Data from field experiments will be presented illustrating the importance of angle of incidence on acoustic to seismic coupling.

10:30

5aPA7. Effect of ground variability on the acoustic-seismic transfer function. Margarita Fokina, Vladimir Fokin, James M. Sabatier, and Ning Xiang (Natl. Ctr. for Physical Acoust., 1 Coliseum Dr., University, MS 38677)

Experimental measurements have shown that the use of a multilayered elastic media is necessary for transfer function numerical modeling. The present work deals with the effect of variability of ground properties (compressional and shear wave speeds, density, attenuation and thickness of the layers) on the acoustic-seismic transfer function (admittance). An analysis is performed on the planes of parameters of the ground in a wide frequency range for all angles of incidence. The Dunking–Throwers matrix approach is used to increase the accuracy of computations. It is revealed that the acoustic–seismic transfer function is sensitive to ground properties and small variations in the shear speed may cause a strong variation in the acoustic–seismic transfer function. Results of outdoor measurements of the acoustic–seismic transfer function are presented and a correlation between high magnitudes of the acoustic–seismic transfer function in certain frequency ranges and moisture content on the surface is revealed. A simple model explaining the correlation between the moisture content in the upper layer, acoustic–seismic transfer function and ground properties is suggested. [Work supported by ONR.]

10:45

5aPA8. Geoacoustic inversion via acoustic-seismic transfer function. Vladimir N. Fokin, Margarita S. Fokina, James M. Sabatier, and Wheeler B. Howard (Natl. Ctr. for Physical Acoust., 1 Coliseum Dr., University, MS 38677)

Inversion methods for estimation of geoacoustic model parameters often use the scattered field data for obtaining the properties of viscoelastic layered media. This work presents a method to retrieve ground properties information using the outdoor acoustic–seismic transfer function (admittance). The resonance method developed in A. Nagle, H. Uberall, and K-B. Yoo [Inverse Prob. **1**, 99–110 (1985)] is extended and used for a preliminary estimation of a set of parameters for a three-layered ground model. The least squares method is used afterward to choose the model with the best fit to experimental data. Results of the reconstruction show good agreement with the experimental data. A description of the resonant technique and the experimental setup are presented. The effect of a finite size of sound source on the acoustic–seismic transfer function is also discussed. [Work supported by ONR.]

11:00

5aPA9. Monitoring surface sealing and crusting of soils and glass beads using acoustic techniques. Craig J. Hickey, Greg Broussard, James M. Sabatier (NCPA Univ. of Mississippi), and David A. DiCarlo (USDA—ARS Natl. Sedimentation Lab.)

When soils dry after a rain they form a crust or seal that changes both the stiffness and hydraulic properties of a thin surface layer which in turn effects the erodability of the soil. This soil surface crusting/sealing is examined by measuring the acoustic to seismic (A/S) transfer function on laboratory samples. An A/S transfer function measurement is a completely noncontact technique using a suspended loud speaker to impinge acoustic energy from the air onto the sample and a laser Doppler vibrometer (LDV) measures the surface particle velocity. Therefore, this technique measures the seismic energy that has been transferred into the soil from the airborne wave. Several soil samples have been investigated. Also, glass beads, being better known samples, are also examined for dry, crusted, and sealed cases. Samples tested show a quantifiable decrease in the seismic energy transferred as the crust forms but is not as sensitive to changes in sealing. An acoustic reflection technique is proposed to measure changes in the hydraulic properties associated with sealing. These measurements are compared to a poroelastic model proposed by Denneman *et al.*

11:15

5aPA10. Poroelastic effects on media with air or water saturation. Hefeng Dong, Jens M. Hovem, and Hans Olav Hygen (Dept. of Telecommunications/Acoust., Norwegian Univ. of Sci. and Technol., N-7491, Trondheim, Norway, hefeng.dong@tele.ntnu.no)

The Biot theory predicts that three body waves can propagate in a porous medium, a shear wave, a fast and a slow compressional wave. This paper reports the result of an analysis in order to answer the importance of these three wave types in connection with underwater sound propagation and in air over a porous ground. Three approaches are applied: an analytical solution and frequency wave number integration technique extended to treat poroelasticity for reflection and transmission coefficients and finite difference modeling for time responses. The reflected and transmitted signals at the interface between a fluid and a porous medium for both air and water saturated cases are calculated and analyzed. This shows that conversion to a transmitted slow compressional wave is important in air saturated cases and can dominate the reflection loss. Conversion to a slow compressional wave in water saturated cases is of less importance. The transmitted slow compressional wave is in both air and water saturated cases diffusive and exists only in a very thin sheet near the interface.

Session 5aSC

Speech Communication: Vocal Fold Physiology, Modeling and Disorders

Roger W. Chan, Chair
Otolaryngology-Head Neck Surgery, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9035

Chair's Introduction—9:00

Contributed Papers

9:05

5aSC1. Constitutive characterization of vocal fold viscoelasticity based on a modified Arruda-Boyce eight-chain model. Roger W. Chan (Otolaryngol.-Head & Neck Surgery, Univ. of Texas Southwestern Medical Ctr., Dallas, TX 75390-9035)

Previous studies have shown that vocal fold tissues exhibit nonlinear viscoelastic behavior under different loading conditions. Hysteresis and strain rate dependence of stress–strain curves have been observed for vocal fold ligament and muscle tissues when subjected to sinusoidal tensile loading. Nonlinear viscoelastic response and tissue failure have also been described for vocal fold mucosa subjected to constant strain-rate tests under large-strain shear. These findings cannot be adequately described by the traditional constitutive formulations of linear and quasilinear viscoelasticity. This study attempts to characterize some nonlinear behavior of vocal fold tissues under tensile loading based on a modified version of the Arruda-Boyce (Bergström-Boyce) hyperelastic model for polymers, which has been shown to adequately predict the rate-dependent behavior of some elastomers and biological tissues. Results indicated that the model was only capable of describing the relatively linear portion of the nonlinear stress–strain curves of the vocal fold muscle (at strain smaller than 20%), while failing to predict the exponential increase of stress at higher strain. However, the model was capable of predicting the dependence of stress on strain rate reasonably well. This finding was consistent with the model's assumptions on the constitutive behavior of the two constituent polymer networks.

9:20

5aSC2. Viscoelastic properties of laryngeal posturing muscles. Fariborz Alipour, Eric Hunter, and Ingo Titze (Dept. of Speech Pathol. & Audiol., Univ. of Iowa, Iowa City, IA 52242, alipour@blue.weeg.uiowa.edu)

Viscoelastic properties of canine laryngeal muscles were measured in a series of *in vitro* experiments. Laryngeal posturing that controls vocal fold length and adduction/abduction is an essential component of the voice production. The dynamics of posturing depends on the viscoelastic and physiological properties of the laryngeal muscles. The time-dependent and nonlinear behaviors of these tissues are also crucial in the voice production and pitch control theories. The lack of information on some of these muscles such as posterior cricoarytenoid muscle (PCA), lateral cricoarytenoid muscle (LCA), and intraarytenoid muscle (IA) was the major incentive for this study. Samples of PCA and LCA muscles were made from canine larynges and mounted on a dual-servo system (Ergometer) as described in our previous works. Two sets of experiments were conducted on each muscle, a 1-Hz stretch and release experiment that provides stress–strain data and a stress relaxation test. Data from these muscles were fitted to viscoelastic models and Young's modulus and viscoelastic constants are obtained for each muscle. Preliminary data indicates that elastics properties of these muscles are similar to those of thyroarytenoid and cricothyroid muscles. The relaxation response of these muscles also shows some similarity to other laryngeal muscles in terms of time constants.

9:35

5aSC3. A direct approach to the investigation of quasisteady approximation in phonation. Jong Beom Park and Luc Mongeau (Ray W. Herrick Labs., School of Mech. Eng., Purdue Univ., West Lafayette, IN 47907-2031)

Unsteady pulsating air jets through the human glottis are the main sound source in phonation. The quasisteady approximation has been frequently used for voice modeling. The time-varying air flow in the vocal tract is modeled using a sequence of comparable steady flows with the same wall position and boundary conditions as for the dynamically varying situation. Previous attempts to verify the quasisteady approximation used an indirect method, based on inverse filtering. In the present study, a direct method of comparing the orifice discharge coefficients of both dynamic and static jet flows was investigated. Time varying orifice coefficients of pulsating jets were derived from the measured instantaneous velocity field and orifice area, and then, compared to those of static configurations for the same orifice areas. It was found that dynamic orifice coefficients coincided with the static ones during closing rather than opening. Leading and trailing vortices of the pulsating jets were hypothesized as the main factor causing the dynamic transient effects.

9:50

5aSC4. Logarithmic temporal axis manipulation and its application for measuring auditory contributions in $F0$ control using a transformed auditory feedback procedure. Ryuichiro Yanaga (Grad. School of Wakayama Univ., Wakayama 640-8510, Japan) and Hideki Kawahara (Wakayama Univ., Wakayama 640-8510, Japan, ATR Kyoto 619-0288, Japan)

A new parameter extraction procedure based on logarithmic transformation of the temporal axis was applied to investigate auditory effects on voice $F0$ control to overcome artifacts due to natural fluctuations and nonlinearities in speech production mechanisms. The proposed method may add complementary information to recent findings reported by using frequency shift feedback method [Burnett and Larson, J. Acoust. Soc. Am. **112** (2002)], in terms of dynamic aspects of $F0$ control. In a series of experiments, dependencies of system parameters in $F0$ control on subjects, $F0$ and style (musical expressions and speaking) were tested using six participants. They were three male and three female students specialized in musical education. They were asked to sustain a Japanese vowel /a/ for about 10 s repeatedly up to 2 min in total while hearing $F0$ modulated feedback speech, that was modulated using an M-sequence. The results replicated qualitatively the previous finding [Kawahara and Williams, Vocal Fold Physiology, (1995)] and provided more accurate estimates. Relations with designing an artificial singer also will be discussed. [Work partly supported by the grant in aids in scientific research (B) 14380165 and Wakayama University.]

10:05

5aSC5. Quantitative measurement of the medial surface dynamics of the vocal folds using high-speed digital imaging. Michael Doellinger, Juergen Neubauer, and David A. Berry (Div. of Head & Neck Surgery, 31-24 Rehab. Ctr., 1000 Veteran Ave., UCLA, Mail Code 179422, Los Angles, CA 90095-1794, mdoellinger@mednet.ucla.edu)

To increase our understanding of pathological and healthy voice production, the quantitative measurement of the medial surface dynamics of the vocal folds is significant, albeit rarely performed because of the inaccessibility of the vocal folds. Hence, an excised hemilarynx procedure is applied, Berry *et al.* recently reported such quantitative measurements along one coronal plane of the left vocal fold of a canine [J. Acoust. Soc. Am. **110**, 2539–2547 (2001)]. The present work extends previous studies by capturing vibrations along the entire medial surface of the left vocal fold. The number of vertical rows of sutures used to demarcate fleshpoints was increased from one to five. An automatic algorithm for tracking vocal fold fleshpoints will be reported, along with calibration techniques, and error estimation. Preliminary results will be reported for both periodic and aperiodic vocal fold vibrations. High-speed digital imaging was performed using a Photron APX machine with a sampling frequency of 4000 Hz, a spatial resolution of 1024×512 pixels, and 256 levels of grayscale.

10:20

5aSC6. Perceptual evaluation and acoustical analyses of 6 kinds of electrolarynges. Noriyo Sugimori, Hiroyoshi Fuyuki (Dept. of Rehabilitation, Tokyo Medical Univ. Hosp., 6-7-1 Nishi-Shinjuku, Shinjuku, Tokyo 160-0023, Japan), Noriko Kobayashi, Satoshi Horiguchi, Minako Koike, and Hajime Hirose (School of Allied Health Sci., Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan)

Six different kinds of electrolarynges, which were commercially available in Japan, were compared with regard to the listeners preference and the acoustical characteristics. Listeners were 4 speech therapists who were familiar with electrolarynx speech and 15 university students who had little experience with the device. Sound samples were obtained for each *E*1 under two conditions: the original sound and the sustained vowels produced by two skilled *E*1 users with larynges. Perceptual evaluations were performed using Visual Analogue Scale (VAS) on intelligibility, acceptability, noisiness, and naturalness. Speech therapists perceived unnaturalness for an *E*1 with peculiar *F*0 declination while naive listeners did not. The lower the perceptual evaluations, the higher the mechanical noise level was found by acoustical analyses. High levels of higher harmonic components were found for the *E*1's which had lower perceptual evaluation due to the noise disturbance. Although naive listeners tended to despise *E*1s with rich higher harmonic components in general, speech therapists did not perceive this as noisiness. Various factors seemed to exist for the listeners preference of *E*1s. [Work supported by Grant-in-Aid for Scientific Research on Priority Areas (B) Prosody and Speech Processing, Ministry of Education and Sciences.]

10:35

5aSC7. Use of loud phonation as a voice therapy technique for children with vocal nodules. Noriko Kobayashi, Hajime Hirose (School of Allied Health Sci., Kitasato Univ., Kanagawa 228-8555, Japan), and Koichiro Nishiyama (School of Medicine, Kitasato Univ., Japan)

For the treatment of vocal nodules, educational programs for vocal hygiene and voice training for acquisition of correct phonation are essential. In the case of children, special considerations are necessary as some of their vocal behaviors and reaction to voice disorders are different from those of adults. In this study, a voice therapy program for child vocal nodules were developed and good results were obtained for six children. They were four boys and two girls (Age: 4–11 yr) and bilateral nodules were found for all of them. In addition to a conventional vocal hygiene program for children, correct production of loud voice (so-called gBeltingh) was the major focus of the voice therapy as the visual inspection of the larynges and perceptual evaluations of the voice revealed inappropriate loud voice production with laryngeal constriction in all children. After 5–24 voice therapy sessions, disappearance of the nodules was found in five children and the reduction of the nodule sizes was found in one child. Improvement of the GRBAS scores, longer maximum phonation time, and extension of vocal ranges were found after the completion of the therapy programs.

FRIDAY MORNING, 14 NOVEMBER 2003 SAN MARCOS ROOM, 9:30 TO 11:00 A.M.

Session 5aSP

Signal Processing in Acoustics: Selected Applications

Donald R. DelBalzo, Chair
Neptune Sciences, Inc., 40201 Highway 190 East, Slidell, Louisiana 70461

Contributed Papers

9:30

5aSP1. Applications of the continuous wavelet transform to architectural acoustics. Zackery B. Belanger (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607)

Wavelet analysis, which utilizes waveforms of limited duration, can reveal a signals time and frequency information without the resolution tradeoff that is typical of the sinusoid-based Fourier analysis. The continuous wavelet transform is an effective tool in the analysis of impulse responses from both in situ and scale model measurements. Simultaneous time and frequency information in architectural acoustics allows both beneficial and problematic reflections to be located and analyzed. A brief conceptual overview of wavelets is given, followed by general application to architectural acoustics and specific application to a few projects.

9:45

5aSP2. Objective sampling with EAGLE to improve acoustic prediction accuracy. Erik R. Rike and Donald R. DelBalzo (Neptune Sci., Inc., 40201 Hwy. 190 E., Slidell, LA 70461)

Some Navy operations require extensive acoustic calculations. The standard computational approach is to calculate on a regular grid of points and radials. In complex environmental areas, this implies a dense grid and many radials (i.e., long run times) to achieve acceptable accuracy and detail. However, Navy tactical decision aid calculations must be timely and exhibit adequate accuracy or the results may be too old or too imprecise to be valuable. This dilemma led to a new concept, OGRES (Objective Grid/Radials using Environmentally-sensitive Selection), which produces irregular acoustic grids [Rike and DelBalzo, Proc. IEEE Oceans (2002)]. Its premise is that physical environmental complexity controls the need for dense sampling in space and azimuth, and that transmission loss

already computed for nearby coordinates on previous iterations can be used to predict that complexity. Recent work in this area to further increase accuracy and efficiency by using better metrics and interpolation routines has led to the Efficient Acoustic Gridder for Littoral Environments (EAGLE). On each iteration, EAGLE produces an acoustic field for the entire area of interest with ever-increasing resolution and accuracy. An example is presented where approximately an order of magnitude efficiency improvement (over regular grids) is demonstrated. [Work sponsored by ONR.]

10:00

5aSP3. Logarithmic temporal axis manipulation and its application for measuring perceptually salient acoustic features of loudspeakers based on multiple observations. Masanori Morise (Wakayama Univ., Wakayama 640-8510, Japan) and Hideki Kawahara (Wakayama Univ., ATR, Kyoto 619-0288, Japan)

A new procedure for measuring impulse responses of loudspeakers in a real world situations were proposed, specifically aiming at extracting perceptually saliant features of the acoustic systems. The procedure consists of two stages. The first stage is an impulse response measurement using specially designed test signals (TSP or M-sequence). Post-processing using a logarithmic transformation of the temporal axis and filtering enables selective elimination of a time-frequency representation of the response that does not consists of components stimulated by the input signal. The second stage is to extract perceptual ''variance'' based on multi-location observations. The procedure was optimized to provide the best inverse filter in terms of the signal duration of the compensated response. Simulations were conducted to test general behavior of the proposed method. Preliminary tests using a robot controlled measuring stage demonstrated feasibility and effectiveness of the proposed method. [Work partly supported by the grant in aids in scientific research (B) 14380165 and Wakayama University.]

10:15

5aSP4. Application of the combined higher order time–frequency method to the identification of wave propogation. Sang-Kwon Lee and Sung-Gyu Ko (Dept. of Mech. Eng., Inha Univ., 253 Yonghyun Dong, Inchon)

The wave propagation in a solid is an interesting topic for mechanical engineering since it often gives an idea for the structure damage. However it is often difficult to measure the wave propagation in beam due to the difference of speed of wave depending on frequency. In order to solve this problem, many time–frequency methods have been developed. Especially Wigner–Ville distribution, wavelet analysis and short time–frequency analysis are recently popular methods in this field. These methods have an advantage and a disadvantage depending on the type of structure. In this paper, the combined higher order time–frequency method is developed and used for the good identification of wave propagation. It can be applied to the damage analysis of the structure in detail. [Work supported by the Nuclear Research Project in Korea.]

10:30

5aSPb. Estimation of acoustic signals in reverberant conditions via the characteristic function. Tran H. Dat (CIAIR, Nagoya Univ., Japan) and Itakura Fumitada (Itakura Lab., Nagoya Univ., Japan)

In many circumstances at reverberant conditions, acoustic signals are mixed by the multiple reflections and scattering phenomena. Recently the independent component analysis (ICA) has gained great advantages for processing this class of signals. Many implementations for both temporal and temporal-frequency, stationary and nonstationary, second order and high order statistics are given. However in these cases, the dynamic structure described by the changing of the time of sources number was not considered. In this paper first we proposed a generalized stochastic model allowing describing both a static and dynamic structure of sources. Then the joint characteristic function is analytically obtained. Second, we examine the estimation of an empirical characteristic function for given models.

10:45

5aSP6. Equalization of a one-dimensional sound field using acoustic energy density. Xi Chen, Timothy Leishman, and Scott Sommerfeldt (Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602, drcxchen@hotmail.com)

For many years, reproduced and reinforced sound fields have been equalized using acoustic pressures measured at discrete field points. In this paper, well-known problems associated with this equalization method are investigated using a one-dimensional model. An optimum equalization scheme is then suggested that is based on the spatially averaged mean-square pressure of the field. A practical method based on the energy density of the field is subsequently discussed and compared to the other methods. Although the energy density is evaluated at a single field point, its spatial uniformity is shown to provide results approaching those of the optimum equalization method.

FRIDAY MORNING, 14 NOVEMBER 2003 SABINE ROOM, 8:00 TO 11:45 A.M.

Session 5aUW

Underwater Acoustics: Noise and Fluctuations

Peter G. Cable, Chair

BBN Technologies, 11 Main Street, Mystic, Connecticut 06355-3641

Contributed Papers

8:00

5aUW1. Bottomed acoustic array gain variability in the South China Sea. Marshall Orr, Bruce Pasewark, Stephen Wolf (4555 Overlook Ave. SW, Washington, DC 20375), James Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), and Ching-Sang Chiu (Naval Postgrad. School, Monterey, CA 93943)

The stratified sound speed field at the shelf break of the South China Sea is constantly perturbed by the tide and the passage of the internal tide, linear and nonlinear internal waves and associated fine structure. An 18-day acoustic propagation experiment was performed at the shelf break in May 2001 as part of the ONR supported AsiaEx experiment. A 32-element 465-m bottomed horizontal array received 300- and 500-Hz center frequency FM acoustic signals from sources moored at a range of ~18.9 km. The temporal and spatial variability of signal coherence, the array gain variability, and residual signal gain variability has been extracted. Temporal and spatial variability of the signal coherence and array gain variability has a degree of correlation with temperature field variability measured near the source and receiver locations. These data as well as histograms of the residual array gain variability in time and space will be presented.

8:15

5aUW2. Directional noise and correlation of signal and noise variability in the ASIAEX South China Sea Experiment. Timothy F. Duda, James F. Lynch, Glen G. Gawarkiewicz (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), Ching-Sang Chiu (Naval Postgrad. School, Monterey, CA 93943), Ruey-Chang Wei (Natl. Sun Yat-Sen Univ., Kaohsiung City, Taiwan), and Phil Abbot (Ocean Acoust. Services and Instrumentation Systems, Inc., Lexington, MA 02421)

As part of the 2001 ASIAEX South China Sea experiment, known acoustic signals and noise were recorded for 16 days with co-located horizontal and vertical receiving arrays in 120-m depth water at the edge of the continental shelf. The location was south of China and southwest of Taiwan. Vertical and horizontal directionality of both signals and noise can be estimated using the two arrays. Directionality can be linked to vertical mode bandwidth. Furthermore, because of mode stripping effects, it can indicate the amount of mode coupling in the vicinity of the receiver. Previously, arrival time and total energy of the signals have been shown to respond to the strong nonlinear internal waves and internal tides of the area. Here, we investigate the response of noise energy and directionality to the waves, and compare with the response of signal energy and directionality. The results can be compared to the Shelfbreak PRIMER experiment in the Mid-Atlantic Bight, where signal and noise parameters recorded with a VLA were correlated. Correlation of signal and noise violates assumptions used in many signal processing algorithms.

8:30

5aUW3. Propagation of uncertainty for acoustic fields using polynomial chaos expansions. Steven Finette (Acoust. Div., Naval Res. Lab., Washington, DC 20375)

The concept of uncertainty is related to both reducible and irreducible errors that can enter a computer simulation through several channels: modeling errors, discretization of the continuous dynamical system, missing knowledge/information concerning boundary and initial conditions, and selection of parameter values. In ocean waveguide modeling, uncertainty inherent in the lack of knowledge of the sound speed distribution, for example, is mapped into the acoustic field. In a sense, it propagates along with the acoustic field to which it is coupled, though the computed pressure field at a point in the waveguide does not normally include an explicit measure of this uncertainty. This presentation will describe a method for explicitly incorporating uncertainty into the dynamics for acoustic wave propagation in ocean waveguides. In order to quantify uncertainty in numerical simulations for underwater acoustics problems, the application of stochastic basis expansions of random processes that effectively map uncertainty in the parameter input space to uncertainty in the system output space is considered. The relationship to Monte Carlo methods is also discussed. [Work supported by ONR.]

8:45

5aUW4. Environmentally associated signal-to-interference variability in low-frequency active sonar. Peter G. Cable (BBN Technologies, 11 Main St., Mystic, CT 06355-3641)

An analysis has been conducted of the spatial and temporal variability of target echo-to-interference measurements made during Area Characterization Test I (ACT I) in 1992 on the West Florida Shelf in the Gulf of Mexico. The measurements were made for five bistatic geometries using explosive sources, a bottom-mounted horizontal receiving hydrophone array and an echo repeater ''target'' in 180 m depth water under downward refracting acoustic conditions. The bottom at the ACT I site was nearly flat of sand-silt-clay composition, and oceanographic and meteorological conditions were very quiescent during the test. In all respects these measurements could be described as having been in a quiet, well-behaved sonar environment. The observed temporal (standard deviation of 1.6 dB) and spatial (standard deviation of 2 dB) variability of the received signal-to-interference will be discussed in terms of physical causes and measurement error. [Work supported by ONR.]

9:00

5aUW5. Intensity fluctuations from deterministic and random azimuthal variability. Laurel K. Reilly-Raska, William L. Siegmann (Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180), James F. Lynch, John A. Colosi, and Timothy F. Duda (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Shallow-water sites for shelfbreak PRIMER and ASIAEX show substantial ocean horizontal variability arising from nonlinear internal wave trains. These relatively coherent structures strongly influence acoustic intensity observations, and deterministic propagation modeling has demonstrated their azimuthal effects. Moreover, random variations in the nonlinear internal waves, as well as processes such as linear internal wave fields, contribute to observed intensity fluctuations. Stochastic modeling is performed to show the azimuthal changes induced in the intensity fluctuations. The principal metric of interest is the scintillation index (SI), the ratio of standard deviation to mean intensities. Limiting cases of the SI are examined for examples including both nonlinear internal waves and random ocean variations. Simulations using a three-dimensional parabolic equation model [Collins, J. Acoust. Soc. Am. **94**, 2269–2278 (1993)] are discussed, along with an extension of a random modal fluctuation approach [Dozier and Tappert, J. Acoust. Soc. Am. **63**, 353–365 (1978)]. [Work supported by ONR.]

9:15

5aUW6. Effects of the propagation environment on coherence of underwater broadband acoustic signals. Rachel Romond, David L. Bradley, and R. Lee Culver (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804, romond@psu.edu)

Acoustic and environmental measurements were collected in 550 m of water approximately 2.5 km east of San Clemente Island, CA over a 4-day period in August 2002. Continuous-wave and frequency-modulated pulses with center frequencies of 20 and 40 kHz were transmitted over a 1 km distance. Four projectors were located between 10 and 70 m depth, and five receivers were positioned between 40 and 220 m. Data from an anemometer, wave rider buoy, acoustic Doppler current profiler and towed conductivity-temperature-depth array were recorded simultaneously over the test period to characterize the marine environment. The extensive set of environmental data provided a basis for boundary condition and water column characterization. The most significant environmental dynamic was found to be tidally related water current. The environmental conditions were input into a propagation model to predict the acoustic signal structure and propagation behavior in the measured environment. Predictions of acoustic propagation parameters such as time fluctuation, amplitude fluctuation, and effect of bandwidth will be compared to results from on-site acoustic measurements. [Work supported by ONR Code 321US under the ARL Program.]

9:30

5aUW7. High frequency intensity fluctuations: Comparison of theory with experiment. Steven D. Lutz, David L. Bradley, and R. Lee Culver (Appl. Res. Lab., Penn State Univ., P.O. Box 30, N. Atherton St., State College, PA 16804-0030, sdl149@psu.edu)

Amplitude fluctuations were measured during August 2002 near San Diego using 20 and 40 kHz cw and fm signals. Source–receiver separation was 1 km; source depths were 10 m to 67 m; receiver hydrophone depths were 44 m to 217 m. A 15-element chain of CTD sensors was towed to measure horizontal temperature and salinity variation with 1m resolution. Theory exists explaining the relationship between amplitude fluctuations and acoustic frequency, source–receiver separation, and index of refraction patch size in the ocean [e.g., Flatte *et al.*, *Sound Transmission Through a Fluctuating Ocean* (Cambridge Press, 1979); Uscinsky *et al.*, ''Intensity Fluctuations. Part 1: Theory,'' J. Acoust. Soc. Am. **74** (1983)]. However, the amplitude fluctuations we observe are much lower than that predicted by theory [J. Acoust. Soc. Am. **113**, 2333 (2003)]. Our frequencies are 1 to 2 orders of magnitude higher than those used in the experiments upon which the theory is based and our propagation ranges are

considerable shorter 1 km versus many km. Results will be discussed from a more detailed analysis of the ray path locations within the water column activity (Langmuir cell formation and internal wave passage).

9:45

5aUW8. Acoustic source model input parameter simulation via hydrodynamic modeling of breaking waves. Steven L. Means and William G. Szymczak (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375)

Low-frequency source models for sound generation from breaking waves based on the collective oscillation of an entrained bubble cloud require a number of input parameters, e.g., void fraction, cloud size, etc. The present effort seeks to obtain simulations of these parameters using a volume of fluid type hydrodynamic model (BUB) to simulate the air entrainment processes occurring within a breaking wave. The constraint that mass is conserved and energy is nonincreasing is a unique feature of the BUB code that allows it to handle liquid collisions and violent free surface motions appropriately. Simulations of breakers of both spilling and plunging type will be presented along with their simulated acoustic source model input parameters. The simulations, in the waves reference frame, will be compared against video of breaking waves obtained in a wave tank at the University of Maryland with a high-speed, wave-following digital camera. Relationships between the simulated acoustically relevant hydrodynamic parameters and characteristics of breaker-generated acoustic signatures measured concurrently within the wave tank will be identified. [Work supported by ONR base funding at NRL.]

10:00–10:15 Break

10:15

5aUW9. Diurnal variability in the anisotropic nature of the ocean's midfrequency (1–10 kHz) noise field. Gerald L. D'Spain, Jeffrey D. Skinner, Galina L. Rovner, and Heidi H. Batchelor (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 93940-0704)

A 131-element, 2-D billboard array with 12-kHz bandwidth per element and interelement spacing equal to half-wavelength at 3.75 kHz was designed and built by the Marine Physical Lab. It recorded data nearly continuously over 8 days in July, 2002. The site was in 175-m water located 2 km from the ''43-Fathom Spot,'' a popular fishing area in the Southern California Bight. Results from combined conventional beamforming in the vertical and adaptive beamforming in the horizontal commonly display a strong azimuthal dependence of the vertical directionality. A noise notch in the horizontal typically exists at many azimuths during daytime hours, sometimes more than 15 dB lower than higher-angle energy. Numerical modeling illustrates the importance of local (within 20-km range) large-scale bathymetry variations. However, at night, the horizontal noise levels can exceed those at higher angles by more than 10 dB. Horizontal beam spectrograms show two broad peaks centered at 1.5 kHz and around 4.5 kHz that appear shortly after sunset, are strongest at midnight, and disappear at sunrise. This horizontal energy arrives from the direction of the 43-Fathom Spot and probably represents a nighttime chorusing behavior of some species of fish or invertebrate. [Work supported by ONR, Code 321(US).]

10:30

5aUW10. Spatial nonuniformity of noise measurements in the Gulf of Mexico LADC test area. Anthony I. Eller (Sci. Applications Intl. Corp., 1710 Saic Dr., McLean, VA 22102, Anthony.I.Eller@saic.com), George E. Ioup, Juliette W. Ioup, and James P. Larue (Univ. of New Orleans, New Orleans, LA 70148)

Long-term, omnidirectional underwater acoustic noise was measured by three buoys in the northeastern Gulf of Mexico during the summer of 2001 as part of the Littoral Acoustic Demonstration Center project, Phase I. The buoys were placed over the coastal slope above the 600-, 800- and 1000-m contours, approximately in a straight line, 18 km from the first to the last. Initial looks at the noise time series, processed in standard one-third-octave bands from 10 to 5000 Hz, show noise levels in close agreement at each buoy except in the lowest bands (roughly 25 Hz and lower) where the noise levels grow higher at the shallower sites. Results presented here address buoy-to-buoy comparisons of long-term temporal variability and the cross-spectral density of fluctuations across the frequency spectrum. [Research supported by ONR.]

10:45

5aUW11. Emergence of the time domain Green's function from ocean noise correlations. Karim G. Sabra, Philippe Roux, W. A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., Univ. of California–San Diego, La Jolla, CA 92093, ksabra@mpl.ucsd.edu), W. S. Hodgkiss, Hee Chun Song, T. Akal (Univ. of California–San Diego, La Jolla, CA 92093), and Mark Stevenson (SACLANT Undersea Res. Ctr.)

The time domain Green's function (TDGF) between two receivers can be extracted from the time-averaged correlation of synchronized recordings of ocean noise [Roux *et al.*, J. Acoust. Soc. Am. **113**, 2218 (2003)]. The deterministic TDGF appears through spatial and temporal integration over the distributed random noise sources in the ocean and corresponds to the correlation of random noise passing through both receivers. We will present an analysis of the convergence time of this correlation process based on ultrasonic noise and high frequency ocean noise recordings. We will investigate how, for a given environment, the emergence of the TDGF depends on (1) the spatial and temporal noise distribution, (2) the correlation window length, and (3) the frequency bandwidth.

11:00

5aUW12. Bubble and bubble cloud excitation. Harish Gopolan, William M. Carey, and Ronald R. Roy (Boston Univ., Boston, MA 02215)

Bubbles, spray, splash, and clouds are formed by breaking waves and have been shown to be responsible for the production of sound. In the low-frequency-region, i.e., 20–500 Hz, the oscillation of the cloud as a whole is thought to be responsible for the radiation of sound. This paper treats the radiation of individual bubbles by first reviewing the nonlinear bubble dynamic equations and the subsequent linearized equations. Transient solutions are then derived for several initial conditions using the Laplace transform method and are compared to individual bubble experiments to determine and bracket the radiated energy. Compact bubble clouds are shown by use of a multipole expansion to be described by a dampened simple harmonic oscillator equation. The compact cloud is first excited by an initial compression and then by subresonant energy contributions from individual bubbles oscillating at much higher resonance frequencies.This paper compares the radiated energy from fundamental excitation mechanisms. We provide a qualitative answer to the question how the bubble cloud is excited and how it radiates sound. The differences between possible excitation mechanisms are shown to be significant.

11:15

5aUW13. Mid- to high-frequency ambient noise anisotropy and notch-filling mechanisms. Patrick A. Ferat and Juan I. Arvelo, Jr. (11100 Johns Hopkins Rd., Laurel, MD 20723)

Mid- to high-frequency noise (>3 kHz) is dominated by sub-surface bubble breaking caused by wind-driven wave action. For a submerged receiver in a downward-refracting environment, a noise notch centered about the horizontal is predicted in the vertical noise directionality. The noise notch width is related to the ratio of the sound speed at the receiver to the sound speed at the surface. The noise notch depth is determined by other noise filling mechanisms. For receiver arrays with vertical beamwidths sufficiently narrow to look in the notch, the characterization of the notch level is important for the prediction of beam noise levels near the horizontal. Three notch-filling mechanisms are described: down-slope

conversion of surface energy due to bathymetry, volume scattering from volume inhomogeneities, and internal waves. A ray-based noise model is used to examine these mechanisms, as well as the sensitivity to uncertainty in environmental inputs.

11:30

5aUW14. Correlation of underwater ambient noise with wind speed in extreme conditions. Joal J. Newcomb (Naval Res. Lab., Stennis Space Center, MS 39529, newcomb@nrlssc.navy.mil), Wesley R. Hillstrom (Naval Oceanogr. Office, Stennis Space Center, MS 39522), and Ralph R. Goodman (The Univ. of Southern Mississippi, Stennis Space Center, MS 39529)

During August 2002, the Littoral Acoustic Demonstration Center (LADC) deployed three Environmental Acoustic Recording System (EARS) buoys, at approximately the same location in the northern Gulf of Mexico, in water depths of 645 m to 1034 m. The hydrophone of each buoy was approximately 50 m from the bottom. During the experiment, Hurricane Lili passed within approximately 116 nautical miles west of the EARS buoys. The proximity of this hurricane to the EARS buoys, in conjunction with wind speed data from three nearby NOAA weather buoys, allows for the direct correlation of underwater ambient noise with wind speed. During hurricane conditions, it is expected that the noise contribution due to local shipping would be absent. Time-series data from the shallowest and the deepest of the EARS buoys were processed to yield calibrated spectral levels. These spectra are compared to corresponding spectra derived from Wenz. In the wind-dependent frequency range, a general consistency of agreement is observed in spectral shape with the Wenz spectra and the calibrated spectral levels. General consistency is also observed with the NOAA buoy data and Beaufort scale estimates. Shallow water noise levels are generally higher than temporally corresponding deep-water levels. [Research supported by ONR.]

FRIDAY AFTERNOON, 14 NOVEMBER 2003 TRINITY A AND B ROOMS, 1:15 TO 3:00 P.M.

Session 5pBBa

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Topical Meeting on Shock Waves in Medicine III: Orthopedic Applications

James A. McAteer, Cochair
Anatomy MS-5035, Indiana University School of Medicine, 635 Barnhill Drive, Indianapolis, Indiana 46202-5120

Michael R. Bailey, Cochair
Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, University of Washington, Seattle, Washington 98105

Invited Papers

1:15

5pBBa1. High energy devices versus low energy devices in orthopedics treatment modalities. Reiner Schultheiss (SWS Shock Wave Systems AG, Wilen 4, 8574 Illighausen, Switzerland, Reiner.Schultheiss@freesurf.ch)

The orthopedic consensus group defined in 1997 the 42 most likely relevant parameters of orthopedic shock wave devices. The idea of this approach was to correlate the different clinical outcomes with the physical properties of the different devices with respect to their acoustical waves. Several changes in the hypothesis of the dose effect relationship have been noticed since the first orthopedic treatments. The relation started with the maximum pressure p+, followed by the total energy, the energy density; and finally the single treatment approach using high, and then the multiple treatment method using low energy. Motivated by the reimbursement situation in Germany some manufacturers began to redefine high and low energy devices independent of the treatment modality. The OssaTron as a high energy, single treatment electro hydraulic device gained FDA approval as the first orthopedic ESWT device for plantar fasciitis and, more recently, for lateral epicondylitis. Two low energy devices have now also gained FDA approval based upon a single treatment. Comparing the acoustic data, differences between the OssaTron and the other devices are obvious and will be elaborated upon. Cluster analysis of the outcomes and the acoustical data are presented and new concepts will be suggested.

1:30

5pBBa2. Treatment of nonunions of long bone fractures with shock waves. Ching-Jen Wang (Chang Gung Memorial Hospital at Kaohsiung, 123 Ta Pei Rd., Niao Sung Hsiang, Kaohsiung, Taiwan 833)

A prospective clinical study investigated the effectiveness of shock waves in the treatment of 72 patients with 72 nonunions of long bone fractures (41 femurs, 19 tibias, 7 humeri, 1 radius, 3 ulnas and 1 metatarsal). The doses of shock waves were 6000 impulses at 28 kV for the femur and tibia, 3000 impulses at 28 kV for the humerus, 2000 impulses at 24 kV for the radius and ulna, and 1000 impulses at 20 kV for the metatarsal. The results of treatment were assessed clinically, and fracture healing was assessed with plain x-rays and tomography. The rate of bony union was 40% at 3 months, 60.9% at 6 months and 80% at 12 months followup. Shock wave treatment was most successful in hypertrophic nonunions and nonunions with a defect and was least effective in atrophic nonunions. There were no systemic complications or device-related problems. Local complications included petechiae and hematoma formation that resolved spontaneously. In the author's experience, the results of the shock wave treatment were similar to the results of surgical treatment for chronic nonunions with no surgical risks. Shock wave treatment is a safe and effective alternative method in the treatment of chronic nonunions of long bones.

1:45

5pBBa3. Overview of orthopedic shock wave devices. Roy Brown (HealthTronics Surgical Services, Inc., 1841 W. Oak Pkwy., Ste. A, Marietta, GA 30062) and Reiner Schultheiss (SWS Shock Wave Systems AG, Wilen 4, 8574 Illighausen, Switzerland)

This presentation will cover the various types of shock wave devices that are known to have completed serious patient studies for orthopedic applications. This will include high energy and low energy instruments, as well as focused and unfocused devices, and the method which each device uses to generate a shockwave. The presentation will not be limited only to those devices that have been FDA approved, but will include those that are currently in trials or are expected to begin trials with the FDA in the near future. The power of the equipment, the applications, and the number of treatment sessions generally used will be examined. Some of the devices which are evaluated in particular are the equipment from HMT, Dornier, Siemens, and Storz.

2:00

5pBBa4. Acoustic and cavitation fields produced by shock wave therapy (SWT) devices with different generating principles. Robin O. Cleveland and Parag V. Chitnis (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215, robinc@bu.edu)

We present measurements from two SWT devices with different generating principles: an electrohydraulic source (EHS) with a focusing reflector (HMT Equitron) and a pneumatic source (PS) with no focusing (EMS Swiss Dolorclast). The acoustic field was measured with a PVDF membrane hydrophone. For the EHS the focal waveform consisted of a shock wave with a short compressive pulse lasting 1 μs (rise time less than 50 ns) and a negative tail of 1.5 μm. The average peak positive pressure at the focus was 20 MPa and showed little variation with energy level. The waveform from the PS had a compressive phase of 2-μm-duration (rise time 1 μs) followed by a complex negative phase of 4-μs duration. The peak positive pressure increased from 2 to 8 MPa with increasing energy setting. A passive cavitation detection (PCD) system was used to monitor the cavitation activity produced by the two devices. The EHS produced measurable cavitation with a life time that varied between 110 to 140 μs. We were unable to detect cavitation from the PS. These data indicate that the two devices studied here have different physical characteristics and this may affect the processes by which they operate clinically.

2:15

5pBBa5. Simultaneous measurements of the cavitation and acoustic fields of an electrohydraulic lithotripter. Parag Chitnis, Javier Van Cauwelaert, and Robin Cleveland (Aerosp. and Mech. Eng. Dept., Boston Univ., 110 Cummington St., Boston, MA 02215, pchitnis@bu.edu)

Simultaneous measurements of the cavitation and acoustic fields were carried out in an electrohydraulic lithotripter. Cavitation was measured with a dual passive cavitation detector (PCD) and acoustic pressure with a fiber optic probe hydrophone. The devices were positioned so that they measured the same volume of space. Measurements were acquired at various locations in the field of the lithotripter. The characteristic time (t_c) was used as a parameter for quantifying the strength of cavitation activity. At the focus we found a linear correlation between the peak negative pressure ($P-$) of the SW and t_c as well as between the area under the negative tail and t_c. The maximum t_c was found to be 500 μs at $P-$ of 10 MPa and the minimum value of t_c was 200 μs at $P-$ of 5 MPa. An acoustic signal produced due to the collapse of cavitating bubbles was also detected by the hydrophone and was consistent with the collapse signal measured by the PCD. [Work supported by NIH.]

2:30

5pBBa6. Stone orientation affects the mechanism of failure in artificial kidney stones subject to shock waves. Javier Van Cauwelaert and Robin O. Cleveland (Boston Univ., 110 Cummington St., Boston, MA 02215)

Micro computed tomography (CT) imaging was used to follow the progressive development of cracks in artificial kidney stones. The artificial stones were made from U30 cement with a cylindrical shape (6.5 mm diameter and 8.5 mm long). The stones were held within a polypropylene vial in one of three orientations: vertical, horizontal, and angled at 45 deg. The stones were treated with an electromagnetic lithotripter and the initiation and growth of cracks was observed using microCT. The images show that the orientation of the stones with respect to the shock changes the dominant mechanism for fragmentation. Vertical stones developed a spall-like crack near the distal surface, which propagated from the surface to the interior of the stone. Initiation of a secondary spall-like crack was observed proximal to the first crack. Little surface damage was observed. Horizontal stones presented pitting in the proximal surface and erosion in lateral faces, indicating the action of cavitation. Angled stones presented both spall-like fracture in either the leading or the distal corners and surface damage (pitting) in the proximal surface. Experiments are being performed to follow the development of cracks in human kidney stones. [Work supported by the Whitaker Foundation.]

2:45–3:00

Panel Discussion

Session 5pBBb

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Topical Meeting on Shock Waves in Medicine IV: Bioeffects and Devices

Michael R. Bailey, Cochair
Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, University of Washington, Seattle, Washington 98105

Pei Zhong, Cochair
Department of Mechanical Engineering and Materials Science, Duke University, 3000 Science Drive, Durham, North Carolina 27708-0300

Invited Papers

3:15

5pBBb1. Overview of biological-barrier permeabilization with ultrasound. Pierre D. Mourad (Appl. Phys. Lab., 1013 NE 40th St., Seattle, WA 98105)

This talk contains an overview of how ultrasound has been used to permeabilize various biological barriers, quoting liberally and with citation the work of others (with some contribution from the present author). These barriers include skin, blood clots, cell membranes, and the blood–brain barrier. In all cases I will try to review the mechanisms by which ultrasound achieves biological barrier permeabilization, although typically the main mechanism is cavitation.

3:30

5pBBb2. Shock wave induced sonoporation and gene transfer. Douglas L. Miller (Dept. of Radiol., Univ. of Michigan, Ann Arbor, MI 48109-0553)

During shockwave (SW) treatment, cavitation activity can be applied for cell killing. A bonus is that some surviving cells appear to be briefly permeabilized, or sonoporated, allowing them to take up large molecules including DNA. *In vitro* research has indicated that as the number of SW increased, survival declined exponentially but the number of sonoporated cells increased to better than 50% of survivors for 1000 SW. *In vivo* tests have demonstrated SW-induced tumor ablation could indeed be accompanied by the transfection of marker plasmids into mouse B16 melanoma tumors *in vivo*. With intratumor injection of plasmid DNA and air bubbles, significant results were obtained for only 400 SW. In a trial of cancer therapy, the effects of 500 SW combined with interleukin-12 immuno-gene therapy was observed on the progression of two mouse tumors, B16 melanoma and RENCA renal carcinoma. The combination of SW and IL-12 plasmid injection provided a statistically significant inhibition of tumor growth relative to SW alone for both tumor models, demonstrating feasibility for this treatment method. In the future, the development of intravenous gene delivery and improved transfection, together with image-guided ultrasound treatment, should lead to the clinical application of ultrasound enhanced gene therapy. [Work supported by NIH Grant No. EB002782.]

3:45

5pBBb3. Overview of ultrasound-induced lung hemorrhage. William D. O'Brien, Jr. (Dept. of Elec. and Computer Eng., Univ. of Illinois, 405 N. Mathews, Urbana, IL 61801), Douglas G. Simpson (Univ. of Illinois, Champaign, IL 61820), Leon A. Frizzell, Michael L. Oelze (Univ. of Illinois, Urbana, IL 61801), and James F. Zachary (Univ. of Illinois, Urbana, IL 61802)

It is well documented that ultrasound-induced lung hemorrhage can occur in mice, rats, rabbits, pigs, and monkeys. Our own experimental studies have focused on mice, rats, and pigs as animal models. The characteristics of the lesions produced in mice, rats and pigs were similar to those described in studies by our research group and others, suggesting a common pathogenesis for the initiation and propagation of the lesions at the macroscopic and microscopic levels. Five experimental *in vivo* studies have been conducted to evaluate whether cavitation is responsible for ultrasound-induced lung hemorrhage. The studies evaluated the dependencies of hydrostatic pressure, frequency, pulse polarity, contrast agents and lung inflation, and the results of each study appeared inconsistent with the hypothesis that the mechanism for the production of a lung hemorrhage was inertial cavitation. Other dependencies evaluated included beam width, pulse repetition frequency, pulse duration, exposure duration, and animal species and age. The thresholds for producing ultrasound-induced lung hemorrhage, in general, were less than the FDA's regulatory limit of a Mechanical Index (MI) of 1.9. Further, the MI does not appear to provide a risk-based index for lung hemorrhage. [Work supported by NIH Grant No. R01EB02641.]

4:00

5pBBb4. Effects of an acoustic diode on the pressure waveform and cavitation bubble dynamics produced by a piezoelectric shock wave generator. Songlin Zhu and Pei Zhong (Dept. of Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708-0300)

High-speed schlieren imaging, combined with fiber optical probe hydrophone (FOPH) and passive cavitation detection (PCD) were used to access the effects of an acoustic diode (AD) on the pressure waveform and associated cavitation activities produced by a piezoelectric shock wave (PSW) generator. Without the AD, a typical pressure waveform at the focus of the PSW generator consists of a leading shock wave, followed by a tensile wave and several oscillation waves (OWs) of gradually reduced amplitudes. When the AD was placed 30 mm in front of the focus, the amplitude of the tensile wave was reduced and the subsequent OWs were removed. The pulse intensity integral of the tensile wave was reduced by 58%, and subsequently, PSW-induced bubble dynamics were altered significantly. Based on PCD data, the collapse time of cavitation bubble(s) was reduced by about 11%. Although intensive collapse of microbubbles was observed in about 10 μs following the shock front of the original PSW, the forced collapse of microbubbles was not observed when the AD was used, presumably due to the removal of the OWs. Theoretical calculation based on the Gilmore model confirmed these experimental observations. [Work supported by the Whitaker Foundation and NIH.]

4:15

5pBBb5. Comparison of acoustic fields produced by the original and upgraded HM-3 lithotripter. Yufeng Zhou, Songlin Zhu (Dept. of Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708), Thomas Dreyer, Marko Liebler (IHE-Akustik, Universitaet Karlsruhe, D-76128 Karlsruhe, Germany), and Pei Zhong (Duke Univ., Durham, NC 27708)

To reduce tissue injury in shock wave lithotripsy (SWL) while maintaining satisfactory stone comminution, an original HM-3 lithotripter was upgraded by a reflector insert to suppress large intraluminal bubble expansion, which is a primary mechanism of vascular injury in SWL. The pressure waveforms produced by the original and upgraded HM-3 lithotripter were measured by using a fiber optical probe hydrophone (FOPH), which was scanned both along and transverse to the lithotripter axis at 1-mm step using a computer-controlled 3-D positioning system. At $F2$, the pressure waveform produced by the upgraded HM-3 lithotripter at 22 kV has a distinct dual-pulse structure, with a leading shock wave of $\sim$45 MPa from the reflector insert and a 4-μs delayed second pulse of $\sim$15 MPa reflected from the uncovered bottom surface of the original HM-3 reflector. The beam sizes of the original and upgraded HM-3 lithotripter are comparable in both axial and lateral directions. The pressure waveforms measured at the reflector aperture will be used as input to the KZK equation to predict the lithotripter shock wave at $F2$. Furthermore, bubble dynamics predicted by the Gilmore model will be compared with experimental observation by high-speed imaging. [Work supported by NIH.]

4:30

5pBBb6. Design of compact piezoelectric transducers for shock wave applications. Thomas Dreyer, Marko Liebler, Rainer E. Riedlinger (Universitaet Karlsruhe, IHE-Akustik, Kaiserstrasse 12, D-76128 Karlsruhe, Germany, thomas.dreyer@ihe.uka.de), and Siegfried Ginter (Richard Wolf GmbH, D-75434 Knittlingen, Germany)

The application of focused intense sound pulses to treat several orthopedic diseases has gained in importance during the past years. Self-focusing piezoelectric transducers known from ESWL are not well suited for this purpose due to their size. Therefore compact transducers have to be designed. This implies an increase of the pressure pulse amplitude generated at the radiating surface. A stacked placement of two piezoelectric layers driven by two high-voltage pulses with an adjustable delay accomplishes this. Several designs are presented here representing transducers of different sizes. In principle piezoelectric transducers have the ability to vary the pressure pulse shape to a wider extent than other shock wave sources. Based on FEM simulations of the transducer the influence of some driving parameters, like a variation of the interpulse delay or shape of the driving voltage, on the resulting focal pressure signal is demonstrated. The results show the feasibility to control some parameters of the signal, for example the peak negative pressure amplitude. This possibility could provide new aspects in basic research as well as in clinical applications.

4:45

5pBBb7. Development and application of compact piezoelectric ESWT devices. Siegfried Ginter, Werner Krau, Edgar Bauer (Richard Wolf GmbH, Pforzheimerstr. 32, 75438 Knittlingen, Germany), Thomas Dreyer, and Marko Liebler (Univ. of Karlsruhe, Germany)

The usage of shockwave devices permanently expands from initially lithotripsy to orthopedics and perhaps to drug-delivery in future applications. This demands more compact and cheaper but still powerful shockwave devices, which should generate reproducible sound fields and should be adjustable in some way to the special requirements of the therapy. Here several self-focusing piezoelectric transducers based on a double-layer piezo technology are presented. Sound field measurement results show that it is possible to obtain maximal peak pressure amplitudes from 58 to 132 MPa from small to medium sized sources with diameters of aperture ranging from 70 to 220 mm. The peak pressure amplitude and the amount of energy delivered to the focal region can be adjusted in a wide range. Clinical results demonstrate the therapeutic effectiveness for the treatment of orthopedic indications. These compact devices have the ability to meet the requirements of an extended range of therapeutic applications.

Session 5pUW

Underwater Acoustics: Underwater Applications

Peter R. Stepanishen, Chair
Department of Ocean Engineering, University of Rhode Island, Narragansett Bay Campus, Narragansett, Rhode Island 02882

Contributed Papers

1:00

5pUW1. Acoustic transients from subsonic high-speed sources using a linear time-varying systems approach. Peter R. Stepanishen, James Hu (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882-1197), and Robert Kuklinski (Naval Undersea Warfare Ctr., Newport, RI)

Several approaches are available to investigate acoustic transient radiation from high-speed subsonic moving multipole sources in a free field. A new impulse response approach will be presented to address the problem for sources in a free field and ducted environments. In contrast to the case of a fixed source which leads to a convolution integral representation for the field, the new approach leads to an integral representation in which the kernel is related to the impulse responses for stationary sources along the path of the source. A brief physically motivated development of the approach will be first presented and then the approach will be used to address several acoustic transient radiation problems of interest. As a baseline for comparison numerical results will be first presented to demonstrate the well known Doppler effects for moving harmonic sources in a free field environment and an unbounded half space. Numerical results will then be presented to illustrate the additional complexity of the transient field for moving sources in an acoustic waveguide representing a long test acoustic test tank. [Work supported by ONR.]

1:15

5pUW2. Analysis of the near field acoustic signature from subsonic high-speed projectiles in an acoustic test tank. Sam Carroll, James Hu, Peter Stepanishen (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882-1197), and Robert Kuklinski (Naval Undersea Warfare Ctr., Newport, RI 02841)

A series of tests to obtain the near field acoustic signature from subsonic high-speed supercavitating projectiles in an underwater acoustic test tank at NUWC/NPT have been recently obtained. The projectiles are approximately 0.1 meters in length and 0.02 meters in diameter and travel at speeds up to 900 m/s in the test tank. Underwater acoustic near field transient pressure data has been obtained with several sensors and a preliminary analysis of the data is reported. Possible noise source mechanisms which have been investigated include the radiated noise of the launcher, the near field hydrodynamic pressure field associated with the motion and shape of the projectile, acoustic multipath contributions from the walls of the tank, oscillations of the trailing cavity convected with the projectile and the collapse of bubbles from the aft end of the cavity. The data has been analyzed using an empirical mode decomposition/Hilbert–Huang transform (EMD/HHT) method and hybrid methods of EMD/ordinary spectral analysis, EMD/wavelet analysis, and EMD/short time Fourier transform (STFT) methods to obtain frequency–time information. Preliminary results for the relative importance of the launcher noise, hydrodynamic component and the bubble collapse are presented and discussed. [Work supported by ONR.]

1:30

5pUW3. Acoustic near field pressures from subsonic high-speed projectiles. Robert Kuklinski, Thomas Gieseke (Naval Undersea Warfare Ctr., Newport, RI 02841), Sam Carroll, Peter Stepanishen, and James Hu (Univ. of Rhode Island, Narragansett, RI 02882-1197)

Preliminary experiments performed to measure the radiated noise from subsonic high-speed underwater projectiles indicated the presence of a characteristic disturbance in the mean pressure field during the approach of the projectile to the location of the sensing hydrophone in the test tank. A numerical model based on the potential flow field surrounding the projectile was developed to explain this behavior. A series of 10 projectile shots were fired at speeds ranging from 700–900 m/s to investigate this phenomenon. The underwater range in the tank was instrumented with 5 PCB narrow span pressure transducers. The transducer pressure ranges and locations in the tank were selected using the potential flow model predictions. Results from these experiments are presented. The mean pressure field disturbances show steep cross range gradients. Passage of the projectile at a fixed location in the tank was determined by observing self-similar features of the pressure field at different transducer locations. Projectile velocity estimates were determined from time delays between transducers and are compared with break screen data. The cross range gradients and velocity measurement techniques will be used to develop an open range system to track underwater projectiles. [Work supported by ONR.]

1:45

5pUW4. Target re-acquisition using acoustic features with an autonomous underwater vehicle-borne sonar. Joseph Edwards and Henrik Schmidt (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

Concurrent mapping and localization (CML) is a technique for unsupervised feature-based mapping of unknown environments, and is an essential tool for autonomous robots. For land robots, CML can be applied using video, laser, or acoustic sensors, while for autonomous underwater vehicles (AUVs) the only effective transducer in most situations is sonar. In the Generic Oceanographic Array Technology Sonar (GOATS) experiment series, CML was effectively demonstrated using a single AUV. A further hurdle in the full implementation of AUV minehunting is to reacquire and identify targets of interest. Target re-acquisition allows other vehicles to be called into a target location to further investigate with adaptive sonar geometries or alternative sensor suites designed for classification. In this work, the features in the CML-generated map are extended from only spatial coordinates to include acoustic features such as spectral response. It is demonstrated that the inclusion of acoustic features aids in

the global positioning within the map, although the fine positioning is still accomplished through standard CML. In addition, areas that are sparsely populated with targets, e.g., a sandy coastline, are shown to be more readily navigable using acoustic features.

2:00

5pUW5. Acoustic sensitivity analysis of optimal asset allocation for ASW. Donald R. DelBalzo and Erik R. Rike (Neptune Sci., Inc., 40201 Hwy. 190 E., Slidell, LA 70461)

In 1946, Koopman [Operations Evaluation Group Report 56, Navy Department] characterized optimal search for a stationary target in terms of arbitrary space-time effort density, with the searcher's effectiveness defined by cumulative detection probability (given by the exponential search formula). Koopman's solution maximizes target exposure at time t, given failure to detect at all previous times. This plan is myopic because it does not look into the future. In 1980, Brown [Operations Research 28, 1275–1289] generalized this result for a moving target. Brown's algorithm allocates effort to maximize target exposure, given failure to detect by all prior and future search efforts. It converges to an optimal allocation, and typically converges rapidly because it simplifies sensor performance to a single factor (sweep rate in each cell), which eliminates the need for detailed sensor performance predictions. Experience has shown that sweep rate is an excellent predictor of search effectiveness, for high-level analyses. This work describes an approach to optimize several searchers simultaneously and investigates optimal asset allocation for coordinated, multiship missions under detailed conditions. The results show that attention to acoustic details can significantly increase overall search effectiveness, especially when using sensors with different capabilities. [Work sponsored by ONR.]

2:15

5pUW6. Determination of optimum response requrements for underwater fiber-optic and electroacoustic transducers. Fred C. DeMetz, Sr. (Sound Path Technologies, LLC, 102 Ballview Ct., Waveland, MS 39576, demetz@bellsouth.net)

To aid in the design of new sensors the methodology of determining the optimum sensitivity, linearity, and dynamic range requirements of underwater acoustic sensors is described. The relationships between the commonly specified ''rated'' sensitivity and its ''absolute'' detectability are explored in terms of the required minimum and maximum detectable signal levels and background noise in the operating environment and signal conditioning equipment. Performance limitations of fiber-optic and electroacoustic sensor designs and cost tradeoffs between the sensor and signal conditioning equipment are compared.

2:30

5pUW7. A new forecasted method of echo stealth effect for underwater target with laying anechoic tiles. Qiong Zhang (Sic. Test & Cont. Tech. Inst., P.O. Box 67, 116013 Dalian, PROC), Fukuan Zhao (Ship Development & Design Inst., 430064 Wuhan, PROC), and Rongqing Wang (Sic. Test & Cont. Tech. Inst., 116013 Dalian, PROC)

Its target strength can be reduced undoubtedly for an underwater target that is possessed of laying anechoic tiles. For evaluating its echo stealth effect, traditional ideas are a numerical evaluation and test of the full-scale target. This paper suggests a new forecasted method named ''The Law of Equivalent Proportions:'' For laying sound-absorbing material target of sound-absorption coefficient α_N at frequency f_N with a 1/N target mold, its target strength is reduced to ΔTS_N at frequency f_N in the anechoic tank test. If a full-scale target is assumed that lay anechoic tiles with sound-absorption coefficient α_0 at frequency f_0, its target strength is reduced to ΔTS_0 at frequency f_0, and $f_N = Nf_0$ as well as $\alpha_N = \alpha_0$, it is $\Delta TS_0 = \Delta TS_N$. It is evident from the elliptic cylinder test in the anechoic tank that this conclusion is correct to put the target in a lateral place.