CODEN: JASMAN

ISSN: 0001-4966

The Journal of the Acoustical Society of America

Vol. 114, No. 6 December 2003

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 114, NO. 6, DECEMBER 2003

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 114, NO. 6, DECEMBER 2003

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 114, NO. 6, DECEMBER 2003

CONTENTS—*Continued from preceding page*

INDEX TO VOLUME 114

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.*

Announcement of the 2004 Election

In accordance with the provisions of the bylaws, the following Nominating Committee was appointed to prepare a slate for the election to take place on 14 May 2004:

William M. Hartmann, *Chair*
James V. Candy
Ronald R. Freiheit
Diane Kewley-Port
Julian D. Maynard
Stephen C. Thompson

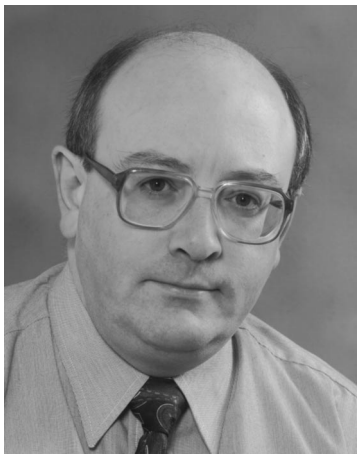
The bylaws of the Society require that the Executive Director publish in the *Journal* at least 90 days prior to the election date an announcement of the election and the Nominating Committee's nominations for the offices to be filled. Additional candidates for these offices may be provided by any Member or Fellow in good standing by letter received by the Executive Director not less than 60 days prior to the election date and the name of any eligible candidate so proposed by 20 Members or Fellows shall be entered on the ballot.

Biographical information about the candidates and statements of objectives of the candidates for President-Elect and Vice President-Elect will be mailed with the ballots.

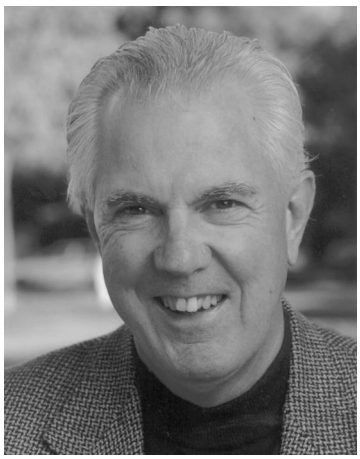
CHARLES E. SCHMID
Executive Director

The Nominating Committee has submitted the following slate:

FOR PRESIDENT-ELECT



Gilles A. Daigle



William A. Yost

FOR VICE PRESIDENT-ELECT

Whitlow W. L. Au

Donna L. Neff

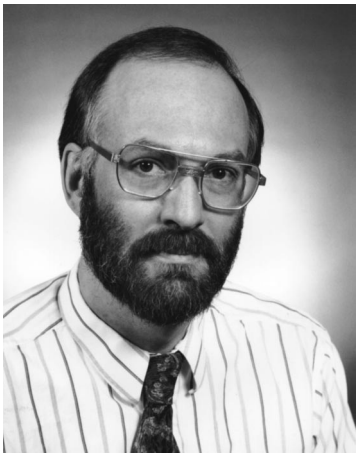
 0001-4966/2003/114(6)/2973/3/$19.00

FOR MEMBERS OF THE EXECUTIVE COUNCIL

Bennett M. Brooks

Judy R. Dubno

E. Carr Everbach



David I. Havelock

K. Anthony Hoover

Victor W. Sparrow

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.

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]. Deadline for receipt of abstracts: 16 January 2004.

3–7 Aug. 8th International Conference of Music Perception and Cognition, Evanston, IL [School of Music, Northwestern Univ., 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).

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.

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: wwwsoc.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)

10–12 **Tenth AIAA/CEAS AeroAcoustics Conference**, Manchester, UK. (Web: www.aiaa.org)

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

June 2004

6–9 **13th International Conference on Noise Control**, Gdynia, Poland. (Web: www.ciop.pl/noise_04)

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)

5–8 ***Eleventh International Congress on Sound and Vibration (ICSV11)**, St. Petersburg, Russia. (Web: www.iiav.org)

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)

30–1 ***Low Frequency 2004**, Maastricht, The Netherlands. (G. Leventhall, 150 Craddlocks Avenue, Ashtead, Surrey KT 21 1NL, UK; Web: www.lowfrequency2004.org.uk)

September 2004

13–16 ***Subjective and Objective Assessment of Sound**, Poznań, Poland. (Institute of Acoustics, Adam Mankiewicz University, Poznań, Poland; Web: www.ia.amu.edu.pl/index.html)

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)

15–17 ***26th European Conference on Acoustic Emission Testing**, Berlin, Germany. (DGZIP, Max-Planck-Str. 26, 12489 Berlin, Germany; Web: www.ewgae2004.de)

October 2004

5–8 **8th Conference on Spoken Language Processing (INTERSPEECH)**, Jeju Island, Korea. (Web: www.icslp2004.org)

November 2004

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

May 2005

16–20 ***149th Meeting of the Acoustical Society of America**, Vancouver, B.C., Canada. (ASA, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502 USA; Fax: +1 516 576 2377; Web: asa.aip.org)

August 2005

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 2005

4–8 ***9th Eurospeech Conference (EUROSPEECH'2005)**, Lisbon, Portugal. (Fax: +351 2131 45843; Web: www.interspeech2005.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 Announcements

August 2005

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

June 2006

26–28 **9th Western Pacific Acoustics Conference (WESPAC 9)**, Seoul, Korea. (Details to be announced later)

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)

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*
IBRAHIM M. HALLAJ, *Wolf Greenfield & Sacks P.C., 600 Atlantic Avenue, Boston, Massachusetts 02210*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
RICHARD S. MCGOWAN, *CReSS LLC, 1 Seaborn Place, Lexington, Massachusetts 02420*
HASSAN NAMARVAR, *Department of BioMed Engineering, University of Southern California, Los Angeles, California 90089*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prairie, 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*

6,356,846

43.20.Tb SYSTEM AND METHOD OF REDUCING MOTION-INDUCED NOISE IN THE OPTICAL DETECTION OF AN ULTRASOUND SIGNAL IN A MOVING BODY OF MATERIAL

Charles C. Habeger, Jr. *et al.*, assignors to Institute of Paper Science and Technology, Incorporated
12 March 2002 (Class 702/40); filed 13 October 1998

Ultrasonic measurements are used in the manufacturing process of sheet materials, such as steel and paper. To ensure quality control, the speed of sound of a test pulse in the material is continually monitored. A laser system scans the surface of the materials to detect the ultrasonic test pulse and make the measurement. The present patent is directed to a problem encountered in rough sheet material such as paper, where noise in the laser detector system due to paper roughness overwhelms the underlying ultrasonic pulse detection. Here, the inventors provide a system where the laser detector moves at high speed along and in the same direction as the moving sheet of paper, thus canceling substantially all relative motion between the laser and the sheet, and therefore generating much less granular surface fiber noise.—IMH

6,343,510

43.20.Ye ULTRASONIC TESTING USING SYNTHETIC IMPULSES

Ian A. Neeson *et al.*, assignors to VN Instruments Limited
5 February 2002 (Class 73/602); filed 22 April 1999

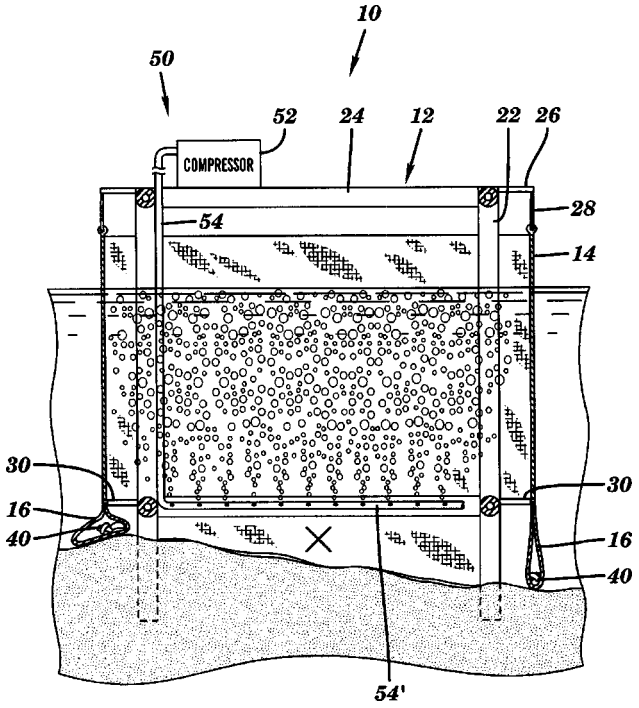
A method for obtaining a property of a test object is given. The method involves delivering a broad-band acoustic pulse towards the object. The pulse is transmitted through or is reflected back by the object. The transmitted or reflected waveform is captured and convolved with a copy of the original broad-band pulse. The resulting convolution provides a very sensitive indication of properties of the object compared to mere examination of the transmitted or reflected waveform.—IMH

6,567,341

43.30.Es BOOM SYSTEM AND ITS USE TO ATTENUATE UNDERWATER SOUND OR SHOCK WAVE TRANSMISSION

Harold B. Dreyer and William F. Gunderson III, assignors to Gunderboom, Incorporated
20 May 2003 (Class 367/1); filed 20 November 2001

The site of some noise generating underwater activity is surrounded by a set of curtains **14** made of flexible material that allows water to flow through them. The curtains are suspended from a support structure **12** containing both vertical **22** and horizontal **24** supports fashioned from rigid pipe of steel or PVC. The vertical supports may penetrate into the floor of the



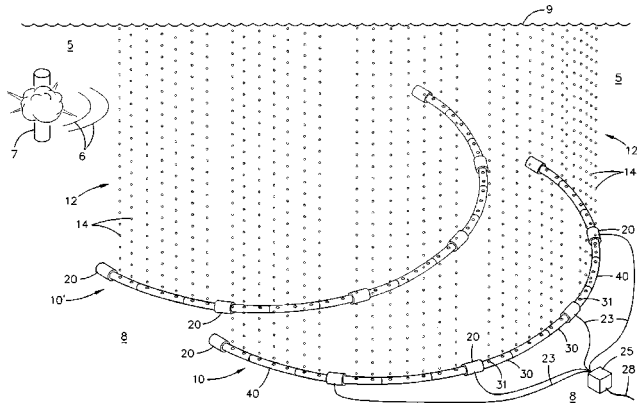
body of water. Skirts **16**, fashioned of a deformable material such as rubberized canvas, hanging from structural elements **30** and containing ballast members **40**, form a seal against the floor of the body of water. Interior to the region bounded by these curtains is a gas injection system **50** consisting of compressor **52** and a set of conduits **54** and **54′**. The latter, which surround the site of the noise generating activity, contain a number of outlets. Compressed gas released through these outlets forms a ''gas curtain'' which attenuates the propagation of the noise thereby containing it within the site. The actual curtains **14** help to prevent the dispersal of the gas curtain by water currents. Alternate curtain construction is discussed.—WT

6,571,906

43.30.Es UNDERWATER SOUND MITIGATION SYSTEM FOR EXPLOSIVE TESTING

William S. Jones *et al.*, assignors to The United States of America as represented by the Secretary of the Navy
3 June 2003 (Class 181/117); filed 20 August 2001

A system to reduce the transmission of acoustic energy from an underwater explosion **5** consists of a set of bottom moored gas generating devices **20** (possibly pyrotechnic or chemically reacting devices although they could simply be manifold structures for receiving pressurized gas through hoses,



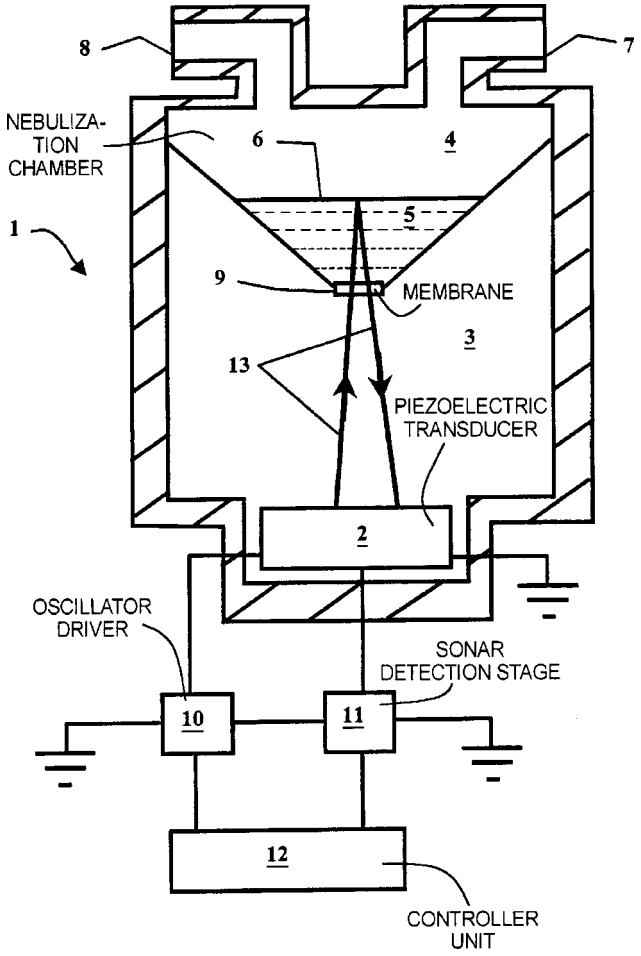
not shown), adjacent heat resistant tubing **30**, and other tubing pieces **40** to form assembly **10** or additional assemblies **10′** of a size deemed sufficient to create an effective barrier. All the tubing pieces are perforated and the resulting, nearly continuous, curtain of rising bubbles mitigates the sound transmission.—WT

6,357,671

43.30.Vh ULTRASONIC NEBULIZER

Göran Cewers, assignor to Siemens Elema AB
19 March 2002 (Class 239/102.2); filed in Sweden 4 February 1999

A nebulizer is a (medical) device that dispenses a fine liquid mist that can be inhaled by a patient. The mist can contain a medication for example. This patent is directed to a nebulizer that monitors the level of liquid in the nebulizing chamber using sonar echo ranging principles. The nebulizing power source is controllable and interruptable to allow an accurate depth of liquid to be obtained by an acoustic transducer placed below the nebulizing chamber and liquid therein. Time of flight measurement provides the measurement of the amount of fluid in the nebulizing chamber. Dosage measurement and indication of low fluid level can then be obtained.—IMH



6,577,557

43.30.Xm WATER COLUMN SOUND SPEED PROFILING SYSTEM

Thomas R. Stottlemyer, assignor to The United States of America as represented by the Secretary of the Navy
10 June 2003 (Class 367/127); filed 14 February 2002

A free-falling source is deployed at a known position and time. It radiates omnidirectional acoustic pulses at predetermined times after deployment. An acoustic receiver at sequential known positions on the surface of the body of water detects these pulses. From knowledge of the time differential between radiations and receptions and of the positions of the receivers and the known rate of descent of the source, the speed of sound can be determined as a function of depth of the water column.—WT

6,567,342

43.30.Yj FLARED WAVE-GUIDE PROJECTOR

Christopher John A. Purcell and Richard Alfred G. Fleming, assignors to Her Majesty, the Queen in right of Canada, as represented by the Minister of National Defence
20 May 2003 (Class 367/176); filed 17 June 2002

In a small refinement of an axial driven resonant pipe projector, the two cuplike hollow, free flooding, end masses of a ''dumbbell'' transducer have flared sides so that the cuplike volume is conical in shape rather than a right circular cylinder. It is suggested that this will affect a better impedance match between the transducer and the surrounding water.—WT

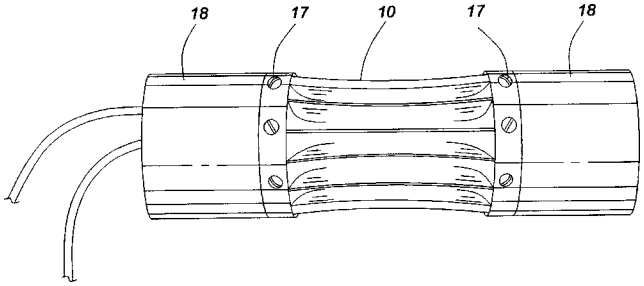
6,567,343

43.30.Yj FLEXTENSIONAL RESONANT PIPE PROJECTOR

Christopher John A. Purcell and Richard A. Fleming, assignors to Her Majesty the Queen in right of Canada, as represented by the Minister of National Defence

20 May 2003 (Class 367/176); filed 17 June 2002

A flextensional transducer consists of a longitudinally vibrating stack of piezoceramic discs with end masses of diameter greater than the discs. That assembly is then shrouded by a one-piece concave aluminum shell **10** with longitudinal folds in the shell defining would-be staves of a typical flextensional shell; hence there are no gaps or openings in the outer surface. This is called a folded shell flextensional projector. The refinement discussed here is to attach an open-ended, cylindrical, aluminum waveguide **18** to the end masses at either or both ends. The length of this waveguide is chosen to yield an organ-pipe resonance at a frequency somewhat higher than the fundamental resonant frequency of the folded shell flextensional projector alone, resulting in a double-peaked, hence broader bandwidth, transmitting voltage response.—WT



6,570,821

43.30.Yj LOW-NOISE TOWED ACOUSTIC LINEAR ANTENNA

Gilles Moresco, assignor to Thales Underwater Systems S.A.S.

27 May 2003 (Class 367/154); filed in France 10 November 1999

The boot of a towed line array is pierced with a number of holes to allow the inner space to free flood. The interior hydrophones are individually enclosed in oil- or gel-filled polyurethane jackets. Buoyancy of the whole structure is adjusted by doping the outer boot with hollow microspheres during manufacture. Much of the interior space is filled with an open-cell foam material. The hydrophones are said to be acoustically decoupled from the array structure itself by this design and hence the self-noise value is said to be significantly lower than that of other designs.—WT

6,580,661

43.30.Yj HYDROPHONE ARRAY

Richard Anton Marschall, New South Wales, Australia *et al.*

17 June 2003 (Class 367/154); filed in Australia 22 December 1998

A towed hydrophone array is discussed which consists of a number of transducer modules, each housed in a hydrodynamically shaped pod and spaced along a cable. The transducer elements are generic and could be PZT, PVDF, PZR, fiber optic, etc. In general they are cylindrical in shape and concentric with the cable.—WT

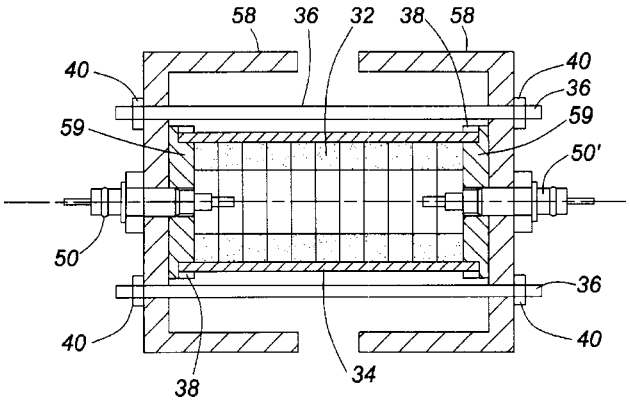
6,584,039

43.30.Yj MULTI-MODE PIPE PROJECTOR

Christopher John A. Purcell and Richard A. Fleming, assignors to Her Majesty the Queen in right of Canada, as represented by the Minister of National Defence

24 June 2003 (Class 367/176); filed 17 June 2002

In a modification of the so-called axial-drive resonant pipe projector, the cuplike end masses **58** of a dumbbell-type longitudinally vibrating transducer are oriented towards one another as indicated in the figure rather than away from one another. The cylindrically shaped space created between the walls of the end masses **58** and the neoprene boot **34** that encloses the stack of piezoceramic discs **32** is free flooding. The longitudinal gap between the two end masses allows for acoustic radiation to the surrounding medium because of motion of that entrained water as, of course, the axial motion of the end masses also does. Items **36** are a set of prestress bolts located externally to the piezoceramic. The dimensions of the cuplike end masses are chosen such that an additional resonance(s) occurs in the water-filled waveguide at a frequency close to the fundamental longitudinal resonant frequency of the transducer itself, thereby resulting in a double humped, but broader bandwidth, transmitting voltage response. In an alternate embodiment, the end masses may be fabricated from a disc and a separate short cylinder rather than being milled from a solid block as implied by this figure.—WT



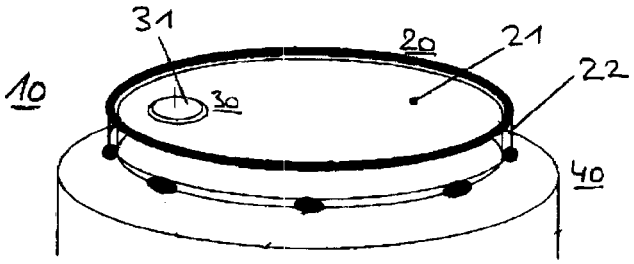
6,561,005

43.35.Pt PROCEDURE AND DEVICE FOR ACOUSTICALLY DETECTING MICROPARTICLES

Martin Müller *et al.*, assignors to Max-Planck-Gesellschaft zur Forderung

13 May 2003 (Class 73/1.74); filed in Germany 10 January 2000

In biotechnology and genetic engineering, minute amounts of solid or liquid substances (microparticles) may be deposited by a dispensing head onto a substrate at defined locations. Calibration of the dispensing head is



difficult because the particles are too small to be checked visually. However, they have enough mass to excite acoustic waves in a suitable sensor film **21**. As the dispensing head is moved across the surface, these waves are picked up by one or more sensors **31** and the resulting electrical signals are processed to determine the relative positions of the impact points.—GLA

6,580,841

43.35.Sx CHIRPED FIBER ACOUSTO-OPTIC BANDPASS FILTER

Duane Anthony Satorius, assignor to The United States of America as represented by the Director of the National Security Agency
17 June 2003 (Class 385/7); filed 8 November 2001

The object of the device described in this patent is to provide an improved chirped fiber acousto-optic band-pass filter that does not require a core block and uses less electrical power. In this filter, one or more bands of optical wavelengths may be selected for further transmission. 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 chirped broadband cladding mode coupler. These cladding mode light waves then enter a narrow-band core mode coupler, in which selected optical bands of wavelengths are recoupled back into the core of the optical fiber. The chirped broadband cladding mode coupler is isolated from the narrow-band core by an acoustic absorber in order to limit acoustic interaction between them.—DRR

6,587,255

43.35.Sx ACOUSTO-OPTIC SCANNING SYSTEM WITH FAST NON-LINEAR SCAN

Nir Davidson *et al.*, assignors to Yeda Research and Development Company Limited
1 July 2003 (Class 359/314); filed in Israel 14 December 2000

The device purports to provide an acousto-optic scanner capable of high scanning speed through the use of two acoustic waves with the same frequency modulation, propagating in opposite directions through one or more acousto-optical media disposed in the path of the beam to be scanned. It is said that this technique completely suppresses any linear frequency chirp, thus enabling the generation of fast nonlinear scans with only a limited reduction in the number of resolvable points as compared to the limited scan case. By changing the phase between the modulating signals, the method also provides fast longitudinal scans of the focal plane.—DRR

6,545,947

43.35.Yb ACOUSTIC MATCHING MATERIAL, METHOD OF MANUFACTURE THEREOF, AND ULTRASONIC TRANSMITTER USING ACOUSTIC MATCHING MATERIAL

Hideki Morozumi *et al.*, assignors to Matsushita Electric Industrial Company, Limited
8 April 2003 (Class 367/152); filed in Japan 12 November 1999

Materials to match the acoustic impedance of solid transducer surfaces to that of air are obtained by using multiple fine pieces, such as small hollow glass spheres, that are bonded to each other at contact points or encapsulated in a matrix.—EEU

6,580,374

43.35.Yb AUDIBLE COMMUNICATION SYSTEM

Martin H. Schrage, Arlington, Massachusetts
17 June 2003 (Class 340/933); filed 3 August 2001

An ultrasonic projector is proposed as an adjunct for safety by placing such devices at crosswalks, dangerous intersections, and the like. According to the inventor, the demodulation of the ultrasonic beam can be accomplished by a windscreen (windshield) of an automobile. The remainder of the patent explores these possibilities.—MK

6,582,369

43.35.Yb METHOD FOR DYNAMIC FOCUS CONTROL

Jing-Jung Huang and Pai-Chi Li, assignors to Computed Ultrasound Global Corporation
24 June 2003 (Class 600/447); filed 2 January 2002

This method provides for dynamic focusing of a coherent array imaging system. Dynamic focusing in ultrasonic array imaging entails real time computations and data communication and can be rather complex for real-time imaging using fully sampled two-dimensional arrays. The argument is made that the use of both spatial and temporal characteristics of the focusing delay patterns provides a significant simplification of the delay control mechanism. The simplification is said to result from (1) grouping adjacent channels into subapertures for the range-dependent focusing component and (2) nonuniform quantization of the delay values.—DRR

6,338,765

43.35.Zc ULTRASONIC IMPACT METHODS FOR TREATMENT OF WELDED STRUCTURES

Efim S. Statnikov, assignor to UIT, L.L.C.
15 January 2002 (Class 148/558); filed 8 April 1999

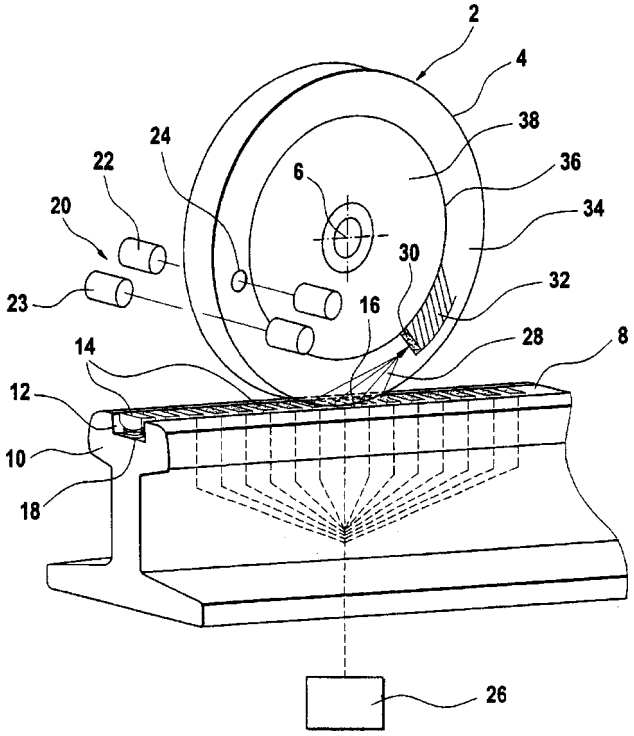
This patent describes a method for improving fatigue strength and life of work product materials having concentrated stress regions. The method relies on heating an exterior surface of the work product by ultrasonic impact of strength, frequency, and phase to induce the desired heating. A tool having needle-shaped transducer extensions is placed near the concentrated stress regions to affect the impact to specific regions. The procedure can be applied at the time of manufacture, during the useful life of a product, or after a failure and repair has taken place.—IMH

6,347,550

43.35.Zc ULTRASONIC TEST DEVICE

Michael Kroening *et al.*, assignors to Fraunhofer-Gesellschaft zur Foerderung der Angewandten Forschung E.V.
19 February 2002 (Class 73/598); filed in Germany 22 July 1998

Multiple ultrasonic probes **14** are disposed in a groove **12** of a track **10**, such as a rail. Acoustic signals are sent to a passing cylindrical object to be tested, such as a railway wheel **2**. The returned signals are sensed and used to determine whether a flaw exists in the test object **2**. The ultrasonic



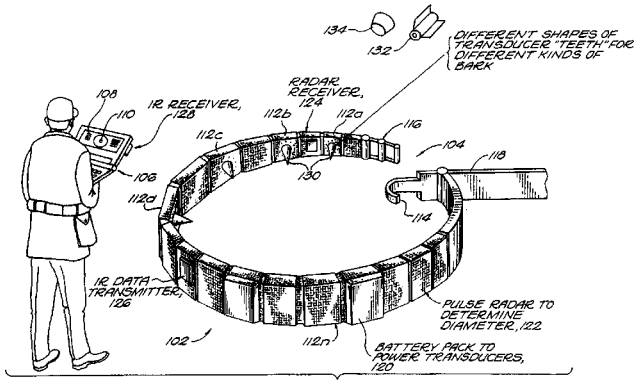
probes are mounted on flexible supports such as springs **18**. Auxiliary systems for monitoring the relative positions of the ultrasonic probes and the track or test object are described.—IMH

6,347,551

43.35.Zc ACOUSTIC TREE AND WOODEN MEMBER IMAGING APPARATUS

Roger M. Turpening *et al.*, assignors to University of Alaska; Massachusetts Institute of Technology
19 February 2002 (Class 73/628); filed 6 November 2000

Trees are checked to make sure they are healthy before killing them. That is, a tree or wooden member, e.g., a log, is acoustically tested for its internal structure using an acoustic array of transceivers **102** surrounding the tree and computed tomography techniques. To obtain good contact, a belt **102** is placed around the tree, and tightened and secured with sharp spikes **130** on the inner face of the belt. Signals are sent into the tree by each of the transceivers and received by each of the other transceivers to form the CT image of a cross section of the tree. A human-wearable version of the device is not mentioned.—IMH



6,347,552

43.35.Zc DYNAMIC NON-INVASIVE DETECTION OF ANALYTICAL CONTAINER FEATURES USING ULTRASOUND

Paul E. Purpura *et al.*, assignors to Bayer Corporation
19 February 2002 (Class 73/633); filed 5 September 2000

In container racks for holding specimens and liquids, it is desired to measure whether a container is present in each slot of the rack, and whether the container is capped, and other information about the contents or level of filling of the containers. Accordingly, the rack and its containers are passed beneath an acoustic source and receiver that use transmitted and received acoustic pulses to determine at least one level of filling or status of each container. Applications include medical laboratory or pharmaceutical manufacturing where vials and containers are placed in racks.—IMH

6,545,945

43.35.Zc MATERIAL CLASSIFICATION APPARATUS AND METHOD

David D. Caulfield, assignor to Ocean Data Equipment Corporation
8 April 2003 (Class 367/87); filed 26 February 2001

Materials in a container (such as a suitcase) or buried in the ground (e.g., explosives or pipelines) are identified by the use of electromagnetic or acoustic sources and receivers. Energy pulses of different frequencies are directed at the area of interest and the received signals are analyzed to determine the rate of change of energy absorption with frequency. This information can then be compared with a database of known objects.—EEU

6,552,803

43.35.Zc DETECTION OF FILM THICKNESS THROUGH INDUCED ACOUSTIC PULSE-ECHOS

Haiming Wang *et al.*, assignors to KLA-Tencor Corporation
22 April 2003 (Class 356/503); filed 17 August 1999

Measurement of the thickness of a film in microelectronics manufacturing is accomplished by directing laser pulses at the film surface. The resulting acoustic pulse is reflected from the interface between the film and the substrate, and is again reflected from the free surface. Heterodyne interferometry is used to measure the time interval between the first and second reflections, from which the film thickness is determined.—EEU

6,554,826

43.35.Zc ELECTRO-DYNAMIC PHASED ARRAY LENS FOR CONTROLLING ACOUSTIC WAVE PROPAGATION

Dana L. Deardorff, assignor to TxSonics-Limited
29 April 2003 (Class 606/27); filed 21 April 2000

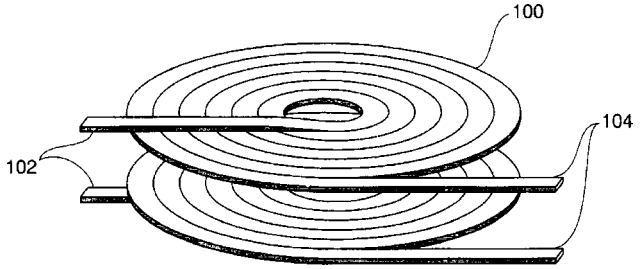
This ultrasonic lens consists of a planar array of cells, each element of which may be controlled separately, so that incident ultrasonic energy can be focused at desired points without the need for changing lenses.—EEU

6,570,819

43.38.Dv LOW FREQUENCY ACOUSTIC PROJECTOR

Joseph A. Clark and Jane Ann Young, assignors to The United States of America as represented by the Secretary of the Navy
27 May 2003 (Class 367/142); filed 8 March 2002

Two separate, parallel, spiral-wound coils of wire **100** are spaced apart within some encapsulating shell (not shown). Voltages applied across the wire ends **102** and **104** are such that currents of opposite polarity flow in the two coils. The magnetic forces thus created result in the two coils being repelled from each other. For a sinusoidal electrical excitation, the resulting



vibration is at twice the drive frequency because the forces are always repulsive regardless of the polarity of the electrical excitation. Some silicon rubber bands (not shown) plus the hydrostatic water pressure provide restorative forces to the two coils.—WT

6,546,106

43.38.Ja ACOUSTIC DEVICE

Henry Azima, assignor to New Transducers Limited
8 April 2003 (Class 381/152); filed in the United Kingdom 2 April 1998

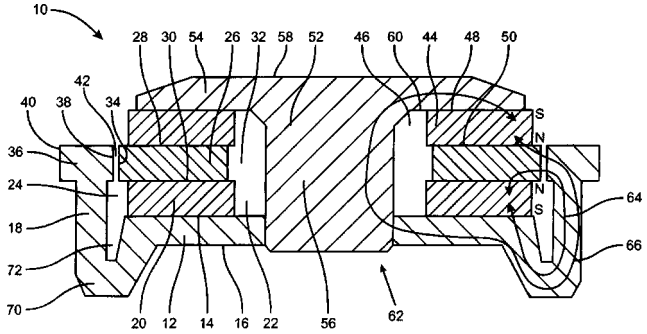
A flat loudspeaker with favorable frequency characteristics is formed from a panel with suitably selected edge restraints and distributed ribs that divide the panel into subpanels of different areas, thicknesses, and damping.—EEU

6,563,932

43.38.Ja MAGNET SYSTEM FOR LOUDSPEAKERS

Paul Cork, assignor to KH Technology
13 May 2003 (Class 381/412); filed in the United Kingdom 16 January 2001

An electrodynamic loudspeaker magnetic circuit has a ring-shaped gap **42** between inner and outer pole pieces **26** and **36**. Typically, the gap would



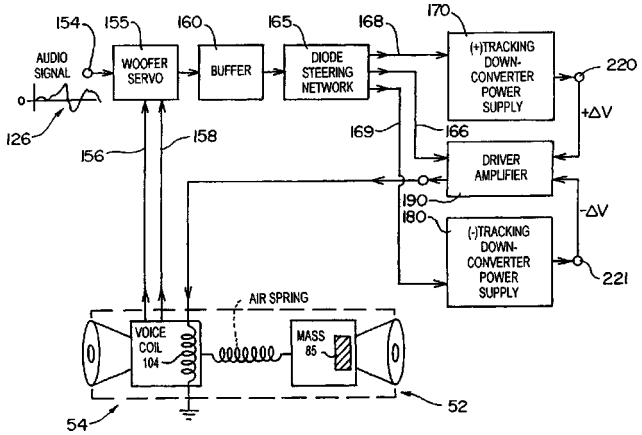
be energized by a single magnet **20**. The addition of complementary magnet **44** is said to overcome deficiencies of prior art in terms of reduced size, improved performance, and ease of assembly.—GLA

6,566,960

43.38.Ja HIGH BACK-EMF HIGH PRESSURE SUBWOOFER HAVING SMALL VOLUME CABINET LOW FREQUENCY CUTOFF AND PRESSURE RESISTANT SURROUND

Robert W. Carver, Snohomish, Washington
20 May 2003 (Class 330/297); filed 12 May 1999

In 1998, the inventor was granted United States Patent 5,748,753 for a ''High Power Audio Subwoofer''—a small, self-powered system made up of a long-throw 12-in. woofer, a passive radiator, and an unusual power amplifier. The combination delivered impressive amounts of acoustic energy



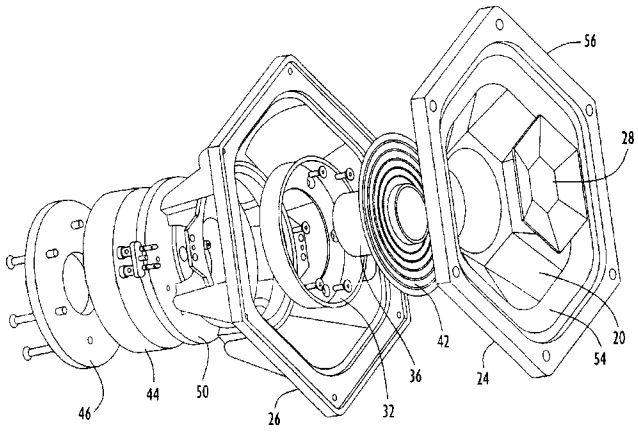
down into the 25-Hz region and achieved immediate commercial success. The patent at hand describes the latest in a series of subsequent improvements. After several pages devoted to setting up and demolishing straw men the patent discloses an interesting amplification circuit capable of swinging more than 100 V into a 4-ohm voice coil.—GLA

6,567,529

43.38.Ja LOW FREQUENCY LOUDSPEAKER DRIVER HAVING A POLYGONAL DIAPHRAGM AND METHOD FOR MOUNTING DRIVERS IN A TIGHTLY PACKED TWO-DIMENSIONAL ARRAY

Richard S. Roark, assignor to Mitek Corporation
20 May 2003 (Class 381/423); filed 28 December 2001

Consider a conventional loudspeaker having a round cone in a round frame. If a number of these are packed into a tight planar array, there is some space left over where the frames don't touch. The patent argues that



this space can be reclaimed by configuring the speaker as a hexagonal cone in a hexagonal frame. True, but an even more efficient design is the well-known round cone in a hexagonal frame. Oh well, 13 patent claims must count for something.—GLA

6,568,504

43.38.Ja MULTI PURPOSE HEADGEAR

John H. Cowgill and Charles W. Elroy, Jr., assignors to Sportniks, Incorporated
27 May 2003 (Class 181/178); filed 26 November 2001

A collapsible watch cap can be extended and used as a megaphone.

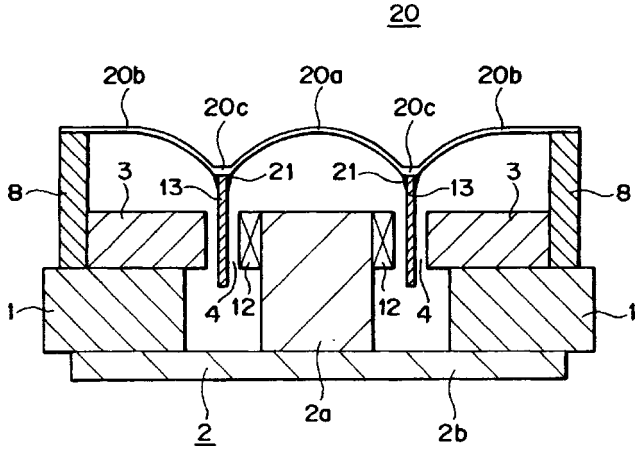
The illustration shows a man literally talking through his hat.—GLA

6,570,995

43.38.Ja SPEAKER DEVICE

Yoshio Ohashi and Masaru Uryu, assignors to Sony Corporation
27 May 2003 (Class 381/430); filed in Japan 14 November 2000

Sony loudspeaker engineers seem to be devoting a good part of their time to designing super super tweeters. These devices are intended to provide smooth response in the range above 20 kHz. The example shown is an induction tweeter. Dome **20a** and outer ring **20b** are formed integrally from a material having an acoustic loss coefficient greater than 0.02 above 20 kHz. Single-turn coil **13** is bonded rigidly to the diaphragm and driven by fixed coil **12**. According to the patent, frequencies higher than 20 kHz are



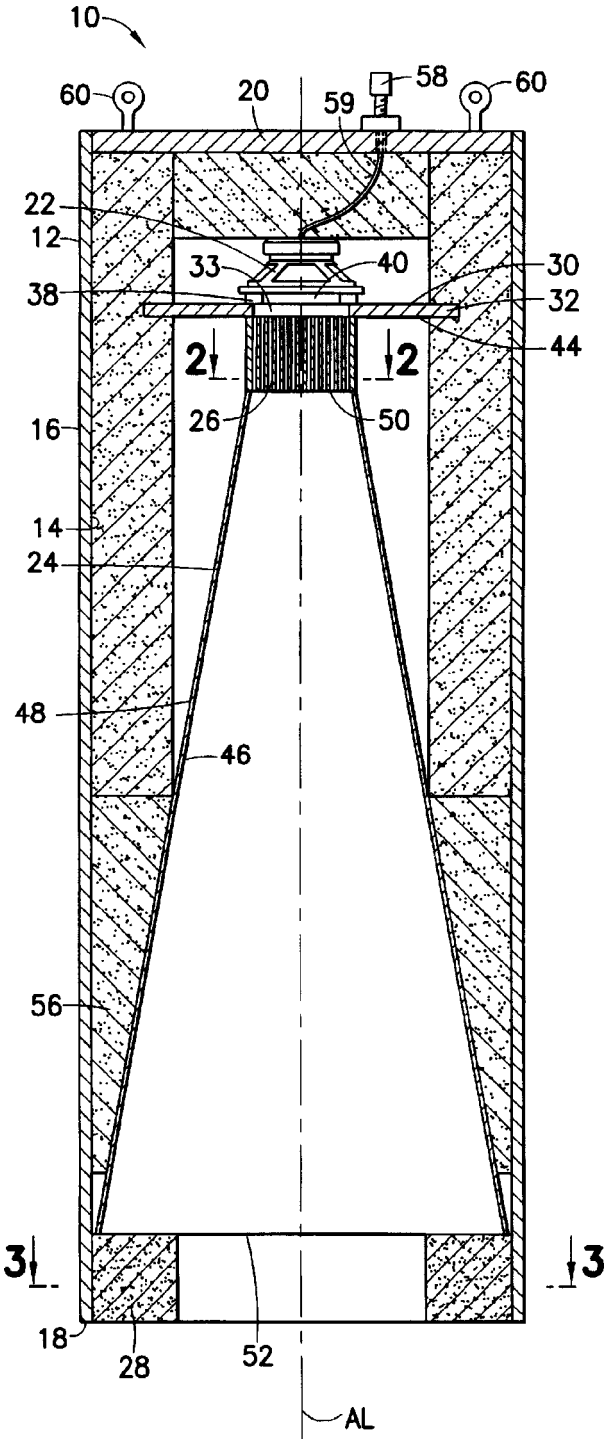
reproduced ''...utilizing split vibrations of said speaker diaphragm,'' but internal losses minimize peaks and dips in response.—GLA

6,574,344

43.38.Ja DIRECTIONAL HORN SPEAKER SYSTEM

David Wiener and Steven Burgess, assignors to Soundtube Entertainment, Incorporated
3 June 2003 (Class 381/343); filed 12 April 2000

This is a continuation of United States Patent 6,055,320 [reviewed in J. Acoust. Soc. Am. **109**(2), 448 (2001)] and describes the same basic configuration: a loudspeaker **22**, mounted on floating panel **32**, drives waveguide **24** which is fitted with ''acoustic trap'' **28** at its mouth. An ''in-line phase plug'' **26** is intended to somehow flatten the wave front. Two variants are described



which provide additional wave-shaping means and also vent the rear chamber through the acoustic trap. ''At the perimeter of the wave guide, the

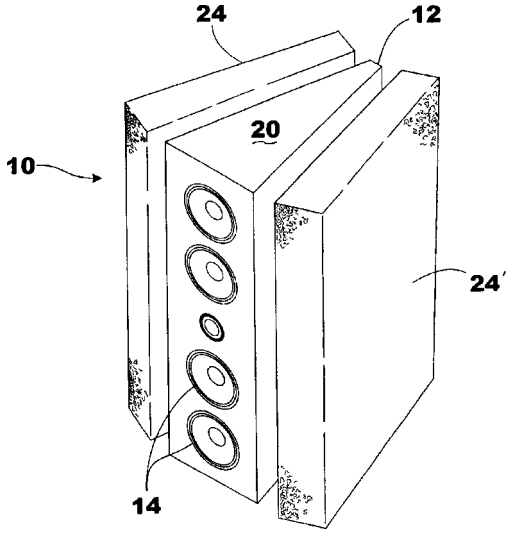
forward and rearward propagated sound waves are theoretically canceled by each other to confine the remaining sound waves and control dispersion.''—GLA

6,571,909

43.38.Ja LOUDSPEAKER ENCLOSURE

Jeffrey Olinger, West Des Moines, Iowa
3 June 2003 (Class 181/199); filed 31 August 2001

Foam blocks **24** are apparently intended to damp the side walls of box **20** and at the same time soak up unwanted diffracted waves. The blocks are



made of polyurethane—the patent claims are specific about this—but no mention is made of open-cell versus closed-cell foam.—GLA

6,574,346

43.38.Ja BASS REPRODUCTION SPEAKER APPARATUS

Shoji Tanaka, assignor to Matsushita Electric Industrial Company, Limited
3 June 2003 (Class 381/421); filed in Japan 26 April 1999

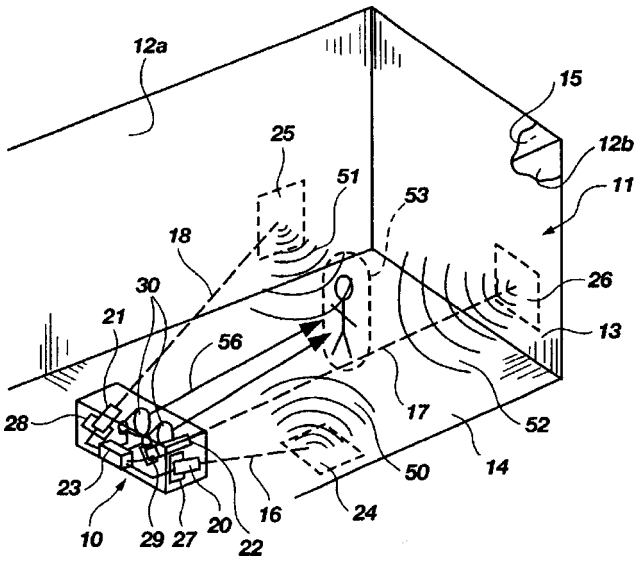
In a closed box loudspeaker system, air stiffness can be largely cancelled by fitting the loudspeaker cone suspension with a negative spring. Unfortunately, an air pump controlled by a servo mechanism is needed to keep the voice coil centered during operation. This arrangement was patented in 1957 (United States Patent 2,810,021). Wouldn't it be easier to include dynamic dc offset in the audio drive signal itself? This patent describes a practical method of doing just that. A skeptic might suspect that overall system efficiency is about the same as if the negative spring were discarded and brute force equalization used instead of dc offset.—GLA

6,577,738

43.38.Ja PARAMETRIC VIRTUAL SPEAKER AND SURROUND-SOUND SYSTEM

Elwood G. Norris and James J. Croft III, assignors to American Technology Corporation
10 June 2003 (Class 381/77); filed 7 May 2001

The inventor continues to amend a basic scheme first set forth in United States Patent 5,889,870. Virtual surround sound sources are created by modulated ultrasonic waves reflected from room boundaries. In the illustration, listener **53** hears direct sound **56** from conventional loudspeakers and surround sound from virtual sources **25** and **26**. According to the patent, this arrangement ''...not only generates an exhilarating sensory response



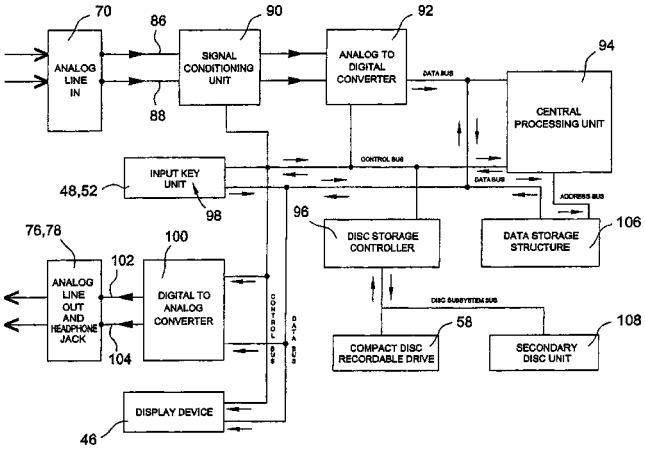
with the listener, but enlarges the experience with a three-dimensional sense of depth within the room.''—GLA

6,587,404

43.38.Ne OPTICAL STORAGE DEVICE CAPABLE OF RECORDING A SET OF SOUND TRACKS ON A COMPACT DISC

Peter J. Keller *et al.*, assignors to Advanced Audio Devices, LLC
1 July 2003 (Class 369/30.06); filed 17 August 2000

The inventors describe a standalone device for ''ripping'' audio inputs onto CD-R blanks. Internally, they propose using a ''PC compatible motherboard'' to control the whole system. If you are going to go that way, why



bother with a standalone system and just use your tower or laptop? It is difficult to believe this was novel when it was filed in 2000.—MK

6,563,869

43.38.Vk DIGITAL SIGNAL PROCESSING CIRCUIT AND AUDIO REPRODUCING DEVICE USING IT

Yuji Yamada, assignor to Sony Corporation
13 May 2003 (Class 375/232); filed in Japan 15 May 1998

A number of ''auralization'' methods have been developed to create a three-dimensional sound environment for a listener wearing headphones.

Many of these use FIR digital filters to replicate the complex sound paths from a given source to the listener's left and right ears. This patent points out that to simulate a typical sound environment, fewer taps are needed for high frequencies than for low frequencies. ''The present invention greatly decreases the number of taps of a digital signal processing circuit, that is, the number of delay circuits and multiplication circuits, by paying attention to the above point.''—GLA

6,574,339

43.38.Vk THREE-DIMENSIONAL SOUND REPRODUCING APPARATUS FOR MULTIPLE LISTENERS AND METHOD THEREOF

Doh-hyung Kim and Yang-seock Seo, assignors to Samsung Electronics Company, Limited
3 June 2003 (Class 381/17); filed 20 October 1998

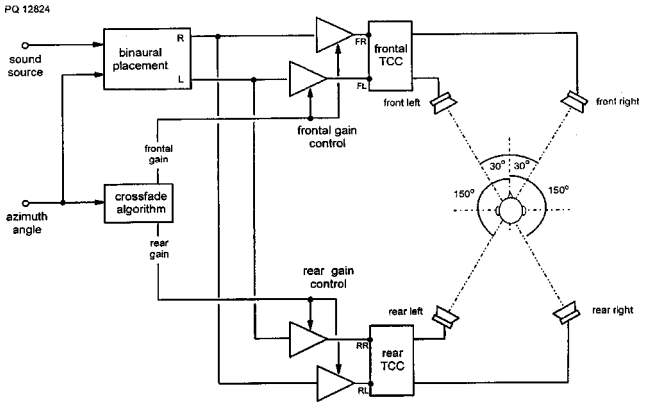
Using head-related transfer functions and some fancy digital filtering, it is possible to create a convincing three-dimensional sound field from two loudspeakers. A major drawback is that the illusion is effectively limited to a single listener at a defined location. We might give the listener a choice of locations by allowing him to select from, say, three filter settings. But suppose that all three settings are selected sequentially at some optimum time interval. Will the result be aural confusion or will the effective sound field be expanded to accommodate multiple listeners? The patent argues for the latter.—GLA

6,577,736

43.38.Vk METHOD OF SYNTHESIZING A THREE DIMENSIONAL SOUND-FIELD

Richard David Clemow, assignor to Central Research Laboratories Limited
10 June 2003 (Class 381/18); filed in the United Kingdom 15 October 1998

A home surround sound system typically uses two or three front speakers plus two rear speakers. When mixing program material for this format, conventional panning techniques work well when moving sound images laterally but cannot create accurate phantom sources between front and rear speakers. Conversely, by making use of head-related transfer functions and



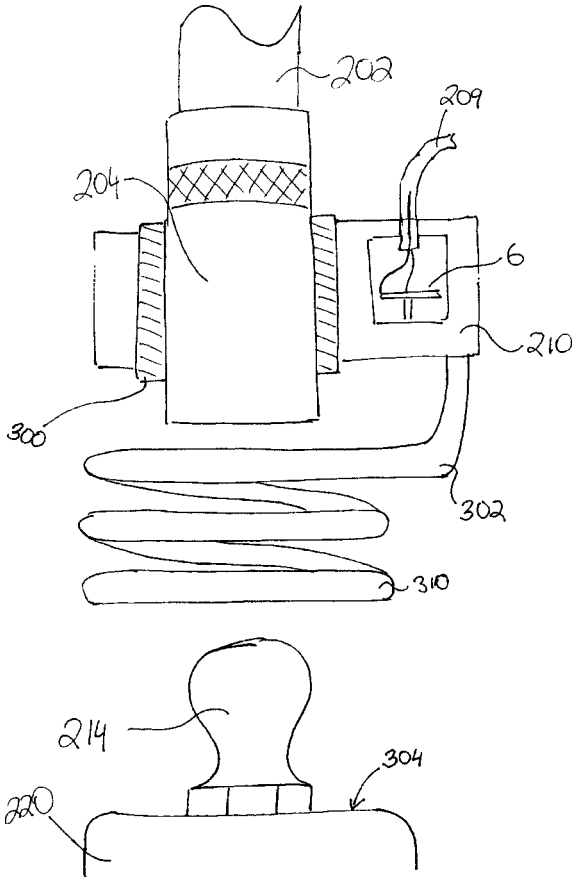
interaural crosstalk cancellation, a two-speaker playback system can theoretically locate sound images almost anywhere if it is designed for one particular listener at a specific location. The inventor has developed an interesting method for combining the best features of both systems. According to the patent, it can easily be combined with conventional multi-channel mixing techniques.—GLA

6,339,961

43.40.Le ULTRASONIC DETECTING LUBRICATION APPARATUS WITH ACOUSTICALLY ISOLATED TRANSDUCER

Mark A. Goodman *et al.*, assignors to UE Systems, Incorporated
22 January 2002 (Class 73/593); filed 9 March 2000

The system shown in the accompanying figure includes a bearing **220** and a transducer **6** that can be acoustically coupled through waveguide **302** to the bearing. An operator uses the transducer to measure the noise levels produced by the bearing as an indication of a need for lubricating the bearing. If an abnormally high level of ultrasound is produced by the bearing, a



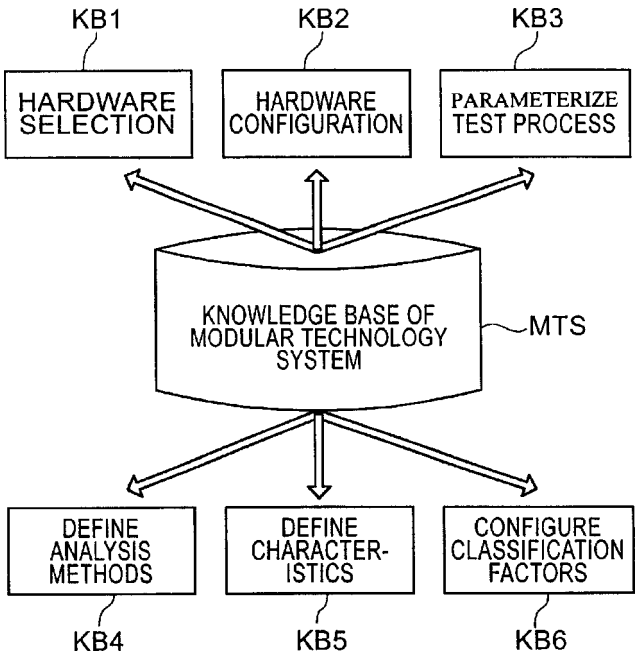
grease gun **202** is used to lubricate the bearing. Elements **204** and **214** are grease fittings to transfer the grease from the gun to the bearing. The transducer is acoustically isolated from the grease gun by an absorbent (e.g., rubber) **300**.—IMH

6,591,226

43.40.Le ACOUSTIC DIAGNOSTIC SYSTEM AND ASSOCIATED METHOD

Doris Hartmann *et al.*, assignors to Siemens Aktiengesellschaft
8 July 2003 (Class 702/183); filed in Germany 17 April 1998

The patent refers to an acoustic diagnostic system and an associated method that are claimed to be suitable for automatic quality control of test objects such as motors or ceramics. The system includes a sensor mechanism to convert mechanical vibrations into electrical signals. The system also features a signal conditioning unit to amplify and low-pass filter the electrical signals and an analog/digital converter to convert the amplified and low-pass filtered signals into discrete digital data. A signal processing unit analyzes the characteristics of the signals and classifies the signals into error classes and/or quality classes which are associated with a respective



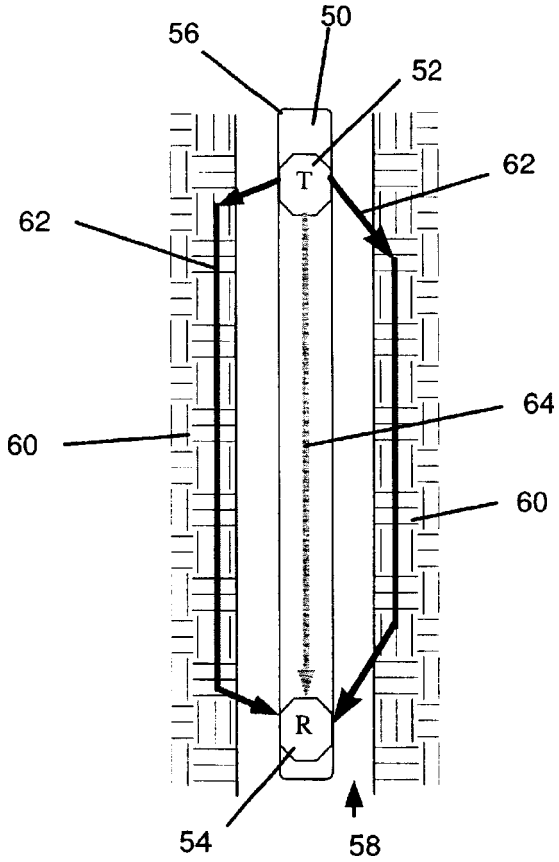
testing objective. The diagnostic system also includes an archive to automatically archive occurrences of the error classes and/or quality classes.—DRR

6,564,899

43.40.Ph METHOD AND APPARATUS FOR ABSORBING ACOUSTIC ENERGY

Abbas Arian *et al.*, assignors to Dresser Industries, Incorporated
20 May 2003 (Class 181/102); filed 24 September 1998

A number of earlier patents deal with acoustic logging tools used during oil well drilling. The illustration shows a known arrangement in which transmitter **52** emits an acoustic signal that travels through strata **60** surrounding well bore **58** and is detected by receiver **54** located at some distance from the transmitter. The transit time can be calculated and used to establish certain characteristics of the formations encountered by the drill bit. It is obvious that vibrations **64** traveling directly through the tool itself will interfere with accurate measurements. The patent describes a number of mechanical isolation techniques designed to attenuate both compressional and flexural vibration modes.—GLA



6,547,049

43.40.Tm PARTICLE VIBRATION DAMPER

Geoffrey R. Tomlinson, assignor to Rolls-Royce plc
15 April 2003 (Class 188/379); filed in the United Kingdom 5 November 1999

One or more chambers are partially filled with multiple particles. When the damper is attached to a vibrating object the particles impact against each other and against the chamber wall, thus removing energy from the primary vibration.—EEU

6,547,294

43.40.Tm SHOCK ABSORBING CONSTRUCTION FOR MOTORCYCLES

Kazuhiro Yamamoto, assignor to Honda Giken Kogyo Kabushiki Kaisha
15 April 2003 (Class 293/105); filed in Japan 27 December 2000

A shock absorber intended to cushion front or rear impacts on a motorcycle consists of a boxlike attachment that is made of plastic and filled with foam. A grillage of ribs within the plastic box provides strength and controlled collapse.—EEU

6,547,526

43.40.Tm ARTICLE HAVING DAMPENING MEMBER INSTALLED INTO AN IMBEDDED CAVITY

Thomas J. Van Daam and Timothy J. Hosking, assignors to The Boeing Company
15 April 2003 (Class 416/190); filed 3 July 2001

This patent pertains primarily to the damping of turbine blades. One or more circumferential cavities extend through the tips of all of the turbine blades and contain loosely fitting rings. Friction between these rings and the cavity surfaces damps the blade vibrations.—EEU

6,550,440

43.40.Tm ACOUSTIC SUPPRESSION ARRANGEMENT FOR A COMPONENT UNDERGOING INDUCED VIBRATION

Gary Allan Vrsek *et al.*, assignors to Ford Global Technologies, LLC
22 April 2003 (Class 123/184.61); filed 17 January 2002

Automotive engine components, such as intake manifolds made from polymeric materials, are friction-damped via elements that are snap-fit onto them.—EEU

6,550,868

43.40.Tm DAMPER PLATE FOR AUTOMOBILE WHEEL

Yuichi Kobayashi and Takumi Sekiguchi, assignors to The Yokohama Rubber Company, Limited
22 April 2003 (Class 301/6.91); filed in Japan 12 December 2000

A disc made of a lamination of steel and elastomeric plates is inserted between an automobile wheel and its axle in order to improve ride comfort and reduce interior noise.—EEU

6,551,057

43.40.Tm DAMPED TORQUE SHAFT ASSEMBLY

Frederic G. Haaser *et al.*, assignors to General Electric Company
22 April 2003 (Class 415/119); filed 22 November 1999

The torque shaft assembly to which this patent pertains is a shaft that is used to adjust the stator vanes on gas turbine engines. This shaft is not part of the rotating assembly, but actuates the stator vanes via connecting elements. In order to limit the vibrations of the torque shaft, it is made in the form of a hollow tube that is filled with a granular material. A 98% by volume filling of steel shot is preferred.—EEU

6,554,112

43.40.Tm VIBRATION-DAMPING DEVICE FOR VEHICLE

Rentaro Kato *et al.*, assignors to Tokai Rubber Industries, Limited
29 April 2003 (Class 188/379); filed in Japan 8 September 2000

The devices described here in essence consist of metal cups in which metal balls can rattle when the assembly is subjected to vibrations. The balls are coated with rubber and so are some of the interior surfaces of the cups. On some cup surfaces the rubber elements are supported only at their edges, so that they tend to act as relatively soft springs, resulting in low natural frequencies for the affected ball rattling.—EEU

6,565,287

43.40.Tm APPARATUS FOR SUPPRESSION OF VORTEX INDUCED VIBRATION WITHOUT AQUATIC FOULING AND METHODS OF INSTALLATION

David Wayne McMillan, Houston, Texas *et al.*
20 May 2003 (Class 405/211.1); filed 19 December 2000

To suppress vortex-induced vibrations of a cylindrical shaped underwater structure, it is surrounded either by a cylindrical shroud or by a set of tubes whose axes are all parallel to the axis of the underwater structure in question. Alternatively this set of tubes may be twisted into a helix about the axis of the underwater structure. All additional components have sufficient copper content to suppress the growth of marine life upon them.—WT

6,571,878

43.40.Tm SMOOTH BUOYANCY SYSTEM FOR REDUCING VORTEX INDUCED VIBRATION IN SUBSEA SYSTEMS

Richard Bruce McDaniel *et al.*, assignors to Shell Oil Company
3 June 2003 (Class 166/367); filed 25 July 2001

To suppress the vortex-induced vibrations of certain underwater structures, they are enclosed within a cylindrically shaped buoyancy element that has an extremely smooth outer surface.—WT

6,543,620

43.40.Yq SMART SCREENING MACHINE

Daryoush Allaei, assignor to Quality Research, Development & Consulting, Incorporated
8 April 2003 (Class 209/365.1); filed 23 February 2001

Distributed electrically controlled transducers are used to vibrate a screen used for separation of materials. This concept is said to result in screens that are smaller and less noisy than conventional ones. It also permits automatic detection of clogging and adjustment of the transducers to provide unclogging.—EEU

6,545,762

43.40.Yq METHOD OF INVESTIGATING VIBRATIONS AND AN APPARATUS THEREFOR

Harvey Lewis and Andrew Rogoyski, assignors to Logica UK Limited
8 April 2003 (Class 356/502); filed in the United Kingdom 17 April 2000

As in other laser vibrometers, light from a laser is split, with part being directed at the test object and part being mixed with light reflected from the object to create an interference signal. The latter is processed to derive a signal representing the object's vibrations. Unlike other laser vibrometers, which do not work in the presence of bulk motion of the test object, the system described in this patent derives an estimate of the frequency of the bulk movement from the observed signal and uses this frequency to reduce the bandwidth of the observation, so as to limit the interference from the bulk motion.—EEU

6,550,574

43.50.Gf ACOUSTIC LINER AND A FLUID PRESSURIZING DEVICE AND METHOD UTILIZING SAME

Zheji Liu, assignor to Dresser—Rand Company
22 April 2003 (Class 181/286); filed 21 December 2000

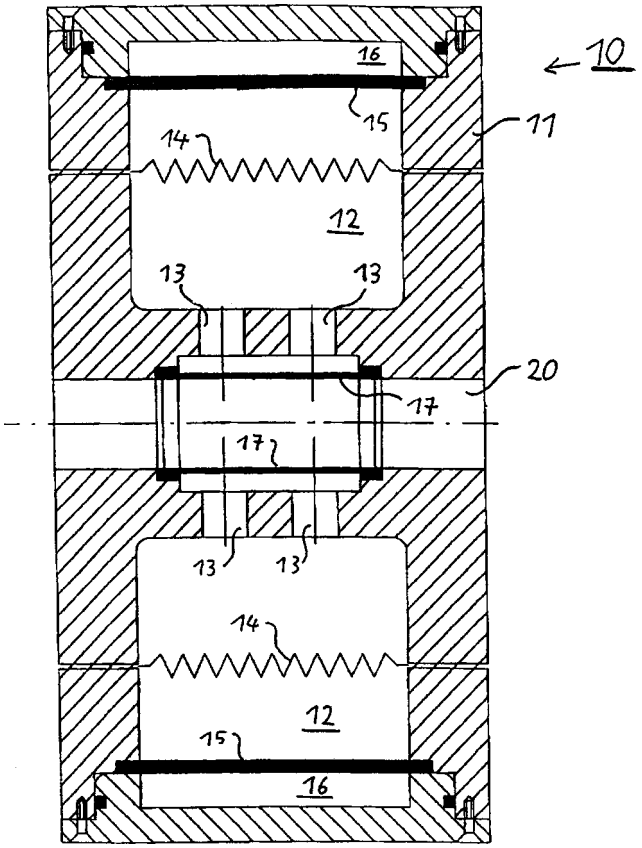
In order to obtain quieter centrifugal compressors a disc-shaped liner consisting of an array of Helmholtz resonators is installed near the compressor's exit volute. This liner, which must be rugged, in essence consists of a plate into which there are drilled a number of large-diameter holes that do not extend entirely through its thickness, with a number of smaller holes drilled through the remaining thickness.—EEU

6,581,722

43.50.Gf ACOUSTIC ABSORBER AND SOUND-ABSORPTION METHOD

Peter Faulhaber and Reinhard Pongratz, assignors to EADS Deutschland GmbH
24 June 2003 (Class 181/250); filed in Germany 24 January 2000

This acoustic absorber attaches to a duct (or other conduit for fluid



flow). The resonant frequency of the absorber can be varied mechanically or thermally depending on the nature and frequency of the source.—CJR

6,589,034

43.50.Gf BACKFLOW ORIFICE FOR CONTROLLING NOISE GENERATED BY A ROTARY COMPRESSOR

Christoph Vorwerk, and Volker Heumann, assignors to Ford Global Technologies, Incorporated
8 July 2003 (Class 418/15); filed in the European Patent Office 21 August 2001

The patent covers the design of backflow orifices for a rotary compressor, e.g., a Roots blower. The fluid conveyed from an inlet to an outlet by two oppositely rotating rotors flows back partially via the backflow orifices. The edges of the backflow orifices vary in height from an inner surface contour of the pump casing. The height variation is said to contribute to a reduction in the noise generated by the compressor.—DRR

6,589,112

43.50.Gf DUCT SILENCER

Evan Ruach, East Peoria, Illinois
8 July 2003 (Class 454/276); filed 21 December 2001

This involves a design for a ventilation duct that contains a number of inner and outer sound absorbing elements. The sound absorbing members lie in separate planes with portions overlapping one another to inhibit the transmission of sound through the duct. The sound absorbing members, each including a longitudinal cavity filled with a sound absorbing material, feature multiple sound transmission passages extending through these members to expose the sound absorption material to incoming sound waves. The members may be shaped to have a reverse megaphone effect to further reduce the sound transmitted across the duct.—DRR

6,543,719

43.50.Nm OSCILLATING AIR JETS FOR IMPLEMENTING BLADE VARIABLE TWIST, ENHANCING ENGINE AND BLADE EFFICIENCY, AND REDUCING DRAG, VIBRATION, DOWNLOAD AND IR SIGNATURE

Ahmed A. Hassan *et al.*, assignors to McDonnell Douglas Helicopter Company
8 April 2003 (Class 244/17.13); filed 24 November 1998

Porous surfaces are provided on aircraft surfaces, such as rotor blades, and are driven with oscillating pressures either continually or when predetermined flight conditions occur. These pressures control the aerodynamics of the surfaces and are claimed, among other things, to be able to produce significant reductions in the blade-vortex-interaction noise of helicopter rotors.—EEU

6,550,332

43.50.Yw REAL-TIME NOISE SOURCE VISUALIZING SYSTEM USING ACOUSTIC MIRROR

Myung-Han Lee, assignor to Hyundai Motor Company
22 April 2003 (Class 73/583); filed in the Republic of Korea 22 November 2000

This system, used in wind tunnel tests of automobiles, employs a microphone located at the focus of a parabolic reflector. The microphone system and a camera connected to it can be traversed to focus on selected areas. A display system provides color-coded sound images.—EEU

6,568,135

43.55.Ev SOUND ABSORBING STRUCTURE

Yoshiaki Yokoyama *et al.*, assignors to Nichias Corporation; Alumu Corporation; Yotsumoto Acoustic Design Incorporated
27 May 2003 (Class 52/145); filed in Japan 22 April 1999

Molded metal fibers are compressed into a sheet. The sheet is then configured as a face plate for a resonant chamber. The plate also incorporates slits that open into preconfigured Helmholtz resonators.—CJR

6,583,193

43.55.Ev POLYOLEFIN FOAM USEFUL FOR SOUND AND THERMAL INSULATION

Chung P. Park *et al.*, assignors to Dow Global Technologies Incorporated
24 June 2003 (Class 521/142); filed 9 March 2001

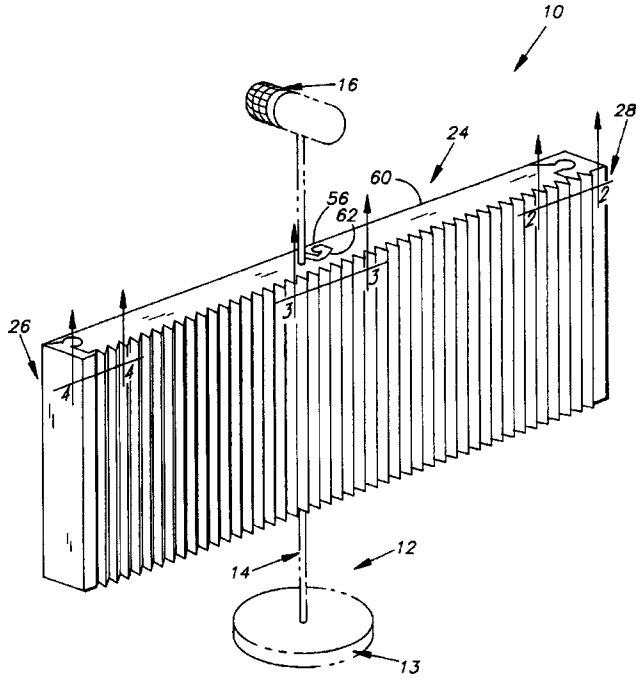
This patent describes an extruded, coalesced foam strand propylene polymer material that is either open celled (hence useful for sound insulation applications) or close celled (hence useful for thermal insulation applications). There is also a description of an apparatus suitable for preparing such foams.—CJR

6,584,736

43.55.Ev STAND-MOUNTABLE FOAM-TYPE ACOUSTIC PANEL

Jeffrey E. Szymanski *et al.*, assignors to Auralex Acoustics, Incorporated
1 July 2003 (Class 52/145); filed 30 March 2001

This foam-type acoustic panel is designed to be mounted on a stand. Also, the formation of the panels allows them to be interlocked and linked to



other panels so that the final arrangement can be best suited for a particular room or space.—CJR

6,590,006

43.55.Ev MACROCELLULAR POLYOLEFIN FOAM HAVING A HIGH SERVICE TEMPERATURE FOR ACOUSTICAL APPLICATIONS

Chung P. Park *et al.*, assignors to Dow Global Technologies Incorporated
8 July 2003 (Class 521/134); filed 13 February 2001

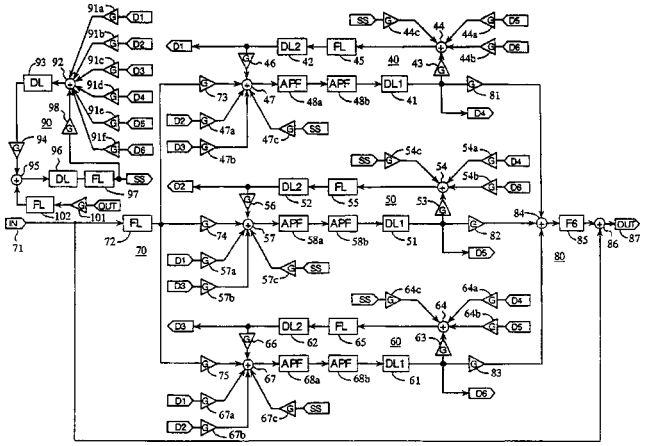
This patent is for sound absorbing and insulating macrocellular polyolefin foams that can withstand temperatures above 110 °C. These foams have a large cell size (greater than 1.5 mm), are either open- or closed-cell foams, and preferably have at least one perforation per square centimeter and/or a density less than 25 kg per cubic meter.—CJR

6,580,796

43.55.Lb SOUND EFFECT IMPARTING APPARATUS

Ryuichiro Kuroki, assignor to Yamaha Corporation
17 June 2003 (Class 381/63); filed in Japan 27 January 1998

You can't tell from the title that the patent covers a reverberation algorithm that emulates the classic sound of a mechanical reverb. The ex-



planation of the signal flow graph is surprisingly coherent but, like all artificial reverberators in the Schroeder class, it is still *ad hoc* in design.—MK

6,581,724

43.55.Ti SOUND ABSORBING CLOSURE PANELS FOR SOUND ISOLATION MODULES

Steven Dutton and John Phillips, assignors to Acoustic Systems, Incorporated
24 June 2003 (Class 181/285); filed 27 April 2001

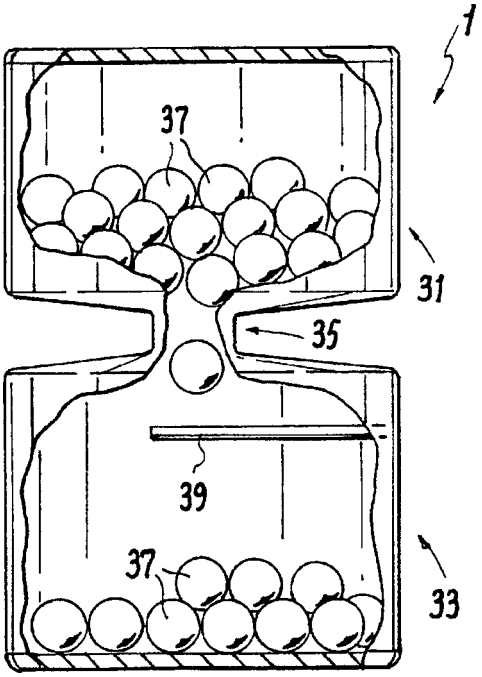
These panels form a secondary (additional) closure around sound isolation modules. This adds a visual closure, improves noise reduction performance, and reduces flanking sound transmission.—CJR

6,341,518

43.58.Hp ULTRASONIC STANDARD

Mark A. Goodman and Betty J. R. Chavez, assignors to U.E Systems, Incorporated
29 January 2002 (Class 73/1.86); filed 10 December 1999

An hourglass-shaped device having two chambers and a neck portion is loaded with uniform objects such as marbles. The device has a cantilevered rod or plate **39** attached to the wall of one of the chambers and makes a reproducible and standard sound whenever one of the objects drops and



strikes the cantilevered rod. The device is to be used as a calibration standard for calibrating sound measuring devices.—IMH

6,580,805

43.58.Ta METHOD AND RECOGNIZING MUSIC AND COMPUTER-READABLE RECORDING MEDIUM HAVING MUSIC-RECOGNIZING PROGRAM RECORDED THEREIN

Seiji Nakano, assignor to Kabushiki Kaisha Kawai Gakki Seisakusho
17 June 2003 (Class 382/100); filed in Japan 26 December 1997

This is a unique optical character recognition patent specifically concerning the recognition of musical scores. The algorithm is threshold based but will only work for very simple scores. Notably missing are references to all of the score reading programs, algorithms and literature of years past.—MK

6,584,300

43.58.Ta OBJECT-ORIENTED SYSTEM FOR SIMULATING SONAR TARGET ACOUSTIC SCATTERING

Gabriel A. Lengua, assignor to The United States of America as represented by the Secretary of the Navy
24 June 2003 (Class 434/6); filed 28 February 2001

A computer system is discussed for simulating sonar targets, such as submarines, which includes a number of target highlight features and which is accurate over a broad frequency band.—WT

6,561,908

43.58.Wc GAMING DEVICE WITH A METRONOME SYSTEM FOR INTERFACING SOUND RECORDINGS

Stephen J. Hoke, assignor to IGT
13 May 2003 (Class 463/35); filed 13 October 2000

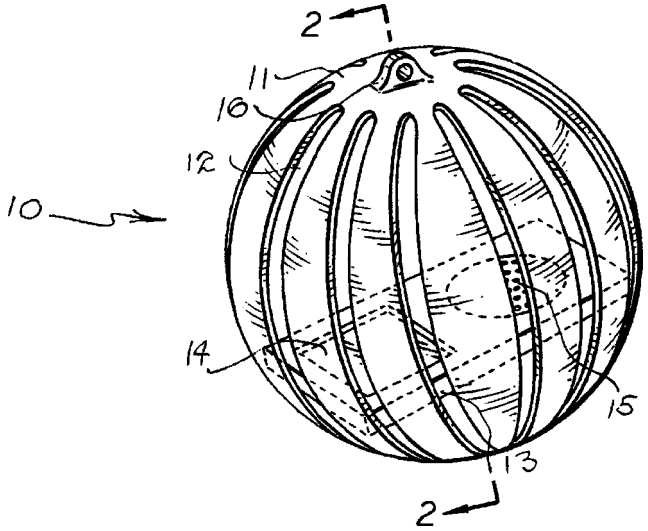
In a computerized slot machine, various musical programs are presented to the player to enhance the ''enjoyment and entertainment.'' The problem is that prerecorded scores should switch at arbitrary time points depending on the player's actions. The inventor describes an internal beat mechanism maintained by the computer. Then the scores can switch and the beat and the gambling can continue, aided and abetted by a new tune.—MK

6,578,527

43.58.Wc SOUND GENERATING PET TOY

Diana Mathers, San Marco, California
17 June 2003 (Class 119/707); filed 13 February 2002

Another entry in the use of the ubiquitous sound chip: the pet toy. Now



authentically squeaky mice can entertain your cat instead of a bell.—MK

6,588,585

43.58.Wc MUSICAL PRESENTATION AND STORAGE BOX

Doris Allen-Carpenter, Durham, North Carolina
8 July 2003 (Class 206/6.1); filed 15 November 2002

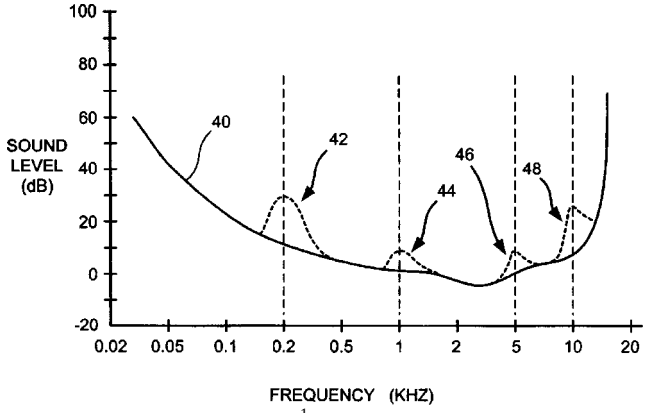
Remember that toy box with the hand that reaches out and grabs the coin and then turns itself off? This ''extends'' this toy to play a tune while hiding the ring or other valuable object.—MK

6,477,496

43.60.Ek SIGNAL SYNTHESIS BY DECODING SUBBAND SCALE FACTORS FROM ONE AUDIO SIGNAL AND SUBBAND SAMPLES FROM DIFFERENT ONE

Eliot M. Case, Denver, Colorado
5 November 2002 (Class 704/272); filed 20 December 1996

This audio signal coder filters the signal into subbands, either by time versus frequency or as frequency, phase, and amplitude, corresponding to



the various MPEG schemes. Each $\frac{1}{3}$-octave subband may be allocated up to eight bits, coded at 0.5 dB per step. These subband signals are perceptually coded, which is to say that any signals below the hearing threshold **40** are dropped. Masking effects, as shown by the dotted lines **42–48**, are also taken into consideration. A presumably novel feature is then introduced in the form of a linear relationship between scale factors of the subbands. This is said to result in easier synthesis and simpler representations for many natural and synthetic sounds.—DLR

6,590,984

43.60.Qv SOUND QUALITY REGULATION APPARATUS

Shigeji Ohama and Hiroyuki Saito, assignors to Rohm Company, Limited
8 July 2003 (Class 381/56); filed in Japan 29 January 2001

This apparatus is designed to control or regulate the quality of sound generated by an audio signal by regulating the amplitudes of individual signals in certain predetermined frequency bands. The sound quality regulation apparatus possesses one attenuation unit and one amplification unit connected in series with the signal line. The output of the attenuation unit is fed to a set of sound regulators, each associated with one of the split frequency bands of the input signal. Each sound regulator includes a switched-capacitor filter and a selector for selectively supplying the output of the sound regulator to either the attenuation unit, the amplification unit, or neither.—DRR

6,353,576

43.60.Sx DETECTOR IN ULTRASONIC HOLOGRAPHY

George F. Garlick and Jerod O. Shelby, assignors to Advanced Diagnostics Systems, Incorporated
5 March 2002 (Class 367/10); filed 8 June 2000

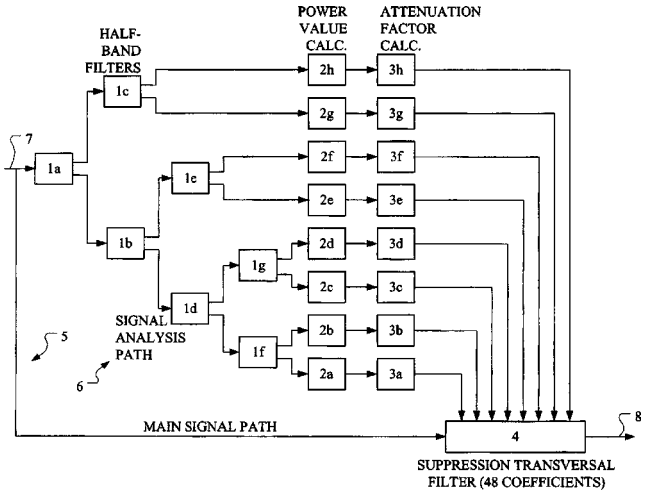
A detector chamber for ultrasonic holograms is described. The chamber has a thick solid bottom (about 5–7.5 mm thick) and rigidly attached walls. The bottom is typically made of plastic. A layer of detection fluid is placed in the chamber to a depth of about 0.2–0.5 mm. The detection fluid preferably has a depth of a $\frac{1}{4}$-wavelength multiple of the ultrasonic wave being used. The detection fluid is preferably a fluorinated organic compound having a kinematic viscosity between 1 and 20 cs and a surface tension between 12 and 19 dynes/cm. An inert gas covers the detection fluid. An optically transparent cover covers the chamber and encloses the fluids.—IMH

6,580,798

43.66.Ts HEARING AID

Arthur Schaub, assignor to Bernafon AG
17 June 2003 (Class 381/312); filed in Switzerland 8 July 1999

Unlike traditional noise suppression hearing aids in which a Fourier analysis is performed in multiple frequency bands, this invention describes suppression of interfering noise in a hearing aid accomplished by a transversal filter placed in series with the main signal path. Attenuation factors for the transversal suppression filter are determined in a signal analysis path



that is parallel to the main signal path. Analysis of the input signal is performed by $\frac{1}{2}$-band transversal filters comprising low-pass filters and high-pass filters in a number of sharply separated frequency bands. Advantages of this approach are said to be low processing delay and moderate computing requirements.—DAP

6,584,207

43.66.Ts MOLDED HEARING AID HOUSING

Robert S. Yoest *et al.*, assignors to Beltone Electronics Corporation
24 June 2003 (Class 381/322); filed 16 July 1999

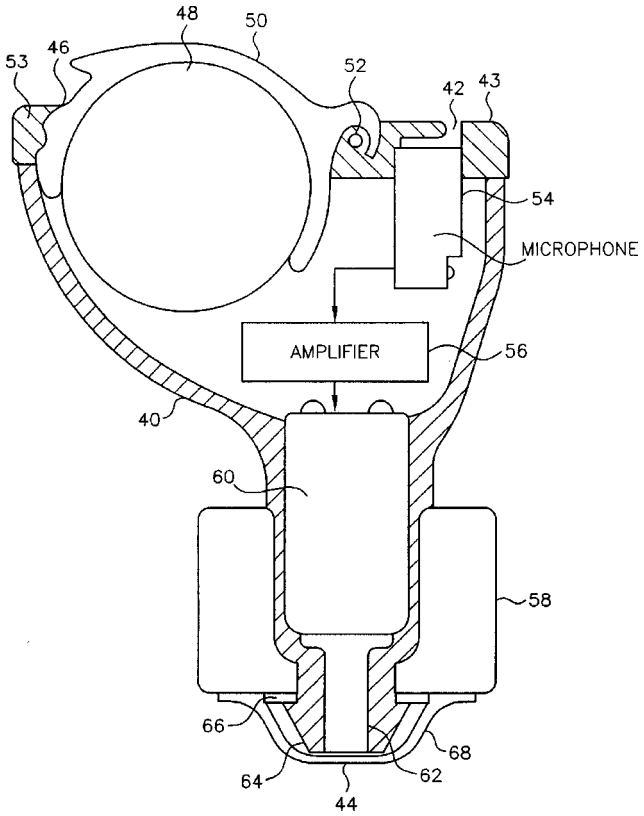
An ear impression of a portion of a hearing aid wearer's ear canal is used to produce a flexible hearing aid housing. A compliant female mold of the shape of at least a portion of the ear canal is created from the ear impression. The same compliant molding material forms a male mold in the cavity remaining in the female mold when the ear impression has been removed. The male mold is cured, removed from the female mold, and covered with a curable plastic such as a UV curable acrylic. The male mold is removed after curing, leaving a rigid plastic shell whose internal volume replicates the shape of the ear impression.—DAP

6,585,075

43.66.Ts HEARING AID HAVING HARD MOUNTED SPEAKER AND ENERGY ABSORBING TIP

Edouard A. Gauthier, Windham, Maine
1 July 2003 (Class 181/135); filed 23 October 2000

A method is described to prevent mechanical and acoustic feedback in hearing aids. An energy absorbing tip surrounds the hearing aid receiver.



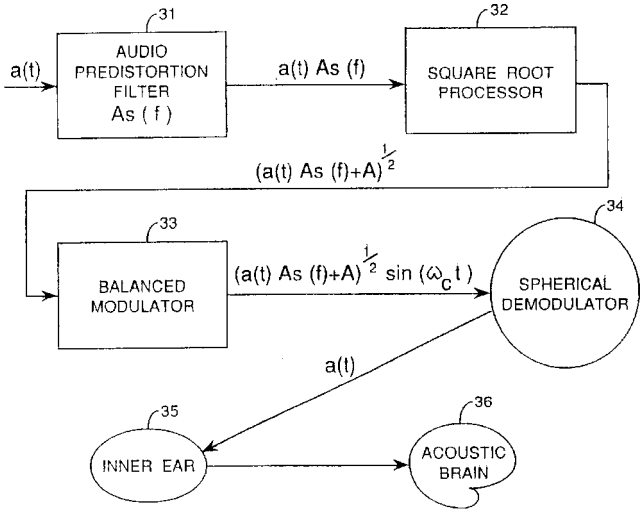
Acoustic and mechanical energy generated by the receiver is channeled to the energy absorbing tip. The tip is held in place with a washer attached to it.—DAP

6,587,729

43.66.Ts APPARATUS FOR AUDIBLY COMMUNICATING SPEECH USING THE RADIO FREQUENCY HEARING EFFECT

James P. O'Loughlin and Diana L. Loree, assignors to The United States of America as represented by the Secretary of the Air Force
1 July 2003 (Class 607/55); filed 24 April 2002

Hearing a speech signal is facilitated via a so-called "rf hearing effect." Preprocessing of the input signal includes low-pass filtering, adding a



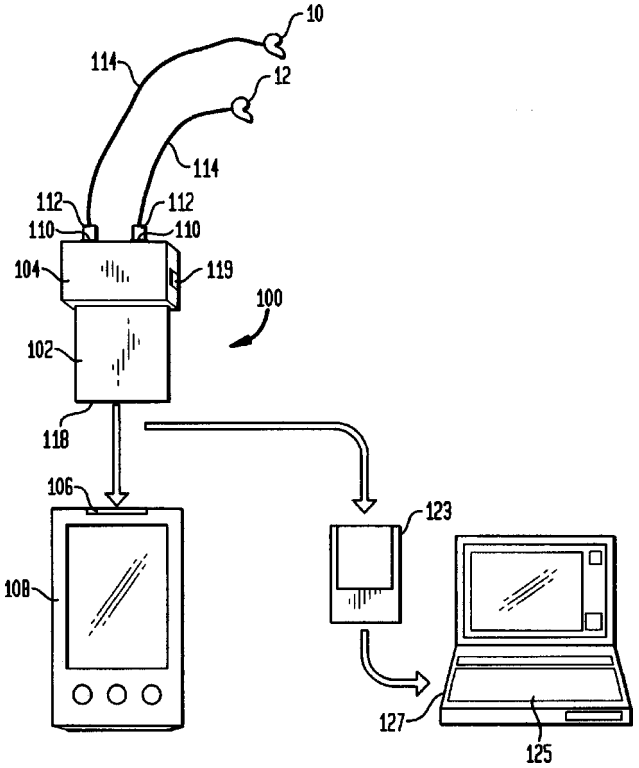
bias level, and taking the root of the predistorted waveform. The resultant signal amplitude modulates a rf carrier with suppressed carrier allowing one or two sidebands. The rf hearing effect is said to result from energy absorption in the head, which causes mechanical expansion and contraction resulting in an acoustical signal.—DAP

6,590,986

43.66.Ts PATIENT-ISOLATING PROGRAMMING INTERFACE FOR PROGRAMMING HEARING AIDS

Joseph D. Fazio, assignor to Siemens Hearing Instruments, Incorporated
8 July 2003 (Class 381/314); filed 12 November 1999

A programming interface is packaged on a PCMCIA card **102** with a compact flash adaptor. The interface electrically isolates the patient from the



computer via opto-isolators. The programming interface is applicable to either hand-held, laptop, or desktop computers. Power is supplied to the programming interface from the host computer.—DAP

6,590,987

43.66.Ts TWO-WIRED HEARING AID SYSTEM UTILIZING TWO-WAY COMMUNICATION FOR PROGRAMMING

David J. Delage, assignor to Etymotic Research, Incorporated
8 July 2003 (Class 381/315); filed 17 January 2001

Hearing aids are programmed with a two-wire protocol using a conventional battery pill, a tri-voltage-level arrangement for normal operation, and two programming modes. Advantages are said to be not having to re-

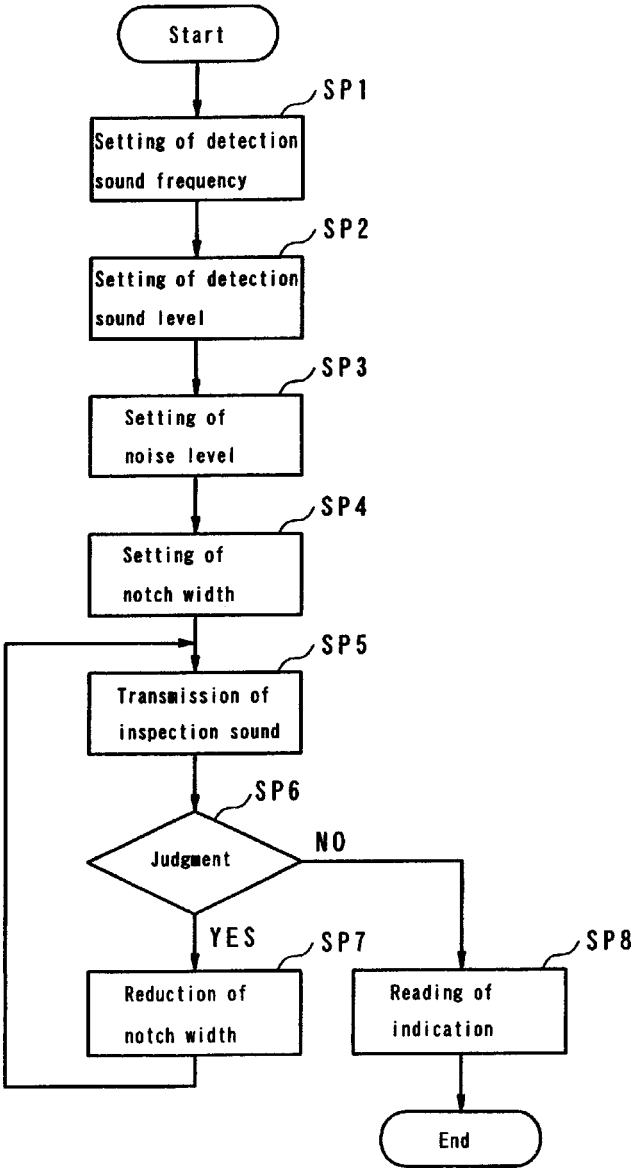
move the battery drawer during programming, faster programming, a simpler and more widely available battery pill, and no additional wiring to the hearing aid circuitry from the faceplate.—DAP

6,582,378

43.66.Yw METHOD OF MEASURING FREQUENCY SELECTIVITY, AND METHOD AND APPARATUS FOR ESTIMATING AUDITORY FILTER SHAPE BY A FREQUENCY SELECTIVITY MEASUREMENT METHOD

Takeshi Nakaichi *et al.*, assignors to Rion Company, Limited
24 June 2003 (Class 600/559); filed in Japan 29 September 1999

The patent refers to a method of measuring frequency selectivity of hearing and estimating the shape of a human auditory filter. The method consists of adding notched noise with a predetermined notch width to a detection sound which sets frequency and sound pressure level at a predetermined value to produce an inspection sound. The inspection sound is



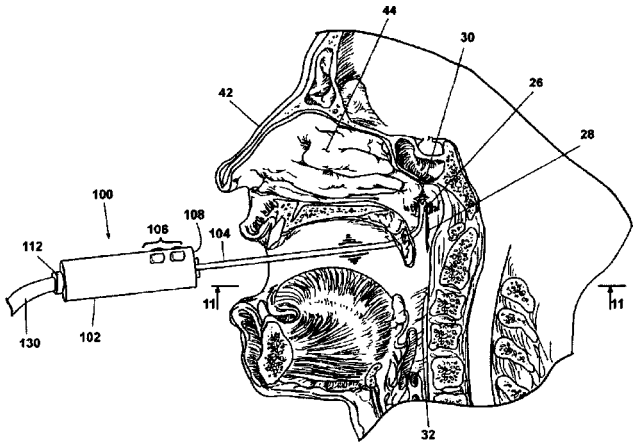
transmitted to a subject. By widening the notch width and obtaining the threshold notch for the subject, a judgment is made as to whether or not the subject can perceive the detection sound separately from the inspection sound.—DRR

6,578,581

43.66.Yw METHOD AND APPARATUS FOR RELIEVING FLUID BUILD-UP IN THE MIDDLE EAR

Siri Nam Khalsa, Amado, Arizona
17 June 2003 (Class 128/898); filed 7 September 2001

The aim of this device is to drain the Eustachian tube and middle ear of fluid when there occurs a blockage of the opening of the Eustachian tube into the nasopharynx region of the throat. The device consists of an attachment to a medical suction unit that consists of a handle with controls and a rigid tube that curves to enable reaching the opening of the Eustachian tube



from the mouth. The semi-rigid tube may either be disposable or nondisposable and can be fabricated in different sizes. The handle incorporates controls for the suction, lighting, a spray device for medications, visualization means, and vibration. One embodiment of the device consists of only a handle with vibration and other features but no suction.—DRR

6,567,775

43.72.Fx FUSION OF AUDIO AND VIDEO BASED SPEAKER IDENTIFICATION FOR MULTIMEDIA INFORMATION ACCESS

Fereydoun Maali and Mahesh Viswanathan, assignors to International Business Machines Corporation
20 May 2003 (Class 704/231); filed 26 April 2000

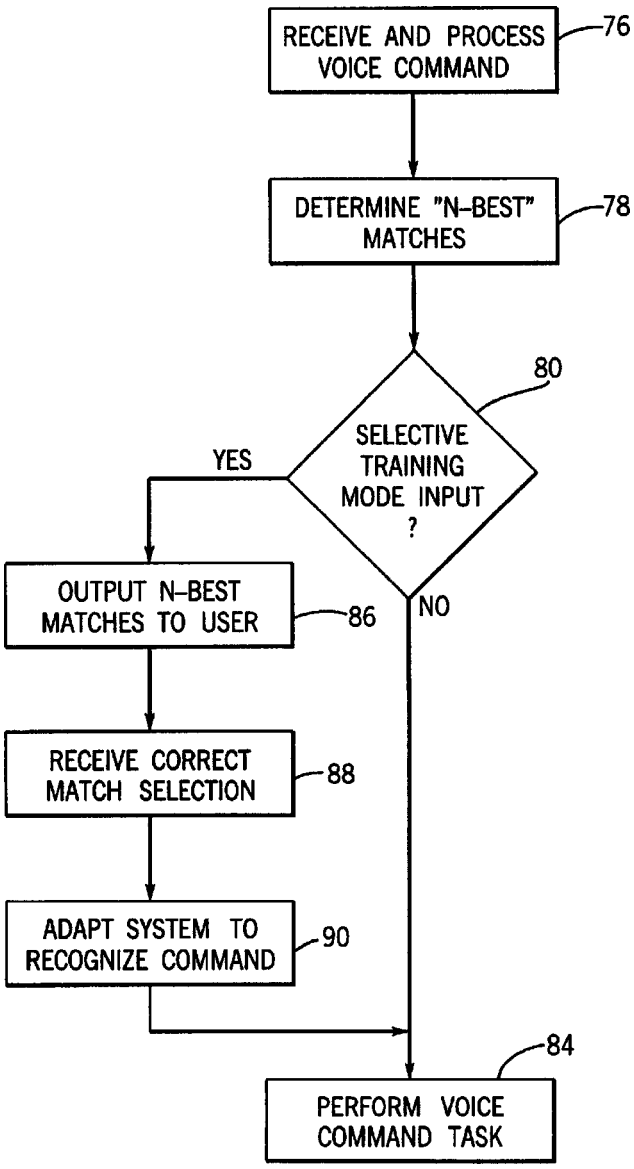
A method is proposed for speaker indentification from an enrolled database using scores derived from both speech and video identifications. The ranked scores from each modality are fitted with lines. The normalized slopes of the two lines are used to provide the weights in combining the scores.—RSM

6,587,824

43.72.Fx SELECTIVE SPEAKER ADAPTATION FOR AN IN-VEHICLE SPEECH RECOGNITION SYSTEM

Charles Allen Everhart and Scott Alan Thompson, assignors to Visteon Global Technologies, Incorporated
1 July 2003 (Class 704/275); filed 4 May 2000

An in-vehicle speech recognition system adapts the system parameters during a training phase to the voice characteristics of a particular speaker



using *N*-best voice commands. The user selects the correct command from the *N*-best list and the system then adapts to the speech characteristics of the selected car command.—HHN

6,477,495

43.72.Ja SPEECH SYNTHESIS SYSTEM AND PROSODIC CONTROL METHOD IN THE SPEECH SYNTHESIS SYSTEM

Nobuo Nukaga *et al.*, assignors to Hitachi, Limited
5 November 2002 (Class 704/268); filed in Japan 2 March 1998

This speech synthesis system includes what might be called a lookup table prosodic parameter generator. A number of representative sentences are stored along with their analyzed prosodic information. When a text item is to be synthesized, the system searches the stored material for similar text patterns. If found, the prosodic structure from the stored material is used as a basis for generating a new prosodic structure for the synthesized output. Noting the recent increases in memory storage capacity, the patent assumes that it will be possible to store an adequate amount of speech material for acceptable quality synthesis. If the subject domain is sufficiently narrow, it might even work.—DLR

6,477,491

43.72.Ne SYSTEM AND METHOD FOR PROVIDING SPEAKER-SPECIFIC RECORDS OF STATEMENTS OF SPEAKERS

Mark Chandler, Show Low, and Daniel Altman, Tucson, both of Arizona
5 November 2002 (Class 704/235); filed 27 May 1999

This speech recognition application automatically generates a transcript of a multispeaker session, such as a business meeting, trial, or hearing, etc. Each speaker uses a specific microphone and each mic feeds a separate channel, which may be a separate computer, with a recognition system tuned to that speaker. Although the authors clearly understand that each person speaking will need to provide training data for the corresponding recognition system, the patent does not address that issue.—DLR

6,477,494

43.72.Ne UNIFIED MESSAGING SYSTEM WITH VOICE MESSAGING AND TEXT MESSAGING USING TEXT-TO-SPEECH CONVERSION

Henry C. A. Hyde-Thomson and Roger Liron, assignors to Avaya Technology Corporation
5 November 2002 (Class 704/260); filed 7 January 2000

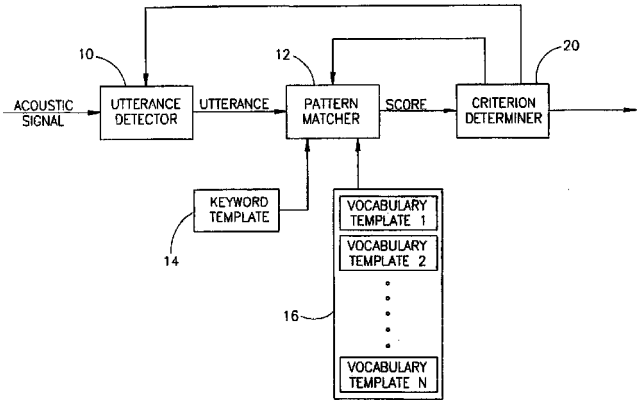
This telephone message system provides for the storage and recovery of both text and voice messages. A text-to-speech facility includes a procedure for testing a stream of text and, on the basis of trigram, or triple letter, frequencies, the text may be synthesized into any of several languages. Conversions in the opposite direction are not contemplated.—DLR

6,591,237

43.72.Ne KEYWORD RECOGNITION SYSTEM AND METHOD

Adoram Erell, assignor to Intel Corporation
8 July 2003 (Class 704/251); filed 13 December 1999

A speaker-dependent keyword recognition system works based on searching all of the training samples, including keywords and a closed vocabulary. Only if the match to the keyword is closer than the matches to all of the vocabulary templates will the utterance then be accepted as a keyword.—HHN

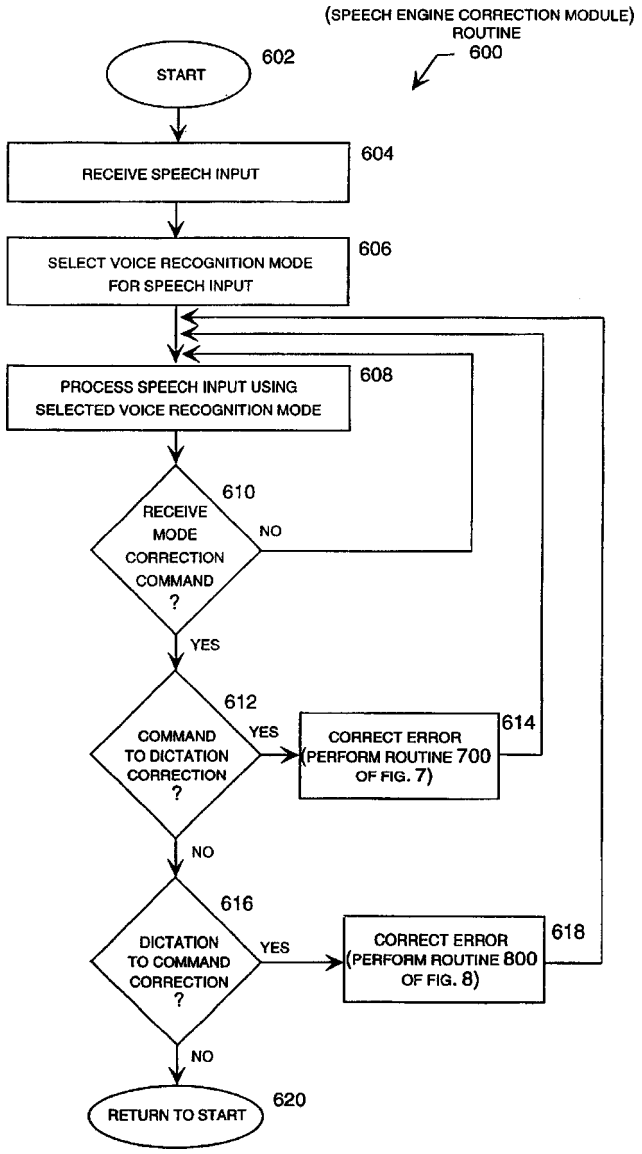


6,581,033

43.72.Ne SYSTEM AND METHOD FOR CORRECTION OF SPEECH RECOGNITION MODE ERRORS

Jeffrey C. Reynar *et al.*, assignors to Microsoft Corporation
17 June 2003 (Class 704/231); filed 19 October 1999

The speech engine correction module is activated by the user when an



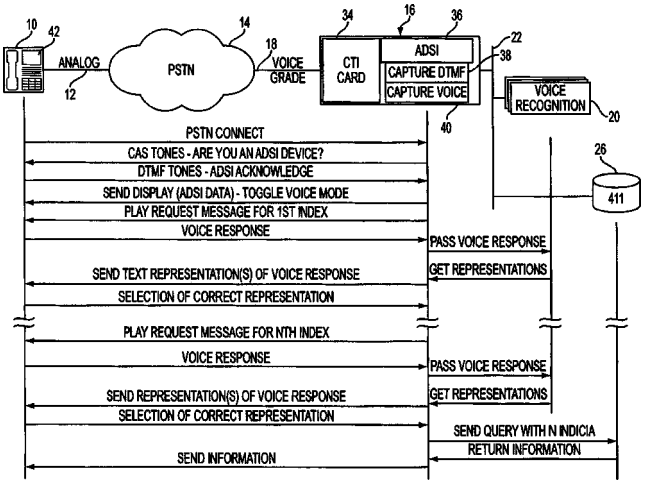
error in recognition is encountered. The engine then provides a list of alternatives for the user, thus it can reduce user time in correcting errors.—HHN

6,584,179

43.72.Ne METHOD AND APPARATUS FOR IMPROVING THE UTILITY OF SPEECH RECOGNITION

Stéphane Fortier *et al.*, assignors to Bell Canada
24 June 2003 (Class 379/88.01); filed 24 October 1997

The purpose of this system is to automate directory services for own-



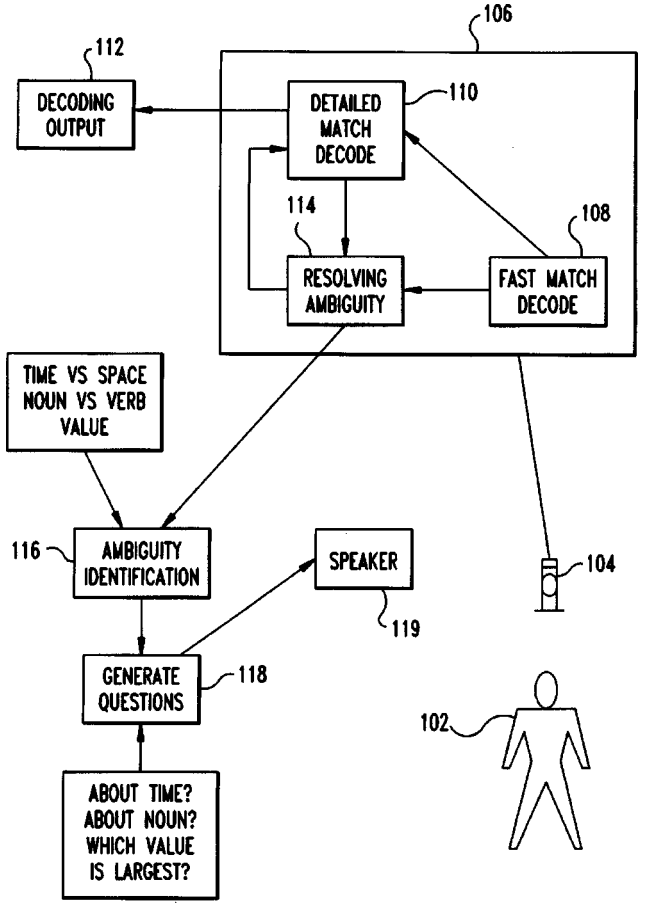
ers of display telephones using a speech recognition engine and an access server which communicates with a display telephone.—HHN

6,587,818

43.72.Ne SYSTEM AND METHOD FOR RESOLVING DECODING AMBIGUITY VIA DIALOG

Dimitri Kanevsky *et al.*, assignors to International Business Machines Corporation
1 July 2003 (Class 704/251); filed 28 October 1999

A dialog based decoding ambiguity system interrupts recognition, intermediate to the decoding process when an ambiguity is encountered, and asks the user to select appropriate response classes. Therefore, recognition



performance and accuracy are improved by reducing the number of final decoding alternatives.—HHN

6,477,500

43.72.Pf TEXT INDEPENDENT SPEAKER RECOGNITION WITH SIMULTANEOUS SPEECH RECOGNITION FOR TRANSPARENT COMMAND AMBIGUITY RESOLUTION AND CONTINUOUS ACCESS CONTROL

Stephane Herman Maes, assignor to International Business Machines Corporation
5 November 2002 (Class 704/275); filed 12 April 2000

This speech and speaker recognition system generates feature vectors from frames of the input speech. The feature vectors are continuously compared to codebook entries from a vector coding scheme, providing continuous evaluation of the speaker identity, as well as the phonetic content, throughout the utterance. This allows the speaker's identity to be confirmed more accurately than is possible with a typical system and can also be done independently of the text content.—DLR

6,570,078

43.75.Bc TACTILE, VISUAL, AND ARRAY CONTROLLERS FOR REAL-TIME CONTROL OF MUSIC SIGNAL PROCESSING, MIXING, VIDEO, AND LIGHTING

Lester Frank Ludwig, Belmont, California
27 May 2003 (Class 84/600); filed 19 March 2001

The inventor seeks to unify eastern and western instruments and traditions. So, in his words: ''What is needed is some reach into the souls (rather than make samples) of deep non-Western and Western instruments, a recasting of the now institutionalized signal processing chains, adaptations of new classes of applicable physical phenomenon, extensions as to the types and forms of meaningful human control, and, in the context of performance, a deeper integration of visual and audio environments.'' And so, the patent is a mishmash of applications of Western miking and amplification to acoustic instruments of other cultures. The list is long and varied. In addition, these many inputs can be used to control video and lighting. Too bad the prior art list was entirely supplied by the patent office; now that would have been interesting reading.—MK

6,563,033

43.75.Gh STRINGED MUSICAL INSTRUMENT WITH APPARATUS ENHANCING LOW FREQUENCY SOUNDS

Louis B. Porzilli, Sparta, New Jersey
13 May 2003 (Class 84/294); filed 31 December 1997

The inventor wants to compensate for the low output of thin acoustic guitars. His proposal: add a passive radiator in the rear plate. Optionally, waveguides may be added internally as well. But if sound radiates to the rear, isn't it being absorbed by the player's clothing?—MK

6,586,667

43.75.Wx MUSICAL SOUND GENERATOR

Toru Morita, assignor to Sony Computer Entertainment, Incorporated
1 July 2003 (Class 84/659); filed in Japan 3 March 2000

Considering it was filed in 2000, this patent is remarkably vacuous. The system described converts MIDI-like events to sound, i.e., a synthesizer. Like the patent claims, the prior art cited is also severely lacking.—MK

6,589,116

43.75.Wx GAME WITH SOUNDS AND SYSTEMS FOR ITS IMPLEMENTATION

Sergei Valerievich Grigoriev *et al.*, assignors to Bluevista Invest and Finance Limited
8 July 2003 (Class 463/35); filed 8 June 2001

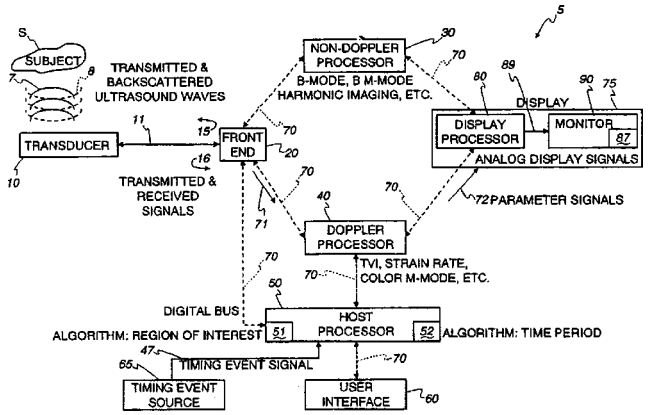
Imagine a computerized musical blocks game in which the blocks can be arranged either by moving them on the screen or by arranging them on a board. The blocks contain musical fragments and can be arranged in any order to create a reference. And the computer can offer hints, too.—MK

6,579,240

43.80.Qf ULTRASOUND DISPLAY OF SELECTED MOVEMENT PARAMETER VALUES

Steinar Bjaerum *et al.*, assignors to GE Medical Systems Global Technology Company, LLC
17 June 2003 (Class 600/447); filed 10 June 2002

This echocardiographic device provides an ultrasound system for generating an image representing moving cardiac structure. Peak values of mean parameter signals are displayed in the form of colors that indicate movement of the structure. A Doppler processor generates a set of parametric signals representing movement within the cardiac structure. A host processor embodies a tracking function and a peak-detection function to generate a set of tracked movement parametric profiles and a set of peak values



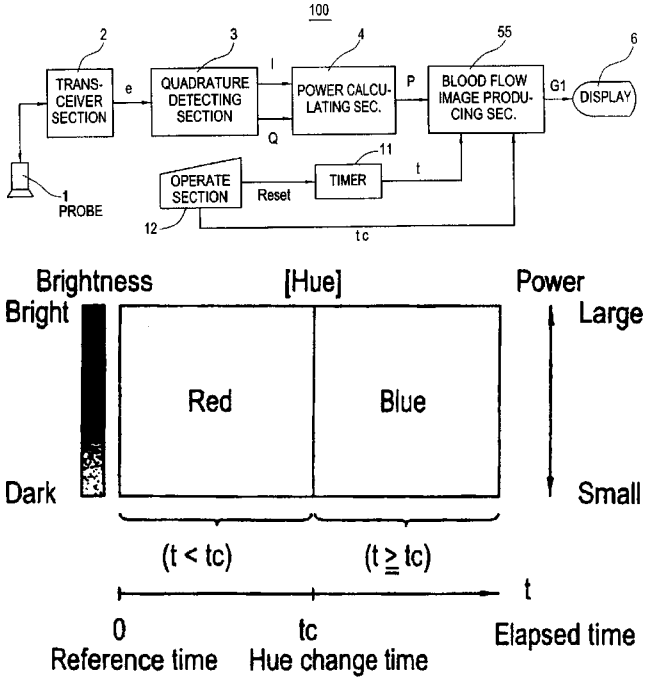
of the movement parameter over a time interval corresponding to anatomical locations within the region of interest. This set of peak values is then mapped into a set of colors which are displayed on a color monitor.—DRR

6,582,370

43.80.Qf BLOOD FLOW IMAGING METHOD, BLOOD FLOW IMAGING APPARATUS AND ULTRASONIC DIAGNOSTIC APPARATUS

Takao Jibiki, assignor to GE Medical Systems Global Technology Company, LLC
24 June 2003 (Class 600/455); filed in Japan 1 February 2001

This is a method for providing images of multiple blood flow streams by which different contrast agent arrival times or blood concentration properties can be discerned. The ultrasonic probe **1** has a transceiver section **2**, a quadrature detection section **3**, and a calculating section **4**. The power **P** of blood flow is calculated and the power peak is compared with a predecessor



peak value $P0$. The larger of the two is output as a peak value P'. A timer **11** measures elapsed time **t** from a reference value. Section **55** produces a power peak hold blood flow image **G1** in which the hue is changed at specified elapsed times and the brightness varies according to the peak power P'. Blood flow in different arteries or veins can be displayed as different colors.—DRR

6,585,647

43.80.Qf METHOD AND MEANS FOR SYNTHETIC STRUCTURAL IMAGING AND VOLUME ESTIMATION OF BIOLOGICAL TISSUE ORGANS

Alan A. Winder, Westport, Connecticut
1 July 2003 (Class 600/437); filed 19 January 2001

This is a system for discerning and imaging body organs. The processor-equipped system generates low-frequency ultrasound signals that are applied to tissue so as to generate a weighted sum of tissue ramp, step, and impulse signatures. The system analyzes these tissue signatures to establish the low-frequency target profile that is used to generate a graphic presentation of the tissue as well as to estimate the volume of tissue and to classify the tissue according to its type and condition using a set of stored tissue data. The classifier may include a neural network and/or nearest neighbor rule processor. The system functions as a noninvasive acoustic measurement and imaging system and it applies synthetic structural imaging techniques to provide information concerning the size and shape of biological structures for classification and visualization of normal and abnormal biological structures.—DRR

6,585,648

43.80.Qf SYSTEM, METHOD AND MACHINE READABLE PROGRAM FOR PERFORMING ULTRASONIC FAT BEAM TRANSMISSION AND MULTILINE RECEIVE IMAGING

Brent Stephen Robinson, assignor to Koninklijke Philips Electronics N.V.
1 July 2003 (Class 600/437); filed 15 November 2002

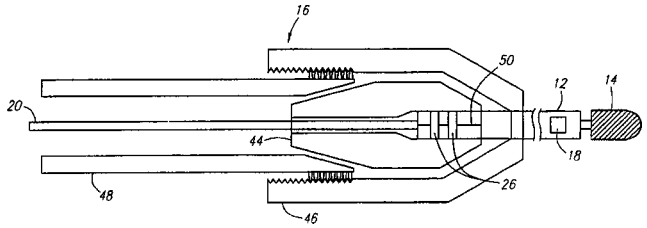
The patent concerns a system for performing ultrasonic imaging by launching a broad beam (Fat TX) having a preselected spatial energy profile at a target and processing energy reflected from the target to determine target characteristics. The launching transmitter includes multiple transducer elements, each capable of producing a waveform of the desired shape. A processor controls the shape of the waveform of each transducer element. A receiver processes the reflected energy to establish the image data.—DRR

6,585,654

43.80.Qf ULTRASOUND IMAGING GUIDEWIRE WITH STATIC CENTRAL CORE AND TIP

David A. White and W. Martin Belef, assignors to Scimed Life Systems, Incorporated
1 July 2003 (Class 600/463); filed 27 April 2001

An ultrasound imaging guidewire that is inserted into a patient's body features a static central core and an imaging guidewire body that function together as an acoustical scanning device. The device can be rotated to



produce 360° acoustical images of a site of interest in the patient's body. The imaging guidewire also includes a connector that allows the guidewire body to be disengaged from the static central core tip so that the body can be axially translated to obtain multi-position imaging.—DRR

6,587,564

43.80.Qf RESONANT CHAMBER SOUND PICK-UP

Ronald Y. Cusson, Newman, California
1 July 2003 (Class 381/67); filed 25 May 1999

An argument is made for a noise-reducing resonant chamber sound pickup for an electronic stethoscope. The resonant chamber is said to acoustically amplify the selected frequencies of heart sounds, in the 50–150-Hz range, and bruits of breath or blood, in the range of 150–500 Hz, while rejecting ambient sounds. A skin contact has elastic foam support, whose rigidity can be varied by changing the applied pressure in order to select the preamplifier band of sounds. The other side of the resonant chamber contains a sound transducer largely encased in heavy malleable metal that absorbs sound not coming from the chamber. A damping vent hole shapes the band of amplified sounds.—DRR

6,589,054

43.80.Qf INSPECTION OF TEETH USING STRESS WAVE TIME NON-DESTRUCTIVE METHODS

Daniel A. Tingley and Kenneth Johnson, both of Corvallis, Oregon
8 July 2003 (Class 433/215); filed 18 July 2001

This is a noninvasive ultrasonic diagnostic device using stress waves for detecting lesions on tooth surfaces or under dental restorations such as gold crowns. According to the method, a tooth under examination is subjected to an acoustic stress wave emanating from a suitable transducer on one side of the structure. The stress wave propagates through the tooth and is received on the other side by a receiving transducer. From the analysis of the transmission time and/or the resulting waveform, diagnostic information is provided regarding the presence of dental disease that may be present on the tooth surface or under dental restorations, such as fillings or metal crowns.—DRR

6,589,182

43.80.Qf MEDICAL DIAGNOSTIC ULTRASOUND CATHETER WITH FIRST AND SECOND TIP PORTIONS

Rickard C. Loftman *et al.*, assignors to Acuson Corporation
8 July 2003 (Class 600/466); filed 12 February 2001

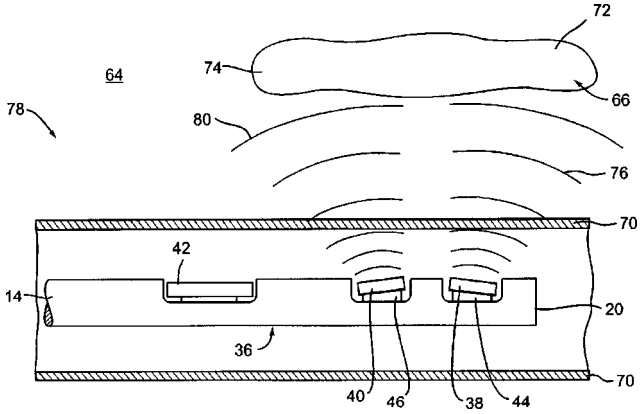
The preferred embodiments of this apparatus provide a medical diagnostic ultrasound catheter equipped with a tip portion carrying an ultrasound transducer. The tip portion is partly in the acoustic path of the ultrasound transducer and partly outside of that path. These two portions carry different acoustic and mechanical properties. The first part can provide desired acoustic properties to the catheter tip, while the mechanical properties of the second part can give the catheter tip sufficient durability to withstand normal use without bowing. This can be especially important for small-diameter catheters with reduced stiffness of the catheter tip.—DRR

6,585,656

43.80.Sh CATHETER WITH MULTIPLE TRANSDUCERS

Donald Masters, assignor to Scimed Life Systems, Incorporated
1 July 2003 (Class 600/466); filed 5 February 2002

This diagnostic/therapeutic catheter provides multiple ultrasound transducers in the distal end of a drive shaft. In one embodiment, the center frequencies of the second and third transducers may be substantially equal to the fundamental and second harmonic, respectively, of the ultrasound waves generated by the first transducer. In another embodiment, the second transducer transmits waves with a center frequency substantially equal to the



center of the first transducer. In a third embodiment, the second transducer has a center frequency that differs from the center frequency of the first transducer and the center frequency of the third transducer may be substantially equal to the difference between the fundamental frequencies of the first and second transducers.—DRR

6,579,244

43.80.Sh INTRAOSTEAL ULTRASOUND DURING SURGICAL IMPLANTATION

Mark R. Goodwin, assignor to Cutting Edge Surgical, Incorporated
17 June 2003 (Class 600/561); filed 24 October 2002

As defined in this patent, intraosteal ultrasound is the use of acoustical energy, viz. ultrasound, to facilitate ''real time'' manipulation and navigation of a device for intraosseous placement of synthetic or biologic implants. Typical applications include placement of implants, such as bone screws, through vertebral pedicles during spinal fusion surgery. The devices include means for creating a lumen or channel into the bone at the desired site in combination with a probe for providing real-time feedback of differences in density of tissue, i.e., differences in acoustical impedance between cancellous and cortical bone. The devices will also usually include means for monitoring the feedback in the format of a screen display showing an image to the surgeon as he creates the channel and/or an audible signal indicating the presence of different signals. The system may also be used for diagnostic purposes.—DRR

6,585,750

43.80.Sh METHOD OF USING ULTRASONIC VIBRATION TO SECURE BODY TISSUE

Peter M. Bonutti *et al.*, assignors to Bonutti 2003 Trust-A
1 July 2003 (Class 606/232); filed 15 February 2002

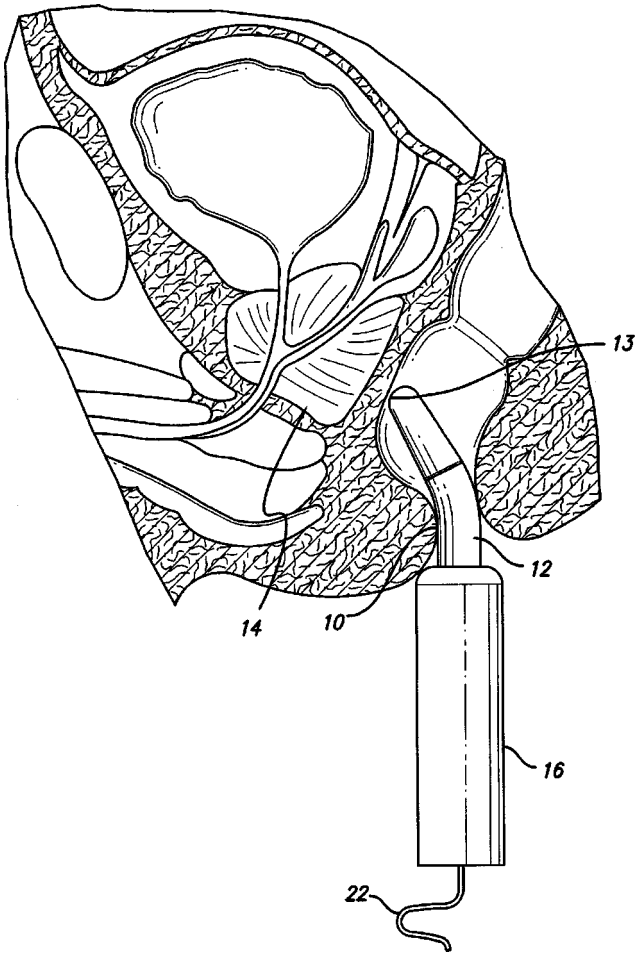
Here a suture and a suture retainer are positioned relative to the body tissue. Ultrasonic vibratory energy is applied to heat the suture retainer and to bond portions of the suture retainer to each other and/or to the suture itself. Portions of the body tissue may be pressed together and held in place by mutual interaction between the suture and suture retainer. The suture retainer may include one or more passages through which the suture extends.—DRR

6,589,173

43.80.Sh ULTRASOUND SYSTEM FOR DISEASE DETECTION AND PATIENT TREATMENT

Samir Mitragotri, assignor to The Regents of the University of California
8 July 2003 (Class 600/437); filed 16 July 2001

A device is described here for performing a diagnostic test and/or treating solid cancer tumor, e.g., prostate cancer. In the diagnostic mode, ultrasound is used to extract a cancer tumor marker from interstitial fluid into a coupling medium. In the therapeutic mode, ultrasound is used to deliver one or more drugs from the coupling medium directly to the site of the tumor. The device consists of an ultrasound transducer and a connecting reservoir. The reservoir is in the form of a cartridge that is contoured for placement against the tissue in the region of the disease to be diagnosed or treated and it contains a coupling medium to transmit ultrasound to the



tissue, thereby producing cavitation for permeabilizing the tissue. The coupling medium may also contain a cavitation enhancer.—DRR

6,589,174

43.80.Sh TECHNIQUE AND APPARATUS FOR ULTRASOUND THERAPY

Rajiv Chopra and Michael J. Bronskill, assignors to Sunnybrook & Women's College Health Sciences Centre
8 July 2003 (Class 600/439); filed 20 October 2000

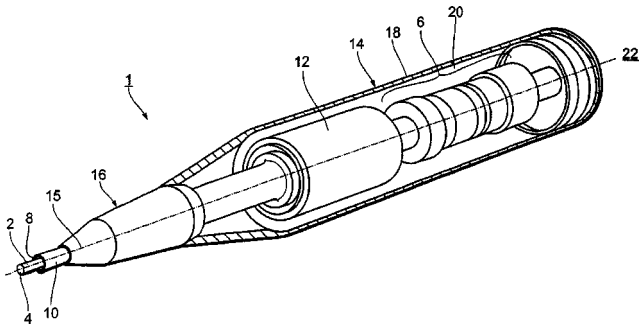
This apparatus provides an ultrasound heating applicator for thermal tissue therapy. The applicator is said to be compatible with imaging (such as MRI) and is also compatible with image-guided interstitial therapy, preferably of benign and malignant tissues. The interstitial ultrasound applicator consists of a preferably planar transducer with multiple acoustic matching layers enabling operation at a range of frequencies for optimal ''control'' of the depth of thermal coagulation. In one embodiment, the applicator has the capability for varying the frequency of each individual element, thereby enabling the tissue temperature to be adjusted both radially and along the length of the application or catheter.—DRR

6,589,191

43.80.Sh MANUALLY ACTUABLE ULTRASONIC DISINTEGRATOR FOR BREAKING UP OR REMOVING HUMAN OR ANIMAL TISSUE

Kai Desinger, Berlin, Germany
8 July 2003 (Class 601/2); filed in Germany 3 May 2000

This device is a manually operated ultrasonic disintegrator for the removal of human or animal tissue. It includes a sonotrode for transmission of sound waves to tissue at its distal end, an ultrasonic transducer that can be



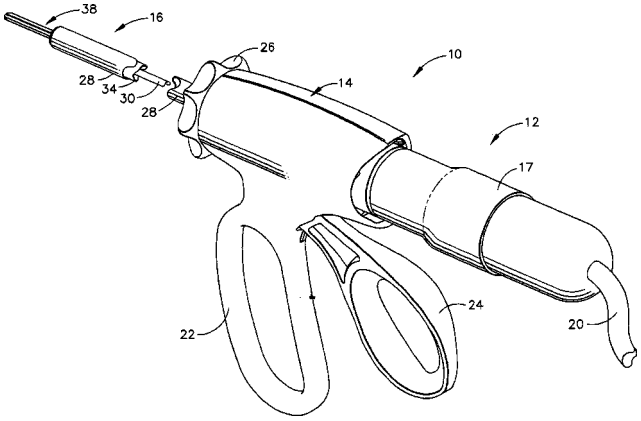
coupled to the sonotrode, and a passage that extends lengthwise along the sonotrode for flushing and/or sucking away disintegrated or ablated tissue. In order to optimize the application of ultrasound into the tissue to be treated, the distal end of the sonotrode is in the form of a full surface outside which the flushing or suction removal passage extends.—DRR

6,589,200

43.80.Sh ARTICULATING ULTRASONIC SURGICAL SHEARS

Richard F. Schwemberger and John C. Wright, assignors to Ethicon Endo-Surgery, Incorporated
8 July 2003 (Class 604/22); filed 16 December 1999

This surgical instrument consists of an ultrasonic transducer, an ultrasonically activated end effector, and a substantially solid wave guide that connects the ultrasonic transducer to the end effector. The wave guide forms a transmission section extending from the transducer to a fixed node and an articulation section extending from the fixed node to a pivoting node. The



end effector features a bifurcated portion. A handle is adapted to contain the transducer. An outer sheath extends from the handle to the end effector and surrounds the wave guide. An actuation trigger is rotatably positioned on the handle and an actuation arm extends from the distal end of the actuation trigger to the pivoting nose. The movement of the trigger performs both articulation and actuation of the end effector.—DRR

6,589,253

43.80.Sh ULTRASONIC ANGIOPLASTY TRANSMISSION WIRE

Wayne E. Cornish and Douglas H. Gesswein, assignors to Advanced Cardiovascular Systems, Incorporated
8 July 2003 (Class 606/128); filed 30 December 1999

The aim of this wire is an improved ultrasonic angioplasty transmission member that allows alteration of the properties of sonic transmission, flexibility, and tensile strength by providing the member with selected layering of different ultrasound transmitting materials and thicknesses of materials.—DRR

6,579,238

43.80.Vj MEDICAL ULTRASONIC IMAGING SYSTEM WITH ADAPTIVE MULTI-DIMENSIONAL BACK-END MAPPING

Constantine Simopoulos *et al.*, assignors to Acuson Corporation
17 June 2003 (Class 600/443); filed 23 February 2001

This system uses spatial variance to identify image regions that correspond to soft tissue and uses a noise frame acquired with no transmitted signal to determine the system mean noise level. From the noise level and tissue regions, back-end maps including gain and dynamic range are locally and adaptably set.—RCW

6,579,239

43.80.Vj SYSTEM AND METHOD FOR AUTOMATIC ADJUSTMENT OF BRIGHTNESS AND CONTRAST IN IMAGES

Gopal B. Avinash *et al.*, assignors to GE Medical Systems Global Technology Company, LLC
17 June 2003 (Class 600/443); filed 5 April 2002

The contrast and brightness of an image are adjusted using an algorithm based on a histogram of pixel intensity. The algorithm is designed to enhance contrast and brightness without affecting the near-field region of an ultrasonic image. The filtering operations may be performed by a dedicated processor or by a fast general-purpose computer.—RCW

6,582,368

43.80.Vj MEDICAL INSTRUMENT SHEATH COMPRISING A FLEXIBLE ULTRASOUND TRANSDUCER

Charles R. Holdaway *et al.*, assignors to Paul F. Zupkas; C. Lowell Parsons
24 June 2003 (Class 600/443); filed 17 April 2001

This patent refers to a sheath for use with a medical transducer. The sheath is constructed from a nonconductive material and configured to be installed over a medical device. The sheath presses a piezoelectric copolymer transducer snugly against the medical device when the ultrasound transducer is installed over the medical device. The transducer converts ultrasound signals into corresponding electric signals. A ground conductor is coupled to one surface of the piezoelectric polymer transducer. A signal conductor is coupled to another surface of the transducer and conducts the corresponding electric signals to a proximal end of the medical device.—DRR

6,582,372

43.80.Vj ULTRASOUND SYSTEM FOR THE PRODUCTION OF 3-D IMAGES

Mckee D. Poland, assignor to Koninklijke Philips Electronics N.V.
24 June 2003 (Class 600/463); filed 22 June 2001

This system uses a two-dimensional array in conjunction with little or no specialized three-dimensional software or hardware to produce images that have depth cues. Multiple planes of data each containing a number of echo lines and points are obtained from the probe by a control unit. The control unit combines points from matched lines across the planes to display an image with the appearance of depth.—RCW

6,582,381

43.80.Vj MECHANICAL POSITIONER FOR MRI GUIDED ULTRASOUND THERAPY SYSTEM

Oded Yehezkeli *et al.*, assignors to TxSonics Limited
24 June 2003 (Class 601/2); filed 31 July 2000

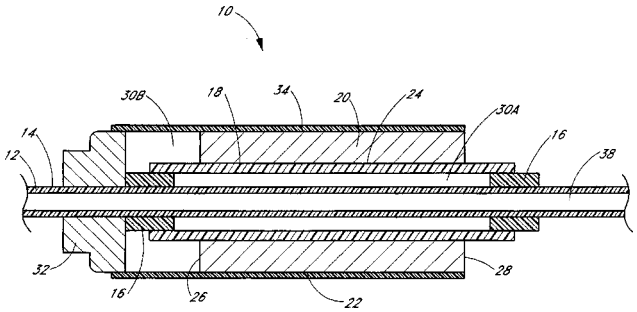
A positioning system is provided here for magnetic resonance imaging (MRI) guidance of a therapy device. The system utilizes a mechanical assembly to locate and aim an ultrasonic transducer. The positioner provides for roll and pitch control as well as control in the lateral and longitudinal directions. Piezoelectric vibrational motors can effectively operate within the field of the MRI system without interfering with its operation. Two sets of position encoders and fine relative encoders are both coupled to the positioning motors to allow for precise control over the position and orientation of the ultrasonic transducer.—DRR

6,582,392

43.80.Vj ULTRASOUND ASSEMBLY FOR USE WITH A CATHETER

Frederick J. Bennett *et al.*, assignors to Ekos Corporation
24 June 2003 (Class 604/22); filed 16 August 1999

This assembly consists of a catheter body having a chamber containing a low acoustic impedance medium. The catheter body includes an elongated body with an external surface and an ultrasound transducer having an exter-



nal side between the two ends of the body. The transducer is positioned over the external surface of the elongated body in such a manner so that one end is adjacent to the chamber.—DRR

6,587,709

43.80.Vj METHOD OF AND IMAGING ULTRASOUND SYSTEM FOR DETERMINING THE POSITION OF A CATHETER

Torsten Solf and Kai Eck, assignors to Koninklijke Philips Electronics N.V.
1 July 2003 (Class 600/424); filed in Germany 28 March 2001

This system produces a three-dimensional ultrasonic image of a region in which a catheter is placed for therapeutic or diagnostic intervention. The catheter includes three ultrasound receivers that are capable of detecting the arrival of pulses transmitted from the imaging system probe. The distance between the imaging probe and the receiver is calculated from the transit time of the pulses to localize the position of the receivers. Localization enables the selection of a plane that contains all three receivers of the catheter. This allows the tip of the catheter to be tracked automatically and displayed on a monitor without manual displacement of the ultrasound transducer.—RCW

6,589,175

43.80.Vj REAL-TIME ARBITRARY MMODE FOR ULTRASONIC IMAGING SYSTEM

David M. Prater *et al.*, assignors to Koninklijke Philips Electronics N.V.
8 July 2003 (Class 600/443); filed 30 April 2001

An arbitrary user-defined curve within an ultrasonic image is obtained from a scan converter and the curve is displayed as a function of time in an m-mode format.—RCW

6,589,176

43.80.Vj ULTRASONIC IMAGE STABILIZATION SYSTEM AND METHOD

James Jago and Lars Jonas Olsson, assignors to Koninklijke Philips Electronics N.V.
8 July 2003 (Class 600/443); filed 5 December 2001

Motion, either of an imaging probe or in the region being imaged, is analyzed and a decision is made to inhibit or allow image stabilization. Images are anatomically aligned on the basis of the motion sensing. The stabilization can be activated either manually or automatically and adaptively.—RCW

6,589,177

43.80.Vj METHOD AND APPARATUS FOR OBTAINING B-FLOW AND B-MODE DATA FROM MULTILINE BEAMS IN AN ULTRASOUND IMAGING SYSTEM

Paul R. Detmer *et al.*, assignors to Koninklijke Philips Electronics N.V.
8 July 2003 (Class 600/443); filed 15 November 2002

Multiple beams are transmitted. Each beam is separated from an adjacent beam by a predetermined distance. Echoes corresponding to the beams are received. The echoes are spaced from the corresponding beam by a predetermined distance that is less than the distance of the beams. The transmitted beams are arranged so that received echoes corresponding to adjacent ultrasound beams substantially overlap. Received echoes are processed to produce *b*-mode data. Additionally, overlapping pairs of received echoes are processed to obtain *b*-flow data. Doppler flow, power, or motion data may also be obtained from the received echoes.—RCW

6,589,179

43.80.Vj THREE-DIMENSIONAL DOPPLER ULTRASONIC IMAGING SYSTEM AND METHOD

Aline Criton and Helen Routh, assignors to Koninklijke Philips Electronics N.V.
8 July 2003 (Class 600/454); filed 27 November 2001

A pair of apertures is aligned along an axis and another pair of apertures is aligned along another axis that is perpendicular to the first axis. Each of the apertures is in a common plane. Signals from the apertures along each axis are processed to obtain two-dimensional Doppler motion vectors. The resulting pairs are then processed to produce three-dimensional motion vectors that are used to produce a three-dimensional Doppler image.—RCW

6,589,180

43.80.Vj ACOUSTICAL ARRAY WITH MULTILAYER SUBSTRATE INTEGRATED CIRCUITS

Kenneth R. Erikson *et al.*, assignors to Bae Systems Information and Electronic Systems Integration, Incorporated
8 July 2003 (Class 600/459); filed 8 March 2002

This device is a high-density, complex, and compact ultrasound transducer array constructed through the use of active integrated circuit devices on various substrates and passive devices with micro-bump interconnections between the substrates and micro-vias configured with conductors extending through the substrates. This allows the use of different integrated circuit technologies arranged on a layer-by-layer basis. The various layers may be assembled with solders of lower reflow temperatures to allow for testing of selected layers and circuits prior to completion.—DRR

The Schultz curve 25 years later: A research perspective[a)]

Sanford Fidell
Fidell Associates, 23139 Erwin Street, Woodland Hills, California 91367

(Received 8 May 2003; accepted for publication 22 September 2003)

The contemporary technical rationale for assessing effects (''impacts'') of transportation noise on communities rests in large part on a purely descriptive dosage-effect relationship of the sort first synthesized by Schultz [J. Acoust. Soc. Am. **64**, 377–405 (1978)]. Although U.S. federal adoption of an annoyance-based rationale for regulatory policy has made this approach a familiar one, it is only one of several historical perspectives, and not necessarily the most useful for all purposes. Last reviewed by the U.S. Federal Interagency Committee on Noise (FICON) 10 years ago, the accuracy and precision of estimates of the prevalence of a consequential degree of noise-induced annoyance yielded by functions of noise exposure leave much to be desired. This tutorial article traces the development of the dosage-effect relationship on which FICON currently relies, in a wider historical context of efforts to understand and predict community response to transportation noise. It also identifies areas in which advances in genuine understanding might lead to improved means for predicting community response to transportation noise.
[DOI: 10.1121/1.1628246]

PACS numbers: 43.10.Ln, 43.50.Ba [ADP] Pages: 3007–3015

I. INTRODUCTION

A quarter of a century ago, the *Journal of the Acoustical Society of America* published what proved to be an influential article on community reaction to transportation noise exposure (Schultz, 1978). Schultz demonstrated that the results of social surveys conducted in disparate cities and languages on the effects of aircraft and surface transportation noise could be interpreted in common terms, and usefully summarized in the form of a dosage-effect relationship. Successors to this relationship are relied upon today to characterize noise impacts for purposes such as planning transportation infrastructure projects, and for determining eligibility for federal funding of large-scale noise mitigation projects.

Schultz's 1978 study was a major work of scholarship and technical insight that began the integration of a scattered world literature on community-level noise effects. It helped to promote a measure of time-weighted average noise exposure as a primary predictor of community reaction to noise, established the current paradigm for analysis of such effects, served as the impetus for considerable subsequent research, and offered the prospect of a much-prized technical rationale for environmental noise regulation.

Although Schultz's approach eventually came to be regarded as the conventional wisdom, his paper remained controversial for years (*cf.* Kryter, 1982). Initially, many took issue with details of his conversions of diverse noise metrics into Day-Night Average Sound Level (abbreviated DNL and expressed symbolically in mathematical expressions as L_{dn}) or found fault with his adoption of self-reported annoyance (rather than speech or sleep interference, or complaints) as the dependent variable of his dosage-effect relationship. Others objected to Schultz's rejection of a measure of central tendency of annoyance as a dependent variable, and to his preference for a single relationship to summarize reaction to both aircraft and surface transportation sources. Schultz and others eventually suggested alternate fitting functions, reanalyzed and updated the corpus of findings available for analysis, identified source-specific dosage-effect relationships, and attempted to develop theory-based underpinnings for Schultz's empirical relationship.

Enough has been learned in the years following publication of Schultz's pioneering work on community reaction to transportation noise to warrant reexamination of the research and regulatory paradigms that followed from it. Before doing so, however, it is helpful to review (1) the context in which Schultz conducted his original analyses, and (2) subsequent research findings, understandings, and practical applications of Schultz's work.

II. CONTEXT OF SCHULTZ'S ANALYSIS

The origins of modern legislative and regulatory concern with transportation noise exposure in the United States can be traced to the introduction of jet aircraft at military bases in the early 1950s, to the start of passenger jet service in 1958, and to development of the national highway network in the 1960s.[1] The higher levels and the distinctive features of the noise emissions of jet aircraft *vis-a-vis* those of propeller-driven aircraft, as well as expansion in numbers of flight operations, elicited strongly adverse reactions in communities near military airbases and civil airports. By the 1970s, increased highway traffic noise led to large-scale studies of relationships among traffic flow parameters, noise emissions, and community reaction.

The U.S. Noise Control Act of 1972 was a legislative acknowledgment of national concern with the effects of residential noise exposure. The Environmental Protection Agency's ''Levels Document'' (EPA, 1974), a product of the Noise Control Act, identified a time-weighted average measure of sound level (eventually standardized as DNL) as a convenient expression of the total environmental noise of

[a)]Review and tutorial paper.

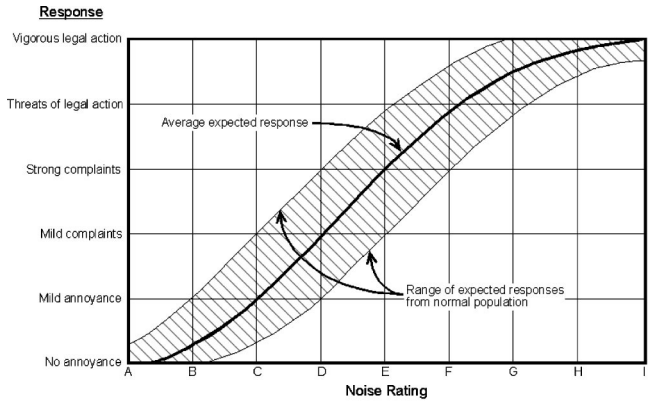


FIG. 1. Relationship between community noise rating and predicted behavioral consequences of environmental noise exposure, adapted from Fig. D-16 of Appendix D of ''Levels Document'' (EPA, 1974).

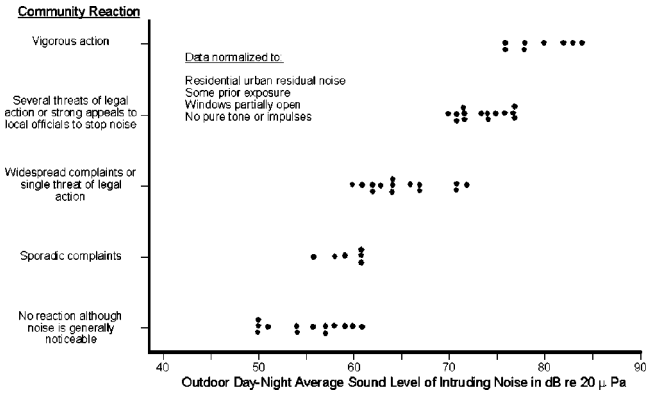


FIG. 2. Community reaction to intrusive noises of many types as a function of the outdoor Day-Night Average Sound Level of the intruding noise. (Adapted from Fig. D-7 of EPA Report 550/9-74-004, ''Information on Levels of Environmental Noise Requisite to Protect Public Health and Welfare with an Adequate Margin of Safety,'' March, 1974.)

communities.[2] Schultz's work started in 1976, under contract to the U.S. Department of Housing and Urban Development. HUD sought to develop consistent criteria for approving federal financial participation in housing projects in neighborhoods with varying degrees of environmental noise exposure.

At the time, the state of the art of assessing the habitability of housing in noisy areas had advanced little from the pioneering work on ''community noise ratings'' conducted throughout the 1950s for the U.S. Air Force and for the Port of New York Authority (*e.g.*, Stevens and Pietrasanta, 1957; Beranek *et al.*, 1959; Galloway and Pietrasanta, 1963). The early approach to characterizing adverse community reaction to aircraft noise focused on prediction of its overt (complaint and similar) behavioral consequences. Rosenblith *et al.* (1953) and Stevens *et al.* (1955) devised a framework for interpreting the findings of 20-odd case studies of community reaction to aircraft noise that characterized community reaction in terms of ''sporadic'' through ''widespread'' complaints, ''threats of community action,'' and ''vigorous community action.'' Figure 1 summarizes the relationship that Rosenblith *et al.* inferred from their case studies.

A ''Community Noise Rating'' (CNR) value was determined by first estimating a ''noise level rank'' from a set of idealized spectral shapes for community noise. These shapes were derived from laboratory findings about the loudness of sounds in different frequency bands. The noise level rank was modified (normalized to standard conditions) by site-specific factors such as ambient noise levels, time of day and year, tonal content, dynamic range of noise intrusions, and novelty of exposure.

CNR-based assessment of community reaction to environmental noise required a detailed case study, involved more-or-less arbitrary judgments about the detailed nature of noise exposure, and made no effort to account for the range of reactions associated with the same rating level (for example, from ''sporadic complaints'' to ''threats of community action'' at rating ''E'') in different communities.[3] The CNR scheme was purely descriptive, and identified no mechanisms by which noise exposure was transformed into complaints.

Despite its limitations, CNR remained influential for two decades after its initial formulation, as is recognizable in Fig. D-7 of EPA's 1974 Levels Document (reproduced above as Fig. 2). CNR evolved in the 1960s into an increasingly simplified Composite Noise Rating (CNR-2), and eventually into a Day-Night Average Sound Level (modeled on California's ''Community Noise Equivalent Level,'' which included 5- and 10-dB evening and nighttime weightings).

As late as EPA's ''Levels Document'' (pp. 20 *et seq.*), non-health-related effects of noise on people were addressed under the rubric of ''Activity Interference/Annoyance.'' The ''activity interference'' portion of this concern referred to masking of communication by environmental noises, as indicated by references to ''listening to a desired sound, such as speech or music'' and ''interference with speech intelligibility.'' The explanatory appendices to EPA's Levels Document are replete with further evidence that annoyance was not the effect of principal concern in identification of protective noise levels. Figures D-7, D-8, and D-16 (of which Figs. D-7 and D-16 are reproduced here as Figs. 2 and 3) of Ap-

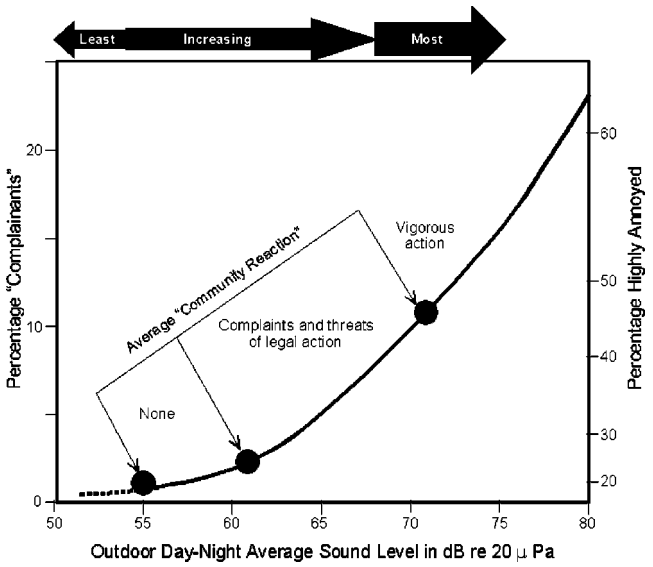


FIG. 3. Illustration of vestigial influence of CNR methodology on assessment of community reaction to aircraft noise exposure in EPA's 1974 ''Levels Document'' (Fig. 16, Appendix D).

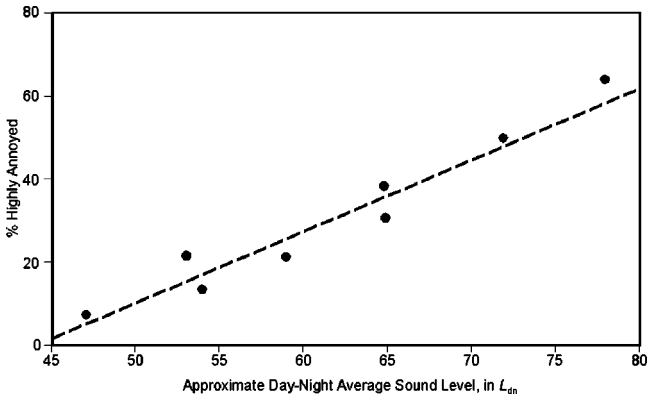


FIG. 4. Early estimate of relationship between cumulative noise exposure and prevalence of aircraft noise-induced annoyance contained in supporting documentation for EPA Levels Document.

pendix D of the Levels Document (''Noise Interference with Human Activities and Resulting Overall Annoyance/Health Effects'') make it clear that EPA's use of the term ''community reaction'' refers to complaints. Figure 2 is merely a reworking of the Rosenblith and Stevens case study complaint information, intended to reduce the variance in each reaction category. Figure 3 is an attempt to relate complaint and annoyance data to noise exposure information through the early dosage-effect relationship shown in Fig. 4. In other words, EPA's 1974 rationale for identifying sound levels requisite to protect public health and welfare was based on speech interference and complaints rather than annoyance:

''Thus, the levels identified [in the Levels Document] primarily reflect results of research on community reaction [*i.e.*, complaints] and speech masking'' (EPA, 1974, p. 21).

The first large-scale social survey that attempted to associate attitudinal factors with noise exposure estimates was conducted in the vicinity of London Heathrow Airport in 1961. The supplementary reports to EPA's Levels Document were cognizant of the results of this and other early social surveys, but not sufficiently swayed by the quantity or interpretability of social survey data to base identification of protective sound levels on this information. Figure 4 illustrates the interpretation afforded to annoyance data at the time of publication of the Levels Document. Several aspects of Fig. 4 remain of interest today: (1) rejection of average annoyance in favor of ''high'' annoyance as the measure of noise effect; (2) reliance upon a fitting function with an assumed form (linear, in this case) to describe the field data; (3) use of the then-newly defined DNL as the predictor of the prevalence of annoyance; and (4) characterization of about a third of the population as highly annoyed by aircraft noise at $L_{\mathrm{dn}} = 65$ dB. It was not until a decade later that Schultz's more extensive work lent enough credibility to such analyses to shift the technical rationale underlying noise regulatory policy from complaints and speech interference to annoyance.[4]

III. SUBSEQUENT EXTENSIONS OF SCHULTZ'S ANALYSIS

The segment of a third-order polynomial function that Schultz used to describe his 11 original clustering surveys (% Highly Annoyed $= 0.8533L_{\mathrm{dn}} - 0.0401L_{\mathrm{dn}}^2 + 0.00047L_{\mathrm{dn}}^3$) was an informal approximation, rather than a relation derived from regression analysis. The limitations of both the data set from which the arbitrary fit was derived and of the fitting function itself were readily apparent. Perhaps the most striking aspect of the data set that Schultz and his successors (*e.g.*, Fields, 1991) assembled is its great variability (*cf.* Schomer, 2002, Fig. 6). Noting the relatively small amounts of variance accounted for by relationships between noise exposure and the prevalence of annoyance in individual studies, Job (1988) inferred that nonacoustic factors that were not reflected in DNL values played a role comparable to exposure itself in determining community reaction to noise.

Schultz recognized the preliminary nature of his original synthesis curve, and did not expect it to remain the final word for long. For example, Fidell *et al.* (1988) modeled the shape of a fitting function on the basis of first principles[5] rather than purely descriptive regression analysis. Green and Fidell (1991) later applied this model to an expanded data set developed by Fidell *et al.* (1991), quantifying the influences of nonacoustic factors on annoyance reports. Harris (Finegold *et al.*, 1994) omitted selected points from the latter data set to derive an ogival fitting function in place of the quadratic form of Fidell *et al.* (1991). CHABA (Fidell, 1996) eventually identified fitting functions for community reaction to high-energy impulsive sounds, while Miedema and Vos (1998) argued for three separate quadratic functions (to fit data from rail, road, and air traffic) in place of a single generalized function for all transportation noise.

IV. PRACTICAL IMPORTANCE OF DOSAGE-EFFECT RELATIONSHIPS IN COMMUNITY NOISE ASSESSMENTS

The U.S. Federal Interagency Committee on Noise (FICON) declared in its 1992 report that annoyance was its preferred ''summary measure of the general adverse reaction of people to noise,'' and that ''the percentage of the area population characterized as 'highly annoyed' by long-term exposure to noise'' was its preferred measure of annoyance. FICON institutionalized the fitting function developed by Harris (*cf.* Fig. 5) for the U.S. Air Force as its preferred dosage-effect relationship. FICON also indicated in Section 3.3.1.2 of its 1992 report that ''the DNL methodology'' (i.e., its preferred dosage-effect relationship) was the basis for its judgments about the acceptability of noise exposure, as expressed in the agency's ''land use compatibility''[6] recommendations.

The canon of community noise policy of U.S. federal agencies is based on FICON's endorsements (1) of annoyance as the primary measure of community reaction[7] to noise exposure, (2) of a particular fitting function as a means for predicting annoyance from cumulative exposure, and (3) of a set of guidelines for the acceptability of annoyance prevalence rates, expressed as ''land use compatibility'' recommendations. Thus, decisions about the award of billions of dollars of federal subsidies to construct airport and highway infrastructure and to mitigate their noise impacts ostensibly rest on the shape of a purely descriptive fitting function,

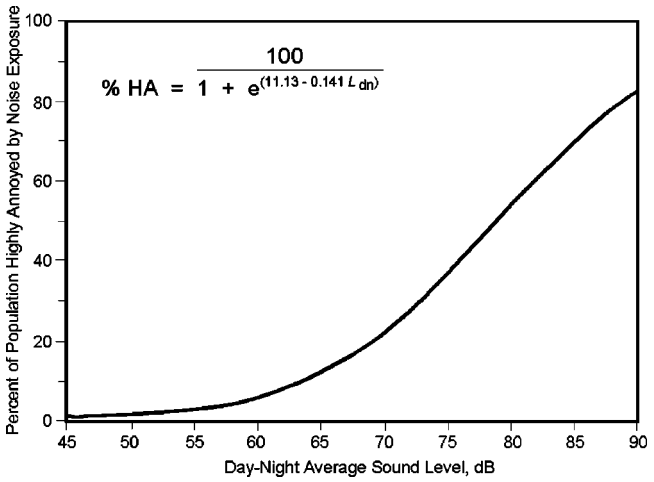


FIG. 5. Fitting function adopted by FICON (1992) as a dosage-effect relationship.

unsupported by quantitative, theory-based, or other systematic understanding of the origins and mechanisms of community reaction to transportation noise.[8]

V. PRAGMATIC LIMITATIONS OF DOSAGE-EFFECT ANALYSIS

Because the dosage-effect relationship seen in Fig. 5 lacks pronounced inflection points, it is not self-interpreting for policy purposes. The slope of the curve varies smoothly from about 1% to 3% highly annoyed per decibel of noise exposure throughout its range, such that the curve itself does not strongly constrain the choice of policy points for regulatory purposes. Historically, such policy points have been identified at 5-dB intervals, in tacit recognition of the uncertainty of measurements of both noise exposure and community reaction. Definition of any particular value of noise exposure as a ''significant'' noise impact is thus inescapably arbitrary, and must be made on nontechnical grounds. At $L_{dn}=65$ dB, the FICON curve seen in Fig. 5 predicts an annoyance prevalence rate of 12.3%, a less than self-evident definition of significance.

Several aspects of FICON's dosage-effect relationship and its application to regulatory policy regularly attract critical comment, even though controversy over its manner of creation has largely subsided. A common criticism of the relationship is that it demonstrably underestimates the prevalence of annoyance due to aircraft noise. Part of this underestimation is due to the functional form of the relationship, and to the range of exposure values over which the relationship was developed. Another source of underestimation is its lack of source-specificity.

The segment of a third-order polynomial fitting function identified by Schultz (1978) was suitable for evaluation only within a restricted range of commonly encountered transportation noise exposure values. The ogival form of the FICON relationship was favored in part for its asymptotic behavior at low and high exposure levels, and in part to control where the broad knee of the curve lies on the abscissa. It is not clear, however, that regulatory policy analyses are as well served by the ogival form of the fitting function as was initially hoped.

The composition and character of community noise differ greatly throughout the enormous (>40 dB) range of exposure levels from which FICON's relationship is derived. In quiet, low population density residential settings, community noise exposure may be governed by relatively small numbers of low level, individually identifiable, discrete noise events that are produced by small numbers of sources. In high-density urban settings, noise exposure is generally created by larger numbers of temporally overlapping, higher level noise events (Fidell *et al.*, 1981). Developing a dosage-effect relationship over this entire range, rather than from data in the vicinity of potential policy points (that is, round-numbered DNL values at which regulatory agencies consider certain actions justifiable), implies a belief that the same processes that give rise to annoyance in quiet rural and suburban settings also give rise to annoyance in noisy urban settings such as those adjacent to airport runways.

The ''equal energy hypothesis''—the notion that the effects of number, duration, and level of noise events are completely equivalent and interchangeable determinants of the annoyance of noise exposure—provides the rationale for including information about community reactions to extremely low and extremely high levels of noise exposure from all sources in a single dosage-effect analysis. Although evidence exists to support the plausibility of the hypothesis, counter-evidence also exists about the unequal influences on annoyance of maximum levels and numbers of noise events.[9] It is for reasons of expedience rather than any conclusive demonstration of causality that DNL intentionally combines into a single index (and thus confounds) all of the primary physical characteristics of noise events that could arguably cause noise-induced annoyance.

The shape of FICON's fitting function is strongly affected by reactions of communities exposed to transportation noise at extreme levels about 20 dB higher and lower than those of practical regulatory interest. Why should a curve intended to inform decisions about tolerable levels of annoyance in common circumstances of noise exposure so strongly reflect information about reactions observed in communities with highly atypical exposure? There can be no realistic expectation that noise-induced annoyance in high population density, motorized society can be limited to that of quiet rural areas, nor that residential uses can freely be made of lands exposed every few minutes, night and day, to high levels of aircraft noise. Forcing the ogival form of FICON's fitting function through the high noise exposure data effectively depresses the broad knee of the curve at more moderate exposure values. This in turn biases the function toward underestimation of the prevalence of annoyance at more commonly occurring exposure levels.[10]

This effect is readily apparent in comparing the means of measured annoyance prevalence rates to the FICON curve in adjacent exposure ranges of practical interest. Figure 6 shows an expanded view of annoyance prevalence rates with aircraft noise exposure levels in the vicinity of $L_{dn}=65$ dB. These data are those of Green and Fidell (1991), supplemented by the findings of subsequent opinion surveys. The

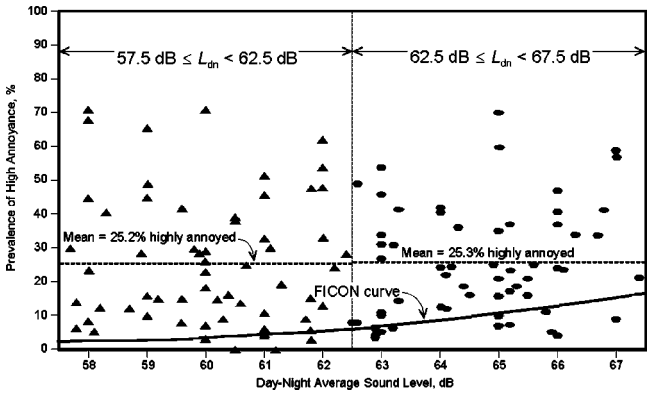


FIG. 6. Expanded view of data on prevalence of aircraft noise-induced annoyance in the vicinity of $L_{dn}=60$ and 65 dB.

triangular data points on the left-hand side of the figure represent observations of the prevalence of annoyance due to aircraft noise in the range of $57.5\,\text{dB} \leqslant L_{dn} < 62.5\,\text{dB}$, while the oval data points on the right-hand side represent observations of the prevalence of annoyance due to aircraft noise in the range of $62.5\,\text{dB} \leqslant L_{dn} < 67.5\,\text{dB}$. The two sets of data points thus represent ranges of ± 2.5 dB around the pragmatically important exposure values of $L_{dn}=60$ and 65 dB, respectively.

The dashed horizontal lines in the lowermost panel of Fig. 6 show the means of the field observations. The curved line is the FICON relationship. It is readily apparent (1) that the FICON relationship underestimates the prevalence of field measurements aircraft noise-induced annoyance, and (2) that the aircraft annoyance data themselves do not compel identification of a DNL value of 65 dB as a self-evidently justifiable or data-driven policy point.

VI. INTERPRETABILITY OF DOSAGE-EFFECT RELATIONSHIPS FOR POLICY PURPOSES

In hindsight, the purely descriptive and exclusively acoustic approach to the problem of predicting community reaction to noise that Schultz pioneered has not been as much of a panacea as once hoped, because the resulting relationships fail to take into account or explain the great variability of community reaction. A less than compelling dosage-effect relationship provides the appearance but not the substance of a systematic basis for policy interpretations which in reality reflect the charters and interests of regulatory agencies at least as much as information about actual noise effects.

A dosage-effect relationship implies that variation in whatever quantity is plotted as the independent variable on the abscissa *causes* variation in whatever quantity is plotted as the dependent variable on the ordinate. When the independent variable is merely an expedient one (such as cumulative noise exposure, an adventitious measure devised for other purposes), and when there is good reason to believe that the dependent variable is strongly influenced by other factors as well, the persuasiveness and utility of a dosage-effect relationship are open to question.

Many of the limitations of the work inspired by Schultz's 1978 relationship stem from its noncausal nature. The most obvious deficiency of the many curve-fitting exercises that followed Schultz's is that none accounts for the better part of the variance in what is now a very large body of social survey data on the prevalence of annoyance associated with environmental noise exposure. This means that no systematic explanations are available for large differences in annoyance prevalence rates in different communities with the same noise exposure. It also means that accurate predictions of the prevalence of annoyance in communities exposed to change in noise levels (for example, from increases in air traffic due to increased market demand, a favorable regulatory climate, over-building of airport infrastructure, or other causes) remain elusive. It further means that prediction of the benefits of costly measures intended to mitigate noise exposure cannot be made with confidence, and that regulatory policies intended to balance conflicting societal interests remain largely arbitrary and poorly supported by technical analysis (Fidell, 1999).

VII. CONSEQUENCES OF FICON'S ENDORSEMENT OF A PREFERRED METHODOLOGY FOR PREDICTION OF COMMUNITY REACTION TO NOISE

FICON'S endorsement of the prevalence of a consequential degree of annoyance as the primary (and for practical purposes, sole) measure of community reaction to noise, and of a particular dosage-effect relationship between noise exposure and annoyance, has undeniably simplified the process of estimating and disclosing transportation noise impacts as mandated by the U.S. National Environmental Policy Act of 1969. This approach errs on the side of oversimplification of the process of predicting community reaction to transportation noise, since (1) noise exposure is neither a necessary nor a sufficient antecedent condition for annoyance, and (2) noise exposure *per se* is not a particularly effective predictor of the prevalence of annoyance. A recent summary by Schomer (2002) has catalogued the various ''adjustments,'' ''corrections,'' and ''normalizations'' to DNL that have been suggested to improve the accuracy of prediction of community reaction from noise exposure measurements. Suggesting *ad hoc* adjustments to exposure measurements construes the problem as one of measurement rather than one of theory, however, and thereby treats the symptoms rather than the disease. Bandaids applied to exposure measurements are akin to the epicycles that Ptolemy's views about the orbits of planets required to account for their otherwise inexplicable retrograde motions. Such patchwork solutions appear helpful in the short run, but only postpone development of more systematic and fundamental explanations.

In the United States, FICON's doctrine has codified the *status quo* in understanding of community reaction to noise as of a quarter century ago, led to repeated misprediction of community reaction to noise exposure, and generally reinforced policies that do not accomplish their own goals. A greater proportion of the population than predicted by FICON is demonstrably highly annoyed by aircraft noise at the *de facto* threshold of federal concern ($L_{dn}=65\,\text{dB}$) (Miedema and Vos, 1998); many airport noise controversies remain inexplicable from the perspective of official recom-

mendations of compatible land use; and vigorous opposition to construction of airport infrastructure is more the rule than the exception.

Overreliance on officially predicted annoyance prevalence rates to assess community reaction to aircraft noise has also created an institutional disconnect between local and federal perspectives. For all practical federal purposes, ''community reaction to noise'' means little more than an annoyance prevalence rate estimated by an assumption-laden fitting function. In the daily experience of airport proprietors and local governments, however, ''community reaction'' generally refers to numbers of recent noise complaints. Imprecise predictions of prevalence rates of covert attitudes have in effect taken precedence over the overt behaviors that were the original focus of Rosenblith *et al.*, and which remain the crux of many aircraft noise controversies.[11]

VIII. ALTERNATIVES FOR IMPROVING ASSESSMENT OF COMMUNITY REACTION TO AIRCRAFT NOISE

According to FICON (1992), Green and Fidell (1991) ''demonstrated how the variability in the data points of the Schultz curve could be significantly reduced by assuming that citizens of the same community tend to share common criteria for deciding when an intruding noise is 'highly annoying'.'' Systematic consideration of the aggregate effect of nonacoustic factors on self-reported annoyance can indeed improve the accuracy and precision of predictions of annoyance prevalence rates. FICON also noted in its 1992 report that ''This work is continuing and may provide a basis for an improved understanding of community response to noise.'' In the decade since publication of FICON's report, however, its successor agency, FICAN, has taken no major action to further improve the accuracy of prediction of the prevalence of noise-induced annoyance in communities.

Furthermore, land use compatibility recommendations (notionally linked to dosage-effect analysis, which in turn relies on cumulative noise exposure as a sole predictor variable) have effectively displaced all other interpretations of transportation noise effects for federal purposes.[12] In particular, FICON (1992) rejects complaint behavior as a basis for interpreting noise effects on the grounds that ''Annoyance can exist without complaints and, conversely, complaints may exist without high levels of annoyance.'' As Schultz and his successors have amply demonstrated, however, it is equally true that high levels of annoyance can exist at low levels of noise exposure, and low levels of annoyance can exist at high levels of noise exposure. The lack of a strong or simple relationship between noise exposure and its effects is neither a consistent nor a persuasive rationale for ignoring noise complaints in policy analyses.

In fact, annoyance prevalence rates and complaint rates may be usefully viewed as two sides of the same coin. Annoyance prevalence rates are estimated from systematically solicited opinions about noise. Complaints are spontaneous (unsolicited) reports of adverse opinions about noise. Complaints and annoyance may differ in gestation period, but understanding of the time course of arousal and decay of annoyance is so imprecise that nothing definitive is known about the terms of exposure that give rise to either annoyance or complaints.

Questionnaire items soliciting self-evaluations of degrees of annoyance necessarily focus on the long term, because it is impractical to administer a social survey in real time to a representative sample of a community about reactions to individual aircraft noise events. Spontaneous self-reports about reactions to aircraft noise often concern egregious individual noise events or periods of exposure. Airports seldom receive complaints on New Year's eve about annual average exposure levels over the course of the preceding calendar year.

In this context, it makes no more sense to ignore complaint behavior because it may or may not be closely related to annoyance than to ignore attitudes of annoyance because they may or may not be closely related to complaints. Both solicited and unsolicited forms of self-report confound ''true'' sensitivity to noise with reporting bias (Green and Fidell, 1991). Biases associated with complaints may strike some as more obvious than biases associated with self-reported annoyance, but neither spontaneous nor solicited forms of expression are free from nonacoustic influences. Neither complaints nor annoyance are any less worthy of consideration because of this confounding, and neither the acoustic nor the nonacoustic determinants of annoyance and complaints can be summarily dismissed by airport proprietors or regulatory agencies.[13]

In reality, noise complaints play a strong, albeit unspoken, role in airport design and operation. Dallas–Ft. Worth International Airport was sited on about 18 000 acres of land in the early 1970s, even though its projected $L_{dn}=65$ dB cumulative noise exposure contour encompassed far less area. Likewise, Denver International Airport was sited on about 29 000 acres, even though its projected $L_{dn}=65$ dB cumulative noise exposure contour was considerably smaller. Both of these greenfield airports have nonetheless attracted tens of thousands of aircraft noise complaints over the years, some from communities many miles from their $L_{dn}=65$ dB cumulative noise exposure contours. Regional airspace use and flight track modification controversies such as the Extended East Coast Plan are typically complaint-driven, and frequently require resolution of noise problems at exposure levels that are inconsequential from the perspective of federal land use compatibility guidelines. Although such adverse community reaction may seem ''wrong'' from the perspective of airport proprietors and regulators, it nonetheless has substantive consequences and obvious implications for the adequacy of cumulative exposure as a sole predictor of community reaction.

Complaints were abandoned as a measure of community reaction to noise at the federal level in the 1970s largely because of the promise that Schultz's relationship seemed to offer. At the time, noise complaints were difficult to process and systematically compare, largely inaccessible to researchers, and generally awkward to interpret. These limitations have lessened over the last decade as computer-based aircraft noise and operations monitoring systems have become commonplace at major airports, and as geo-information system

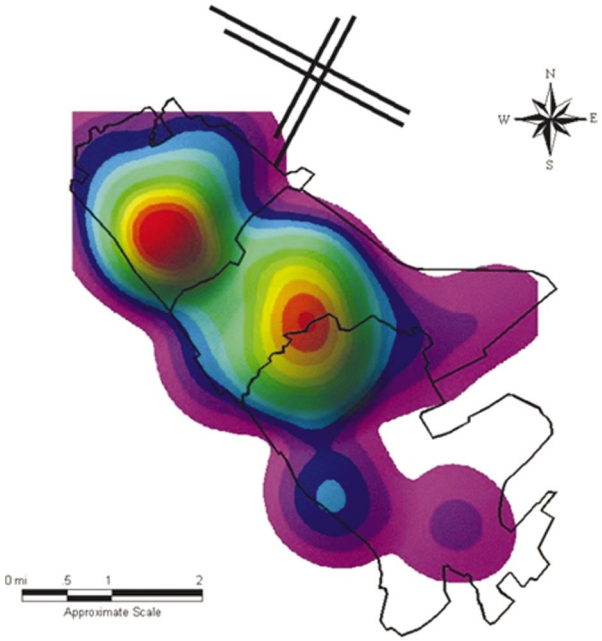


FIG. 7. Rendering of complaint density pseudo-terrain as redundantly color-coded false elevation behind main departure runways at San Francisco International Airport.

software has come of age. Larger airports now routinely maintain well-organized, long-term files of geo-coded noise complaints. These are proving more tractable to interpretation than previously believed (Fidell and Howe, 1998).

Perhaps the most common remaining complaints about complaints as a measure of community reaction to noise are (1) that they are not obviously related to cumulative noise exposure, and (2) that most aircraft noise complaints are received from geographic areas outside the $L_{dn}=65$ dB noise exposure contour at most airports (GAO, 2000). These circular concerns are misplaced, given that cumulative noise exposure is itself a far from perfect predictor of annoyance. Complaint rates are sometimes also denigrated as emphasizing the views of small numbers of frequent complainants, even though analysis of very large, computer-maintained aircraft noise complaint files shows that mean and modal numbers of complaints per complainant are quite small (Fidell and Howe, 1998).

One example of the ready interpretability of complaint information is evident in the geographic pattern of noise complaints associated with start of takeoff roll noise at San Francisco International Airport. An airport-sponsored analysis (Pearsons *et al.*, 2000) of noise complaints lodged over a period of 6 years was conducted by geo-coding street addresses of complainants to contour complaint densities. Figure 7 shows these complaint densities coded as false elevation. The peaks of the pseudo-terrain correspond to two concentrations of complaints, located behind and roughly 45° to the sides of the extended centerlines of the airport's primary departure runways. These locations correspond to the lobes of the directivity patterns of jet engine exhaust noise of aircraft departing on these runways. The complaint concentrations are well beyond the airport's $L_{dn}=65$ dB cumulative noise exposure contour.

The unusually great low-frequency content of start of takeoff roll (''backblast'') noise is an arguably special case in which the degree of adverse reaction to noise is underestimated by the A-weighting of cumulative exposure (cf. Fidell *et al.*, 2002). Limitations of the A-weighting network do not account for similar findings about the geographic distribution of complaints with respect to DNL contours at airports elsewhere, however. An airport-sponsored complaint analysis conducted at Naples Municipal Airport in Florida documents a mismatch between overt community reaction to aircraft noise and land use compatibility recommendations premised on annoyance prevalence rates. Figure 8 shows two ''mountains'' in complaint density (rendered as false elevation) along the extended centerline of the primary departure runway at the airport. The contour draped over the complaint density pseudo-terrain that encompasses the bulk of the high ground is the 95 dB maximum A-level contour. (The airport's $L_{dn}=65$ dB contour closes much nearer to the end of the runway.)

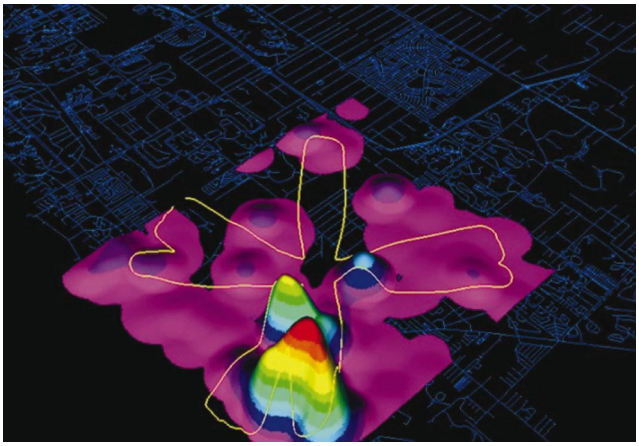


FIG. 8. Rendering of complaint density pseudo-terrain as redundantly color-coded false elevation near Naples Municipal Airport, with 95 dB maximum A-weighted aircraft noise contour superimposed in yellow.

Noise complaints at Naples Municipal Airport were dominated by a very small number of unscheduled operations by an unusually noisy aircraft. Noise emissions from the fleet operating at Hanscom Field, however, are less influenced by such small numbers of operations of especially

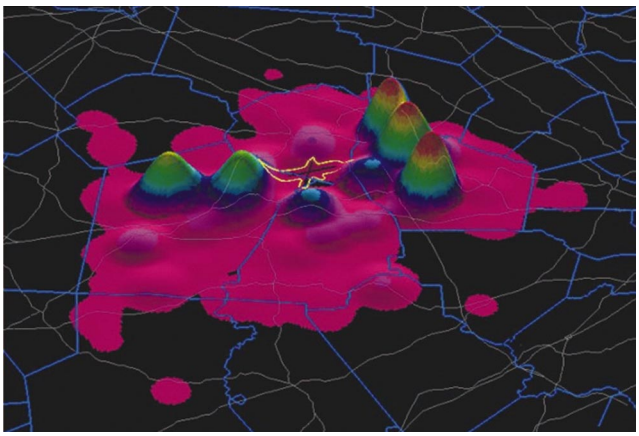


FIG. 9. Rendering of complaint density pseudo-terrain as redundantly color-coded false elevation near Hanscom Field, with $L_{dn}=65$ dB noise exposure contour superimposed in yellow.

noisy aircraft. Nonetheless, Fig. 9 shows that peaks of complaint density remain well outside of the $L_{dn}=65$ dB contour that supposedly distinguishes airport-compatible from airport-incompatible residential land uses.

The geographic distributions of noise complaints with respect to runway ends, flight tracks, and directivity of aircraft noise sources are more consistent with proximity to flight tracks and directivity of noise sources than with current regulatory policy for assessment of transportation noise impacts. The increased interpretability of noise complaints made possible by computer-based record keeping and geoinformation system software suggests a more prominent role in the future for complaint rate information in the design of aircraft noise mitigation projects and impact assessments. Ironically, such a role would be reiminiscent of that which complaints played in community reaction assessments prior to Schultz's 1978 synthesis work.

IX. CONCLUSIONS

A quarter of a century of follow-up work to Schultz's 1978 synthesis is sometimes cited as establishing credibility for assessment of environmental noise impacts exclusively on the basis of DNL values. It is apparent in retrospect, however, that a point of diminishing returns has been passed in dosage-effect analysis, and that the impetus to research and policy analysis that Schultz's work provided has run its course without yielding further major improvements in systematic understanding of causes and mechanisms of community reaction to transportation noise. An administratively convenient partial solution to a vexing societal problem may suffice for some nontechnical policy purposes. Expedient but incomplete solutions do not constitute genuine understanding of community reaction to noise, however, and can not serve in lieu of theory development and research to improve understanding in this field.

A thorough review of the technical rationale for FICON's decade-old endorsement of dosage-effect analysis would be a useful initial step toward improved understanding of community reaction to transportation noise. The following issues are among those that warrant scrutiny in light of what has been learned since Schultz's 1978 work:

(i) the effects on policy analyses of poor correlation between annoyance prevalence rates predicted by the fitting function preferred by FICON and rates actually observed in communities;
(ii) analysis of the logic and effects on noise impact interpretations of the range of exposure values over which the fitting function is developed, and of its form;
(iii) computation of error bounds and confidence intervals for predicted annoyance prevalence rates, and frank exploration of their effects on land use compatibility recommendations;
(iv) adoption of a data-driven rationale for selection of policy points rather than an imprecise predictive function;
(v) systematic, quantitative, and theory-based consideration of nonacoustic factors as codeterminants of the annoyance of transportation noise; and
(vi) formal recognition of geographic distributions of noise complaints as an alternate indication of actual community reaction to transportation noise.

ACKNOWLEDGMENTS

The author is grateful for comments and discussion of draft versions of this manuscript by Leo Beranek, Karl Pearsons, Steven Pflaum, Perry Rosen, and Paul Schomer, and for suggestions made by two anonymous reviewers.

[1]The first urban noise survey (Fletcher *et al.*, 1930) had been conducted in New York City three decades earlier, not long after the development of portable electronic sound measurement instruments made such work possible. Societal interest in environmental noise effects remained minimal through the intervening decades of economic depression and world war, however.

[2]The reasoning that led to EPA's embrace of DNL is described in great detail by von Gierke (1973) in supplementary reports prepared in support of the Levels Document.

[3]The U.S. Air Force later developed a set of numeric equivalents for the original CNR letter categories A though M (Stevens and Pietrasanta, 1957), in which the equivalent level of the 300–600 Hz octave band of aircraft noise was substituted for the original ''level rank'' curves. (This spectral region is a reasonable predictor of the ability of aircraft noise to interfere with speech.) The final development of this ''Composite Noise Rating'' by Galloway and Pietrasanta (1963) substituted perceived noise levels for equivalent levels in the 300–600 Hz octave band. A CNR value of 100 is equivalent to $L_{dn}=65$ dB. By the early 1970s, the Composite Noise Rating scale had evolved into the ''Noise Exposure Forecast'' (NEF) scale in which the earliest aircraft noise exposure contours were expressed.

[4]The initially controversial nature of characterizing community reaction to noise in terms of annoyance is apparent from written comments by the Boeing Commercial Airplane Group (von Gierke, 1973, p. III-C-18): ''The selection of 60 dBA as a goal appears to be founded on arbitrary conclusions about the relationship between cumulative noise exposure and the highly subjective concept of 'public annoyance'.'' The ''subjective'' nature of annoyance still grates on some who regret that people do not respond to noise exactly as do sound level meters. If Schultz's work has accomplished nothing else, it has demonstrated the futility of attempts to treat community reaction to noise as an exclusively physical process.

[5]Taking a normative rather than a descriptive approach, Fidell *et al.* (1988) hypothesized that the shape of a dosage-effect relationship should be governed by the rate of increase of annoyance with effective (duration-adjusted) loudness. They also attributed deviations from the hypothesized rate of growth of annoyance to the effects of nonacoustic factors. These nonacoustic factors translate (''bias'') the prediction function along the abscissa, but do not alter its form or shape.

[6]It is important to recall that the asset that ''land use compatibility'' guidelines protect is public investment in airport facilities. Surrounding land uses are defined as compatible with an airport when they do not jeopardize or constrain the airport's continued operation and expansion.

[7]Since the goal of assessing community reaction to noise exposure changed from predicting overt group action to annoyance prevalence rates, the term has become something of a misnomer. ''Community reaction'' today means little more than a prevalence rate of a consequential degree of annoyance among individuals.

[8]Decisions about the conduct of large civil works projects are of course influenced to a greater degree by political, economic, and pragmatic considerations than by their noise impacts. Nonetheless, challenges to such projects based on state and federal environmental disclosure statutes often turn on issues of noise policy and interpretation. Thus, the lack of explicit or systematic linkage between FICON's fitting function and its policy interpretations of ''land use compatibility'' guidelines underscores the arbitrariness of such recommendations.

[9]Exposure, the logarithmic sum of numbers and levels of individual noise events (commonly normalized to 1-s durations), is obviously highly correlated with both numbers and levels of noise events. Given this high correlation, as well as the influences of inevitable nuisance variables, a critical

experiment to determine whether numbers of noise events, levels of noise events, or their product are more closely related to the prevalence of annoyance would require unreasonably large sample sizes and precision of measurement.

[10]The parabolic fit of Miedema and Vos (1998) more closely reflects the mean annoyance prevalence rate for aircraft noise in the vicinity of $L_{dn} = 65$ dB, but, like the FICON curve, fails to account for the better part of the variance in the data set.

[11]A large part of the rationale for reducing the 1981 budget of EPA's Office of Noise Abatement and Control to zero was that, like politics, noise controversies are inherently local (Shapiro, 1991). This rationale is inconsistent, however, with the preemptive disconnect between federal and local perspectives on transportation noise impacts.

[12]FICON acknowledges in principle the limited utility of noise metrics such as integrated time in excess of a threshold level and maximum sound level for supplementary analytic purposes, but expresses its formal interpretations of ''land use compatibility'' only in terms of cumulative exposure.

[13]Some argue that complaints ought not inform regulatory decisions because small numbers of them could have disproportionate influence on such decisions. Given that a subjective judgment about the ''significance'' of noise exposure is *not* a scientific nor a technically justifiable decision in the first place, there is no technical basis for determining whether the virtues of representative sampling (in the case of quantifying the long-term attitude of annoyance) outweigh the value to public officials of spontaneous reports by their constituents of adverse reactions to noise (in the case of the immediate behavior of complaining).

Beranek, L., Kryter, K., and Miller, L. (**1959**). ''Reaction of people to exterior aircraft noise,'' Noise Control, Acoust. Soc. Am., September, pp. 23–32.

Environmental Protection Agency (**1974**). ''Information on Levels of Environmental Noise Requisite to Protect Public Health and Welfare with an Adequate Margin of Safety,'' U.S. Environmental Protection Agency, EPA/ONAC 550/9-74-004, Washington, DC.

Federal Interagency Committee on Noise (FICON) (**1992**). ''Federal Agency Review of Selected Airport Noise Analysis Issues,'' Report for the Department of Defense, Washington, DC.

Fidell, S. (ed.) (**1996**). ''Assessment of Community Response to High Energy Impulsive Sounds,'' National Research Council, National Academy Press, Washington, DC.

Fidell, S. (**1999**). ''Assessment of the effectiveness of aircraft noise regulation,'' Noise & Health **3**, 17–25, ISSN 1463–1741.

Fidell, S., and Howe, R. (**1998**). ''Use of Airport Noise Complaint Files to Improve Understanding of Community Response to Aircraft Noise,'' NASA Contractor Report 207650, NASA Langley Research Center, Hampton, VA.

Fidell, S., Barber, D., and Schultz, T. J. (**1991**). ''Updating a dosage-effect relationship for the prevalence of annoyance due to general transportation noise,'' J. Acoust. Soc. Am. **89**, 221–233.

Fidell, S., Horonjeff, R., and Green, D. M. (**1981**). ''Statistical analyses of urban noise,'' Noise Control Eng. J. **16**, 75–80.

Fidell, S., Schultz, T. J., and Green, D. (**1988**). ''A theoretical interpretation of the prevalence rate of noise-induced annoyance in residential populations,'' J. Acoust. Soc. Am. **84**, 2109–2113.

Fidell, S., Pearsons, K., Silvati, L., and Sneddon, M. (**2002**). ''Relationship between low-frequency aircraft noise and annoyance due to rattle and vibration,'' J. Acoust. Soc. Am. **111**, 1743–1750.

Fields, J. M. (**1991**). ''An updated catalog of 318 social surveys of residents' reactions to environmental noise (1943–1989),'' NASA Contractor Report 187553, Contract NAS1-19061.

Finegold, L., Harris, C. S., 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**, 25–30.

Fletcher, H., Beyer, A. H., and Duel, A. B. (**1930**). ''Noise Measurement,'' in *City Noise, Report of the Noise Abatement Commission*, Department of Health, City of New York.

Galloway, W., and Pietrasanta, A. (**1963**). ''Land use planning with respect to aircraft noise,'' AFM 86-5, TM 5-365, NAVDOCS (also published by FAA as TR-821).

Government Accounting Office (GAO) (**2000**). ''Aviation and the environment: results from a survey of the nation's 50 busiest commercial service airports,'' GAO/RCED-00-222, Washington, DC, pp. 35–36.

Green, D. M., and Fidell, S. (**1991**). ''Variability in the criterion for reporting annoyance in community noise surveys,'' J. Acoust. Soc. Am. **89**, 234–243.

Job, R. F. S. (**1988**). ''Community response to noise: A review of factors influencing the relationship between noise exposure and reaction,'' J. Acoust. Soc. Am. **83**, 991–1001.

Kryter, K. D. (**1982**). ''Community annoyance from aircraft and ground vehicle noise,'' J. Acoust. Soc. Am. **72**, 1222–1242. See also ''Rebuttal by K. D. Kryter to comments by T. J. Schultz,'' J. Acoust. Soc. Am. **72**, 1253–1257.

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

Pearsons, K., Fidell, S., Silvati, L., Sneddon, M., and Howe, R. (**2000**). ''Study of the Levels, Annoyance and Potential Mitigation of Backblast Noise at San Francisco International Airport,'' BBN Report 8257.

Rosenblith, W., Stevens, K. N., and the staff of Bolt Beranek and Newman (**1953**). *Handbook of Acoustic Noise Control, Volume II, Noise and Man*, WADC Technical Report 52-204, Wright Air Development Center, OH.

Schomer, P. (**2002**). ''On normalizing DNL to provide better correlation with response,'' Sound Vib. 14–23.

Schultz, T. J. (**1978**). ''Synthesis of social surveys on noise annoyance,'' J. Acoust. Soc. Am. **64**, 377–405.

Shapiro, S. (**1991**). ''The dormant Noise Control Act and options to abate noise pollution,'' Administrative Conference of the United States, Washington, DC.

Stevens, K., and Pietrasanta, A. (**1957**). ''Procedures for estimating noise exposure and resulting community reactions from air base operations,'' WADC TN-57-10, Wright-Patterson Air Force Base, OH.

Stevens, K. N., Rosenblith, W., and Bolt, R. (**1955**). ''A community's reaction to noise: Can it be forecast?'' Noise Control **1**, 63–71.

von Gierke, H. (**1973**). ''Report on Impact Characterization of Noise Including Implications of Identifying and Achieving Levels of Cumulative Noise Exposure,'' prepared for Task Group 3 of the Environmental Protection Agency's Aircraft/Airport Noise Report Study.

On derivation of differential equations of coupled-mode propagation from the reciprocity principle (L)

Oleg A. Godin[a)]
CIRES, University of Colorado and NOAA/Environmental Technology Laboratory, 325 Broadway, Boulder, Colorado 80305-3328

(Received 14 November 2002; revised 2 August 2003; accepted 25 August 2003)

In a Letter to the Editor, McDaniel [J. Acoust. Soc. Am. **114**, 3020–3023 (2003)] criticizes the derivation of equations of coupled-mode propagation [O. A. Godin, J. Acoust. Soc. Am. **103**, 159–168 (1998)] based on the reciprocity principle and proposes another derivation. A close examination of McDaniel's arguments shows that her criticism is unwarranted. The derivation presented by McDaniel does not apply to acoustic waveguides with sloping boundaries and/or interfaces, which are the subject of analysis reported in the 1998 article. Two alternative derivations are discussed and shown to give the same mode-coupling equations as originally obtained by Godin.
 [DOI: 10.1121/1.1619979]

Pages: 3016–3019

PACS numbers: 43.30.Bp, 43.20.Bi, 43.20.Mv [WLS]

I. INTRODUCTION

In a 1998 paper,[1] Godin pointed out that there was a significant conflict between coupled-mode formulations proposed by different authors to describe wave propagation in acoustic waveguides with sloping interfaces and/or rigid boundaries. With the same set of local modes, the difference in the rate of mode coupling due to sloping interfaces, as predicted by various theories, could be arbitrarily close to 100%, regardless of the interface slope or fluid parameters contrast across the interface.[1] Some of the published derivations of mode-coupling equations were found to be in error for waveguides with sloping boundaries and/or interfaces, and the error was traced back to unjustified term-by-term differentiation of the expansion of acoustic pressure p in terms of the local modes f_m. The series representing the expansion is slowly convergent when the full acoustic field and local modes obey distinct boundary conditions. This is the case when there are sloping rigid boundaries or interfaces between fluids of distinct densities in a waveguide. The slow convergence prevents one from term-by-term differentiation when calculating the first derivative of p with respect to range or depth at an interface, as well as the second derivatives of p.

Mode-coupling equations were derived in Ref. 1 from the reciprocity principle, and it was demonstrated that the (correct) mode-coupling equations (*i*) are an exact reformulation of the reduced wave equation and the boundary conditions and (*ii*) ensure energy conservation and acoustic reciprocity in horizontally inhomogeneous waveguides with or without sloping boundaries and interfaces. Mode-coupling coefficients were found in terms of environmental gradients, that is, range derivatives of sound speed and density as well as slopes of interfaces and boundaries. The mode-coupling equations derived in Ref. 1 are consistent with results obtained from different considerations by Brekhovskikh and Godin,[2] Chiu *et al.*,[3] Odom *et al.*,[4] and, for a special case, by Fawcett.[5] The matrix of coupling coefficients obtained by Godin[1] was recently rederived by Abawi.[6]

Although the validity of any results reported in Ref. 1 is not disputed, McDaniel[7] argues that the symmetry of coupling coefficients, which was used in the derivation[1] together with the reciprocity principle, is an unjustified assumption. In Sec. II, we discuss the symmetry, demonstrate its origin, and justify its application. In Sec. III, we consider alternative derivations of mode-coupling equations for waveguides with generic piecewise continuous sound speed and density and compare ensuing mode-coupling equations with those derived in Ref. 1. Section IV summarizes the results and concludes with remarks on the history of the development of coupled-mode theory in acoustic waveguides.

II. SYMMETRY OF MODE-COUPLING COEFFICIENTS

For the 2-D problem of sound propagation in a range-dependent waveguide, mode-coupling equations have been obtained[1] in terms of dependent variables F_n^+ and F_n^-, which are defined as

$$F_m^{\pm}(x) = 0.5(p \pm \omega\rho\nu/\xi_m, f_m). \tag{1}$$

In a range-independent segment of the waveguide, $F_m^+(F_m^-)$ has meaning as the amplitude of the mth normal mode traveling to the right (to the left). Here, $p(x,z)$ is acoustic pressure, $\nu = (i\omega\rho)^{-1}\,\partial p/\partial x$ is the x-component of particle velocity, ω is wave frequency, and $\xi_m(x)$ and $f_m(z;x)$ are the propagation constant and normalized shape function of the local mode of the order m. The inner product used in Eq. (1) is defined for arbitrary, piecewise continuous functions ψ and χ as follows:

$$(\psi,\chi) = \left(\int_0^{h_1} + \int_{h_1}^{h_2} + \cdots + \int_{h_N}^{H} \right) \frac{dz}{\rho(x,z)}\, \psi(x,z)\chi(x,z). \tag{2}$$

Away from the interfaces $z = h_j(x)$, $j = 1,2,...,N$, sound speed $c(x,z)$ and density $\rho(x,z)$ are smooth functions of the coordinates.

Because the acoustic field in a source-free region of the waveguide is uniquely determined by radiation conditions at

a)Electronic mail: oleg.godin@noaa.gov

infinity and p and ν profiles given at a vertical cross section, the derivatives dF_n^+/dx and dF_n^-/dx are functions of F_m^+ and F_m^-. Due to the superposition principle,[2] the dependence of the derivatives on F_m^+ and F_m^- is linear and can be written as

$$dF_n^+/dx - i\xi_n F_n^+ = \sum_m (b_{nm}^+ F_m^+ + b_{nm}^- F_m^-);$$
$$dF_n^-/dx + i\xi_n F_n^- = \sum_m (c_{nm}^+ F_m^+ + c_{nm}^- F_m^-). \tag{3}$$

It was argued in Ref. 1 that invariance of the mode-coupling equations with respect to a reference frame transformation imposes the symmetry requirement

$$b_{nm}^{\pm} = c_{nm}^{\mp} \tag{4}$$

on the coupling coefficients. The physical meaning of the symmetry is very simple: Eq. (4) states that (*i*) coupling between the nth and the mth modes propagating to the right is the same as between the nth and the mth modes propagating to the left, and (*ii*) coupling between the nth mode propagating to the right and the mth mode propagating to the left is the same as between the nth mode propagating to the left and the mth mode propagating to the right.

McDaniel[7] considers a different set of dependent variables, namely,

$$M_n^{\pm} = F_n^{\pm} \pm (i/2\xi_n) \sum_m B_{nm}(F_m^+ + F_m^-), \tag{5}$$

where $B_{nm} = (f_n,\ \partial f_m/\partial x)$. For $M_n^{\pm}$, she obtains coupling equations similar to Eq. (3) and notes that coupling coefficients in those equations do not have the symmetry analogous to Eq. (4). On these grounds, McDaniel declares the symmetry, Eq. (4) used in Ref. 1, an unjustified, *a priori* assumption.

McDaniel is correct in noting that the symmetry does not necessarily hold for an arbitrary definition of mode amplitudes. However, the symmetry has been discussed and used[1] in case of mode amplitudes defined by Eq. (1) only, rather than for an arbitrary splitting of the field into waves propagating in opposite directions. Thus, McDaniel's analysis of equations for $M_n^{\pm}$ is not pertinent to the subject of the derivation[1] validity. More importantly, a physics-based definition (1) of mode amplitudes is unique in a sense. It was explained[1] [see text after Eq. (9) OG] that the symmetry in Eq. (4) stems directly from the facts that the coupling coefficients in Eq. (3) vanish in range-independent segments of the waveguide, are linear functionals of environmental gradients, and, together with the environmental gradients, change their sign when the direction of the x axis is inverted. (Here and below, for brevity, specific equations from Ref. 1 and 7 are referred to by their number followed by letters OG and SM, respectively.)

To formally justify the symmetry (4) of the coupling coefficients in the first-order coupling equations (3) for a waveguide with sloping interfaces, note that the symmetry was independently established by Brekhovskikh and Godin[2] (Sec. 7.1.3) for waveguides without interfaces. As any interface within a fluid can be considered as a limit of smooth but rapid variations of sound speed and density, and the symmetry holds exactly, regardless of the rate of parameter variation, the symmetry of coupling coefficients should necessarily hold in the limit when an interface (discontinuity in sound speed and density) appears. It is crucial for this reasoning that the symmetry was established[2] (Sec. 7.1.3) for both sound speed and density dependent on x and z, thus making a transition to the limit of a sloping interface possible.

Ultimately, the symmetry (4) owes its existence to primitive equations for the acoustic field,

$$\frac{\partial p}{\partial x} = i\omega\rho\nu, \quad i\omega\frac{\partial \nu}{\partial x} = -\frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial p}{\partial z}\right) - \frac{\omega^2}{\rho c^2}p, \tag{6}$$

being of first order in range coordinate, when written in terms of p and ν. Mode-coupling equations (1) for $F_n^{\pm}$ arise as a result of projecting Eqs. (6) on the basis of local modes, i.e., calculating inner products of each of Eqs. (6) and f_n. Obviously, only first derivatives of the environmental parameters with respect to range, $\partial\rho/\partial x$, $\partial c/\partial x$, and $\partial h_j/\partial x$, may arise in calculating the inner products.

McDaniel[7] contrasts "lengthy analysis" in Ref. 1 with "a direct application of the reciprocity principle." In fact, the derivation in Sec. 3B[7] is nothing more than a 2-D rendition of a transformation described in Ref. 1 for the 3-D case; see Eqs. (22) OG and (23) OG. It was also explained[1] [see the text between Eq. (23) OG and the end of Sec. I] why reciprocity considerations and the requirement of coordinate-system independence are not sufficient to determine uniquely the coupling coefficients in the second-order coupling equations, thus making it necessary to resort to the lengthier analysis of the first-order coupling equations.

III. ALTERNATIVE DERIVATIONS OF MODE-COUPLING EQUATIONS

As mentioned in the Introduction, there exist numerous approaches to derive mode-coupling equations for the general case of a fluid with piecewise continuous sound speed and density and sloping boundaries and/or interfaces. The transformations outlined in Sec. IV in Ref. 7 do not qualify as such an approach since sloping boundaries and interfaces are not considered. These results are nevertheless interesting because they explicitly demonstrate consistency of the propagation models based on continuously coupled modes, which are the subject of the present work, with the models based on discretely- (or stepwise-) coupled modes,[8] where a 2-D waveguide is approximated by a sequence of range-independent segments which join at vertical interfaces, and mode coupling occurs at the vertical interfaces only.

In a waveguide with continuous range dependence, mode coupling occurs continuously and is expressed through differential equations for mode amplitudes. To provide an alternative derivation of the mode-coupling equations for a generic waveguide with continuous range dependence, let us calculate inner products of each of Eqs. (6) with a mode shape function f_n. Caution needs to be exercised to avoid term-by-term differentiation of slowly convergent series which represent developments of p and ν into local normal modes. Using integration by parts, from the first of Eqs. (6) we obtain

$$\frac{\partial}{\partial x}(p,f_n)=i(\omega\rho\nu,f_n)+\left(p,\frac{\partial f_n}{\partial x}+\rho f_n\frac{\partial}{\partial x}\frac{1}{\rho}\right)-\sum_{j=1}^{N}\frac{\partial h_j}{\partial x}f_n\frac{p}{\rho}\bigg|_{z=h_j-0}^{z=h_j+0}. \quad (7)$$

After an integration by parts, from the definition of B_{nm}, it follows

$$B_{nm}+B_{mn}=\sum_{j=1}^{N}\frac{\partial h_j}{\partial x}\frac{f_nf_m}{\rho}\bigg|_{z=h_j-0}^{z=h_j+0}-\left(f_m,\rho f_n\frac{\partial}{\partial x}\frac{1}{\rho}\right). \quad (8)$$

Invoking the definition (1) of modal amplitudes, from Eqs. (7) and (8) we find

$$\frac{dF_n^+}{dx}+\frac{dF_n^-}{dx}=i\xi_n(F_n^+-F_n^-)-\sum_m B_{nm}(F_m^++F_m^-). \quad (9)$$

Quite similarly, from the second of Eqs. (6) we have

$$\frac{\partial}{\partial x}(\omega\rho\nu,f_n)=\left(\omega\rho\nu,\frac{\partial f_n}{\partial x}\right)+i\xi_n^2(p,f_n)+i\sum_{j=1}^{N}\left(\frac{p}{\rho}\frac{\partial f_n}{\partial z}-\frac{f_n}{\rho}\frac{\partial p}{\partial z}+i\omega\nu f_n\frac{\partial h_j}{\partial x}\right)\bigg|_{z=h_j-0}^{z=h_j+0}. \quad (10)$$

Boundary conditions for local modes consist in continuity of f_n and $\rho^{-1}\partial f_n/\partial z$ at the interfaces. For the full acoustic field, the pressure p and the normal component of particle velocity, $\nu\ \partial h_j/\partial x+i(\omega\rho)^{-1}\partial p/\partial z$, are continuous. Therefore, each term of the sum on the right-hand side of Eq. (10) equals zero. In terms of the modal amplitudes (1), Eq. (10) now becomes

$$\frac{d}{dx}[\xi_n(F_n^+-F_n^-)]=i\xi_n^2(F_n^++F_n^-)+\sum_m \xi_mB_{mn}(F_m^+-F_m^-). \quad (11)$$

Resolving the linear equations (9) and (11) with respect to the derivatives $dF_n^{\pm}/dx$, one arrives at the mode-coupling equations (3) with coupling coefficients

$$b_{nm}^{\pm}=c_{nm}^{\mp}=\mp\delta_{nm}\frac{\partial\xi_n/\partial x}{2\xi_n}\pm\frac{\xi_m}{2\xi_n}B_{mn}-\frac{1}{2}B_{nm}. \quad (12)$$

The result agrees with Eq. (19) OG obtained in Ref. 1 from different considerations.

The length of a derivation of mode-coupling equations depends on initial assumptions and a desired level of mathematical rigor to be achieved. There is no controversy about mode-coupling equations in a waveguide without sloping boundaries and interfaces. If one is prepared to consider interfaces as a limit of continuous variations of sound speed and density,[7] then by far the simplest way to derive the mode-coupling equations for the case of media with sloping interfaces is to consider the limit of the equations established for the case without interfaces. In particular, for a range-dependent waveguide with arbitrary smooth dependence of sound speed and density on coordinates and without sloping boundaries and interfaces, mode-coupling equations are known[2] (Sec. 7.1.3) to have the form (3) with coupling coefficients (12). Piecewise continuous variation of sound speed and density in a medium with sloping interfaces can be considered as a limiting case of smooth variation when the thickness of narrow transition layers around the interfaces tends to zero. As long as the coupling coefficients remain finite in the limit, mode-coupling equations must retain the form (3) and (12) in the presence of sloping interfaces. The same argument equally applies to the second-order coupling equations and to the 3-D problem. An inspection shows that, as expected, in all the cases, the results obtained through the limit process agree with the mode-coupling equations derived in Ref. 1 directly for media with sloping boundaries and interfaces.

To explicitly separate contributions due to sloping interfaces in the mode-coupling coefficients (12), one can start from the expressions [Ref. 2, Eqs. (7.1.17) and (7.1.36)]

$$B_{nm}=\frac{1}{\xi_m^2-\xi_n^2}\int_0^H dz\left[f_nf_m\frac{\partial}{\partial x}\left(\frac{k^2}{\rho}\right)-\xi_m^2f_nf_m\frac{\partial}{\partial x}\left(\frac{1}{\rho}\right)+\frac{1}{\rho^2}\frac{\partial\rho}{\partial x}\frac{\partial f_n}{\partial z}\frac{\partial f_m}{\partial z}\right],\quad m\neq n, \quad (13)$$

$$\frac{\partial\xi_n}{\partial x}=\frac{1}{2\xi_n}\int_0^H dz\left[f_n^2\frac{\partial}{\partial x}\left(\frac{k^2}{\rho}\right)-\xi_n^2f_n^2\frac{\partial}{\partial x}\left(\frac{1}{\rho}\right)+\frac{\partial\rho}{\partial x}\left(\frac{1}{\rho}\frac{\partial f_n}{\partial z}\right)^2\right], \quad (14)$$

derived assuming smooth variation of sound speed and density. Let $K_0(x,z)$ be a continuous function, and a continuously differentiable function $K_1(x,z,\epsilon)$ tends at $\epsilon\to0$ to a bounded function $K_2(x,z)$, which is continuously differentiable at $0<z<h_1$, $h_1<z<h_2,\ldots,h_N<z<H$. The function K_2 may be discontinuous at $z=h_j(x)$, $j=1,2,...,N$. Then it is straightforward to demonstrate that for continuously differentiable $h_j(x)$

$$\lim_{\epsilon\to0}\int_0^H dz\,K_0\frac{\partial x_1}{\partial x}K_1-\sum_{j=1}^{N}\frac{\partial h_j}{\partial x}K_0K_2\bigg|_{z=h_j-0}^{z=h_j+0}+\left(\int_0^{h_1}+\int_{h_1}^{h_2}+\cdots+\int_{h_N}^{H}\right)dzK_0\frac{\partial}{\partial x}K_2. \quad (15)$$

Taking into account the boundary conditions for local modes and applying the identity (15), we obtain from Eqs. (13) and (14)

$$B_{nm}=\frac{1}{\xi_n^2-\xi_m^2}\left[\sum_{j=1}^{N}\frac{\partial h_j}{\partial x}\left(\frac{1}{\rho}\frac{\partial f_n}{\partial z}\frac{\partial f_m}{\partial z}+\frac{k^2-\xi_m^2}{\rho}f_nf_m\right)\bigg|_{z=h_j-0}^{z=h_j+0}-\left(\frac{1}{\rho}\frac{\partial\rho}{\partial x}\frac{\partial f_n}{\partial z},\frac{\partial f_m}{\partial z}\right)-\left(\left(\frac{\partial k^2}{\partial x}+\frac{\xi_m^2-k^2}{\rho}\frac{\partial\rho}{\partial x}\right)f_n,f_m\right)\right],\quad n\neq m, \quad (16)$$

$$\frac{\partial \xi_n}{\partial x} = -\frac{1}{2\xi_n}\sum_{j=1}^{N}\frac{\partial h_j}{\partial x}\left[\frac{1}{\rho}\left(\frac{\partial f_n}{\partial z}\right)^2 + \frac{k^2-\xi_n^2}{\rho}f_n^2\right]\Bigg|_{z=h_j-0}^{z=h_j+0}$$
$$+\frac{1}{2\xi_n}\left(\frac{1}{\rho}\frac{\partial \rho}{\partial x}\frac{\partial f_n}{\partial z}, \frac{\partial f_n}{\partial z}\right)$$
$$+\frac{1}{2\xi_n}\left(\left[\frac{\partial k^2}{\partial x}+\frac{\xi_n^2-k^2}{\rho}\frac{\partial \rho}{\partial x}\right]f_n, f_n\right). \quad (17)$$

Note that no limitation has been imposed on the contrast of fluid parameters across the interfaces. Together with the relation

$$B_{nn} = \left(\frac{f_n}{2\rho}\frac{\partial \rho}{\partial x}, f_n\right) + \sum_{j=1}^{N}\frac{\partial h_j}{\partial x}\frac{f_n^2}{2\rho}\Bigg|_{z=h_j-0}^{z=h_j+0}, \quad (18)$$

which follows immediately from integrating by parts the definition of B_{nm}, Eqs. (12), (16), and (17) fully determine the mode-coupling coefficients in a general range-dependent acoustic waveguide in terms of environmental gradients, including slopes of the interfaces. These results agree with those obtained by a somewhat lengthier derivation in Ref. 1 [Eqs. (29)–(31) OG] without relying on the passage to a limit.[9]

IV. CONCLUSION

In summary, McDaniel's criticism of our article[1] is unjustified. The statement that mode-coupling coefficients are symmetric for an arbitrary definition of mode amplitudes, which McDaniel[7] disproves, was neither used nor made in our article.[1] The derivation of mode-coupling equations[1] is mathematically consistent and applies to waveguides with sloping boundaries and interfaces and arbitrary smooth dependencies of sound speed and density on the coordinates in layers between the interfaces. The results obtained[1] are corroborated by alternative, independent derivations. The mode-coupling equations derived in Ref. 1 are an exact reformulation of the reduced wave equation and boundary conditions for general horizontally inhomogeneous acoustic waveguides.

We feel compelled to use this opportunity to make a brief note on the history of the coupled-mode theory. McDaniel[7] traces back the origin of the coupled-mode approach to the work of Pierce[10] and Milder.[11] To our knowledge, mode-coupling theory was first introduced to acoustics by Stevenson in his 1951 paper[12] on horns with variable cross sections. The theory was developed and extended to various types of acoustic waveguides by Katsenelenbaum[13] and Shevchenko[14] in 1961. Both Stevenson and Katsenelenbaum built on their previous work on guided propagation of electromagnetic waves. Katsenelenbaum[13] found coupling coefficients in terms of parameters of the waveguide rather than range derivatives of shape functions of local modes. Shevchenko[14] was the first to consider the effects of density variation on mode coupling and, in particular, calculated the contribution of a sloping interface of fluids of distinct density to coupling coefficients.

Without doubt, the JASA publications by Milder[11] and especially by Pierce[10] were seminal contributions that had a lasting effect on applying modal concepts to modeling and understanding sound propagation in horizontally inhomogeneous ocean and atmosphere. The strength of the analyses by Pierce[10] and Milder[11] was not in treating boundaries and interfaces. In fairness to the pioneers of the mode-coupling theory in acoustics, it should be noted that Stevenson,[12] Katsenelenbaum,[13] and Shevchenko[14] had complete understanding of and fully accounted for the problems in convergence of modal expansions which arise when local modes obey different boundary conditions than the entire acoustic field. Such an understanding was missing from Pierce,[10] Milder,[11] Rutherford and Hawker,[15] and much of the subsequent work in the 1970s–1990s, which led to the misconceptions and contradictions later addressed in Godin.[1] When discussing acoustic mode coupling at sloping boundaries and interfaces, the early insights by Stevenson,[12] Katsenelenbaum,[13] and Shevchenko[14] should not be overlooked.

ACKNOWLEDGMENTS

This work was supported in part by the U.S. Office of Naval Research through Contract No. 00014-01-F-0317. The author is grateful to Dr. S. T. McDaniel for the cause to elaborate on important issues that, in hindsight, should have been discussed in more detail in our original article[1] and to an anonymous reviewer for helpful comments.

[1] O. A. Godin, "A note on differential equations of coupled-mode propagation in fluids," J. Acoust. Soc. Am. **103**, 159–168 (1998).
[2] L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media. 2: Point Sources and Bounded Beams*, Springer Ser. Wave Phenom., Vol. 10 (Springer-Verlag, Berlin, 1992), Sec. 7.1.
[3] C.-S. Chiu, J. H. Miller, and J. F. Lynch, "Forward coupled-mode propagation modeling for coastal acoustic tomography," J. Acoust. Soc. Am. **99**, 793–802 (1996).
[4] R. I. Odom, M. Park, J. A. Mercer, R. S. Crosson, and P. Paik, "Effects of transverse isotropy on modes and mode coupling in shallow water," J. Acoust. Soc. Am. **100**, 2079–2092 (1996).
[5] J. A. Fawcett, "A derivation of the differential equations of coupled mode propagation," J. Acoust. Soc. Am. **92**, 290–295 (1992).
[6] A. T. Abawi, "An energy-conserving one-way coupled mode propagation model," J. Acoust. Soc. Am. **111**, 160–167 (2002).
[7] S. T. McDaniel, "On the application of the reciprocity principle to coupled-mode propagation," J. Acoust. Soc. Am. **114** , 3020–3023 (2003).
[8] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, New York, 1994), Sec. 5.9.
[9] Misprints (missing factors $1/2\xi_n$ in front of second and third terms on the right-hand side) in Eq. (29) OG are corrected in Eq. (17). Also, there is a misprint in Eq. (B3) OG, in which there should be $\nu=\pi(N+\frac{1}{2})/\alpha$ rather than $\nu=(N+\frac{1}{2})/\alpha$.
[10] A. D. Pierce, "Extension of the method of normal modes to sound propagation in an almost-stratified medium," J. Acoust. Soc. Am. **37**, 19–27 (1965).
[11] D. M. Milder, "Ray and wave invariants for SOFAR channel propagation," J. Acoust. Soc. Am. **46**, 1259–1263 (1969).
[12] A. F. Stevenson, "Exact and approximate equations for waves propagating in acoustic horns," J. Appl. Phys. **22**, 1461–1463 (1951).
[13] B. Z. Katsenelenbaum, "On the theory of irregular acoustic waveguides," Sov. Phys. Acoust. **7**, 159–164 (1961).
[14] V. V. Shevchenko, "Irregular acoustic waveguides," Sov. Phys. Acoust. **7**, 392–397 (1961).
[15] S. R. Rutherford and K. E. Hawker, "Consistent coupled mode theory of sound propagation for a class of nonseparable problems," J. Acoust. Soc. Am. **70**, 554–564 (1981).

On the application of the reciprocity principle to coupled-mode propagation (L)

Suzanne T. McDaniel[a)]
4035 Luxury Lane, Bremerton, Washington 98311

(Received 8 May 2002; revised 30 August 2002; accepted 25 August 2003)

The reciprocity principle is applied to the derivation of coefficients describing the coupling of acoustic normal modes in waveguides characterized by range-dependent density gradients and discontinuities. The requirement of reciprocity alone is found insufficient to determine the mode-coupling coefficients. The stronger requirement of continuity of pressure and particle velocity at horizontal discontinuities of the waveguide is shown to provide coupling coefficients in agreement with those of Fawcett [J. Acoust. Soc. Am. **92**, 290–295 (1992)]. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1619980]

PACS numbers: 43.30.Bp, 43.20.Bi, 43.20.Mv [WLS] Pages: 3020–3023

I. INTRODUCTION

Coupled-mode theory is useful for studying acoustic propagation in range-dependent ocean environments.[1,2] Pierce[3] derived equations governing the coupling of normal modes in fluid media by expressing the acoustic field as a sum of local solutions to the depth eigenvalue problem. Milder[4] obtained relationships between Pierce's coupling coefficients and sound-speed gradients, thus providing a method of readily implementing the theory. Graves *et al.*[5] subsequently extended the theory to treat scattering from irregular pressure-release boundaries. Efforts[6] have also been made to extend the theory to treat range-dependent density discontinuities. Maupin[7] resolved this difficulty for propagation in layered fluid-elastic waveguides. Abawi[8] discusses and applies the methods used to treat mode coupling in the geophysical literature. Fawcett[1] derives mode-coupling equations for layered fluid waveguides and notes that Pierce's derivation, which is based on term-by-term differentiation of the modal sum, is not valid at sloping interfaces between fluids of differing density.

Fawcett provides a heuristic development of the mode-coupling equations for the case of range-dependent density discontinuities, and outlines a lengthy more rigorous derivation. Godin[9] presents a derivation of the mode-coupling coefficients for a two-dimensional wave-guide based on the reciprocity principle. His analysis is also lengthy and is based on an assumption that cannot be justified at the outset of the derivation.

In this study, it is shown that the reciprocity principle alone is not sufficient to completely determine the mode-coupling coefficients. By applying the constraints of continuous pressure and particle velocity to the limiting case of weak horizontal discontinuities in a waveguide, mode-coupling coefficients consistent with the findings of Fawcett are obtained. This analysis provides a concise derivation of coupling equations in the presence of range-dependent density gradients and discontinuities.

Section II of this paper presents a brief review of coupled-mode theory. In Sec. III, Godin's application of reciprocity to mode-coupling is analyzed, and a direct application of the reciprocity principle is presented. In Sec. IV, mode-coupling equations are derived by considering propagation in a discontinuous waveguide. Section V summarizes the findings of this study.

II. COUPLED-MODE THEORY

Consider propagation in a range-dependent two-dimensional waveguide in which the pressure $p(x,z)$ obeys the wave equation

$$\rho(z;x)\nabla_2\cdot\left[\frac{\nabla_2 p(x,z)}{\rho(z;x)}\right]+k^2(z;x)p(x,z)=0, \tag{1}$$

in each layer of the waveguide, with $\nabla_2=\hat{\mathbf{x}}\partial/\partial x+\hat{\mathbf{z}}\partial/\partial z$. A solution to (1) may be obtained as an expansion in terms of range-dependent modes

$$p(x,z)=\sum_m F_m(x)f_m(z;x), \tag{2}$$

where the depth-dependent functions $f_m(z;x)$ are solutions of

$$\frac{\partial}{\partial z}\left[\frac{\partial f_m(z;x)/\partial z}{\rho(z;x)}\right]+\frac{k^2(z;x)-\kappa_m^2(x)}{\rho(z;x)}f_m(z;x)=0, \tag{3}$$

subject to the continuity of $f_m(z;x)$ and $\rho^{-1}(z;x)\partial f_m(z;x)/\partial z$ at interfaces. With an appropriate choice of boundary conditions at $z=0$ and $z=L$, and $k(z;x)$ real, the $f_m(z;x)$ form an orthonormal set of functions obeying

$$\int_0^L\frac{dz}{\rho(z;x)}f_m(z;x)f_n(z;x)=\delta_{m-n}, \tag{4}$$

where $\int_0^L dz$ denotes the sum of integrals over each layer of the waveguide $\int_0^{Z_1^-}+\cdots+\int_{Z_j^+}^L dz$. The $f_m(z;x)$ also obey the completeness relation

$$\sum_n f_n(z;x)f_n(z';x)=\rho(z;x)\,\delta(z'-z). \tag{5}$$

a)Electronic mail: ohmcdanl@pacific.telebyte.com

In the absence of density gradients and discontinuities, Pierce[3] finds that the modal amplitudes $F_n(x)$ obey

$$\frac{d^2F_n}{dx^2}+\kappa_n^2F_n=\sum_m\left[\alpha_{nm}F_m+\beta_{nm}\frac{dF_m}{dx}\right]. \tag{6}$$

III. APPLICATION OF THE RECIPROCITY PRINCIPLE

A. Godin's approach

To obtain the mode-coupling coefficients α_{nm} and β_{nm} of (6) for the case of range-dependent density variations, Godin introduces two functions $F_n^+(x)$ and $F_n^-(x)$ corresponding to the change of variables

$$F_n=F_n^++F_n^-, \tag{7}$$

$$\frac{dF_n}{dx}=i\kappa_n(F_n^+-F_n^-)-\sum_m B_{nm}(F_m^++F_m^-), \tag{8}$$

where

$$B_{nm}=\int_0^L\frac{dz}{\rho}f_n\frac{\partial f_m}{\partial x}. \tag{9}$$

In range-independent segments of the waveguide, $F_n^+(x)$ is the amplitude of a mode propagating in the positive x direction, and $F_n^-(x)$ that of a mode propagating in the negative x direction.

Godin then assumes that coupling coefficients $b_{nm}^{\pm}(x)$ exist such that $F_n^+(x)$ and $F_n^-(x)$ obey the coupled differential equations

$$\frac{dF_n^+}{dx}-i\kappa_nF_n^+=\sum_m\left[b_{nm}^+F_m^++b_{nm}^-F_m^-\right], \tag{10}$$

and

$$\frac{dF_n^-}{dx}+i\kappa_nF_n^-=\sum_m\left[b_{nm}^+F_m^-+b_{nm}^-F_m^+\right]. \tag{11}$$

Godin argues that the symmetrical coupling assumed in (10) and (11) is required for the invariance of the pair of equations under a reversal of the x axis. To see that this is not true, consider the simple change of variables

$$F_n=F_n^++F_n^-, \tag{12}$$

$$\frac{dF_n}{dx}=i\kappa_n(F_n^+-F_n^-). \tag{13}$$

Differentiating (12) with respect to x and equating the result to (13) yields

$$\frac{dF_n^+}{dx}+\frac{dF_n^-}{dx}=i\kappa_n(F_n^+-F_n^-). \tag{14}$$

A second relationship may be obtained by differentiating (13) with respect to x, and substituting the result, and (12) and (13), into (6):

$$\frac{dF_n^+}{dx}-\frac{dF_n^-}{dx}=i\kappa_n(F_n^++F_n^-)-\frac{1}{\kappa_n}(F_n^+-F_n^-)\frac{d\kappa_n}{dx}+(\kappa_n)^{-1}\sum_m\left[\kappa_m\beta_{nm}(F_m^+-F_m^-)-i\alpha_{nm}(F_m^++F_m^-)\right]. \tag{15}$$

Equations (14) and (15) may be solved simultaneously for $dF_n^{\pm}/dx$,

$$\frac{dF_n^+}{dx}-i\kappa_nF_n^+=\frac{1}{2\kappa_n}\sum_m\left[(c_{nm}-i\alpha_{nm})F_m^++(-c_{nm}-i\alpha_{nm})F_m^-\right], \tag{16}$$

$$\frac{dF_n^-}{dx}+i\kappa_nF_n^-=\frac{1}{2\kappa_n}\sum_m\left[(c_{nm}+i\alpha_{nm})F_m^-+(-c_{nm}+i\alpha_{nm})F_m^+\right], \tag{17}$$

where $c_{nm}=\kappa_m\beta_{nm}-\delta_{n-m}d\kappa_n/dx$. It is evident that, for (16) and (17) to display the same symmetry as (10) and (11), α_{nm} must vanish.

Invariance of the form of (6), with a reversal of the x axis, requires that $\beta_{nm}\rightarrow-\beta_{nm}$ and that α_{nm} remain invariant under this transformation. Noting that $d\kappa_n/dx\rightarrow-d\kappa_n/dx$ under this transformation, it follows that (16) and (17) are interchanged upon reversing the x axis: The pair of equations is independent of the choice of the coordinate system, as a consequence of the definition of the $F_n^{\pm}$ and the invariance of (6).

If the same procedure used to obtain (16) and (17) is performed for Godin's change of variables, the result obtained is of considerably greater complexity, but is also independent of the choice of the coordinate system. In this case, to obtain the symmetry of (10) and (11), the assumption that

$$\alpha_{nm}=\sum_j(\beta_{nj}+B_{nj})B_{jm}-\frac{dB_{nm}}{dx} \tag{18}$$

is required. Note that this expression is equivalent to (21). At the outset of the derivation, α_{nm} and β_{nm} are unknown, so that the *a priori* assumption of symmetrical coupling in (10) and (11) cannot be justified.

Godin next proceeds to impose reciprocity,

$$\frac{d}{dx}\int_0^L\frac{dz}{\rho}\left(p_1\frac{\partial p_2}{\partial x}-p_2\frac{\partial p_1}{\partial x}\right)=0, \tag{19}$$

on (10) and (11), where p_1 and p_2 are any two acoustic fields. As noted by Godin, with the choice $p_2=p_1^*$, (19) expresses conservation of energy. By requiring that (7), (8), (10), and (11) be consistent with the reciprocity principle, Godin represents the coupling coefficients $b_{nm}^{\pm}(x)$ as function of $d\kappa_n/dx$ and B_{nm} of (9):

$$b_{nm}^{\pm}=\mp\frac{\delta_{n-m}}{2\kappa_n}\frac{d\kappa_n}{dx}\pm\frac{\kappa_m}{2\kappa_n}B_{mn}-\frac{1}{2}B_{nm}. \tag{20}$$

After algebraic manipulation of (10), (11), and (20), the coefficients α_{nm} and β_{nm} in (6) may be identified as

$$\alpha_{nm}=\sum_j B_{jn}B_{jm}-\frac{dB_{nm}}{dx}, \tag{21}$$

$$\beta_{nm}=B_{mn}-B_{nm}. \tag{22}$$

On correcting the sign of the interfacial terms obtained by Fawcett,[1] (21) and (22) are in agreement with his results.

B. Direct application of reciprocity

Assuming the validity of (6), the reciprocity requirement of (19) may be directly applied to that equation. Following the procedure used by Godin, let

$$p_1(x,z)=\sum_n F_n f_n, \quad p_2(x,z)=\sum_m E_m f_m. \tag{23}$$

By first substituting (23) into (19), and then using (6), the mode-coupling coefficients are found to obey

$$\sum_{n,m}\left[(F_nE_m-F_mE_n)(\alpha_{nm}-U_{nm})+\left(F_n\frac{dE_m}{dx}-E_n\frac{dF_m}{dx}\right)\times(\beta_{nm}-V_{nm})\right]=0, \tag{24}$$

upon interchanging the order of summation. For a single interface $z=Z(x)$ between fluids of different density, the coefficient U_{nm} and V_{nm} in (24) are given by

$$U_{nm}=\int_0^L\frac{dz}{\rho}f_n\left(-\frac{\partial^2 f_m}{\partial x^2}+\frac{1}{\rho}\frac{\partial f_m}{\partial x}\frac{\partial\rho}{\partial x}\right)+\frac{dZ}{dx}\left(\frac{f_n}{\rho}\frac{\partial f_m}{\partial x}\right)_{Z^-}^{Z^+}, \tag{25}$$

$$V_{nm}=\int_0^L\frac{dz}{\rho}f_n\left(-2\frac{\partial f_m}{\partial x}+\frac{f_m}{\rho}\frac{\partial\rho}{\partial x}\right)+\frac{dZ}{dx}\left(\frac{f_nf_m}{\rho}\right)_{Z^-}^{Z^+}. \tag{26}$$

The mode amplitudes F_n and E_m and their derivatives in (24) may be arbitrarily chosen. This equation must vanish irrespective of their choice, so that

$$\alpha_{nm}=U_{nm}+A_{nm}, \tag{27}$$

$$\beta_{nm}=V_{nm}, \tag{28}$$

where A_{nm} is an arbitrary symmetrical matrix. To obtain agreement with the results of Fawcett, the assumption that $A_{nm}=0$ is needed. Thus, one concludes that the reciprocity principle is insufficient to determine the modal coupling coefficients.

IV. CONTINUITY OF PRESSURE AND PARTICLE VELOCITY

Consider propagation in a waveguide that has horizontal discontinuities of sound speed and density at x, but is otherwise independent of range, as depicted in Fig. 1. The continuity of pressure and particle velocity at x requires

$$\sum_n P_n^{(1)}(x)f_n^{(1)}=\sum_n P_n^{(2)}(x)f_n^{(2)}, \tag{29}$$

$$\sum_n \kappa_n^{(1)}Q_n^{(1)}(x)f_n^{(1)}/\rho^{(1)}=\sum_n \kappa_n^{(2)}Q_n^{(2)}(x)f_n^{(2)}/\rho^{(2)}, \tag{30}$$

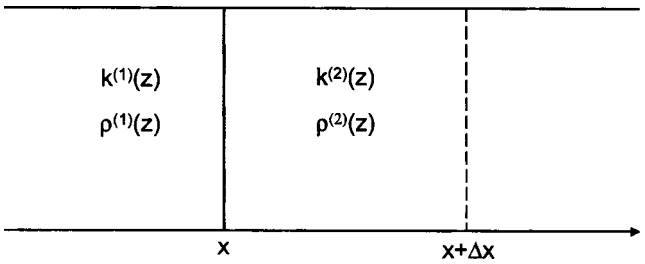


FIG. 1. Waveguide assumed for continuity considerations.

where $P_n=F_n^+ + F_n^-$, $Q_n=F_n^+ - F_n^-$, and the superscripts indicate evaluation of the parameters in regions 1 and 2 of Fig. 1.

Using (4), (29) and (30) may be solved to find

$$P_m^{(2)}(x)=\sum_n U_{mn}^{(2,1)}P_n^{(1)}(x), \tag{31}$$

$$Q_m^{(2)}(x)=\sum_n V_{mn}^{(2,1)}Q_n^{(1)}(x), \tag{32}$$

where it is useful to express the matrices $U_{mn}^{(2,1)}$ and $V_{mn}^{(2,1)}$ in the form

$$U_{mn}^{(2,1)}=\delta_{m-n}-\int_0^L\frac{dz}{\rho^{(2)}}f_m^{(2)}(f_n^{(2)}-f_n^{(1)}), \tag{33}$$

$$V_{mn}^{(2,1)}=\delta_{m-n}+\int_0^L\frac{dz}{\rho^{(1)}}f_n^{(1)}\left(\frac{f_m^{(2)}}{\kappa_m^{(2)}}-\frac{f_m^{(1)}}{\kappa_m^{(1)}}\right)\kappa_n^{(1)}. \tag{34}$$

As Δx vanishes

$$P_n^{(2)}(x+\Delta x)\approx P_n^{(2)}(x)+i\kappa_n^{(2)}\Delta x Q_n^{(2)}(x), \tag{35}$$

$$Q_n^{(2)}(x+\Delta x)\approx Q_n^{(2)}(x)+i\kappa_n^{(2)}\Delta x P_n^{(2)}(x). \tag{36}$$

Substituting (31) and (32) into (35) and (36) yields

$$\frac{P_m^{(2)}(x+\Delta x)-P_m^{(1)}(x)}{\Delta x}-i\kappa_m^{(2)}Q_m^{(2)}(x)\approx\sum_n\frac{U_{mn}^{(2,1)}-\delta_{m-n}}{\Delta x}P_n^{(1)}(x), \tag{37}$$

$$\frac{Q_m^{(2)}(x+\Delta x)-Q_m^{(1)}(x)}{\Delta x}-i\kappa_m^{(2)}P_m^{(2)}(x)\approx\sum_n\frac{V_{mn}^{(2,1)}-\delta_{m-n}}{\Delta x}Q_n^{(1)}(x). \tag{38}$$

In the limit as Δx vanishes, and $F_n^{\pm}$, κ_n, and f_n become continuous, these expressions reduce to the sum and difference of (10) and (11) with the coefficients $b_{nm}^{\pm}$ given by (20). Thus, the mode-coupling coefficients of (21) and (22) are recovered, and agreement with Fawcett's findings is obtained.

V. SUMMARY

Godin's derivation of mode-coupling equations using reciprocity has been shown to rely on an assumption that is *a priori* unjustifiable. The reciprocity principle was then di-

rectly applied to the mode-coupling problem with the finding that reciprocity alone is not sufficient to completely determine the coupling coefficients. Namely, the symmetrical elements of the coupling matrix α_{nm} were undetermined.

Mode-coupling equations in agreement with those of Fawcett were derived by considering propagation in a waveguide having sound-speed and density discontinuities. Requiring the stronger constraints of continuity of pressure and particle velocity, in the limiting case of weak horizontal discontinuities, was sufficient to fully determine the mode-coupling coefficients. The derivation presented is simple, concise, and complements that of Fawcett.

[1] J. A. Fawcett, "A derivation of the differential equations of coupled-mode propagation," J. Acoust. Soc. Am. **92**, 290–295 (1992).
[2] B. E. McDonald, "Bathymetric and volumetric contributions to ocean acoustic mode coupling," J. Acoust. Soc. Am. **100**, 219–224 (1996).
[3] A. D. Pierce, "Extension of the method of normal modes to sound propagation in an almost stratified medium," J. Acoust. Soc. Am. **37**, 19–27 (1965).
[4] D. M. Milder, "Ray and wave invariants for SOFAR channel propagation," J. Acoust. Soc. Am. **46**, 1259–1263 (1969).
[5] R. D. Graves, A. Nagl, H. Überall, and G. L. Zarur, "Range-dependent normal modes in underwater sound propagation: Application to the wedge-shaped ocean," J. Acoust. Soc. Am. **58**, 1171–1177 (1975).
[6] S. R. Rutherford and K. E. Hawker, "Consistent coupled mode theory of sound propagation for a class of nonseparable problems," J. Acoust. Soc. Am. **70**, 554–564 (1981).
[7] V. Maupin, "Surface waves across 2-D structures: A method based on coupled local modes," Geophys. J. **93**, 173–185 (1988).
[8] A. T. Abawi, "An energy-conserving one-way coupled mode propagation model," J. Acoust. Soc. Am. **111**, 160–167 (2002).
[9] O. A. Godin, "A note on differential equations of coupled-mode propagation in fluids," J. Acoust. Soc. Am. **103**, 159–168 (1998).

Relative importance of temporal envelope and fine structure in lexical-tone perception (L)

Li Xu[a)]
School of Hearing, Speech and Language Sciences, Ohio University, Grover Center, Athens, Ohio 45701, and Kresge Hearing Research Institute, Department of Otolaryngology, University of Michigan, Ann Arbor, Michigan 48109

Bryan E. Pfingst
Kresge Hearing Research Institute, Department of Otolaryngology, University of Michigan, Ann Arbor, Michigan 48109

(Received 21 April 2003; revised 5 September 2003; accepted 15 September 2003)

The relative importance of temporal envelope and fine structure in speech and music perception was investigated by Smith *et al.* [Nature (London) **416**, 87–90 (2002)] using ''auditory chimera'' in which the envelope from one sound was paired with the fine structure of another. Smith *et al.* found that, when 4 to 16 frequency bands were used, recognition of English speech was dominated by the envelope, whereas recognition of melody was dominated by the fine structure. In the present study, Mandarin Chinese monosyllables were divided into 4, 8, or 16 frequency bands and the fine structure and envelope of one tone pattern were exchanged with those of another tone pattern of the same monosyllable. Five normal-hearing native Mandarin Chinese speakers completed a four-alternative forced-choice tone-identification task. In the vast majority of trials, subjects based their identification of the monosyllables on the fine structure rather than the envelope. Thus, the relative importance of envelope and fine structure for lexical-tone perception resembled that for melody recognition rather than that for English speech recognition. Delivering fine-structure information in cochlear implant stimulation could be particularly beneficial for lexical-tone perception. [DOI: 10.1121/1.1623786]

PACS numbers: 43.66.Ts, 43.66.Hg, 43.71.Hw [PFA] Pages: 3024–3027

I. INTRODUCTION

The time waveform of signals can be described mathematically as slowly varying envelope (modulation) and rapidly varying fine-time structure (carrier). In a recent report, Smith *et al.* (2002) constructed a set of acoustic stimuli, each having the envelope of one sound and the fine structure of another. They termed the hybrid sounds ''auditory chimera.'' The operation was accomplished with a Hilbert transform that allowed the investigators to extract the envelope and the fine structure of the sounds. For example, a speech–speech chimera can be synthesized so that it contains speech information from one sentence in the envelope and the speech information from another sentence in the fine structure. The auditory chimeras provide a way to study the relative importance of the envelope and the fine structure in speech perception and pitch perception. Knowledge of the relative importance of the envelope and the fine structure might help in determining what stimulus features should be delivered to auditory prostheses and therefore might guide the design of future auditory prostheses.

Smith *et al.* (2002) found that in speech–speech chimeras with 4 to 16 frequency bands, the words represented in the envelope were identified correctly much more frequently than the words represented in the fine structure. This result was consistent with the fact that English-speaking cochlear-implant users have achieved high speech recognition scores using the continuous interleaved sampling (CIS) stimulation strategy (Wilson *et al.*, 1991), in which the speech envelope is delivered to a limited number of frequency channels and the fine structure is absent. Also, Shannon *et al.* (1995) and others have shown that normal-hearing subjects can achieve good word recognition (>85% correct) using a four-band noise vocoder simulating the cochlear implant stimulation.

In contrast to the results with speech–speech chimeras, Smith *et al.* (2002) found that when melody–melody chimeras were synthesized to exchange the envelope and fine structure of two different melodies, subjects almost always reported hearing the melody represented by the fine structure, with up to 16 frequency bands. Data for number of frequency bands >16, should be interpreted with caution due to the artifacts caused by ringing of the very narrow digital filters, as pointed out by Zeng *et al.* (2003).

The question that we address in the present study is the relative importance of the envelope and the fine structure for lexical-tone perception. In tone languages, such as Mandarin Chinese, Cantonese, Thai, and Vietnamese, the tonality of syllables conveys lexical meaning. The number and the pattern of tones vary among different tone languages. In Mandarin Chinese, a language spoken by more people than any other language, there are four tone patterns (tone 1 through tone 4) defined by the contour of the fundamental frequency: (1) flat and high; (2) rising; (3) falling and then rising; and (4) falling (see Fig. 1 in Xu *et al.*, 2002). Typographically, these tone patterns are usually represented as ''¯,'' ''/,'' ''V,'' and ''\,'' respectively.

a)Electronic mail: xul@ohio.edu

Does the relative importance of the envelope and the fine structure for lexical-tone perception resemble that for English speech recognition because tone perception is an aspect of speech perception? Or, does it resemble that for music recognition because tone patterns and music both show variations in the fundamental frequency and harmonics? The hypothesis that we set forth for the present study was that the relative importance of the envelope and the fine structure for tone recognition resembles that for music recognition. We used 4-, 8-, and 16-frequency-band auditory chimera (Smith *et al.*, 2002) to test this hypothesis. In the present experiments, chimeric Mandarin syllables that had one tone pattern represented in the envelope and a different tone pattern represented in the fine structure were presented to native Mandarin-speaking subjects for identification. The range of 4 to 16 frequency bands was chosen (1) to provide a situation that represents the number of frequency channels that a current multichannel cochlear prosthesis can provide; (2) to avoid the digital filter ringing which can occur when a large number of frequency bands (e.g., >16) with very narrow filter bandwidth are used (Zeng *et al.*, 2003); and (3) to overcome the natural recovery of narrow-band envelopes by cochlear filtering when the number of bands is small (e.g., 1 or 2 bands) (Ghitza, 2001). The problem with using a small number of bands was shown in a counterintuitive result in signal processing of speech signals by Ghitza (2001). He first bandpass filtered a speech signal by a critical-band filter and then low-pass filtered the envelope at 16 Hz. When the speech signal with a smoothed envelope was passed through a critical-band filter again, he found that the output had a nonsmoothed envelope that was similar to that of the original signal (Ghitza, 2001). This model illustrates a potential problem that would result if one or two wide bands were used.

II. METHODS

A. Subjects

Five normal-hearing native-Mandarin-Chinese speakers (three females and two males) were recruited from the student and staff population at the University of Michigan. Subjects ranged from 36 to 41 years of age (37.6 ± 2.1, mean and s.d.). All subjects had received at least a college-level education in China. The pure-tone air-conduction thresholds for all subjects were $\leqslant$20 dB HL at octave frequencies between 250 and 8000 Hz. The use of human subjects in this study was reviewed and approved by the University of Michigan Medical School Institutional Review Board.

B. Original and chimeric speech materials

The original speech test materials consisted of ten Mandarin monosyllables. The *pinyin* (i.e., the phonemic spelling system for Mandarin Chinese) of these syllables were ''fu,'' ''ji,'' ''ma,'' ''qi,'' ''wan,'' ''xi,'' ''xian,'' ''yan,'' ''yang,'' and ''yi,'' and each syllable had four tone patterns. The resulting 40 combinations of consonant, vowel, and tone were real words in Chinese. A male and a female speaker were used to record the speech materials. Tokens in which the durations of the four tone patterns of each syllable were equal were selected (Xu *et al.*, 2002). All speech test materials were digitized at a sampling rate of 22 050 Hz and stored in a 16-bit format. These equal-duration speech materials were used in a previous tonal-speech perception study (Xu *et al.*, 2002) and were used in the present study to create chimeric speech materials as described below.

The chimeric stimuli were generated using the methods described by Smith *et al.* (2002). Briefly, two sounds, in this case a single monosyllable with two different tone patterns (e.g., tone 1 and tone 2 of ''ma''), were divided into a number of frequency bands (4, 8, or 16) whose center frequencies were chosen to reflect equal length along the basilar membrane (Greenwood, 1990). The overall frequency range for chimera synthesis was from 80 to 8820 Hz. The bandpass filters used in the present study were identical to the ones in Smith *et al.* (2002). In the case of 8 bands, the cutoff frequencies were 80, 205, 405, 724, 1236, 2055, 3366, 5463, and 8820 Hz, and the range of frequencies common to adjacent bands was about 30 Hz. Hilbert transforms were then used to extract the envelope and the fine structure in each band. Finally, the envelopes and fine structures of the two syllables with the same phonemes (e.g., ''ma 1'' and ''ma 2'') were exchanged to create two chimeric stimuli (e.g., one with envelope of ''ma 1'' and fine structure of ''ma 2'' and the other with envelope of ''ma 2'' and fine structure of ''ma 1''). In a similar fashion, a total of 12 chimeric stimuli were generated for the four tone patterns of each syllable. All signal processing was performed in MATLAB (Mathworks, Natick, MA).

C. Procedures

A custom graphical user interface (GUI) was created, using MATLAB, to present the stimuli and to collect the responses. The GUI displayed on an LCD monitor the typographical representations of the four tone patterns of a syllable (i.e., ''¯,'' ''/,'' ''V,'' and ''\'') and the four associated Chinese characters. The program presented to the loudspeaker one of the 12 chimeric stimuli for that particular syllable. The subject's task was to select, using a computer mouse, the response button associated with the tone pattern of the sound that he or she heard. After each response, the GUI would refresh the screen, display a new set of four Chinese characters with four tone patterns, and present the next acoustic stimulus. The test order of the number of bands (4, 8, or 16) was randomized. Each test contained 240 randomized stimulus presentations (2 voices$\times$10 syllables $\times$12 chimeras) and was administered five times. A total of 3600 responses (3 numbers of bands$\times$240 presentations $\times$5 times) was collected from each subject.

The stimuli were presented through a loudspeaker mounted 1 m away from the subject at 0° azimuth inside an Acoustic Systems (model RE2 242S) double-walled sound-attenuating booth. Stimulus level was roved randomly between 50 to 70 dB A in 5-dB steps.

III. RESULTS

There was no difference in tone recognition between male and female voices and therefore results reported below were pooled from results for both voices. Data obtained with

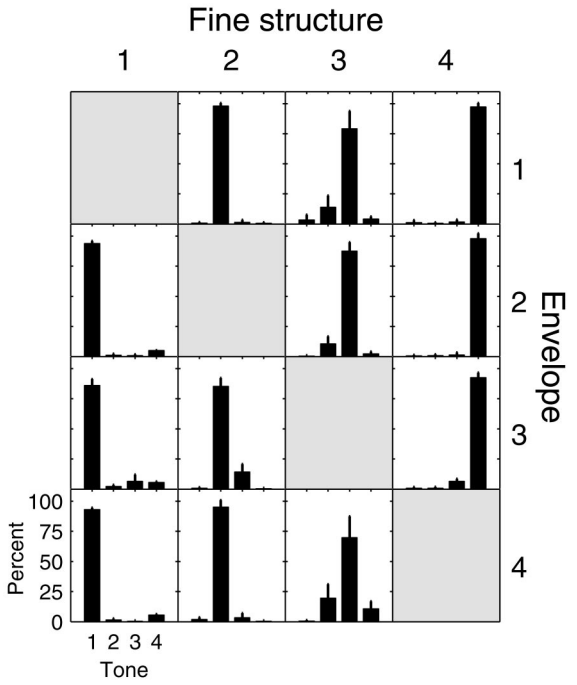


FIG. 1. Means and standard deviations (s.d.'s) of the percentages of the tone-recognition responses across the group of five subjects using 8-band auditory chimera. The 16 small panels represent 16 combinations of four fine structures (column-wise) and four envelopes (row-wise). When the fine structure and envelope were from the same tone patterns, the stimuli were the original sounds and were not tested for tone recognition, as plotted in gray. Results from the remaining 12 combinations of fine structures and envelopes are shown in a histogram format in which the means and s.d.'s of the percentages of the responses as tone 1, 2, 3, or 4 are represented by the bars and the whiskers from left to right, as labeled in the lower left panel.

8 frequency bands are shown in Fig. 1. Data for 4 and 16 bands were very similar to those of 8 bands and were thus omitted from the figure. In Fig. 1, there are 12 histograms representing 12 combinations of fine structure (column-wise) and envelope (row-wise). Each histogram plots the mean and standard deviation (s.d.) of the percentages of the tone-recognition responses across the group of five subjects. Each bar represents a percentage of the responses as tone 1, 2, 3, or 4 as indicated by the labels in the lower left corner. The subjects almost always reported the tone patterns that were consistent with the fine structures of the stimuli, regardless of the envelopes.

Figure 2 summarizes the tone-recognition performance using 4, 8, and 16 frequency bands. Tone-recognition responses that were consistent with the fine structure of the stimuli are shown with the black bars using the left ordinate. Over the range of numbers of bands tested, the vast majority of the responses of the subjects was consistent with the fine structure of the stimuli. On average, 90.8%, 89.5%, and 84.5% of the responses were consistent with the fine structure for 4, 8, and 16 bands, respectively. The differences among the percentages were small but statistically significant (ANOVA, $p<0.05$). Significant differences ($p<0.05$) were further found to be between 4 and 16 bands and between 8 and 16 bands in *post hoc* tests with Bonferroni correction. On the other hand, only 4.3%, 5.0%, and 8.9% of the responses were consistent with the envelope of the stimuli for

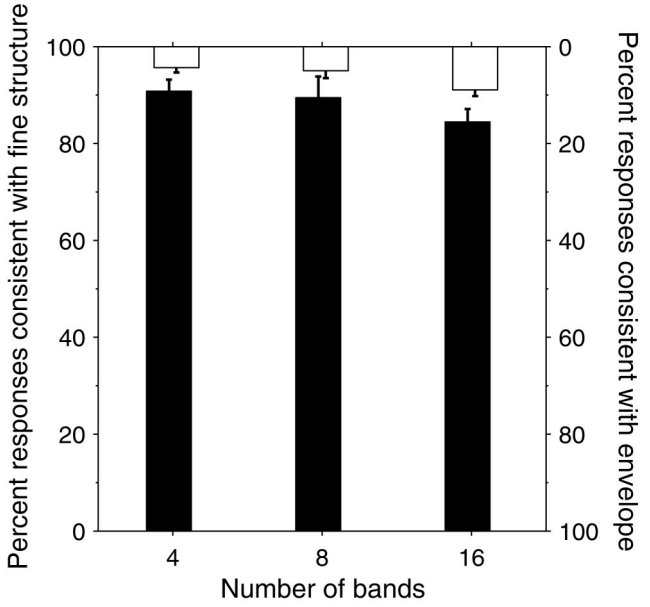


FIG. 2. Means and s.d.'s of tone-recognition performance for 4, 8, and 16 frequency bands across the group of five subjects. The percentage of responses that were consistent with the fine structure of the chimeric stimuli are shown with black bars using the left ordinate. The percentage of responses that were consistent with the envelope of the chimeric stimuli are shown with open bars using the right ordinate. The error bars indicate the s.d.'s.

4, 8, and 16 bands, respectively, as shown by the open bars using the right ordinate in Fig. 2. A small fraction of the responses was not consistent with either the envelope or the fine structure (Figs. 1 and 2). Such errors or confusions occurred at an average of 4.9%, 5.5%, and 6.6% of the responses for 4, 8, and 16 frequency bands, respectively (Fig. 2). Noticeably, the confusions between tone 2 and tone 3 occurred most often (Fig. 1).

IV. DISCUSSION AND CONCLUSIONS

Our results on lexical-tone recognition using auditory chimera were largely consistent with the results of Smith *et al.* (2002) on melody recognition for the range of numbers of frequency bands tested. Lexical-tone recognition depended on the fine structure but not on the envelope when the number of frequency bands was between 4 and 16. Lin (1988) studied tone recognition of 14 Mandarin-speaking subjects with synthesized Mandarin syllables in which the envelopes of tone patterns 1 through 4 were used to amplitude modulate the fine structure of syllables with tone 1. Those sounds could be considered equivalent to the chimeric sounds used in this study except that the number of frequency bands was fixed at 1 in that study. Lin's (1988) results showed that about 92% of the time the subjects reported that they heard tone 1, that is, the tone that contained the fine structure. Those data were comparable to our data with the number of frequency bands between 4 and 16 (Fig. 2). However, as pointed out by Zeng *et al.* (2003), when only one band is used, the ability of the cochlear filters to recover the narrow-band envelopes of the original signals (Ghitza, 2001) might hamper the interpretation of relative importance of the envelope and fine structure.

In the present study, syllables of equal duration were used to create chimeric sounds. It has been shown that syllable duration can be a strong cue for *isolated* syllables (Whalen and Xu, 1992; Xu *et al.*, 2002). However, syllable duration might not be a reliable cue for everyday speech (Xu *et al.*, 2002). Duanmu (2002) has summarized a few studies on duration and tones and has found that the differences in duration for Mandarin tones are rather small, all within 10% among different tones. The use of equal-duration syllables in the present study was a requirement of the chimera technique. In addition, it removed a potential cue for tone recognition, thus eliminating a confounding variable (duration) and allowing us to focus on the main issue of our study, which was lexical-tone perception.

Multichannel cochlear prostheses have achieved high levels of speech perception for English in patients with profound hearing loss (Skinner *et al.*, 1994). However, music perception is still generally poor in these patients (Gfeller *et al.*, 1997, 2002). Current cochlear prostheses can deliver up to 22 frequency channels. However, research data have suggested that the patients cannot functionally use more than 8 frequency channels (Fishman *et al.*, 1997; Fu *et al.*, 1998; Friesen *et al.*, 2001). The present study tested 4 to 16 frequency bands, which should have encompassed the number of functional channels that the majority of cochlear implant patients can use. Current speech processing strategies in the cochlear prostheses emphasize the presentation of envelope information. Our results confirm that fine structure is relatively more important for pitch perception than is envelope (Smith *et al.*, 2002). Therefore, modifying cochlear implant processors to deliver fine-structure information might improve the pitch perception of implant patients, thus facilitating music perception and lexical-tone perception.

One potential way to deliver fine-structure information might be to use analog stimulation or variable pulse rates. However, it is uncertain whether or not patients would be able to use the pitch information in the fine structure of the electrical signal. Some studies have indicated that subjects can use rate-pitch information only up to about 300 pps, or 300 Hz for sinusoids (Shannon, 1983; McKay *et al.*, 1994; Zeng, 2002), although others have reported that some subjects can discriminate rate-pitch changes up to 1000 pps (Townshend *et al.*, 1987; Pijl and Schwarz, 1995). A limited upper boundary of temporal code in electrical hearing would make it difficult for patients to utilize fine structure information. Future research is warranted to determine why some patients seem to have better rate-pitch perception than others and to identify ways to overcome the apparent 300-Hz boundary.

ACKNOWLEDGMENTS

We are grateful to Dr. Bertrand Delgutte for providing the software to create the auditory chimeras and for many useful discussions. Dr. Fan-Gang Zeng, an anonymous reviewer, and Associate Editor, Dr. Peter Assmann, provided useful comments on earlier versions of the manuscript. This research was supported by NIH Grants F32-DC00470 and RO1-DC03808.

Duanmu, S. (**2002**). *The Phonology of Standard Chinese* (Oxford University Press, Oxford).

Fishman, K. E., Shannon, R. V., and Slattery, W. H. (**1997**). ''Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor,'' J. Speech Lang. Hear. Res. **40**, 1201–1215.

Friesen, L., 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.

Fu, Q. J., Shannon, R. V., and Wang, X. (**1998**). ''Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing,'' J. Acoust. Soc. Am. **104**, 3586–3596.

Gfeller, K., Witt, S., Woodworth, G., Mehr, M. A., and Knutson, J. (**2002**). ''Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal,'' Ann. Otol. Rhinol. Laryngol. **111**, 349–356.

Gfeller, K., Woodworth, G., Robin, D., Witt, S., and Knutson, J. (**1997**). ''Perception of rhythmic and sequential pitch patterns by normally hearing adults and adult cochlear implant users,'' Ear Hear. **18**, 252–260.

Ghitza, O. (**2001**). ''On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception,'' J. Acoust. Soc. Am. **110**, 1628–1640.

Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Lin, M. C. (**1988**). ''The acoustical properties and perceptual characteristics of Mandarin tones,'' Zhongguo Yuwen **3**, 182–193.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1994**). ''Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees,'' J. Acoust. Soc. Am. **96**, 2664–2673.

Pijl, S., and Schwarz, D. W. F. (**1995**). ''Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through simple cochlear implant electrodes,'' J. Acoust. Soc. Am. **98**, 886–895.

Shannon, R. V. (**1983**). ''Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics,'' Hear. Res. **11**, 157–189.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.

Skinner, M. W., Clark, G. M., Whitford, L. A., Seligman, P. M., Staller, S. J., Shipp, D. B., Shallop, J. K., Everingham, C., Menapace, C. M., Arndt, P. L., Antogenelli, T., Brimacombe, J. A., Pijl, S., Daniels, P., George, C. R., McDermott, H. J., and Beirer, A. L. (**1994**). ''Evaluation of a new Spectral Peak coding strategy for the Nucleus 22 Channel Cochlear Implant System,'' Am. J. Otol. **15** (Suppl. 2), 15–27.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). ''Chimaeric sounds reveal dichotomies in auditory perception,'' Nature (London) **416**, 87–90.

Townshend, B., Cotter, N., Van Compernolle, D., and White, R. L. (**1987**). ''Pitch perception by cochlear implant subjects,'' J. Acoust. Soc. Am. **82**, 106–115.

Whalen, D. H., and Xu, Y. (**1992**). ''Information for Mandarin tones in the amplitude contour and in brief segments,'' Phonetica **49**, 25–47.

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, L., Tsai, Y., and Pfingst, B. E. (**2002**). ''Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses,'' J. Acoust. Soc. Am. **112**, 247–258.

Zeng, F.-G. (**2002**). ''Temporal pitch in electric hearing,'' Hear. Res. **174**, 101–106.

Zeng, F.-G., Nie, K.-B., Stickney, G., Liu, S., Rio, E. D., Kong, Y.-Y., and Chen, H.-B. (**2003**). ''Facts and artifacts in auditory chimaeras,'' Assoc. Res. Otolaryngol. Abstr. **26**, 213.

Are conical segments useful for vocal-tract simulation?[a] (L)

Hans Werner Strube
Drittes Physikalisches Institut, Universität Göttingen, Bürgerstrasse 42-44, 37073 Göttingen, Germany

(Received 11 July 2003; accepted for publication 15 September 2003)

A time-domain model for nonuniform tubes is considered which is of similar simplicity as the Kelly–Lochbaum model but based on conical segments. Using the bilinear transform or wave digital filters for representing the reflections at the segment boundaries, stability is proved and numerical instabilities can be overcome. The transfer functions are modeled better than with cylindrical segments but about equally well as with half-length cylinders, requiring about the same computational expense. Thus, in most cases there is no real advantage in using cones. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1623789]

PACS numbers: 43.70.Bk, 43.72.Ja [AL] Pages: 3028–3031

I. INTRODUCTION

Efficient time-domain simulations of the vocal tract or other nonuniform tubes with low damping and relatively hard walls are most often based on a one-dimensional chain of uniform segments. At the area steps between these, the forward- and backward-traveling waves are scattered (Kelly and Lochbaum, 1962). Although this representation is simple and fast, occasionally approximations by chains of conical segments also have been considered (Välimäki, 1995; Kob, 2002). Much more literature exists in the context of simulating wind instruments (Benade, 1988; Martínez *et al.*, 1988; Ducasse, 2002), where the models consist of few conical segments of given length, the sampling frequency is usually chosen high, and sophisticated interpolation methods have been invented to overcome a fixed relation between segment length and sampling frequency. In order to prevent instabilities, some approaches compute impulse responses of whole segments by complicated techniques and then convolve with them (Ducasse, 2002). (Since some of these involved approaches even use frequency-domain methods, it might appear preferable to represent all linear parts of the total system as a frequency-domain multiport with final conversion to the time domain rather than to use a time-domain description at the segmental level.)

Unlike instruments, a vocal-tract model usually consists, because of its variable shape, of many small segments of equal lengths with relatively low sampling frequency. Although interpolation methods have been considered, too (Välimäki, 1995; Strube, 2000), we disregard them here for reasons of efficiency, likewise the above-mentioned complicated convolution techniques. Rather, a discretization is investigated that is of similar simplicity as the Kelly–Lochbaum model and can be well compared with the latter.

II. THEORETICAL BASIS

Although the theory is described in the above references, it will briefly be outlined here in a form suitable for our purposes. The following approximative assumptions are made: one-dimensional propagation of plane waves, hard-walled tube; for reasons of clarity, no damping (can be added approximatively); area function $A(x)$ (time variation neglected here). Then with sound pressure p, volume velocity $q=vA$, average density ρ_0, speed of sound c, and acoustic characteristic impedance $Z=\rho_0 c/A$:

$$-\partial_x p=(Z/c)\partial_t q, \quad -\partial_x q=(1/Zc)\partial_t p. \tag{1}$$

Rather than p and q, pressure waves can be used:

$$p^{\pm}=(p\pm Zq)/2, \quad p=p^{+}+p^{-}, \quad q=(p^{+}-p^{-})/Z. \tag{2}$$

In uniform segments, these propagate without scattering; at the boundaries they are scattered (segment numbers $n-1$ and n):

$$p_{n-1}^{-}=r_n p_{n-1}^{+}+(1-r_n)p_n^{-}, \tag{3}$$

$$p_n^{+}=-r_n p_n^{-}+(1+r_n)p_{n-1}^{+}, \tag{4}$$

$$r_n=(Z_n-Z_{n-1})/(Z_n+Z_{n-1}). \tag{5}$$

For more general segment shapes, ''reduced'' field quantities are advantageous:

$$\psi=p/\sqrt{Z}, \quad \phi=q\sqrt{Z}. \tag{6}$$

With

$$W=Z'/2Z=-\sqrt{A}'/\sqrt{A} \tag{7}$$

(prime denotes ∂_x), Eq. (1) becomes

$$-\partial_x\psi=(1/c)\partial_t\phi+W\psi, \tag{8}$$

$$-\partial_x\phi=(1/c)\partial_t\psi-W\phi. \tag{9}$$

A further transformation

$$\chi=\phi+Wc\int\psi dt, \tag{10}$$

turns Eqs. (8) and (9) into

$$-\partial_x\psi=(1/c)\partial_t\chi, \tag{11}$$

$$-\partial_x\chi=(1/c)\partial_t\psi+Vc\int\psi dt, \tag{12}$$

[a]Partially presented as German conference paper ''Simulation inhomogener Rohre mit Segmentketten ohne Querschnittssprünge,'' in *Fortschritte der Akustik—DAGA'03*, edited by M. Vorländer (DEGA, Oldenburg, 2003), pp. 784–785.

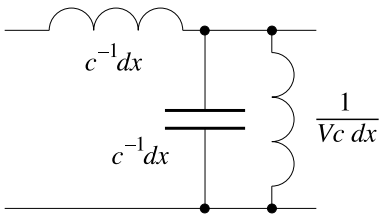


FIG. 1. Infinitesimal transmission-line element for Eqs. (11), (12).

where

$$V = W^2 - W' = \sqrt{A}''/\sqrt{A}. \tag{13}$$

In electrical analogy (ψ=voltage, χ=current), the last term in Eq. (12) can be interpreted as a load with shunt inductances (more exactly, with an inverse-inductance density), Fig. 1. From Eqs. (11) and (12), a Klein–Gordon equation for ψ follows, with V as potential

$$c^{-2}\partial_t^2\psi - \partial_x^2\psi + V\psi = 0. \tag{14}$$

For constant V, this transmission line is formally uniform again, but dispersive if $V \neq 0$. However, we will only consider the nondispersive case $V=0$, which corresponds to a conical tube, $\sqrt{A} = a + bx$. [If desired, a hyperbolic tube can be described in the same way, interchanging $A \leftrightarrow A^{-1}$, $\psi \leftrightarrow \phi$.] Then, in this formal line, forward and backward waves travel without scattering, which are now given by

$$\psi^{\pm} = (\psi \pm \chi)/2, \tag{15}$$

in analogy to Eq. (2). Where two conical segments are connected continuously, but with different slopes, χ becomes discontinuous and V a delta function. In the electrical picture, this means a discrete shunt inductance $1/\Delta W c$, where ΔW is the change of W. Thus, a chain of cones becomes a uniform line with inductive shunts at the segment boundaries, cf. Benade (1988), van Walstijn and Campbell (2003). Where these inductances are negative, instabilities can locally occur, which should, however, compensate each other in the overall system.

The scattering equations between conical segments $n-1$ and n (with boundary area A_n, segment lengths D_{n-1}, D_n) are in the Laplace-frequency domain

$$\psi_{n-1}^{-} = R_n(s)\,\psi_{n-1}^{+} + (1 + R_n(s))\,\psi_n^{-}, \tag{16}$$

$$\psi_n^{+} = R_n(s)\,\psi_n^{-} + (1 + R_n(s))\,\psi_{n-1}^{+}; \tag{17}$$

$$R_n(s) = -\tau_n^{-1}/(\tau_n^{-1} + s), \tag{18}$$

$$\tau_n^{-1} = \frac{\sqrt{A_{n+1}/A_n} - 1}{2D_n/c} + \frac{\sqrt{A_{n-1}/A_n} - 1}{2D_{n-1}/c} \tag{19}$$

(τ_n is twice the shunt inductance). At the ends of the whole line, one has to transform between p, q and ψ, χ.

III. TIME DISCRETIZATION, STABILITY

For cylinder segments of length D the scattering equations (3), (4) can be discretized directly, if the sampling frequency is $f_s = Nc/2D$ ($N = 1,2,\ldots$). For conical segments, however, in Eqs. (16) and (17), $R_n(s)$ from Eq. (18) must be represented by a digital filter. This is a first-order recursive system with exponentially decaying or increasing impulse response. We use the bilinear z transform or, equivalently, a wave-digital-filter implementation (Fettweis, 1986; van Walstijn and Campbell, 2003) of the shunt inductance with a three-port parallel adaptor with two port impedances=1 (at the line) and one port impedance $= f_s\tau_n$ (at the inductance). If s_n is the state of the inductance, the scattering equations read [with all $D_n = D$ and $f_s = c/D$; τ_n^{-1} from Eq. (19)]

$$\psi_{n-1}^{-\,\mathrm{new}} = r_n\psi_{n-1}^{+} + (1 + r_n)\,\psi_n^{-} + 2r_n s_n, \tag{20}$$

$$\psi_n^{+\,\mathrm{new}} = r_n\psi_n^{-} + (1 + r_n)\,\psi_{n-1}^{+} + 2r_n s_n, \tag{21}$$

$$s_n^{\mathrm{new}} = (1 + r_n)(\psi_{n-1}^{+} + \psi_n^{-}) + (1 + 2r_n)s_n; \tag{22}$$

$$r_n = -\tau_n^{-1}/(\tau_n^{-1} + 2f_s). \tag{23}$$

The segment chain always contains negative inductances. In spite of this, stability is preserved, which can be shown as follows. Each conical segment, whose length is minimal ($D = c/2f_s$) without loss of generality, is embedded between two (if necessary, infinitesimal) cylinder segments. The elements of the scattering matrix of such a cone with area transition $A_1 \rightarrow A_2$ have, in the continuous-time case, the common denominator

$$(1 + s\tau_1)(1 + s\tau_2) - \exp(-s\tau), \tag{24}$$

$\tau = -\tau_2 - \tau_1 = 2D/c = 1/f_s$, $\tau_2/\tau_1 = -\sqrt{A_2/A_1}$. In the discrete-time case, the delay term $\exp(-s\tau)$ becomes z^{-1} directly; the first term, however, which is due to the shunt inductances, is transformed bilinearly: $s = 2f_s(z-1)/(z+1)$. Thus, we can continue using s but write the delay term as $z^{-1} = (2f_s - s)/(2f_s + s)$. Let $s\tau = a + ib$, a, b real. At a zero of Eq. (24), the imaginary part of (24) fulfills in the continuous-time and discrete-time case, respectively,

$$1 + 2a\tau_1\tau_2/\tau^2 = \begin{cases} e^{-a}(\sin b)/b, & \text{contin.,} \\ 4/((2+a)^2 + b^2), & \text{discrete.} \end{cases} \tag{25}$$

Hence, $a < 0$ follows, i.e., the poles of the scattering matrix lie in the left half of the s plane; thus, stability holds. The continuous-time case was already derived by Ducasse (2002), including loss terms.

But, the states of the inductances can assume large values, and small differences of large numbers can occur. Hence, the system does become unstable after some time, as also found by van Walstijn and Campbell (2003). As a simple workaround, a small negative constant (say, 0.0001) can be added to all r_n.

IV. COMPARISON OF THE TUBE MODELS

For equal segment number, the cone chains have a higher filter order than the cylinder chains, since each scattering point needs a state quantity. Thus, the transfer function is no more symmetric about $c/4D$ even for real termination impedance. But, the number of free parameters is equal for both models and corresponds to the number of formant frequencies and bandwidths in the frequency range up to $c/4D$, so that above this the behavior cannot be independent. As sampling frequency, c/D should be chosen, but only the range up to about $c/4D$ (abscissa 0.25 in the transfer-

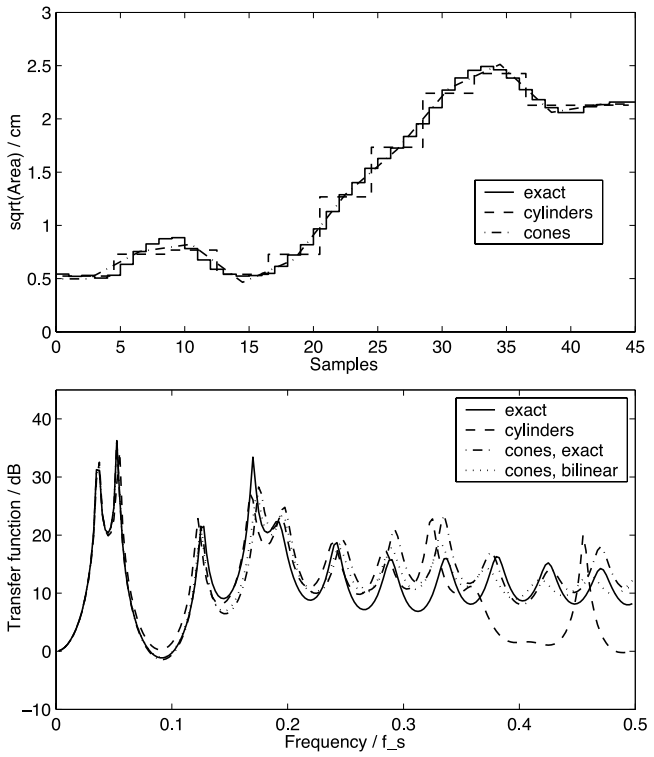


FIG. 2. Top: Area functions, bottom: transfer functions of tube models (''exact,'' cylinders, cones) up to $f_s/2$, vowel [ɑ].

function plots below) should really be used, because otherwise the ''bilinear'' representation of the shunt inductances is too inaccurate.

A. Methods of comparison

For comparing the volume-velocity transfer functions (hard source; radiation impedance: resistance parallel inductance), the first and last segments were chosen cylindrical with length $D/2$, in order to align the scattering points of the cone chain with the segment centers of the cylinder chain. This is preferable since the area samples refer to the segment centers for cylinders but to the segment boundaries for cones. The transfer functions were computed in the frequency domain by means of chain matrices. Compared were: cylinder chain, cone chain with exact inductances, cone chain with bilinear inductances; frequency up to $f_s/2 = c/2D$.

As templates, vowel areas from Story *et al.* (1996) were used, with a sampling interval of 0.396 cm, also with sliding smoothing of $\sqrt{A}$ over five segments each. The tubes were downsampled 4:1, either pointwise or averagingly. In the latter case the cylinder chains were obtained by interval averages (with respect to $\sqrt{A}$) and the cone chains by fitting a straight-line train to $\sqrt{A}$. A cylinder chain with full resolution defined the ''exact'' transfer function, which was compared to the downsampled models.

B. Results

Figure 2 displays an example for vowel [ɑ] (smoothed, averagingly downsampled). Root-mean-square (rms) deviation of the (log) transfer function from the exact one in the frequency range up to $f_s/4$ (199 frequencies) is shown in Table I for three vowels with smoothed tubes. Here, the cone chain was more exact in most cases with averaging downsampling, which however did not hold for every formant (vowel-dependent). But, with pointwise downsampling, there was no clear advantage of either model, and for unsmoothed area functions (not shown in the table), the cylinder chains were 0 to 2 dB advantageous. Surprisingly, the cone chains with bilinear inductances (''Con.bi.'') performed somewhat better than those with exact inductances (''Con.ex.''). For smoothed tubes, the cone chain may even represent formants above $f_s/4$ rather fairly. The difference between the continuous-time and discrete-time transfer functions of the cone chain is small below $f_s/4$ (''Ex.-bi.'').

TABLE I. rms error between logarithmic ''exact'' and model transfer functions, dB. Frequency range up to $f_s/4$. ''av:'' sampling by averaging, ''pt:'' sampling pointwise (see the text); smoothed area functions. Columns, see the text.

Samp.	Vowel	Cyl.	Con.ex.	Con.bi.	Ex.-bi.
av	ɑ	2.51	2.98	2.21	1.11
	i	5.12	4.74	3.15	2.61
	u	9.60	3.79	4.55	4.38
pt	ɑ	2.10	2.72	2.80	1.03
	i	4.33	4.36	3.76	2.19
	u	8.02	6.62	7.61	3.40

C. Comparison with half-length cylinders

Starting from an exact cone chain as model, one can compare whether the discrete-time cone chain requires no more expense than a cylinder chain of finer segmentation. This is indeed so, if each cone is replaced by two cylinders of half length, with their $\sqrt{A}$ linearly interpolated at 1/4 and 3/4 between the values at the cone boundaries, Fig. 3. The first and last segments were again cylindrical and of half length. In this case, the discrete-time transfer functions are surprisingly almost identical, which can also be shown theoretically for $A_n/A_{n-1} \approx 1$; but, there is a slight deviation from the transfer function of the continuous-time cone chain. Even with independent averaging fits of chains of cones and half-length cylinders to a smoothed vowel tube, the transfer functions up to $f_s/4$ differ only by 1 to 2.7 dB rms, with a slight advantage of the cylinders. Figure 4 shows an example.

D. Other time-discretization methods

Instead of transforming the $R_n(s)$ from Eq. (18) to discrete time binlinearly, an impulse invariant design might be used (Välimäki, 1995) or the method of Martínez *et al.*

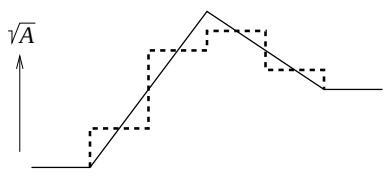


FIG. 3. Comparison of cone chain with cylinder chain of half segment length (principle).

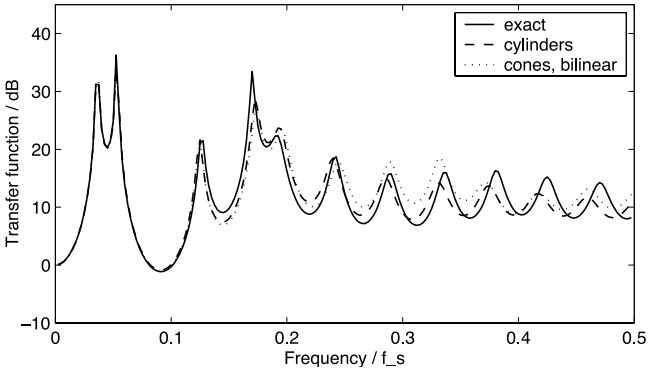


FIG. 4. Comparison of "exact" tube with bilinear cones and cylindrical sections of half length.

(1988) according to their Eq. (21), with linear interpolation of the input signal. These two approaches correspond to

$$R_n(z) = -b_n/(1-z_n z^{-1})$$

or

$$R_n(z) = -[(1/f_s\tau_n) - b_n(1-z^{-1})]/(1-z_n z^{-1}),$$

respectively, with $z_n = \exp(-1/f_s\tau_n)$, $b_n = 1-z_n$. In these cases, the deviation from the ideal transfer function is usually larger, especially at low frequencies, Fig. 5 (here, the radiation impedance was always of continuous-time form). This was already reported by Kob (2002). Moreover, stability remains to be shown.

E. Remarks on damping and time variation

As in the Kelly–Lochbaum model, damping can only be represented in a highly simplified way by exponential decay factors, which does not allow frequency dependence and implicitly assumes equal size of the viscous and heat-conduction losses. In the conical model, the states of the shunt inductances must decay at the same rate as the waves in the segments to preserve stability.

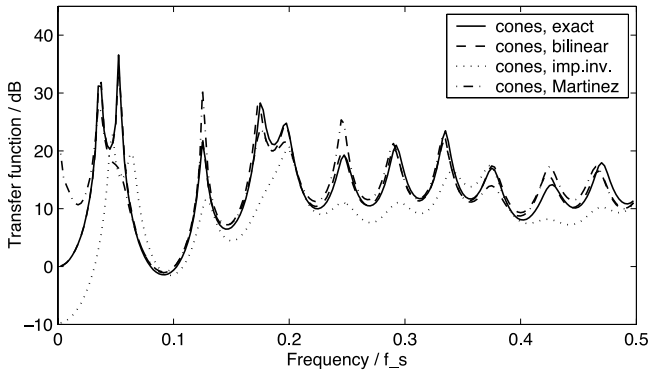


FIG. 5. Continuous time ("exact") versus three discretization methods: bilinear, impulse invariant, Martínez.

In the time-variant case, the model as such would require a different formulation. Whereas this is possible for cylinder segments (Strube, 1982, 2000), it seems to be difficult for conical segments and has not been attempted yet. But, vocal-tract motion is usually sufficiently slow for a quasi-time-invariant treatment. Stability is not well defined in this case, because the changing model parameters can cause parametric amplification of the signal. At first sight, the cancellation of unstable local responses would seem to fail in the present approach. Surprisingly, some informal tests with stepwise or random-walk time changes of the area function did not reveal any instabilities, but for stepwise changes, large rapidly decaying transients occurred.

V. CONCLUSIONS

Efficient time-domain approximation of nonuniform tubes by means of wave quantities in conical rather than cylindrical segments is possible. Time derivatives should be discretized by the bilinear transform or (equivalently) by wave digital filters. Then, in spite of locally unstable filters, the discrete-time cone chain is theoretically stable. Computational inaccuracies cause instability after some time, though, which can be suppressed by small additional terms. But, a clear advantage with regard to cylinder segments shows up only if there are really long approximately conical intervals or the area function is rather smooth. However, cylinder chains of half segment length with equal sampling frequency perform at least as well. Time-variant tubes have not been investigated in depth, but no instabilities were found in first tests.

Benade, A.H. (**1988**). "Equivalent circuits for conical waveguides," J. Acoust. Soc. Am. **83**, 1764–1769.

Ducasse, E. (**2002**). "An alternative to the traveling-wave approach for use in two-port descriptions of acoustic bores," J. Acoust. Soc. Am. **112**, 3031–3041.

Fettweis, A. (**1986**). "Wave digital filters: theory and practice," Proc. IEEE **74**, 270–327.

Kelly, J.L., and Lochbaum, C.C. (**1962**). "Speech synthesis," in Proc. Fourth Int. Congr. Acoust., Paper G42, 1–4; reprinted in *Speech Synthesis*, edited by J.L. Flanagan and L.R. Rabiner (Dowden/Wiley, Stroudsburg, 1973).

Kob, M. (**2002**). "Physical Modeling of the Singing Voice," Doctoral thesis, TH Aachen (Logos, Berlin).

Martínez, J., Agulló, J., and Cardona, S. (**1988**). "Conical bores. Part II. Multiconvolution," J. Acoust. Soc. Am. **84**, 1620–1672.

Story, B.H., Titze, I.R., and Hoffman, E.A. (**1996**). "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am. **100**, 537–554.

Strube, H.W. (**1982**). "Time-varying wave digital filters for modeling analog systems," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-30**, 864–868.

Strube, H.W. (**2000**). "The meaning of the Kelly–Lochbaum acoustic-tube model," J. Acoust. Soc. Am. **108**, 1850–1855.

Välimäki, V. (**1995**). "Discrete-time modeling of acoustic tubes using fractional delay filters," Ph.D. thesis, Helsinki.

van Walstijn, M., and Campbell, M. (**2003**). "Discrete-time modeling of woodwind instrument bores using wave variables," J. Acoust. Soc. Am. **113**, 575–585.

Bridging automatic speech recognition and psycholinguistics: Extending Shortlist to an end-to-end model of human speech recognition[a] (L)

Odette Scharenborg,[b] Louis ten Bosch, and Lou Boves
A^{2}RT, Department of Language and Speech, University of Nijmegen, The Netherlands

Dennis Norris
Medical Research Council Cognition and Brain Sciences Unit, Cambridge, United Kingdom

(Received 10 December 2002; accepted for publication 25 August 2003)

This letter evaluates potential benefits of combining human speech recognition (HSR) and automatic speech recognition by building a joint model of an automatic phone recognizer (APR) and a computational model of HSR, *viz.*, Shortlist [Norris, Cognition **52**, 189–234 (1994)]. Experiments based on ''real-life'' speech highlight critical limitations posed by some of the simplifying assumptions made in models of human speech recognition. These limitations could be overcome by avoiding hard phone decisions at the output side of the APR, and by using a match between the input and the internal lexicon that flexibly copes with deviations from canonical phonemic representations. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1624065]

PACS numbers: 43.71.An, 43.71.Es, 43.72.Ne [DOS] Pages: 3032–3035

I. INTRODUCTION

In this letter, we address speech recognition by making a bridge between two disciplines that have little overlap with respect to theoretical framework and experimental paradigms. One discipline is automatic speech recognition (ASR), which studies the automatic transformation of a speech signal into a sequence of discrete ''recognition tokens'' (commonly words). The main goal in ASR research is to minimize the number of recognition errors on a certain test set under specific testing conditions. The second discipline is the area of human speech recognition (HSR). In HSR, the conversion from an acoustic signal to (a string of) words is studied with a focus on understanding the psychological processes underlying human word recognition, e.g., the word perception process *per se*.

In HSR experiments, the usual stimuli are carefully spoken utterances recorded in noiseless environments. On the basis of theories of HSR, several computational models have been developed to simulate data from experiments on human speech perception. These models compute word activations as the input unfolds over time, where activation can be related to the speed and accuracy with which human listeners can recognize words. However, the existing computational models of HSR model only parts of the human speech recognition process. Typically, one of the missing parts is a module that converts the acoustic speech signal into a representation that forms an appropriate input for the models, which almost invariably assume some kind of *symbolic* representation of the speech signal.

Most experimental studies of HSR are based on read speech; however, in the last few years, the focus is shifting towards (more) spontaneous speech. Much more than read speech, spontaneous speech is affected by articulatory processes such as assimilation and reduction. Since listeners are sensitive to this type of subtle subphonemic information (e.g., Gow, 2002; see Cutler, 1998, for an overview), and to durational differences in the input (Davis *et al.*, 2002), HSR models are now challenged to address the question of how the speech signal is mapped onto lexical representations in more detail. This is an area where established techniques from ASR could be useful in informing future research. Nearey (2001) suggests combining dynamic pattern recognition techniques from ASR with HSR models in order to be able to use ''detailed phonetic models [...] as front ends for reasonable models of lexical access.'' Nearey doubts that existing HSR models ''will work as advertised when attached to real phonetic transduction systems.''

The present letter presents the results of experiments that put Nearey's conjecture to the test by attempting to make a bridge between the two research areas by studying a combined ASR–HSR model (henceforth referred to as ''joint model'') that can be regarded as an end-to-end model of human speech recognition. The input for the computational model of HSR is provided by an automatic phone recognizer (APR). This HSR model is tested with input consisting of extemporaneous, ''real-life'' speech.

II. THE JOINT MODEL

The proposed joint model is a first step in the development of an end-to-end model of HSR. From the available computational models for human word recognition, we have

[a]Earlier results and parts of the research presented in this article are published in the Proceedings of the ISCA Tutorial and Research Workshop on Pronunciation Modeling and Lexicon Adaptation for Spoken Language Technology, Estes Park, Colorado, 2002, and in the proceedings of the 7th International Conference on Spoken Language Processing, Denver, Colorado, USA, 2002.

[b]Electronic mail: o.scharenborg@let.kun.nl

chosen Shortlist (Norris, 1994) to use in the joint model, because it has been successfully applied to a wide range of data from studies of HSR.

The joint model works as follows. The APR decodes a speech signal into a sequence of phone symbols; Shortlist takes this sequence as input and generates a sorted word list. These processes are discussed in more detail below.

A. Automatic phone recognizer (APR)

For the APR, we trained 36 context-independent (hidden Markov) phone models, one silence model, one model for hesitations such as ''uh,'' and one noise model (Scharenborg *et al.*, 2002a). The APR decoding is based on a phone loop with optional silence preceding and following each phone, and is guided by a phone bigram. The APR output is a purely phonemic representation of the acoustic signal—without word boundaries.

B. Shortlist

In its present implementation, Shortlist itself is a two-stage model. In the first stage, the input (i.e., a sequence of phone symbols) is processed from left to right and an exhaustive search of the internal lexicon yields a shortlist of word candidates (max. 30 per phone position[1]) that roughly match the phonemic input processed so far. The ''activation'' of the words in the short list is determined by the ''degree of fit'' between the phones in the input and the string of phones specified in Shortlist's internal lexicon. For each phone in the input that matches the lexicon representation of a word, the word's activation is increased by 1; otherwise, the activation is reduced by the mismatch parameter (default value is 3). In the second stage—the competition stage—the candidates in the shortlist enter into a network where time-overlapping candidates compete with each other. The output consists of (a sequence of) the most activated word(s).

III. MATERIAL

A. Acoustic data

For training the APR, data from a Dutch telephone corpus (the Dutch Directory Assistance Corpus, DDAC) were used (Sturm *et al.*, 2000). DDAC contains telephone calls to the Dutch 118 Directory Assistance service. Most utterances consist of either one Dutch city name or ''ik weet het niet'' (''I don't know'') pronounced in isolation. Others may also contain disfluencies and longer connected speech fragments. From this corpus, an independent test set (DDAC-test) of 10 510 utterances comprising 11 523 words was selected.

B. Lexicons

The *baseline* lexicon of Shortlist consists of 2392 city names and ''ik weet het niet'' (''I don't know''). For each word in the lexicon, one unique ''canonical'' phonemic representation was available.

The psycholinguistic theory underlying Shortlist makes no claim about the manner in which humans cope with pronunciation variation. Specifically, there is nothing in the theory that promotes the exclusive use of citation forms in the mental lexicon. Therefore, in order to deal with pronunciation variation, we created a second lexicon (''*PronVar*'') with on average 2.6 pronunciation variants per word (Scharenborg *et al.*, 2002a).

IV. EXPERIMENT I: BASELINE

We investigated the performance of the joint model in a baseline experiment using the baseline lexicon. The input for Shortlist consists of the speech utterances of DDAC-test transcribed by the APR. The parameter settings of Shortlist are identical to those used in Norris (1994). The ''performance'' of the joint model was tested in terms of the ASR benchmarking method of recognition errors, rather than on the psycholinguistic benchmark of similarity to human performance. Thus, the performance measure in this study is word accuracy: the percentage of utterances for which the reference words (in DDAC-test) receive the highest activation value at the output of Shortlist.

With an accuracy of 23.5%, the performance of the joint model in this baseline experiment appears to be quite poor. Since the performance of Shortlist on *canonical* phone representations is close to 100%, this result shows that recognizing real-life speech is more difficult than recognizing ''perfect'' phonemic transcriptions. An error analysis reveals that the model has great difficulty in dealing with reduced forms: the APR output mostly comprises fewer (and sometimes also different) phones than the canonical representation stored in Shortlist's lexicon.

Two follow-up experiments were carried out. The aim of the experiments was to study the possible improvement of the joint model's baseline performance using two strategies: using a lexicon that accounts for pronunciation variation (experiment II), and adjusting the value of the mismatch parameter in Shortlist (experiment III).

V. EXPERIMENT II: ACCOUNTING FOR PRONUNCIATION VARIATION

The second experiment is identical to experiment I, except that the PronVar lexicon (including pronunciation variations) was used. Using PronVar, Shortlist's performance as a speech recognizer—reported in terms of word accuracy—increases substantially with 16.2% absolute to 39.7%. An error analysis reveals that there are few cases where the correct word is in the shortlist, but that a competitor receives a higher final activation. This finding suggests that, in the case of noncanonical input, the selection of correct lexical candidates into the shortlist is problematic. This problem is addressed in experiment III.

VI. EXPERIMENT III: ADJUSTING THE MISMATCH PARAMETER

Listeners are highly sensitive to any mismatch between input phones and the phonological representations of words; a mismatch of a single phonological feature can eliminate all signs that a word has been activated (e.g., McQueen *et al.*, 1999). Because of these findings, Shortlist weights mismatching information much more heavily than matching information. However, a high value of the mismatch parameter

TABLE I. Effect of $M=3.0$ and $M=0.0$ measured in terms of the accuracy and the percentage of utterances for which the correct word was present in the shortlist (% In shortlist). Two lexicons are used, *viz.* baseline and PronVar.

	Baseline lexicon		*PronVar lexicon*	
Mismatch	Accuracy (%)	In shortlist (%)	Accuracy (%)	In shortlist (%)
3.0	23.5	24.3	39.7	42.3
0.0	32.5	59.5	54.1	76.5

could actually impair recognition of real-life speech considerably, as even quite small deviations from the expected lexical representation might make a word unrecognizable.

In experiment III, we investigated the effect of ''canceling'' the mismatch penalty (M) by setting $M=0.0$ in a test with both lexicons (for a complete account of the experiment, see Scharenborg *et al.*, 2002b). Table I shows the results in terms of the percentage of utterances for which the correct word is present in the shortlist ('In shortlist'). In addition, we report the word accuracy of the joint model on the word recognition task.

The first row of Table I shows the results of experiment II for reference. As can be seen in Table I, using $M=0.0$ increases the model's performance with both lexicons compared to the default value $M=3.0$.

VII. GENERAL DISCUSSION

The aim of the research described in this letter is to build and evaluate an end-to-end computational model of HSR—based on a joint model of an APR and Shortlist—that takes acoustic recordings of real-life speech as input. Real-life speech is characterized by pronunciation variation, which leads to noncanonical phonemic representations. In order to study the effects of noncanonical input to Shortlist, we carried out three experiments. Experiment I was the baseline experiment. In short, experiment II showed that including pronunciation variants in the internal lexicon of Shortlist improves the ability of the joint model to deal with real-life input. Experiment III showed that the combination of a mismatch parameter value of 0.0 and the use of the lexicon containing pronunciation variants is best able to deal with the reduced phonemic forms encountered in real-life speech. This combination yields a recognition accuracy of 54.1%, which is more than twice the baseline performance.

The experiments show that a straightforward combination of an APR and Shortlist does not yield an end-to-end model of HSR that can deal satisfactorily with real-life input, despite the fact that the APR and Shortlist each perform well in their own domains. Apparently, one cannot take for granted that a combination of the best models of two sides yields the best overall end-to-end model. Perhaps this is not too surprising, since neither system was designed with the intention of being interfaced with the other. Nevertheless, these experiments illustrate the consequences of some of the simplifying assumptions made in Shortlist and other HSR models, and show the extent to which these assumptions need to be revised to produce genuine end-to-end models that will be able to deal with the pronunciation variation present in spontaneous speech.

One shortcoming of the joint model is that it makes ''hard'' decisions both at the level of input phones, and in the goodness-of-fit metric used in the search process. Shortlist requires a single string of phone symbols as input. This implies that the APR is forced to make hard decisions about the segmental representation of the speech signal based only on the acoustic information. Also for HSR (e.g., Gaskell *et al.*, 1998; McQueen *et al.*, 1999), data from experiments indicate that human listeners do not make hard decisions prior to lexical selection. This problem with Shortlist has been addressed in the Merge model (Norris *et al.*, 2000), which is derived from Shortlist. However, the present implementation of Merge can handle only very small lexicons. One can eliminate hard decisions in the input by representing the speech signal as a segment-based lattice containing multiple segment-string hypotheses. The subsequent word search or activation algorithm should make the final decision which phones were present by reranking the activated words or taking the first best.

The second level of hard decisions involves the word search process in Shortlist. This search matches input phone strings to the phone strings stored in the lexicon in a way that it is intolerant of deviations from the canonical form of words. This is exactly the problem highlighted by Nearey (2001) and is certainly an area where more flexible pattern-matching techniques (such as dynamic programming as commonly used in ASR) could play an important role in refining computational HSR models. Of course, the resulting refined model should still be able to simulate actual data of HSR experiments.

An important question to be borne in mind when assessing the results of our experiments is whether our conclusion would have been radically different had we been able to drive Shortlist with the output of a human ''phone recognizer'' rather than the APR or with the output of an APR optimized on the task. Cucchiarini *et al.* (2001) showed that automatically generated transcriptions of read speech are very similar to manual phonetic transcriptions created by expert phoneticians. Such transcriptions are to a large extent also noncanonical. Thus, transcriptions created by human expert transcribers would cause similar problems for HSR models. In Scharenborg *et al.* (2002b), it is shown that optimizing the APR settings in order to improve the balance between generating an input phone sequence that is close to the signal and at the same time meets the input criteria of Shortlist, does not improve the performance of the joint model. So, while our experiments may not provide a precise quantitative measure of the extent of the problems faced by Shortlist, the problems are real nonetheless.

Finally, we would like to raise an additional point.[2] A human being is able to identify a nonlexical token as a nonword. However, the joint model is not able to classify any input as a nonword, since it simply activates the nearest known word. Identification of a nonword could be made possible by using an activation *threshold*: when no lexical token exceeds the threshold, the system identifies a nonword. This is one topic for further research.

VIII. CONCLUSION

This letter describes a coupling of an automatic phone recognizer and a computational model of human word recognition, *viz.* Shortlist. The coupling helped to identify aspects of the two components of the joint model that need to be improved in order to build a comprehensive end-to-end computational model of HSR that is able to deal with real-life speech. One of the future research directions is extending the representation of the speech signal from a single linear input phone string to a probabilistic phone graph. This allows, in a natural way, the postponement of a hard decision to a point later in the word search process, which we believe is desirable. A second possibility of improvement lies in changing the current word search in Shortlist into a search algorithm based on dynamic programming techniques. By doing so, deviations from the canonical representations can be dealt with in a natural way.

ACKNOWLEDGMENTS

The authors would like to thank Anne Cutler, James McQueen, and Roel Smits for fruitful discussions about this research and their comments on earlier versions of this letter. Furthermore, the authors would like to thank the four anonymous reviewers for raising additional interesting issues and giving their useful comments on an earlier version of this letter.

[1]The number 30 is arbitrarily chosen; the exact value does not have a large effect on the performance of the model (Norris, 1994).

[2]This issue was raised by one of the anonymous reviewers of an earlier version of this letter.

Cucchiarini, C., Binnenpoorte, D. M., and Goddijn, S. M. A. (**2001**). ''Phonetic Transcriptions in the Spoken Dutch Corpus: How to combine efficiency and good transcription quality,'' Proceedings of Eurospeech, pp. 1679–1682.

Cutler, A. (**1998**). ''The Recognition of Spoken Words with Variable Representations,'' Proceedings of the ESCA Workshop on Sound Patterns of Spontaneous Speech, Aix-en-Provence, pp. 83–92.

Davis, M. H., Marslen-Wilson, W. D., and Gaskell, M. G. (**2002**). ''Leading up the lexical gardenpath: Segmentation and ambiguity in spoken word recognition,'' J. Exp. Psychol. Hum. Percept. Perform. **28**, 218–244.

Gaskell, M. G., and Marslen-Wilson, W. D. (**1998**). ''Mechanisms of phonological inference in speech perception,'' J. Exp. Psychol. Hum. Percept. Perform. **24**, 380–396.

Gow, Jr., D. W. (**2002**). ''Does English coronal place assimilation create lexical ambiguity,'' J. Exp. Psychol. Hum. Percept. Perform. **28**, 163–179.

McQueen, J. M., Norris, D., and Cutler, A. (**1999**). ''Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches,'' J. Exp. Psychol. Hum. Percept. Perform. **25**, 1363–1389.

Nearey, T. M. (**2001**). ''Towards modelling the perception of variable-length phonetic strings,'' Proceedings of the SPRAAC Workshop, Nijmegen, pp. 133–138.

Norris, D. (**1994**). ''Shortlist: A connectionist model of continuous speech recognition,'' Cognition **52**, 189–234.

Norris, D., McQueen, J. M., and Cutler, A. (**2000**). ''Merging information in speech recognition: Feedback is never necessary,'' Behav. Brain Sci. **23**, 299–370.

Scharenborg, O., and Boves, L. (**2002a**). ''Pronunciation Variation Modelling in a Model of Human Word Recognition,'' in Proceedings of Workshop on Pronunciation Modeling and Lexicon Adaptation, Estes Park CO, pp. 65–70.

Scharenborg, O., Boves L., and de Veth, J. (**2002b**). ''ASR in a Human Word Recognition Model: Generating Phonemic Input for Shortlist,'' Proceedings of ICSLP, pp. 633–636.

Sturm, J., Kamperman, H., Boves, L., and den Os, E. (**2000**). ''Impact of speaking style and speaking task on acoustic models,'' Proceedings of ICSLP, pp. 361–364.

Preceding phonetic context affects perception of nonspeech[a] (L)

Joseph D. W. Stephens[b] and Lori L. Holt
Psychology Department and Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

(Received 10 December 2002; revised 19 September 2003; accepted 30 September 2003)

A discrimination paradigm was used to detect the influence of phonetic context on speech (experiment 1a) and nonspeech (experiment 1b) stimuli. Results of experiment 1a were consistent with the previously observed phonetic context effect of liquid consonants (/l/ and /r/) on subsequent stop consonant (/g/ and /d/) perception. Experiment 1b demonstrated a context effect of liquid consonants on subsequent nonspeech sounds that were spectrally similar to the stop consonants. The results are consistent with findings that implicate spectral contrast in phonetic context effects.

[DOI: 10.1121/1.1627837]

PACS numbers: 43.71.An, 43.71.Es, 43.71.Pc [PA] Pages: 3036–3039

I. INTRODUCTION

Perception of speech is highly dependent on surrounding phonetic context. For example, Mann (1980) found that ambiguous consonant–vowel (CV) syllables varying perceptually between /ga/ and /da/ were identified as /ga/ more often when preceded by /al/ than when preceded by /ar/. With this effect, and other phonetic context effects (e.g., Mann and Repp, 1980, 1981), speech identification is shifted in a direction opposite that of the acoustic assimilation caused by coarticulation in speech production, apparently ''compensating for coarticulation.'' The close correspondence between speech production and perception has led theorists to posit gestural origins for phonetic context effects, such that they arise either from listeners' implicit representations of articulatory gestures (Mann, 1980) or from direct perceptual recovery of articulatory gestures (Fowler *et al.*, 1990). Recent evidence, however, suggests that phonetic context effects may arise from perceptual interactions among the spectral characteristics of adjacent sounds. Lotto and Kluender (1998) found that nonspeech sounds lacking gestural information were sufficient to shift identification of subsequent speech: ambiguous CVs between /ga/ and /da/ were identified more often as /ga/ when preceded by a tone at the third formant (F3) offset frequency of /al/. Lotto *et al.* (1997) observed phonetic context effects in Japanese quail: quail trained to peck to /ga/ pecked more when stimuli were preceded by /al/ than when preceded by /ar/. These findings were interpreted as arising from spectrally contrastive perceptual mechanisms at a precategorical level.

If phonetic context effects arise from general perceptual interactions among spectral characteristics, then both speech and nonspeech sounds should elicit context effects. Nonspeech sounds have been shown to affect perception of subsequent speech (Holt, 1999; Holt *et al.*, 2000; Lotto and Kluender, 1998; Lotto *et al.*, 2003), but to date no effect of speech on subsequent nonspeech has been reported. Moreover, gesture-based theories of speech perception do not predict that such an effect will occur because phonetic context effects arise from information specifying articulatory gestures rather than auditory characteristics (e.g., Fowler *et al.*, 2000). The current study assessed the influence of preceding phonetic context on the perception of nonspeech sounds. Because the nonspeech stimuli were unfamiliar sounds for which participants had no labels, the current experiments used a discrimination paradigm in which labeling was unnecessary (modified from Mann and Liberman, 1983).

II. METHOD

A. Participants

Fifteen and 17 undergraduates at Carnegie Mellon University participated in experiments 1a and 1b, respectively. All reported normal hearing, were native English speakers, and received course credit for participation.

B. Stimuli

Two ten-member series of target stimuli were created for use in experiments 1a and 1b. For experiment 1a, target stimuli were CV syllables ranging perceptually from /ga/ to /da/. Target syllables consisted of 80-ms linear formant transitions followed by a 170-ms steady-state vowel. For experiment 1b, target stimuli were nonspeech sounds consisting only of the 80-ms F2 and F3 transitions from the CV syllables used in experiment 1a. Two syllables, /al/ and /ar/, were synthesized for use as precursors in both experiments. The precursors were 250 ms in duration and consisted of a 100-ms steady-state vowel followed by 150-ms linear formant transitions. Formant frequencies for all stimuli were identical to those used by Lotto and Kluender (1998). Stimuli were synthesized with 12-bit resolution and sampled at 10 kHz, using the cascade branch of the Klatt (1980) synthesizer for speech stimuli and the parallel branch of the synthesizer for nonspeech stimuli.

[a]Portions of this work were presented in ''Effect of preceding speech on nonspeech sound perception,'' at the 143rd Meeting of the Acoustical Society of America, Pittsburgh, PA, June 2002.
[b]Electronic mail: jds2@andrew.cmu.edu

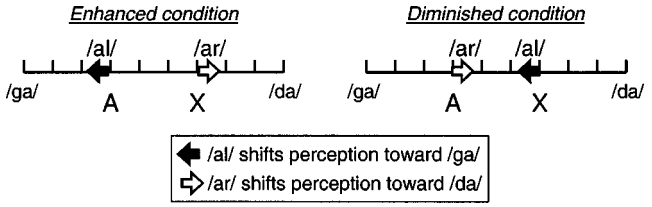


FIG. 1. Design of experiments. On each trial, participants heard two target stimuli and decided whether they were the same or different. Discrimination conditions exploited a known context effect to increase or decrease perceptual distance between target (''enhanced'' and ''diminished'' conditions). In experiment 1b, target syllables from /ga/ to /da/ were replaced by nonspeech sounds that modeled their spectral characteristics.

Each of the target stimuli was matched in root mean square (rms) amplitude and combined with each of the two precursors to yield 40 precursor-target combinations. A silent interval of 50 ms was inserted between precursors and targets, resulting in an overall stimulus duration of 550 ms in experiment 1a and 380 ms in experiment 1b. Stimuli were converted from digital to analog and presented by TDT System II hardware (Tucker-Davis Technologies) over Sennheiser HD-265 linear headphones at 65–70 dB SPL.

C. Procedure

Experiments 1a and 1b each consisted of two phases. In the first phase, participants performed AX discrimination of target stimuli presented in phonetic context. On each trial, two precursor-target combinations were presented with an interstimulus interval of approximately 750 ms. Participants were told to attend to the target stimuli and indicate whether the two targets were different using buttons labeled ''SAME'' and ''DIFFERENT.'' Participants were told that target stimuli would always sound similar and that they should respond ''SAME'' only if they thought the targets were *exactly* the same. On each trial, one target stimulus was preceded by /al/ and the other was preceded by /ar/. Target stimuli either were identical (catch trials) or differed by three steps along the ten-step series (discrimination trials).

The effect of context was tested by comparing two conditions defined by the arrangement of precursors and targets in each trial. In the ''enhanced'' condition, target stimuli with lower F3 onset were preceded by /al/ and target stimuli with higher F3 onset were preceded by /ar/. In the ''diminished'' condition, the opposite arrangement was used. Based on the effect of preceding liquids on stop consonant identification (Mann, 1980), the discrimination of target pairs was expected to be more accurate in the enhanced condition than in the diminished condition. The experimental design is illustrated in Fig. 1. The within-trial order of precursor-target pairs was counterbalanced to yield 28 unique discrimination trials and 20 unique catch trials. All 48 trials were presented in a single block and there were eight repetitions of the trial block, for a total of 384 trials. Order of presentation was random within each trial block. Participants were given a short break half-way through the task.

The second phase of each experiment was an identification task in which participants heard target stimuli one at a time, in isolation, and indicated whether each stimulus sounded like /ga/ or /da/ by pressing buttons labeled ''GA'' and ''DA'' (in experiment 1b participants were asked, ''For each non-speech sound, *if it were a syllable*, would it be /ga/ or /da/?''). All ten targets were presented in each of ten randomly ordered blocks, for a total of 100 trials.

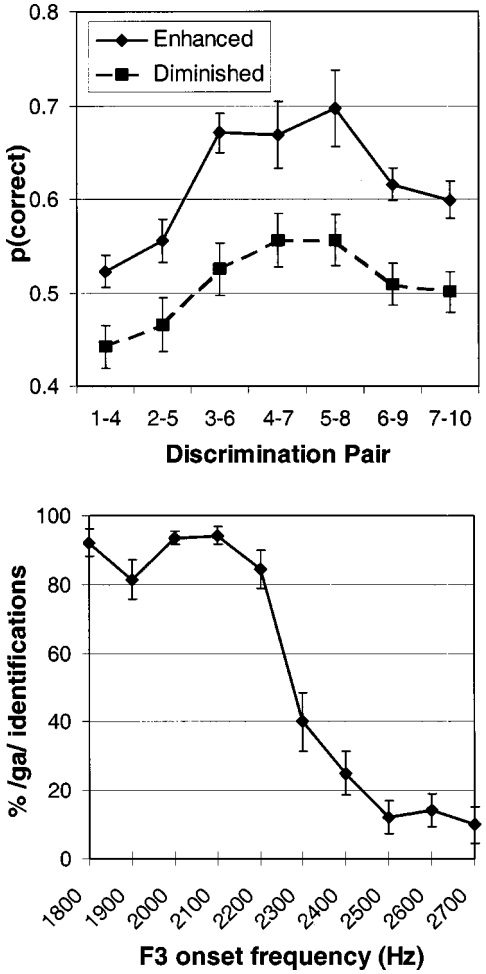


FIG. 2. Results of experiment 1a (speech targets). Mean discrimination (top panel) was better in the enhanced condition than in the diminished condition. Mean identification responses (bottom panel) revealed a typical category boundary along the /ga/-/da/ series. Error bars reflect standard error of the mean.

III. RESULTS

A context effect of preceding liquid on target perception was observed for both speech and nonspeech targets. Data from the identification tasks provided evidence that the target stimuli in experiment 1b were not perceived as speech.

Averaged results of experiments 1a and 1b are shown in Figs. 2 and 3, respectively. Discrimination performance was evaluated by calculating an unbiased measure of proportion correct for each discrimination pair, in each condition, for each participant.[1] Proportion correct was calculated according to Eq. (5.6) of Macmillan and Creelman (1991). Identification responses were evaluated by computing the percentage of trials in which each target stimulus was identified as ''GA.''

For the discrimination task in experiment 1a, an analysis of variance (ANOVA) on the proportion=correct data revealed main effects of enhanced versus diminished condition, $F(1,14)=33.8$, $p<0.001$, and discrimination across tar-

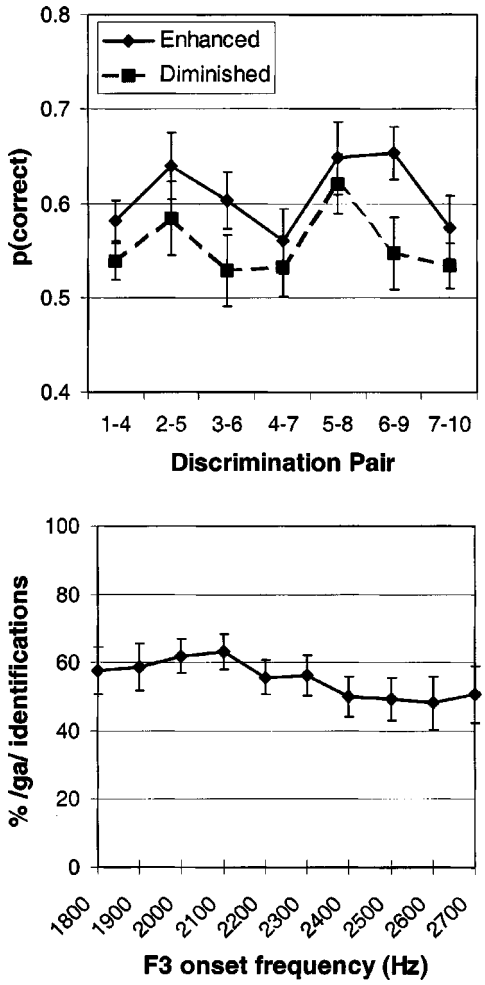


FIG. 3. Results of experiment 1b (nonspeech targets). Mean discrimination (top panel) was better in the enhanced condition than in the diminished condition. Mean identification responses (bottom panel) indicated that participants were unable to correctly assign speech labels to nonspeech stimuli. Error bars reflect standard error of the mean.

get stimulus pair, $F(6,84)=8.05$, $p<0.001$. Target stimulus pairs were more accurately discriminated in the ''enhanced'' condition than in the ''diminished'' condition, indicating that the perceptual distance between target pairs was increased or decreased depending on the arrangement of precursors. Discrimination performance varied significantly across discrimination pairs, with better performance in the middle of the series than at the ends. The identification data from one participant in experiment 1a were discarded due to computer error. Identification data for the remaining participants exhibited a categorical pattern typical of stop-consonant series. An ANOVA on the identification data revealed a significant effect of target stimulus, $F(9,117)=68.5$, $p<0.001$.

The data from one participant in experiment 1b were excluded from analysis due to incorrect execution of the discrimination task (only five ''same'' responses in 384 trials). For the discrimination task in experiment 1b, an ANOVA on the proportion=correct data revealed a significant main effect of enhanced versus diminished condition, $F(1,15)=12.0$, $p<0.005$. Thus, phonetic context effectively influenced discriminability of nonspeech targets. A main effect of discrimination across target stimulus pairs did not reach significance, $F(6,90)=2.16$, $p=0.054$. An ANOVA on the identification data revealed no effect of target stimulus on identification responses, indicating that, on average, participants were unable to assign consistent category labels to the stimuli.

Inspection of identification data from individuals in experiment 1b revealed that some participants *were* able to assign labels to the stimuli. Such response patterns could indicate either that listeners heard the nonspeech stimuli as speech sounds, or that they simply managed to consistently assign arbitrary labels as a function of the F3 onset cue. Therefore, additional analyses were performed to determine whether the results of the discrimination task depended on participants' ability to assign speech labels the nonspeech stimuli. Participants were sorted into three groups: those whose labeling was consistent with the analogous speech categories ($N=7$); those whose labeling was categorical, but opposite to the analogous speech categories ($N=3$); and those whose labeling was not categorical ($N=6$). A participant's identification responses were considered ''categorical'' if the average of his or her ''GA'' responses to the first three members of the stimulus series and the last three members of the stimulus series differed by at least 20 percent. An ANOVA was performed on the discrimination data of experiment 1b, including labeler type as a between-subjects variable. There was no interaction of labeler type and condition, indicating that the effect of phonetic context on nonspeech targets was not related to participants' labels for the nonspeech stimuli. An additional ANOVA was performed on just the discrimination data of the six ''noncategorical'' listeners. An influence of speech precursors upon nonspeech discrimination was observed, $F(1,5)=7.64$, $p=0.04$.

Additional analyses were conducted to compare the results of experiments 1a and 1b. An ANOVA that included data from both experiments revealed a significant interaction of experiment and condition, $F(1,29)=5.54$, $p=0.025$, reflecting the larger effect of context for speech versus nonspeech stimuli. The interaction of experiment and target stimulus pair was also significant, $F(6,174)=3.89$, $p=0.001$, indicating that the pattern of discrimination performance across the target stimulus series differed for speech sounds compared to nonspeech sounds. An ANOVA comparing discrimination data from the seven ''speechlike'' labelers of experiment 1b to discrimination data from experiment 1a also revealed a significant interaction of experiment and target stimulus pair, $F(6,120)=2.84$, $p=0.013$. Thus, even the experiment 1b participants with speechlike identification exhibited a significantly different pattern of discrimination across targets from listeners in experiment 1a.

IV. DISCUSSION

The present results indicate that phonetic context affects the perception of nonspeech sounds, as predicted by a spectral contrast account of phonetic context effects (Lotto and Kluender, 1998). However, some alternative interpretations of the data should be considered.

Fowler *et al.* (2000) proposed that speech and nonspeech context effects originate from different mechanisms, with nonspeech influences upon phonetic perception arising

from masking of the target F3 by nonspeech precursor sounds. Citing Moore (1988), Fowler *et al.* noted that ''acoustic masks tend to reduce sensitivity to frequencies including and surrounding their own; the range of frequencies affected increases with the amplitude of the mask.'' (p. 881) If masking causes nonspeech context effects, then these effects could be attributed to peripheral auditory mechanisms, which, in the account of Fowler *et al.*, are presumably unrelated to the perceptual mechanisms involved in phonetic context effects. However, due to the construction of stimuli used in experiment 1b, masking is unlikely to be responsible for the current results. The acoustically reduced, nonspeech target stimuli were matched in rms amplitude to the acoustically rich precursor syllables, so that the F3 frequencies of targets had greater energy than the F3 frequencies of precursors. Additionally, recent work has ruled out a peripheral masking account for effects of nonspeech precursors on speech targets. Lotto *et al.* (2003) found effects of preceding nonspeech on consonant perception when precursors and targets were presented to opposite ears, and when delays of up to 175 ms were inserted between precursors and targets. Identical experiments using only speech stimuli (Holt and Lotto, 2002) yielded very similar results, consistent with the hypothesis that common mechanisms underlie both effects.

The interpretation of the current results also depends on the validity of the assumption that the perception of nonspeech sounds involves general auditory mechanisms. Due to the overlearned nature of speech perception, ''speech-specific'' perceptual mechanisms might generalize somewhat to acoustically similar nonspeech sounds. Thus, the context effect observed in experiment 1b might have resulted from speechlike processing of the nonspeech target stimuli. Although most listeners did not label the nonspeech stimuli according to analogous speech categories, the identification task could simply have been less sensitive to subtle effects of speech-specific processing than the discrimination task. However, the pattern of discrimination responses in experiment 1b also provides evidence that speech mechanisms were not involved in the perception of the nonspeech stimuli. Discrimination performance did not vary across the nonspeech stimulus series, whereas discrimination of speech stimuli was significantly better in the middle of the stimulus series than at the ends [consistent with the typical finding of a discrimination peak near a category boundary (Liberman *et al.*, 1957)]. Mann and Liberman (1983) interpreted the absence of a discrimination peak for F3 chirps in duplex perception as evidence that chirps were perceived in a nonspeech ''mode.'' It is therefore reasonable to conclude that the results of experiment 1b are not due to speech-specific processes. A replication of the current results in nonhuman listeners would further support for this conclusion, as it has for other context effects (Lotto *et al.*, 1997).

An additional aspect of interest in the current data is that speech precursors had a greater effect on discrimination of speech targets than nonspeech targets. This difference might result from participants' tendency to rely on category labels when discriminating speech sounds. Due to the presence of a category boundary along the speech target series, an increase or decrease in perceptual distance between speech targets should increase or decrease, respectively, the probability that those targets fall into different categories. Thus, category membership could exaggerate the perceptual context effect for speech stimuli compared to nonspeech stimuli.

In summary, the current findings are consistent with the hypothesis that phonetic context effects result from general perceptual processes sensitive to spectral characteristics. They are more difficult to reconcile with theories (Fowler *et al.*, 1990, 2000) that attribute phonetic context effects to implicit knowledge of articulatory dynamics or direct recovery of articulatory gestures from speech input.

ACKNOWLEDGMENTS

The authors thank Christi Adams, Monica Datta, and Camilla Kydland for help with data collection and Andrew Lotto, Randy Diehl, and James Sawusch for helpful comments. This work was supported by an NDSEG Fellowship to JDWS, by a research grant from NOHR to LLH, and by the Center for the Neural Basis of Cognition.

[1] *d′* was not computed due to the complexity of the ''roving'' same–different design, in which multiple discrimination pairs are presented within a single block (see Chap. 6 of Macmillan and Creelman, 1991). Proportion correct for an unbiased observer is an adequate measure of sensitivity for the purposes of the current study.

Fowler, C. A., Best, C. T., and McRoberts, G. W. (**1990**). ''Young infants' perception of liquid coarticulatory influences on following stop consonants,'' Percept. Psychophys. **48**, 559–570.

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.

Holt, L. L. (**1999**). ''Auditory constraints on speech perception: An examination of spectral contrast,'' unpublished doctoral dissertation, University of Wisconsin, Madison.

Holt, L. L., and Lotto, A. J. (**2002**). ''Behavioral examinations of the level of auditory processing of speech context effects,'' Hear. Res. **167**, 156–169.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2000**). ''Neighboring spectral content influences vowel identification,'' J. Acoust. Soc. Am. **108**, 710–722.

Klatt, D. H. (**1980**). ''Software for a cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 971–995.

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (**1957**). ''The discrimination of speech sounds across phoneme boundaries,'' J. Exp. Psychol. **54**, 358–368.

Lotto, A. J., and Kluender, K. R. (**1998**). ''General contrast effects of speech perception: Effect of preceding liquid on stop consonant identification,'' Percept. Psychophys. **60**, 602–619.

Lotto, A. J., Kluender, K. R., and Holt, L. L. (**1997**). ''Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*),'' J. Acoust. Soc. Am. **102**, 1134–1140.

Lotto, A. J., Sullivan, S. C., and Holt, L. L. (**2003**). ''Central locus for non-speech context effects on phonetic identification,'' J. Acoust. Soc. Am. **113**, 53–56.

Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge U.P., New York).

Mann, V. A. (**1980**). ''Influence of preceding liquid on stop-consonant perception,'' Percept. Psychophys. **28**, 407–412.

Mann, V. A., and Liberman, A. M., (**1983**). ''Some differences between phonetic and auditory modes of perception,'' Cognition **14**, 211–235.

Mann, V. A., and Repp, B. H. (**1980**). ''Influence of vocalic context on perception of the [ʃ]-[s] distinction,'' Percept. Psychophys. **28**, 213–228.

Mann, V. A., and Repp, B. H. (**1981**). ''Influence of preceding fricative on stop consonant perception,'' J. Acoust. Soc. Am. **69**, 548–558.

Moore, B. (**1988**). *An Introduction to the Psychology of Hearing* (Academic, London).

Characterization of chaotic dynamics in the vocalization of *Cervus elaphus corsicanus* (L)

Angelo Facchini,[a] Simone Bastianoni,[b] and Nadia Marchettini[c]
Department of Chemical and Biosystems Sciences, University of Siena, Via della Diana 2/A 53100 Siena, Italy

Mauro Rustici[d]
Department of Chemistry, University of Sassari, Via Vienna 2-I 07100 Sassari, Italy

(Received 23 November 2002; revised 28 August 2003; accepted 15 September 2003)

Chaos, oscillations, instabilities, and intermittency represent only some nonlinear examples apparent in the natural world. These phenomena appear in any field of study, and advances in complex and nonlinear dynamic techniques bring about opportunities to better understand animal signals. In this work an analysis method is suggested based on the characterization of the vocal-fold dynamics by means of the nonlinear time-series analysis, and by the computations of the parameters typical of chaotic oscillations: Attractor reconstruction, spectrum of Lyapunov exponents, and maximum Lyapunov exponent were used to reconstruct the dynamic of the vocal folds. Identifying a sort of vocal fingerprint can be useful in biodiversity monitoring and understanding the health status of a given animal. This method was applied to the vocalization of the *Cervus elaphus corsicanus*, the Sardinian red deer. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1624071]

PACS numbers: 43.80.Ka, 43.25.Rq, 43.72.Ar [WA] Pages: 3040–3043

I. INTRODUCTION

The physical and physiological mechanisms of sound production are important to understanding mammal vocalization, which ranges from periodic vocal-fold vibrations to completely aperiodic vibration and atonal noise. Between these two extremes, many phenomena have been observed and reported:[1,2] biphonation, cycles, subharmonic, and chaotic behavior. These behaviors can be predicted by theoretical models. For example, the two-mass model (the most accepted for mammal apparatus of phonation) can exhibit irregular oscillations.[3,4] The apparatus of phonation can be investigated through the characterization of the animal vocalization, where vocal nonlinearity can be used. According to Tokuda[5] the nonlinear analysis of human speech signal has been carried out extensively, while nonlinear characteristics for animal voice signals have not yet been investigated.

Using the methods of nonlinear time-series analysis, we wish to understand the mechanics of the vocal folds starting from the vocalization time series. The characterization of the vocal signal as a chaotic time series can give important information on the health status of the animal, since the oscillation modes are related to the status of the throat tissues and to the strength of the animal. Furthermore, tissues shapes of vocal apparatus are different among the animals and the characterization of several chaotic signals can be used in the monitoring of biodiversity.

The last remaining populations of a subspecies of the red deer: the Sardinian deer (*Cervus elaphus corsicanus*), are found in the well-preserved evergreen forest of *Monte Arcosu* in Sardinia (a protected area owned by WWF Italy). The *Cervus elaphus* is the largest and most phylogenetically advanced species of *Cervus*. Head and body length is 1.65–2.65 m, tail length is 0.11–0.27 m, height at the shoulder is 0.75–0.15 m, and weight is 75–340 kg. The largest and strongest male generally has the largest harem. In order to maintain this position of superiority he must constantly keep his distance from rival males by bellowing out, and chasing off potential rivals who come near his females. After vocalizing, the largest remaining males size each other up, and if antler and body size are comparable, they battle for the females. Their antlers lock and each male attempts to forcefully push the other away. The strongest and most powerful male wins and secures a harem (group) of females for mating. In this work an extensive characterization of the vocalization of *Cervus elaphus corsicanus* is presented by means of Lyapounov exponents of the chaotic oscillation evidence of registered sounds.

II. MATERIAL

A number of different signals corresponding to different sound emissions were considered. Only clear and low noise sound emissions have been analyzed, in order to focus exclusively on meaningful vocalizations, and to avoid spurious effect. The vocalizations were recorded from adult males in their natural environment and digitized with a sampling frequency of 22 050 Hz. Figure 1(b) shows a small portion of the analyzed signal and in Fig. 1(d), the spectrogram (512 points FFT) of the signal is shown.

Discrete Fourier transform [Fig. 1(c)] was used to perform a preliminary spectral analysis on vocalization units. The presence of regions with high density of unresolved frequencies is a necessary, even if not sufficient, condition for the occurrence of chaotic dynamical regimes.[6] Nonlinear dy-

[a]Electronic mail: a.facchini@unisi.it
[b]Electronic mail: bastianoni@unisi.it
[c]Electronic mail: marchettini@unisi.it
[d]Author to whom correspondence should be addressed. Electronic mail: rustici@uniss.it

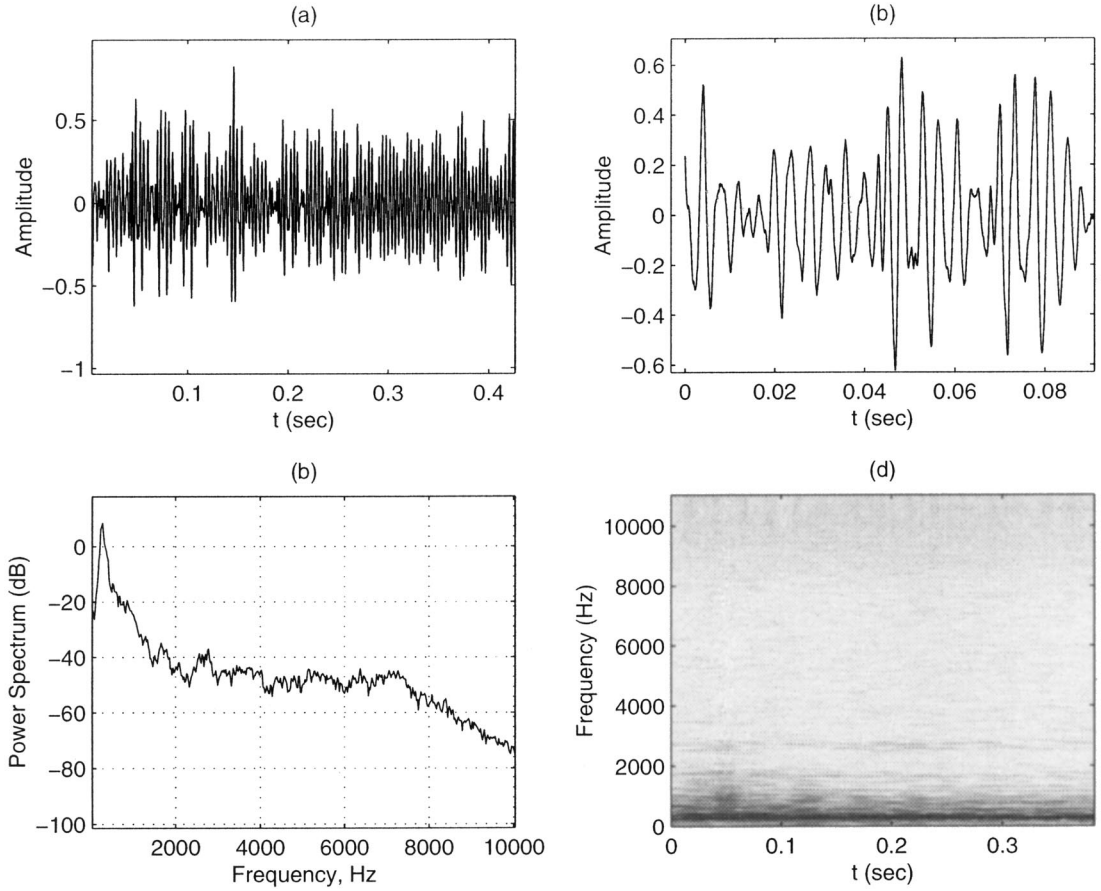


FIG. 1. (a) The analyzed signal. (b) A portion of the signal showing the irregular nature of the vocalization. (c) The power spectrum shows the typical spectral contents of an irregular signal: a broadband and a continuous one. (d) The spectrogram of the signal showing a fundamental frequency of 70 Hz and other frequencies until 2000 Hz.

namics analysis, therefore, was limited to signal units characterized by broadband features in the frequency domain. Results reported in the present work refer to a single signal 0.420 s long. The time series examined consists of 9455 points sampled at 22 050 Hz.

III. COMPUTATIONAL METHODS

The analysis of the time series was performed using the software package TISEAN[7] (TIme SEries ANalysis),[8] valued as the most well-known and robust algorithm set for nonlinear time-series analysis. Typical steps are attractor reconstruction from time series and the characterization of the chaotic dynamic by means of Lyapunov exponents and maximum Lyapunov exponent (MLE).

A. Attractor reconstruction

The attractor of underlying dynamics has been reconstructed in phase space by applying the time delay vector method.[6,9]

Starting from a time series $s(t)=[s_1,\ldots,s_N]$ the system dynamic can be reconstructed using the delay theorem by Takens and Mañe. The reconstructed trajectory $\mathbf{X}$ can be expressed as a matrix where each row is a phase-space vector

$$\mathbf{X}=[X_1,X_2,\ldots,X_M]^T, \tag{1}$$

where $X_i=[s_i,s_{i+T},\ldots,s_{i-(D_E-1)T}]$ and $M=N-(D_E-1)T$.

The matrix is characterized by two key parameters: The *embedding dimension* D_E and the *delay time* T. The embedding dimension is the minimum dimension at which the reconstructed attractor can be considered completely unfolded and there is no overlapping in the reconstructed trajectories. If the chosen dimension is lower than D_E the attractor is not completely unfolded and the underlying dynamics cannot be investigated. Higher dimension was not used due to the increase in computational effort.

The algorithm used for the computation of D_E is the method of *false nearest neighbors*.[10] A false neighbor is a point of trajectory intersection in a poorly reconstructed attractor. As the dimension increases, the attractor is unfolded with greater fidelity, and the number of false neighbors decreases to zero. The first dimension with no overlapping points is D_E.

The delay time T represents a measure of correlation existing between two consecutive components of D_E-dimensional vectors used in the trajectory reconstruction. Following a commonly applied methodology, the time delay T is chosen in correspondence to the first minimum of the average mutual information function.[11]

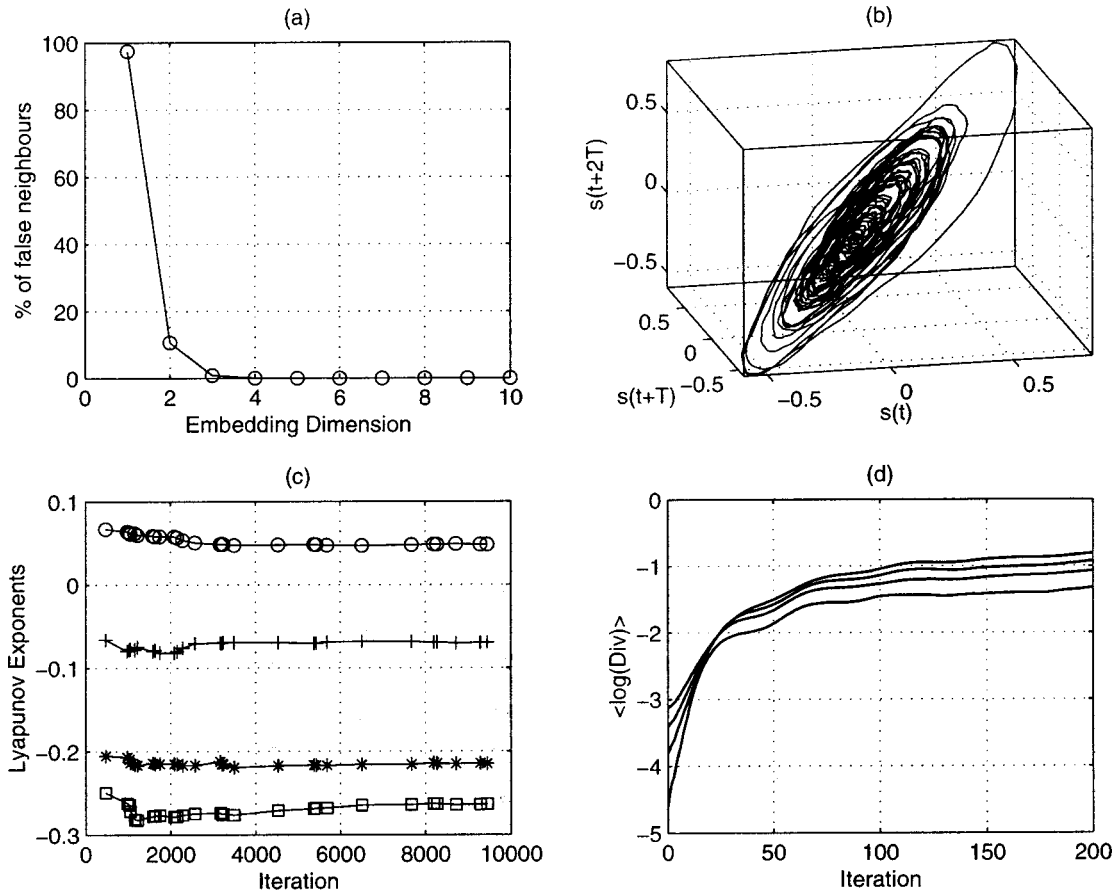


FIG. 2. (a) Computation of the embedding dimension by the false nearest neighbor method applied to the time series. The fraction of false neighbors decrease to zero at a reconstruction dimension $D_E=4$. (b) The attractor reconstructed by the method of delays: This highly structured trajectory indicates that the signal is chaotic and that the irregular motion is not a noisy one. (c) The spectrum of Lyapunov exponent showing the presence of a positive Lyapunov exponent and three negative exponents. (d) Computation of the maximum Lyapunov exponent by the Rosenstein–Kantz algorithm. The value of λ_1 is obtained by a linear regression of the curves in the zone between 0 and 20 iterations. The value $\lambda_1=0.48$ was found.

B. Lyapunov exponents

Chaotic systems display a sensitive dependence on initial conditions. Such a property deeply affects the time evolution of trajectories starting from infinitesimally close initial conditions, and Lyapunov exponents are a measure of this dependence. These characteristic exponents give a coordinate-independent measure of the local stability properties of a trajectory. If the trajectory evolves in an N-dimensional state space there are N exponents arranged in decreasing order, referred to as the *spectrum of Lyapunov exponents (SLE)*

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n . \tag{2}$$

Conceptually, these exponents are a generalizations of eigenvalues used to characterize different types of equilibrium points.

A trajectory is chaotic if there is at least one positive exponent; the value of this exponent, called the *maximum Lyapunov exponent (MLE)*, gives a measure of the divergence rate of infinitesimally close trajectories and of the unpredictability of the system, and gives a good characterization of the underlying dynamics.

Starting from the reconstructed attractor $\mathbf{X}$, it is possible to compute with the method of Sano and Sawada[12,13] the SLE consisting of exactly $n=D_E$ exponents. This method is a qualitative one, and in the presence of a positive exponent, λ_1, a more accurate method is necessary for the computation.

The method of Rosenstein–Kantz[14,15] is used to compute the MLE from the time series. This method measures in the reconstructed attractor the average divergence of two close trajectories in the time $d_j(i)$. This can be expressed as:

$$d_j(i) = C_j e^{\lambda_1 (i\Delta t)}, \tag{3}$$

where C_j is the initial separation. By taking the logarithm of both sides we obtain

$$\ln d_j(i) = \ln C_j + \lambda_1 (i\Delta t). \tag{4}$$

This is a set of approximately parallel lines (for $j=1,2,...,M$), each with a slope roughly proportional to λ_1. The MLE is easily calculated using a least-squares fit to the average line defined by

$$y(i) = \frac{1}{\Delta t}\langle \ln d_j(i) \rangle, \tag{5}$$

where $\langle \cdot \rangle$ denotes the average over all values of j. Figure 2(d) shows a typical plot of $\langle \ln d_j(i) \rangle$: after a short transition there is a linear region that is used to extract the MLE.

TABLE I. Results of the analysis performed on the vocalization signal. A positive Lyapunov exponent and the value of Kaplan–Yorke dimension indicates the chaotic nature of the signal.

Parameter	Value
Delay Time **T**	8
Embedding dimension $\mathbf{D_E}$	4
Maximum Lyapunov exponent λ_1	0.48
Kaplan–Yorke dimension $\mathbf{D_L}$	2.58

IV. RESULTS AND DISCUSSION

The signal considered was characterized by highly complex patterns in which different transients with both periodic and apparently aperiodic features were identified. The apparently random behavior of the numerical series, easily detectable with a simple visual inspection of the sound pattern, was confirmed by the power spectrum and spectrogram. Three different regions were put into evidence: at low frequencies, between 0 and 70 Hz, a first distribution of unresolved peaks is present, a sharp peak is also present at 450 Hz, while a broad band of frequencies, ranging between 850 and 1500 Hz, is easily detectable.

The chaotic characterization was performed calculating the embedding dimension D_E by the false nearest-neighbor method, and in Fig. 2(a) the result of the computation is shown. The figure reports the fraction of false neighbors with respect to the embedding dimension, and a value of $D_E=4$ was found. The delay time was considered as the first minimum of the mutual information function, and the value $T=8$ was found.

Starting from the time series the attractor was reconstructed using the delay method, and in Fig. 2(b) a three-dimensional projection of the attractor is shown. The structure of the attractors, related to the chaotic oscillation of the vocal folds, demonstrated that the irregular behavior observed in the time series was not due to noise.

In order to completely characterize the chaotic nature of the vocalization, the spectrum of Lyapunov exponents and the maximum Lyapunov exponent λ_1 were evaluated. In Fig. 2(c) values of the four exponents are reported and the presence of a positive exponent was detected. The accurate value of the MLE was computed by the Rosenstein–Kantz method, and a value of $\lambda_1=0.48$ was found by a linear regression of the curves in the region between 0 and 20 iterations.

The Kaplan–Yorke fractal dimension D_L of the attractor,[16] equal to $D_L=2.58$, confirms the high-dimensional fractal qualities of the strange attractor. See Table I

V. CONCLUDING REMARKS

The analysis method proposed in this letter was applied to the vocalization of an adult male of *Cervus elaphus corsicanus* and put in evidence the chaotic behavior of the irregular oscillations in the signal considered. A full characterization by means of attractor reconstruction, spectrum of Lyapunov exponents, and maximum Lyapunov exponent was performed. A positive value of MLE was found. Future work aimed at identifying different individuals through the discussed parameters will consist of the analysis of other vocalizations looking for a *vocal fingerprint* that may be useful in biodiversity monitoring.

ACKNOWLEDGMENTS

The authors are thankful to Dr. Carlo Murgia (Director of Oasi di monte Arcosu Sardegna) for providing the vocalizations.

[1] I. Wilden, H. Herzel, G. Peters, and G. Tembrock, "Subharmonics, biphonation, and deterministic chaos in mammal vocalization," Bioacoustics **9**(3), 171–196 (1998).

[2] I. Steinecke and H. Herzel, "Bifurcations in an asymmetric vocal fold model," J. Acoust. Soc. Am. **97**(3), 1874–1884 (1995).

[3] H. Herzel, D. Berry, I. Titze, and I. Steincke, "Nonlinear dynamics of the voice: Signal analysis and biomechanical modeling," Chaos **5**(1), 30–34 (1995).

[4] 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).

[5] I. Tokuda, T. Reide, J. Neubauer, M. J. Owren, and H. Herzel, "Nonlinear analysis of irregular animal vocalizations," J. Acoust. Soc. Am. **111**(6), 2908–2919 (2002).

[6] E. Ott, *Chaos in Dynamical Systems* (Cambridge University Press, Cambridge, UK, 1993).

[7] The TISEAN software package is publicy available at http://www.mpipks-dresden.mpg.de/~tisean/TISEAN_2.1/index.html

[8] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, UK, 1997).

[9] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer, Berlin, 1996).

[10] H. D. I. Abarbanel and M. B. Kennel, "Local false nearest neighbors and dynamical dimensions from observed chaotic data," Phys. Rev. E **47**, 3057 (1993).

[11] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," Phys. Rev. A **33**, 1134 (1986).

[12] J. M. Greene and J. S. Kim, "The calculation of Lyapunov spectra," Physica D **24**, 213–225 (1987).

[13] M. Sano and Y. Sawada, "Measurement of the Lyapunov spectrum from a chaotic time series," Phys. Rev. Lett. **55**, 1082 (1985).

[14] M. T. Rosenstein, J. J. Collins, and C. Luca, "A practical method for calculating largest Lyapunov exponent from small data set," Physica D **65**, 117 (1993).

[15] H. Kantz, "A robust method to estimate the maximal Lyapunov exponent of a time series," Phys. Lett. A **185**, 77 (1994).

[16] J. Kaplan and J. A. Yorke, *Chaotic Behaviour of Multidimensional Difference Equation*, Vol. 730 of Lecture Notes in Mathematics (Springer Verlag, Berlin, 1979).

Sound focusing in rooms. II. The spatio-temporal inverse filter

Sylvain Yon, Mickaël Tanter,[a] and Mathias Fink
Laboratoire Ondes et Acoustique, Université Paris VII/ESPCI, CNRS UMR 7587, 10 rue Vauquelin, 75231 Paris, Cedex 05, France

(Received 31 December 2003; revised 12 September 2003; accepted 20 September 2003)

The potential of time reversal processing for room acoustics has been extensively investigated in the companion of this paper [J. Acoust. Soc. Am. **113**(3), 1533–1543 (2003)]. In particular, a simple implementation of a loudspeaker time reversal antenna able to take advantage of the multiple reflections in reverberating rooms demonstrates its potential for audible range acoustics while improving focusing both in space and time. However, loss of information (e.g., sound absorption in walls or nonequalized bandwidths of the loudspeakers) during a time reversal experiment degrades the quality of time reversal focusing. In this paper, a more sophisticated technique called spatio-temporal inverse filtering is investigated that achieves time and space deconvolution of the propagation operator between the loudspeakers antenna and a set of microphones embedded inside the insonified volume. Theoretical and experimental comparisons between time reversal and inverse filter focusing are presented. Finally, advantages and limitations of both focusing approaches are highlighted. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1628247]

PACS numbers: 43.20.El, 43.60.Gk, 43.38.Hz [RLW] Pages: 3044–3052

I. INTRODUCTION

One of today's most challenging applications in room acoustics is the ability to control or focus the sound field in a predefined area of the propagation environment. Ideally, this sound control should be achieved with loudspeakers and microphones not necessarily located in the vicinity of the region of interest. This situation is common to various cases such as noise attenuation, compensation of room acoustic characteristics, or immersion in virtual sound environments (i.e., auralization).

This problem of sound control has been widely studied in terms of signal processing. Indeed, the propagation in a room between the set of emitting loudspeakers and the set of control points[1] can be assimilated to a multi-dimensional filter. One solution is then to obtain an inverse for this multi-dimensional filter and use it as a set of filters for emission.[2–4] From this ''signal processing'' point of view, the propagation in the medium is considered as a ''black box.'' From another more ''physical'' point of view, the spatial and temporal control of sound can be linked to the problem of focusing the acoustical energy on one (and possibly many) focal points.

In the companion paperd,[5] it has been shown that a technique such as time reversal could be very useful in such a context. Time reversal[6] provides an elegant way to focus acoustical energy spatially and temporally, even in complicated environments such as strongly reverberating rooms. Time reversal consists in using the time-reversed version of the Green's function associated with the desired focal point as a filter during emission. Thus, time reversal provides both spatial[7] and temporal[8] adaptive matched filtering of the medium. In some applications, this adaptive matched filter is equivalent to an inverse filter of the propagation medium. However, loss of information can appear during propagation (e.g., absorption losses, diffraction outside the recording surface,...). These information losses break the equivalence between a matched filter (time reversal) and an inverse filter of the propagation.[7,9] Indeed, this irreversible loss breaks the time reversal invariance and, consequently, time reversal does not ensure an optimal focusing.

In the context of audible range acoustics, it is interesting to compare these two different approaches (matched filtering and inverse filtering) of sound control. In order to study more specifically the problem in terms of propagation, we will introduce a spatio-temporal inverse filtering technique slightly different from the one proposed by Kirkeby *et al.*[3] This technique uses a singular value decomposition of the propagation operator, and has been proposed and studied as an adaptive focusing technique for a complex propagation environment by Tanter *et al.*[7,9] in the context of medical ultrasound. In particular, this approach allows one to link the singular value decomposition of the propagation operator to the plane wave decomposition of the wavefield.[9]

This paper is divided into three main sections: First, a matrix formalism is introduced to describe the slight differences between time reversal and inverse filtering. Next, experimental results are presented for the problem of focusing sound inside a room. These experiments show to what extent the quality of focusing is directly linked to the quality of reverberation in the room. In particular, it is shown that these sound control techniques can be used even if propagation conditions are very complex. For example, focusing can be achieved when loudspeakers are in one room and the area to control is in another one. Finally, we experimentally demonstrate how the multiple reverberations of sound in walls can be used to achieve a super-resolution focusing of sound. Such a super-resolution effect has already been previously explained by our group for multiple scattering and strongly reverberating media[10–12] and numerically studied by

[a]Author to whom correspondence should be addressed. Electronic mail: michael.tanter@espci.fr

Blomgren *et al.*[13] We now consider its application to the domain of audible range acoustics.

II. INVERSE FILTER AND TIME REVERSAL

A. The propagation operator $\{h_{mj}\}(t)$

We define the linear operator relating the J elements of the transducer array to the set of M control points located in the medium. We define an impulse response $h_{mj}(t)$ for each couple (m,j) comprising a control point and a loudspeaker. This impulse response $h_{mj}(t)$ is the signal received at the mth control point when a temporal delta function is applied on the jth loudspeaker. This response includes all the propagation effects in the considered medium, as well as the acousto-electric responses of the loudspeakers and microphones. This set of $M \times J$ temporal functions characterizes the *propagation operator* describing both the propagation environment and the transducer configuration.

Let $e_j(t)$, $1 \leqslant j \leqslant J$, be the J input signals for the loudspeakers. The output signals $f_m(t)$, $1 \leqslant m \leqslant M$, received by each microphone are

$$f_m(t) = \sum_{j=1}^{J} h_{mj}(t) \underset{t}{*} e_j(t), \quad 1 \leqslant m \leqslant M, \tag{1}$$

where $\underset{t}{*}$ is the temporal convolution operator. A temporal Fourier transform leads to the relation

$$\mathbf{F}(\omega) = \mathbf{H}(\omega)\mathbf{E}(\omega) \quad \forall \omega, \tag{2}$$

where $\mathbf{E}(\omega) = \{E_j(\omega)\}_{1 \leqslant j \leqslant J}$ is the column vector of the Fourier transform of the transmitted signals and $\mathbf{F}(\omega) = \{F_m(\omega)\}_{1 \leqslant m \leqslant M}$ is the column vector of the Fourier transform of the received signals. The transfer matrix $\mathbf{H}(\omega) = \{H_{mj}(\omega)\}_{1 \leqslant m \leqslant M, 1 \leqslant j \leqslant J}$ is the temporal Fourier transform of $\{\mathbf{h_{mj}}\}(t)$.

B. Time reversal and inverse filtering

As explained in the Introduction, the time reversal focusing technique consists in playing backwards the signals received at each loudspeaker location after emission of a single impulse at the desired focal spot location. Thus, the emitted signals can be written

$$e_j(t) = h_{m_0 j}(-t) = \{h_{mj}\}(-t) \underset{t}{*} \boldsymbol{\delta}_{m_0}(t), \tag{3}$$

where $\boldsymbol{\delta}_{m_0}(t)$ is null everywhere, except for $m = m_0$ and $t = 0$ [$\boldsymbol{\delta}_{m_0}(t)$ is a temporal and spatial delta function distribution]. After propagation in the medium, the signals received at the control points can be written

$$\mathbf{f}_m(t) = \left\{ \sum_{j=1}^{J} \mathbf{h}_{mj}(t) \underset{t}{*} \mathbf{h}_{mj}(-t) \right\} \underset{t}{*} \boldsymbol{\delta}_{m_0}(t). \tag{4}$$

In the frequency domain, Eq. (4) can be written for each frequency component ω:

$$\mathbf{F}(\omega) = \mathbf{H}(\omega)\mathbf{H}^{\dagger}(\omega)\Delta_{m_0}, \tag{5}$$

where † denotes the transpose conjugate operation, and $\Delta_{m_0} = [0 \cdots 010 \cdots 0]^{\dagger}$. Thus, the signals $\mathbf{F}(\omega)$ are linked to the description function Δ_{m_0} and the operator $\mathbf{HH}^{\dagger}$ known in the literature as the time reversal operator.[14]

The inverse filtering technique is based on an approximation of the inverse of $\mathbf{H}$. The problem of the inversion of the matrix propagator $\mathbf{H}$ is of course ill-conditioned. First, noise introduced during the acquisition of $\mathbf{H}$ would produce very large errors in the reconstructed results. However, even if the measurement was perfect, $\mathbf{H}$ is not necessarily invertible. Many techniques for the regularization of this problem have been described:[3] the regularization is achieved here using the singular value decomposition technique: $\mathbf{H}$ can be decomposed as

$$\mathbf{H} = \mathbf{UDV}^{\dagger}, \tag{6}$$

where $\mathbf{D}$ is a diagonal matrix containing the singular values of $\mathbf{H}$. The matrix inversion is only applied to the physically relevant singular vectors of $\mathbf{H}$, which gives a noise filtered approximation of $\mathbf{H}^{-1}$:

$$\hat{\mathbf{H}}^{-1} = \mathbf{V}\hat{\mathbf{D}}^{-1}\mathbf{U}^{\dagger} = \mathbf{V} \begin{bmatrix} \lambda_1^{-1} & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & & \vdots \\ \vdots & 0 & \lambda_{N(\omega)}^{-1} & & \vdots \\ \vdots & & & 0 & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix} \mathbf{U}^{\dagger}, \tag{7}$$

where N is the number of physically relevant singular values at the frequency ω. Thus, for a given field distribution objective $F_0(\omega)$, we can calculate the set of emission signals $E(\omega)$ that should give rise to $F_0(\omega)$ after propagation in the medium. After emission of $E(\omega)$ by the loudspeakers and propagation in the medium, the field obtained at the set of control points is

$$F_{pr}(\omega) = \mathbf{H}\hat{\mathbf{H}}^{-1}F_0(\omega). \tag{8}$$

It appears from Eqs. (7) and (8) that accurate reconstruction of the target field depends significantly on the number of relevant singular values. For a given frequency, the number of relevant singular values can be predicted for simple cases from the geometry of the problem. As explained by Tanter *et al.*,[7] in the case of free space propagation between two linear parallel arrays, this number N is given by the relation

$$N = 2\frac{L}{\lambda}\sin\left(\tan^{-1}\frac{D}{2F}\right), \tag{9}$$

where D and L are the width of the emitting and receiving arrays, respectively, and F is the distance between the two antennas. The number of degrees of freedom for the propagation operator, N, can be seen as the number of independent focal spots (lateral width $\sim \lambda F/D$) that can be created at the control points aperture L. It also represents the capacity given by the propagation medium and the geometrical disposition of the antenna to recreate a complicated sound field at these control points: if the projection of this desired field on the physically relevant singular vectors does not result in significant loss of information, the system will be able to create the desired sound field with a good accuracy. Thus, projecting on the main singular vectors the wavefield distri-

bution that we would like to achieve at the control points is a good way to predict the wavefield that will be obtained experimentally. Equation (9) relates N to the parameters of the problem in a free space environment. In the case of reverberating rooms, previous works[5,10] indicate that an increase in the number of degrees of freedom due to the increase of the emitting antenna apparent diameter leads to an improvement of the focusing pattern obtained by time reversal. The same result will be presented in this paper for room acoustics.

It has been shown in ultrasound that time reversal tends to achieve an inverse filter of the propagation when no loss of information (e.g., absorption effects, diffraction of the wavefield outside the array aperture, limited spatial directivity and bandwidth of the emitters and receivers) appear during the first step of a time reversal experiment.[7,9] In our case, such information losses are of particular interest as, for example, sound absorption occurs in walls. In addition, even high end loudspeakers have irregular transfer functions, leading to a degradation of the bandwidth that is not compensated and even amplified by time reversal techniques. Thus, in audible range acoustics, there should be some differences between the focusing quality obtained by time reversal and the spatio-temporal inverse filter. By introducing a time and space deconvolution process, the spatio-temporal inverse filtering process provides a way to compensate partially, sometimes completely, for electronics limitations but also spatial setup limitations. In Sec. III, experimental comparisons between time reversal and inverse filtering illustrate this purpose.

III. EXPERIMENTAL RESULTS IN ROOMS

A. Experimental setup

The first series of experiments have been carried out in an empty room with reverberant walls, floor and ceiling. The considered room dimensions are $5.75\times3.45\times3$ m^3 and the Sabine reverberation time (the standard reverberation time has been defined as the time for the sound to die away to a level 60 dB below its original level) is quite long due to the absence of acoustical energy absorbers such as furniture: $T_{60}=1{,}56$ s. In order to study the propagation operator, an antenna made of 16 Audax AX34 loudspeakers is placed near one of the walls (see Fig. 1). The set of control points consists of 25 points located along an axis parallel to the loudspeaker antenna. The impulse responses are acquired successively using a microphone translated to each control point position. The first step of each experiment consists in acquiring the propagation operator $h_{mj}(t)$. In order to obtain a good signal-to-noise ratio, the process is as follows: first, the microphone is placed in position j; then, a chirp signal is emitted from loudspeaker m, over the usable frequency range of the system (400–4000 Hz). The signal measured with the microphone is then correlated with the original chirp in order to obtain the impulse response between loudspeaker m and control point j.

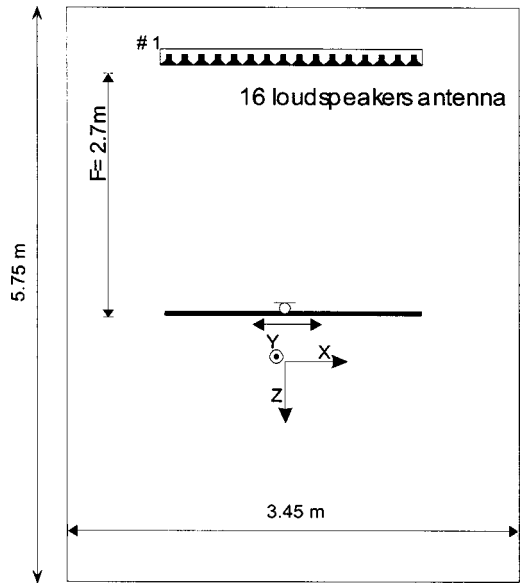


FIG. 1. Setup for the experiments in a reverberating room.

B. The propagation operator

The propagation operator corresponding to the setup described above is first acquired: To measure the impulse response $h_{mj}(t)$, the mth loudspeaker emits a linear chirp in the band 400 Hz to 4 kHz. The corresponding signals are then acquired at each of the J control points. Signals are then correlated with the initial chirp to obtain $h_{mj}(t)$ for each control point. This process is then iterated for each loudspeaker.

After adequate windowing and filtering, the frequency domain representation of propagation operator $H_{mj}(f)$ is obtained. In order to be able to invert matrix $\mathbf{H}$ for each frequency, we must identify which part of $\mathbf{H}$ is related to experimental noise, and which part is related to relevant propagation information. As explained above, only singular vectors corresponding to high singular values should be used during inversion. For this reason, it is important to be able to determine a simple method for differentiating the physically relevant singular values from the ones corresponding to noise. Such a differentiation is made easy by studying the distribution of the singular values plotted in decreasing order as a function of frequency (referred to as the singular value space). Figure 2 gives an example of this representation.

As explained previously, the number of relevant singular values of the propagation operator corresponds to the number of available degrees of freedom for the reconstruction of the desired field. In Fig. 2, the singular value space is calculated in two configurations: in the first configuration, the direct wavefront is detected in the temporal propagation operator, and this single wavefront is used to calculate the singular value space. In others words, no reverberations are taken into account in the free space propagation operator. This is an easy way to obtain a good approximation of the propagation in free space, with the same configuration for emitting transducers and control points. Figure 2(a) presents the singular values distribution of this free space propagation operator. In the second configuration, the complete propagation operator (direct propagation+reverberations) is taken into account

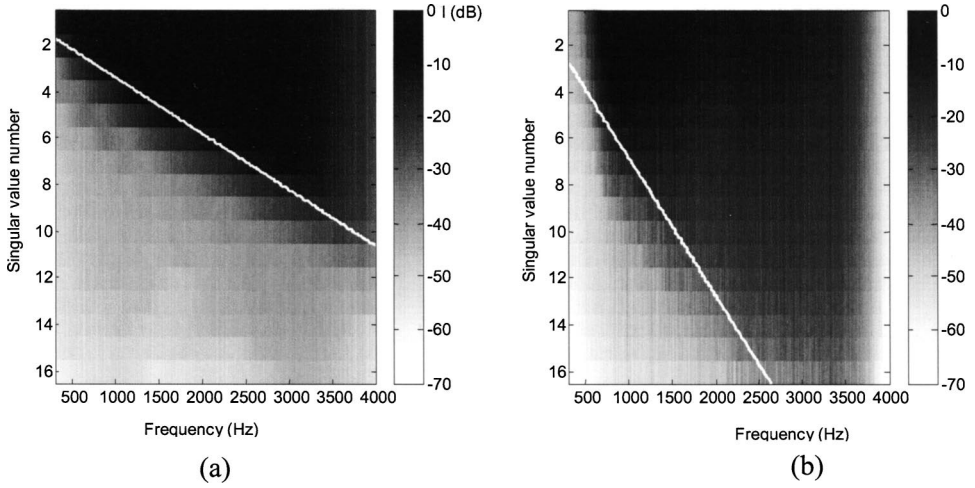


FIG. 2. Singular values space (a) for the propagation operator in free space (after cancellation of the reverberations) and (b) in the reverberating room.

during the singular value decomposition. The singular value space in Fig. 2(b) corresponds to the entire propagation operator when propagation occurs in the reverberant room. As one can notice by comparing Figs. 2(a) and (b), the fact that propagation occurs in a reverberating environment strongly increases the number of degrees of freedom available for each frequency.

This increase of the number of freedom degrees can be explained from a ''diffraction'' point of view: in the case of free space propagation, the finite aperture of the emitting antenna limits the number of independent focal spots (with a lateral resolution $\lambda F/D$) that can be created in the control points aperture L with respect to Eq. (9). Moreover, according to Eq. (9), N should increase linearly with frequency:

$$N = 2f\frac{L}{c}\sin\left(\tan^{-1}\frac{D}{2F}\right), \tag{10}$$

where c is the sound speed. This linear frequency dependence on N is clearly found in the singular value space of the experimental propagation operator in Fig. 2(a). The white line in Fig. 2(a) corresponds to the theoretical Eq. (10) and clearly fits the experimental separation between noise space and singular value space. In the case of a reverberating room, the numerous reflections against the walls create a set of virtual sources corresponding to the mirror images of real loudspeakers in regards to each wall. An observer located at the focal point does not only hear the real loudspeakers but also their images through the wall mirrors, as shown in Fig. 3. Thus, reflections on boundaries allow one to create a virtually infinite array of emitting transducers in every direction. In this case, the focal spot size should be only limited by the classical diffraction minimal width $\lambda/2$. Consequently, the number of degrees of freedom of the propagation operator should reach its maximum value $N = 2L/\lambda = 2fL/c$. Of course, in these experiments, walls are not perfectly reflecting and the amplitude of signals decreases progressively with the number of walls reflections. However, as it results in a slight and progressive loss of amplitude, only the relative weight of each eigenvalue is affected at a given frequency. Thus, Eq. (10) remains valid even in experimental conditions as it deals with the total number of relevant eigenvalues at a given frequency. The theoretical line $N(f) = 2fL/c$ is plotted in Fig. 2(b) and shows a good agreement between the theoretical description and real measurements. Moreover, it confirms that we should be able to create spots of $\lambda/2$ diameter at the set of control points by using correctly the walls reverberations.

The results given in Fig. 4 provide confirmation of the above analysis. The spatial frequency spectrum of the different singular vectors is plotted for $f_0 = 1500$ Hz. The dashed lines separate the singular vectors corresponding to noise from the ones corresponding to physically relevant data. It appears that the spatial frequency bandwidth of physically relevant singular vectors is larger in the case of reverberant rooms, which includes the higher spatial frequencies corresponding to plane wave angles of arrival that cannot be attained when considering only the emitter array in free space. Thus, working in a perfectly reverberant environment allows us to reconstruct a sound field with a large spatial frequency spectrum. In practical applications, it means that such an environment makes it possible to simulate a sound field with any direction of arrival regardless of the positions of the emitters. Of course, this description corresponds to a case where no attenuation occurs during the reflections. In practical systems, however, attenuation implies that the spatial frequencies corresponding to a significant number of reflections

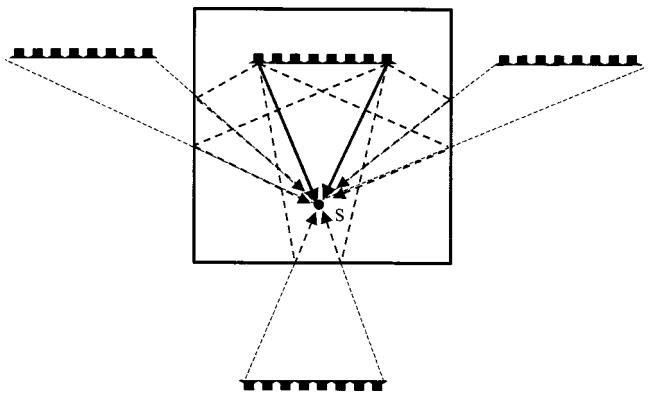


FIG. 3. Principle of the mirror images: an observer S located at the focal point does not only hear the real loudspeakers but also their images through the wall mirrors.

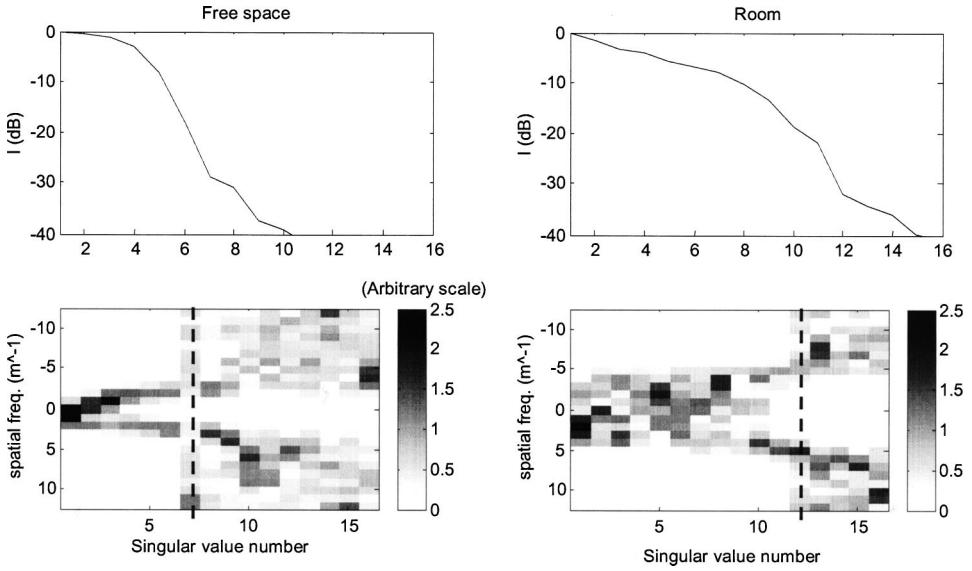


FIG. 4. Singular values weights and corresponding spatial Fourier transform in the emitting plane of the singular vectors of the propagation matrix for the frequency $f_0 = 1500$ Hz.

are sometimes not available. In that case, the spatial positions of the emitters play an important role.

To be able to invert the propagation operator, a separation must be established between the physically relevant part and the noise part of the propagation operator. This separation is easily established from the results presented in Figs. 2 and 4. Indeed, there is a brutal variation in the intensity of the singular values from 0–10 dB to less than −30 dB and this variation corresponds to the appearance of noise in the singular vectors. Thus, in our series of experiments, a limit value of −25 dB will be used for considering that singular values are physically relevant, and may be used in the inversion.

C. Focusing and sound control capacity

Once the propagation operator has been acquired and a good approximation of its inverse form has been calculated as explained above, emission filters $e_j(t)$ are easily calculated for a given field distribution objective $f_{m0}(t)$. Though the desired field can be whatever is needed, our initial interest is to obtain a single focal spot, since any field distribution can be deducted from this simple one.

Through the inverse filtering process and the backpropagation operation [Eq. (8)], we obtain a least mean squares approximation of the desired sound field $f_{m0}(t)$. No specific condition is imposed by the algorithm on this sound field, but physical characteristics of the propagation medium and of the electronic emission/reception system have to be taken into account. Thus, $f_{m0}(t)$ must be chosen according to those characteristics, otherwise the algorithm will use too much energy in trying to recreate a nonphysical field, resulting in an increased noise level. To take into account these conditions for the case of propagation in the room, two points must be respected: (1) The frequency content of the sound field must be included in the bandwidth of the electronic system and (2) the focal spot width cannot be less than $\lambda/2$, where λ is the wavelength corresponding to the minimal frequency in the desired sound field.

Diagrams presented in Fig. 5 show the spatial focusing quality that can be achieved when the conditions explained above are respected. The temporal objective is chosen to be a pulsed signal of bandwidth of 1–3 kHz, and the desired focal spot width is fixed to 17 cm. We also consider the time reversal focusing which is obtained by performing a time reversal experiment with an initial signal identical to the one used as the objective for the inverse filter technique. Figure 5 clearly shows the improvement that can be achieved with inverse filtering compared to time reversal: the spatial sidelobe level is at least 5 dB lower for the spatio-temporal inverse filtering process.

Until here, we have considered a field objective consisting of 25 control points. In these conditions, the inverse filter and time reversal processing are not fairly compared: the quantity of information used to obtain the presented results is 25 times more important for the inverse filtering method. However, this inverse filtering process can also be achieved using fewer control points. This aspect is presented in Fig. 6 describing the evolution of the sidelobes level obtained with both techniques versus the bandwidth and the number of

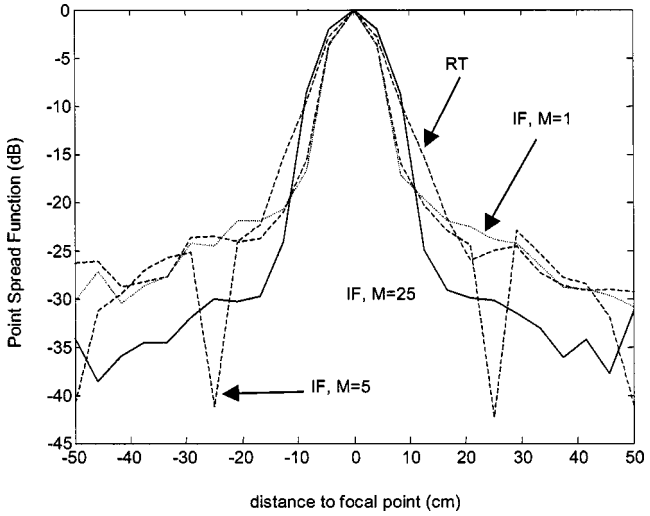


FIG. 5. Directivity patterns obtained with time reversal and inverse filtering for different numbers M of control points.

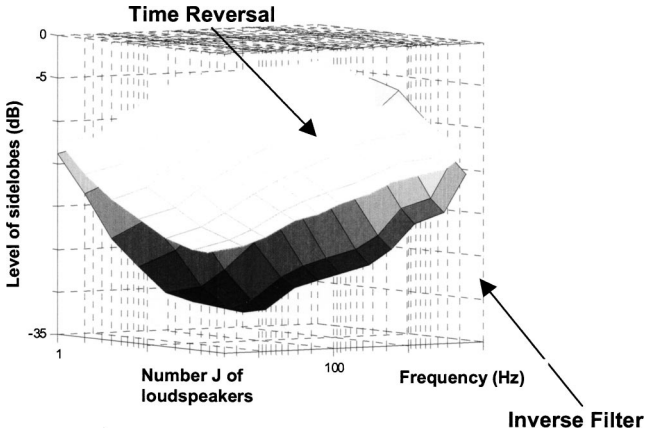


FIG. 6. Evolution of spatial sidelobe level $SL_{(dB)}$ as a function of the number of used transducers J and of the bandwidth BW, for time reversal and inverse filtering. Measurements are averaged over 16 repetitions.

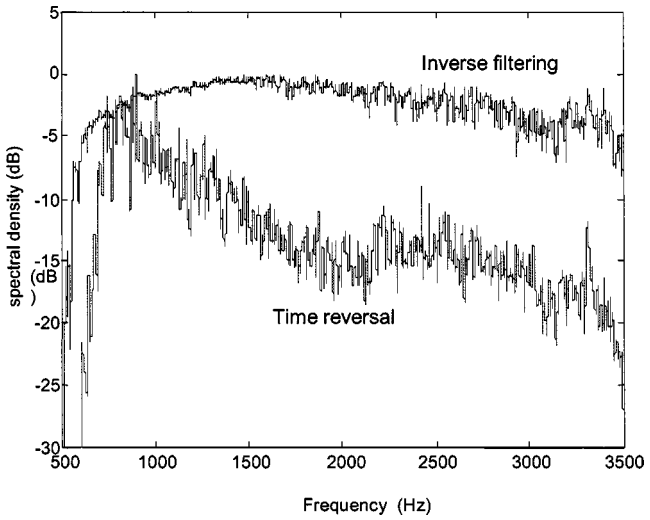


FIG. 7. Spectral content at the focal point after time reversal and inverse filtering.

transducers used. The evolution of the sidelobes level was extensively discussed for the time reversal approach in the companion of this paper[5] and we focus here on the differences between the time reversal and inverse filtering methods. Figure 6 shows that time reversal and the inverse filter give globally similar results in terms of spatial focusing for the same quantity of information. However, as it is presented in Fig. 5 for $M=5$, one of the advantages of spatio-temporal inverse filtering is that it allows us to force certain points to have a null field, with an attenuation up to -40 dB. These results correspond to signals that use the whole available frequency bandwidth and this frequency diversity is essential to obtain such spatial focusing results. Figure 6 clearly illustrates this point. It represents the level of the sidelobes obtained by time reversal and inverse filtering as a function of the used frequency bandwidth and as a function of the number J of loudspeakers emitting the time reversed or inversed signals. The improvement of the level of the sidelobes by inverse filter focusing compared to time reversal is only significant when the frequency bandwidth of the emission signal becomes important. Unfortunately, common signals such as speech are characterized by the fact that in practice only a few frequencies (corresponding to formants) are emitted. As explained in Ref. 5, such signals are quasi-monochromatic from the point of view of diffraction, and lead to similar point spread functions whatever technique is used. Thus, in terms of spatial distribution of the acoustic energy, focusing of speech (quasi-monochromatic) signals is almost identical when using the time reversal or inverse filtering methods.

However, important differences between the time reversal and inverse filtering methods occur in the time domain. As seen in Fig. 7, the frequency spectrum of the signal obtained at focus is quite different for both focusing methods. Contrary to time reversal, the inverse filter focusing technique introduces a frequency bandwidth equalization by achieving a space and time domain deconvolution. This deconvolution corrects for the non-flat amplitude response of the loudspeakers but also for the frequency dependent absorption effects occurring during propagation. One should notice in Fig. 8 that this space and time deconvolution achieved by the inverse filtering technique is not equivalent to the combination of time reversal with bandwidth equalization (in others words, a inverse filter technique with a single control point located at focus). At focus, the temporal sidelobes are almost 20 dB lower for the inverse filtering technique than for time reversal (Fig. 8). The focal signal shape obtained by inverse filtering is sharper, thus permitting attainment of a very good temporal approximation of the desired impulse signal.

D. Robustness with respect to changes in the acoustic environment

In practical systems, sound control needs to be robust, especially in the frequency domain of potential applications:

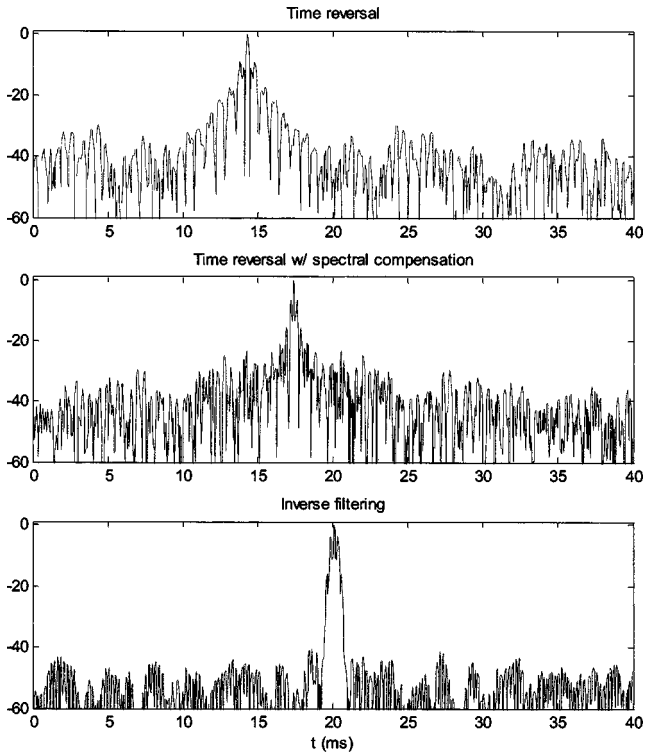


FIG. 8. Temporal sidelobes and temporal recompression observed at the focal spot, for the different techniques. The signal amplitude is presented in a dB scale.

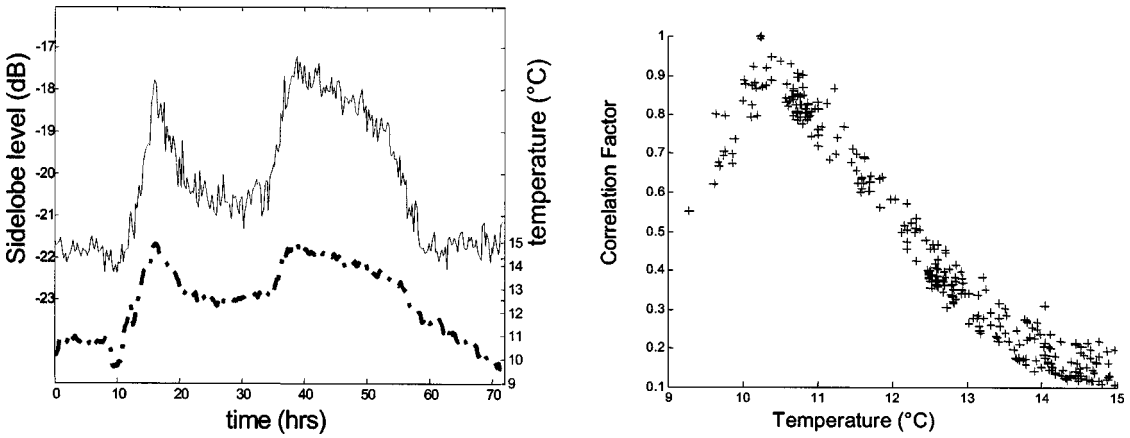


FIG. 9. Robustness of focusing versus temperature and time: initial time $t=0$ corresponds to the acquisition of the propagation operator. The solid line corresponds to the evolution of the level of the sidelobes while focusing every 15 min at the same focus location. The dashed line corresponds to the evolution of temperature operating at the same time.

sound control has to be achieved in media where objects will eventually move and where propagation conditions can vary in time. The influence of such an evolution has been measured in two ways: first, the quality of focusing was measured with different perturbations in the room, such as moving persons, or the introduction of sound absorbing objects. These tests did not show any significant variation of the focusing in terms of the spatial and temporal sidelobes level or the focal spot width.

In a second experiment, the focusing robustness regarding slow variations of the propagation medium is tested over a long time period (48 h). To study this evolution, an initial propagation operator is acquired at a time $t=0$ (room temperature $T=11\,^{\circ}\mathrm{C}$) using the technique described above. From this operator, a set of emission signals is calculated by inverse filtering for a sound field objective $f_{m0}(t)$ such as described in Fig. 5 (desired focal spot width is λ). This set is then reemitted every 15 min in order to study the evolution of the sidelobes level versus time. A preliminary experiment shows that during 48 h, no important degradation of the focusing occurs. However, the inverse filtering technique, like time reversal, is quite sensitive to changes of the medium characteristics, such as sound speed, that induce a cumulative errors during wave propagation. Figure 9 presents the results obtained when this experiment is conducted during 48 h, with variations of the temperature of the room (i.e., variations of the sound speed, $\delta c/\delta t \sim 0.6$ m.s.$^{-1}$.°C^{-1}). An interesting point is that the focusing quality described by the level of the sidelobes when focusing at a given location returns back to its initial value when temperature is equal to the temperature of the initial measurement ($T=11\,^{\circ}\mathrm{C}$). This point can be also illustrated by calculating the correlation coefficient between the initial impulse response relating one loudspeaker to the microphone located at focus and the same impulse response acquired later during the 24-h experiment. As one can notice, this correlation coefficient reaches 1 when the temperature reaches the same value as the one corresponding to the initial acquisition of the impulse response. Thus, the problem of sensitivity to sound speed variations can easily be overcome by acquiring different data banks corresponding to different temperature conditions in the room.

E. Example of sound field shaping

One of the possible applications of sound control techniques is to recreate an approximation of a desired complex sound field on all the control points. In this context, the inverse filtering technique should provide better results since a least mean square approximation is made over the whole set of control points, whereas when time reversal is used, the approximations are calculated separately for each control point.

For this reason, the inverse filtering process is expected to be much more efficient than time reversal in such a context. Figure 10(a) corresponds to the signals received in the room at the control points when loudspeaker 1 emits a short pulsed signal. As one can notice, a lot of successive reverberations are recorded at the control points after arrival of the direct path signal. Results shown in Fig. 10(b) correspond to the sound field observed at the control points in the reverberating room, when trying to recreate the effect of a virtual pulsed sound source located at the position of loudspeaker 1

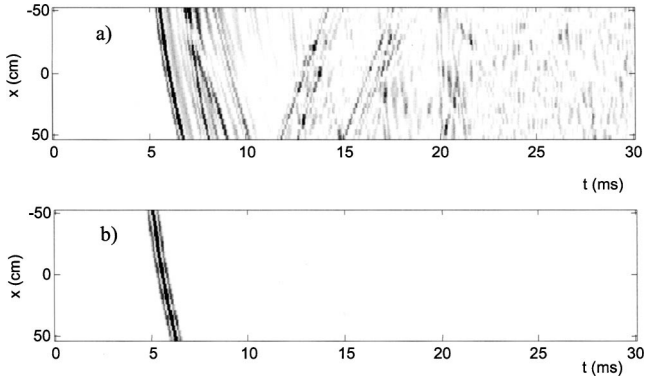


FIG. 10. Bscan representation (temporal signals received at the set of control points) (a) when loudspeaker 1 emits in the reverberating room and (b) when the inverse filter is used on the array of loudspeakers in order to recreate at the control points the virtual sound field corresponding to loudspeaker 1 emitting in free space.

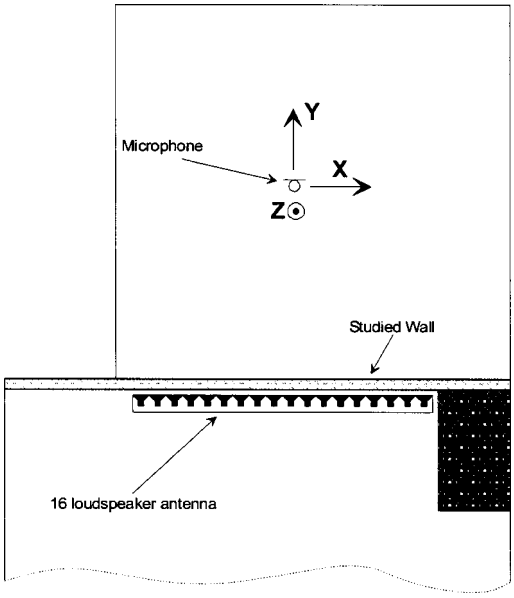


FIG. 11. Experimental setup for focusing through a wall.

(see Fig. 1) and emitting in free space. In other words, the aim is here to use the array of loudspeakers to cancel reverberations in a chosen time domain at the set of control points. As seen in Fig. 10(b), in the vicinity of the control points (vertical axis), the loudspeaker seems to emit in free space, without any reverberation. The horizontal axis corresponds to time and the signal amplitude is represented in a gray scale. Figure 10(b) is obtained with the inverse filtering technique. The reverberations induced in the room have been completely cancelled and the wavefront received at the control points seems to come from a source in a free space environment: the level of the temporal noise is more than 40 dB below the maximal amplitude received at each control points. However, when trying to recreate a similar sound field with a time reversal technique, temporal sidelobe level is here more important than those observed for a single focal spot, and the pulsed signal is wider in time. These effects, due to the fact that time reversal does not provide a way to consider the sound field as a whole, make the inverse filtering concept an interesting technique for such applications.

IV. EXPERIMENTAL RESULTS THROUGH WALLS

A. Experimental setup

In the previous section, focusing was achieved in a reverberant room. This medium is complex by itself, because of reflections and diffraction effects on every object in the room, but attenuation is not very strong. Thus, the situation corresponds to the conditions needed to achieve time reversal focusing with good precision, and differences obtained in terms of quality of focusing between the inverse filtering approach and time reversal focusing are not very important. In the companion paper,[5] it was shown that time reversal focusing could also be achieved through a wall in an experimental setup described in Fig. 11. This situation proved to be a lot more challenging for time reversal, so we try here to compare it to the optimal results by the inverse filtering technique.

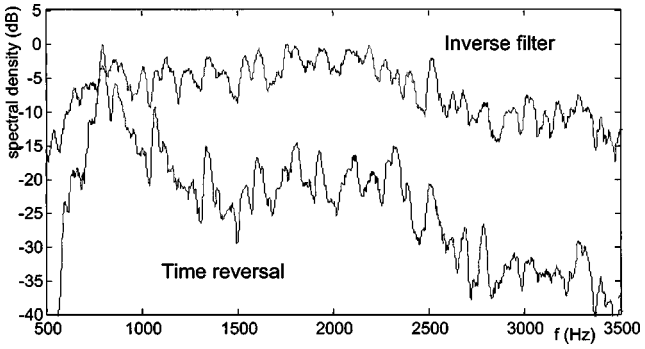


FIG. 12. Spectral content of the signal received at focus after time reversal or inverse filtering (with 25 control points).

The acoustic attenuation[15] between the central loudspeaker of the antenna and the central position of the microphone is 20 ± 3 dB. The 8-cm-thick wall by itself is made of hollow bricks and plaster. However, such a wall provides a normalized acoustic attenuation of $\sim$41 dB. Thus, the most important part of acoustical energy that manages to reach the focal point comes from ''leaks,'' probably via doors and windows, but also from lateral propagation through walls.

These complex conditions for propagation lead also to important variations of attenuation with frequency as shown in Fig. 12. So, inverse filtering could be efficient in such a case, whereas time reversal will be handicapped because it does not compensate for those variations. Experiments presented below have been achieved with exactly the same electro-acoustical and electronic devices as those used for the experiments in the room.

B. Results

As in the case of the reverberating room, relative performance of time reversal and inverse filter can be measured when trying to recreate a single focal spot, such as the ones described in Sec. III C. Theoretically, results obtained with an inverse filter should be very similar to those obtained in rooms, whereas those obtained with a time reversal technique are degraded because of attenuation during propagation. Results shown in Figs. 13 and 14 prove to be quite different from theoretical expectations.

Results obtained with an inverse filtering technique when focusing through walls prove to be as remarkable as results of Sec. III C, when considering temporal compression and noise level of obtained signals. Indeed, as it is shown in Figs. 13 and 14, inverse filtering allows one to obtain a very

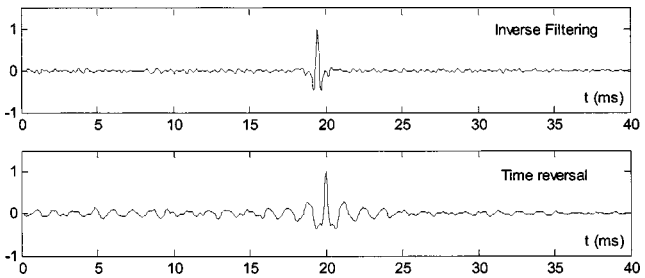


FIG. 13. Temporal recompression at the focal spot using time reversal or inverse filtering (with 25 control points).

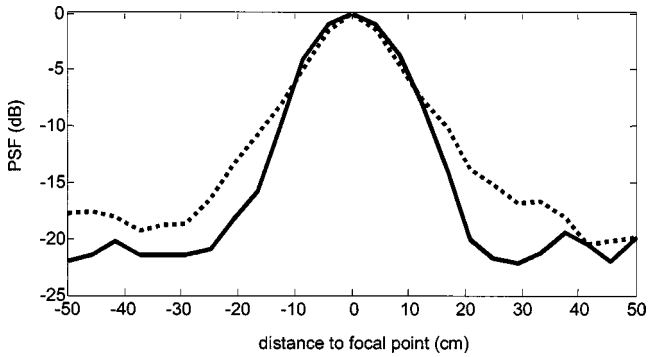


FIG. 14. Directivity patterns obtained using inverse filtering (solid line) and time reversal (dotted line) through the wall.

good temporal approximation of the desired impulsive signal at the focal spot. Specifically, the spectral density can be corrected with good accuracy over the whole bandwidth.

Figure 14 shows, however, that only a slight improvement is provided by inverse filtering compared to time reversal in terms of spatial focusing. This surprising result is only due to a technical limitation of our electronics: time reversal is a matched filter and so maximizes the energy received at the focus[7,8] whereas the inverse filter corresponds to a least mean squares approximation of the desired sound field. Thus, energy obtained at the focus for inverse filtering is lower than the one obtained with time reversal, and electronic limits of the emitting and recording devices used in the experiment were unfortunately reached. However, with a new emission/reception system providing a better dynamic range, inverse filtering should reach through the wall, similar to performances obtained in the room. Indeed, the measured sidelobe levels do not correspond to a conceptual limit, but are only due to an electronic noise limited value. Thus, while using a system with a better dynamic range, the improvement of the level of the sidelobes by inverse filtering compared to time reversal should be about 20 dB.

V. CONCLUSION

Inverse filtering is based on the knowledge of the entire propagation operator between a set of emitters and a set of control points located in the medium. For a given target sound field at the control points, this technique provides theoretically the best approximation of the field in a least mean square sense by achieving both space and time deconvolutions of the propagation operator. However, time reversal remains an interesting technique to approximate such a field because of its simplicity of implementation.

In audible range acoustics, it appears that the main advantage of inverse filter over time reversal is its capacity to compensate for the variations of the spectral densities of the transfer functions. Noise reduction at focus can reach more than 20 dB.

The spatio-temporal inverse filter should be even more adapted to complicated environments suffering propagation losses and to systems relying on low-quality electro-acoustical devices with, however, the cost of much more signal processing and the need for more amplification power.

[1] Control points correspond to a set of measurement points whose signals allow a reconstruction of the sound field in the zone of interest.

[2] V. Martin, P. Vignassa, and B. Peseux, "Numerical Vibro-acoustic modelling of aircraft for the active acoustic control of interior noise," J. Sound Vib. **176**(3), 307–332 (1994).

[3] O. Kirkeby and P. A. Nelson, "Digital Filter Design for Inversion Problems in Sound Reproduction," J. Audio Eng. Soc. **47**(7/8), 583–595 (1999).

[4] Y. Kahana, P. A. Nelson, O. Kirkeby, and H. Hamada, "A multiple microphone recording technique for the generation of virtual acoustic images," J. Acoust. Soc. Am. **105**, 1503–1516 (1999).

[5] S. Yon and M. Fink, "Sound focusing in reverberating rooms and through walls: the time reversal approach," J. Acoust. Soc. Am. accepted in November 2002.

[6] M. Fink, "Time-reversed Acoustics," Phys. Today **50**(3), 34–40 (1997).

[7] M. Tanter, J.-L. Thomas, and M. Fink, "Time reversal and the Inverse filter," J. Acoust. Soc. Am. **108**, 223–234 (2000).

[8] C. Dorme and M. Fink, "Focusing in transmit-receive mode through inhomogeneous media: the time reversal matched filter approach," J. Acoust. Soc. Am. **98**, 1155–1162 (1995).

[9] M. Tanter, J.-F. Aubry, J. Gerber, J.-L. Thomas, and M. Fink, "Optimal focusing by spatio-temporal inverse filter, Part I: Basic principles," J. Acoust. Soc. Am. **101**, 37–47 (2001).

[10] P. Roux and M. Fink, "Time reversal in a waveguide: Study of the temporal and spatial focusing," J. Acoust. Soc. Am. **107**, 2418–2429 (2000).

[11] A. Derode, P. Roux, and M. Fink, "Robust acoustic time reversal with high-order multiple scattering," Phys. Rev. Lett. **75**(23), 4206–4209 (1995).

[12] J.-F. Aubry, M. Tanter, J. Gerber, J.-L. Thomas, and M. Fink, "Optimal focusing by spatio-temporal inverse filter. Part II. Experiments," J. Acoust. Soc. Am. **101**, 48–58 (2001).

[13] P. Blomgren, G. Papanicolaou, and H. Zhao, "Super-resolution in time reversal acoustics," J. Acoust. Soc. Am. **111**, 230–248 (2000).

[14] C. Prada and M. Fink, "Eigenmodes of the time reversal operator: a solution to selective focusing in multiple target media," Wave Motion **20**, 151–163 (1994).

[15] Acoustic attenuation D_T is defined as $D_T = L_1 - L_2$, where L_1 and L_2 are the mean sound levels measured in each room, for a white noise.

Transient fluid–structure interaction of elongated bodies by finite-element method using elliptical and spheroidal absorbing boundaries

S. K. Bhattacharyya[a] and R. Premkumar
Department of Ocean Engineering, Indian Institute of Technology, Madras 600 036, India

(Received 12 September 2002; revised 4 July 2003; accepted 2 September 2003)

In a domain method of solution of exterior scalar wave equation, the radiation condition needs to be imposed on a truncation boundary of the modeling domain. The Bayliss, Gunzberger, and Turkel (BGT) boundary dampers, which require a circular cylindrical and spherical truncation boundaries in two-(2D) and three-(3D)-dimensional problems, respectively, have been particularly successful in the analysis of scattering and radiation problems. However, for an elongated body, elliptical (2D) or spheroidal (3D) truncation boundaries have potential to reduce the size of modeling domain and hence computational effort. For harmonic problems, such extensions of the first- and second-order BGT dampers are available in the literature. In this paper, BGT dampers in both elliptical and spheroidal coordinate systems have been developed for transient problems involving acoustic radiation as well as fluid–structure interaction and implemented in the context of finite-element method based upon unsymmetric pressure-displacement formulation. Applications to elongated radiators and shells are reported using several numerical examples with excellent comparisons. It is demonstrated that significant computational economy can be achieved for elongated bodies with the use of these dampers. [DOI: 10.1121/1.1621010]

PACS numbers: 43.20.Px, 43.20.Tb, 43.40.Rj [ANN] Pages: 3053–3064

I. INTRODUCTION

When a vibrating structure is submerged in a fluid medium whose extent in all directions is large, i.e., several times its characteristic dimension and acoustic wavelengths involved, the medium may be considered infinite or unbounded. Radiation and scattering by such structures are called exterior acoustic problems. When their elastic behavior is also accounted for, it leads to fluid-structure interaction (FSI) problems. Application of FSI cuts across many engineering disciplines, some examples being: radiated noise from ships, scattering of shock waves from an explosion by structures, detection by ship and submarine mounted sonars, noise shielding by towed sonar arrays, acoustic transducers, etc.

In applying a domain method (finite-difference method or FDM and finite-element method or FEM) for unbounded domain problems, one has to restrict the modeling domain to a finite one by a truncation surface or boundary ''sufficiently'' away from the structure. Then, one has to impose a suitable approximate absorbing (also called ''nonreflecting,'' ''radiation,'' and ''transmitting'') boundary condition on this boundary to simulate the asymptotic radiation condition at far field. A variety of boundary and domain methods for acoustics in unbounded domain is available. In the FE context, one of the successful approximations of nonreflecting boundary conditions had been the well-known BGT (Bayliss, Gunzburger, and Turkel, 1982) conditions, also called boundary dampers, which are local in both space and time. These conditions are well established for circular radiation boundary for 2D and spherical radiation boundary for 3D problems (Bayliss and Turkel, 1980; Bayliss *et al.*, 1982). But, for highly elongated bodies, the use of circular or spherical radiation boundary surrounding the structure makes the computational domain large with attendant penalty of computational time. If this boundary is chosen as elliptical (in 2D) or spheroidal (in 3D), the size of computational domain can be reduced significantly (Fig. 1). Grote and Keller (1995) have done this for both first- and second-order BGT dampers on elliptical and spheroidal absorbing boundaries in frequency domain, and used them in harmonic acoustic scattering problems using FDM. However, their formulation only caters to ''head-on'' incident waves, i.e., the formulation is accurate only for a specific direction. Bhattacharyya *et al.* (2001a) developed first- and second-order elliptical BGT dampers in frequency domain to study water wave diffraction problems using FEM. In the process, they generalized the Grote and Keller formulation so that the damper is omnidirectional. The idea of elliptical, spheroidal, and ellipsoidal boundaries has also found application in infinite elements (Burnett and Holford, 1998a, b; Cipolla and Butler, 1998) and wave envelope method (Astley, 1998a, b) to represent the far field. Treatments of more general convex radiation boundaries are also in evidence (Tezaur *et al.*, 2002; Djellouli *et al.*, 2000; Antoine and Barucq, 1999; Meade *et al.*, 1995) in acoustics and electromagnetics.

However, there seems to be no attempt to use the computationally attractive elliptical and spheroidal absorbing boundaries, suitable for elongated bodies, in transient FSI problems in unbounded domain. The present work is an attempt towards this direction. In this paper, finite-element approximations of the first-order BGT-type elliptical and spheroidal dampers are developed, implemented, and used for studying transient FSI of deeply submerged shells and vali-

[a]Electronic mail: skbh@iitm.ac.in

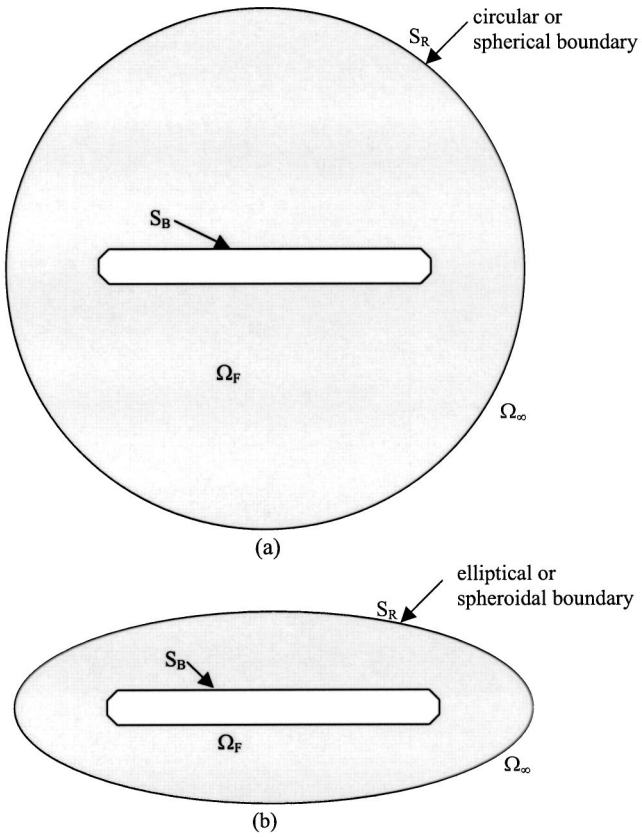


FIG. 1. Definition sketch: (a) Circular or spherical radiation boundary; (b) Elliptical or spheroidal radiation boundary.

dated using numerical examples, including a few for which results are not available in the literature.

II. GOVERNING EQUATIONS AND BOUNDARY CONDITIONS

The exterior transient acoustic problem (Fig. 1) can be modeled as a boundary value problem governed by the wave equation in computational fluid domain (Ω_F)

$$\nabla^2 p = \frac{\ddot{p}}{c^2} \text{ in } \Omega_F, \tag{1}$$

with Neumann (on S_{B1}) and Dirichlet (on S_{B2}) boundary conditions on the body boundary S_B ($=S_{B1}\cup S_{B2}$) as

$$\frac{\partial p}{\partial n} = -\rho_F \overline{\ddot{u}_n} \quad \text{on } S_{B1}, \tag{2a}$$

and

$$p = \overline{p} \quad \text{on } S_{B2}, \tag{2b}$$

subject to the radiation condition at the far field (S_R) given by

$$r^{(h-1)/2}\left(\frac{\partial p}{\partial r} + \frac{\dot{p}}{c}\right) = \mathrm{O}\left(\frac{1}{r}\right) \quad \text{on } S_R. \tag{3}$$

In the above, p is acoustic pressure, t is time, c is sound speed in fluid medium, r is outward radial distance from a suitable reference point with respect to which waves are assumed to be symmetric, ρ_F is density of acoustic fluid, $\overline{\ddot{u}_n}$ is prescribed acceleration normal to body surface S_{B1}, $\overline{p}$ is prescribed pressure on body surface S_{B2}, n is outward normal on S_B, and $h=2$ (2D) or 3 (3D). The infinite acoustic fluid domain Ω_∞ outside S_R is assumed homogeneous, isotropic, and unloaded. The radiation condition in Eq. (3) ensures outgoing radiated or scattered waves as $r\to\infty$.

III. BOUNDARY DAMPERS

Bayliss and Turkel (1980) and Bayliss *et al.* (1982) have proposed a hierarchy of boundary dampers, which is used in the present work as an alternative to the limiting condition, Eq. (3), since it cannot be readily accommodated in FE analysis. These are

$$B_m = \prod_{J=m}^{1}\left(\partial_r + \frac{1}{c}\partial_t + \frac{2J-J}{r}\right), \tag{4}$$

where $\partial_r = \partial/\partial r$, $\partial_t = \partial/\partial t$, and $L=3/2$ (2D) or 1 (3D). The boundary condition (with higher accuracy with increasing m) can be stated as

$$B_m p = 0. \tag{5}$$

In FE context, Eq. (5) has to be transformed to an equivalent Neumann condition of general form using FE damper operator S_m as

$$\nabla p \cdot \mathbf{n} = S_m p \quad \text{on } S_R. \tag{6}$$

IV. FINITE-ELEMENT FORMULATION

Define the functional

$$I(p) = \int_{\Omega_F} \nabla p \cdot \nabla p \, d\Omega + \frac{1}{c^2}\int_{\Omega_F} \dot{p}^2 d\Omega + \rho_F \int_{S_{B1}} \overline{\ddot{u}_n} p \, ds. \tag{7a}$$

The Hamilton's principle for the problem can be written as

$$\int_{t_1}^{t_2}\left(\delta I - \int_{S_R} S_m p \, \delta p \, ds\right) dt = 0. \tag{7b}$$

Carrying out the variations as indicated in this equation, we get

$$\int_{\Omega_F} \nabla p \cdot \delta(\nabla p) d\Omega + \frac{1}{c^2}\int_{\Omega_F} \dot{p}\,\delta\dot{p}\, d\Omega + \rho_F \int_{S_{B1}} \overline{\ddot{u}_n}\,\delta p\, ds - \int_{S_R} S_m p\,\delta p\, ds = 0. \tag{8}$$

The prescribed pressure, $\overline{p}$, on S_{B2} is a kinematic boundary condition and therefore does not enter the variational equation, Eq. (8). Using FE approximation for p as

$$p = \sum_e N_F^e p_e, \tag{9}$$

where p_e is the vector of nodal pressures of an element (e denotes element), N_F^e is the shape function vector of a fluid element, and the summation is implied over all elements and substituting this approximation in Eq. (8), the various terms can be written as

$$\int_{\Omega_F} \nabla p\, \delta(\nabla p)\, d\Omega = \sum_e \delta p_e^T \left(\int_{\Omega_F^e} \nabla N_F^{eT} \nabla N_F^e \, d\Omega \right) p_e, \tag{10a}$$

$$\int_{\Omega_F} \dot{p}\, \delta\dot{p}\, d\Omega = \sum_e \delta p_e^T \left(\frac{1}{c^2} \int_{\Omega_F^e} N_F^{eT} N_F^e \, d\Omega \right) \ddot{p}_e, \tag{10b}$$

$$\rho_F \int_{S_{B1}} \overline{\ddot{u}_n}\, \delta p\, ds = \sum_e \delta p_e^T \left(\rho_F \int_{S_{B1}^e} N_F^{eT} \overline{\ddot{u}_n}\, ds \right). \tag{10c}$$

Equation (10b) has been obtained using by-parts integration and then dropping one term due to initial conditions. The last term in Eq. (8) has to be expanded according to the type of radiation damper adopted as presented in the next two sections.

V. TWO-DIMENSIONAL ELLIPTICAL DAMPER

For $m=1$, the condition given by Eq. (5), using Eq. (4), can be written as

$$B_1 p = \left(\partial_r + \frac{1}{c}\partial_t + \frac{1}{2r} \right) p = 0. \tag{11}$$

In the FEM context, this condition is recast in the form of Neumann condition as

$$B_1 p = \frac{\partial p}{\partial r} - S_1 p = 0, \tag{12}$$

where S_1 is the FE damper operator. Comparing Eqs. (11) and (12), we obtain

$$S_1 p = -\frac{p}{2r} - \frac{\dot{p}}{c}. \tag{13}$$

Equations (11) and (13) are valid for a circular S_R. If this boundary is elliptical, then we work in the elliptical coordinate system (u,v) related to the Cartesian system as

$$x = h \cosh u \cos v; \quad y = h \sinh u \sin v; \qquad 0 \leqslant u < \infty, \;\; 0 \leqslant v \leqslant 2\pi, \tag{14}$$

where the constant u curves are a set of confocal ellipses with

$$a = h \cosh u, \;\; b = h \sinh u, \;\; e = 1/\cosh u, \tag{15}$$

and foci at $(x,y)=(\pm h,0)$ and the constant v curves are a set of confocal hyperbolas with same foci. In Eq. (15), a, b, h, and e are the semimajor axis, semiminor axis, semi-interfocal distance, and eccentricity of the ellipse, respectively. The equations of the family of confocal ellipses and hyperbolas are

$$\left(\frac{x}{h \cosh u} \right)^2 + \left(\frac{y}{h \sinh u} \right)^2 = 1 \quad (1 < u < \infty), \tag{16a}$$

and

$$\left(\frac{x}{h \cos v} \right)^2 - \left(\frac{y}{h \sin v} \right)^2 = 1 \quad (1 < v < \pi), \tag{16b}$$

respectively. For an ellipse with constant a, as $e \to 0$, we have

$$u \to \infty \;\; (\text{since } \cosh u = 1/e), \qquad h \to 0 \;\; (\text{since } \; h = ae), \;\; \text{and} \;\; h \sinh u \to h \cosh u \to r. \tag{17}$$

In other words, the ellipse approaches a circle of radius $r = a$ and the foci coalesce to the origin of the circle $x = y = 0$. The coordinate r in the elliptical system is

$$r = \sqrt{x^2 + y^2} = h \sqrt{\frac{\cosh 2u + \cos 2v}{2}}. \tag{18}$$

Hence

$$\frac{\partial r}{\partial u} = \frac{h \sinh 2u}{\sqrt{2(\cosh 2u + \cos 2v)}},$$

and from which, neglecting the variation with respect to the v coordinate, one obtains

$$\frac{\partial}{\partial r} = \frac{\sqrt{2(\cosh 2u + \cos 2v)}}{h \sinh 2u} \frac{\partial}{\partial u}. \tag{19}$$

Using Eqs. (18) and (19) in Eq. (11), the elliptical damper condition is obtained as

$$B_1 p = \frac{\sqrt{2(\cosh 2u + \cos 2v)}}{h \sinh 2u} \frac{\partial p}{\partial u} + \frac{1}{c} \frac{\partial p}{\partial t} + \frac{p}{h \sqrt{2(\cosh 2u + \cos 2v)}} = 0. \tag{20}$$

The arc length dn defined along the constant v curve (i.e., normal to the ellipse) is

$$dn = \sqrt{\left(\frac{\partial x}{\partial u} \right)^2 + \left(\frac{\partial y}{\partial u} \right)^2}\, du = l_1 du, \tag{21}$$

which gives

$$\frac{\partial}{\partial n} = \frac{1}{l_1} \frac{\partial}{\partial u}. \tag{22}$$

From Eq. (14)

$$\frac{\partial x}{\partial u} = h \sinh u \cos v, \;\; \frac{\partial y}{\partial u} = h \cosh u \sin v,$$

using which Eq. (21) gives

$$l_1 = h \sqrt{\cosh^2 u - \cos^2 v}. \tag{23}$$

Using Eq. (22) in Eq. (20), we get

$$\frac{\partial p}{\partial n} = -\frac{h^2 \sinh 2u}{2rcl_1} \frac{\partial p}{\partial t} - \frac{h^2 \sinh 2u}{4r^2 l_1} p = -C_1 \frac{\partial p}{\partial t} - C_2 p, \tag{24}$$

where r and l_1 are given by Eqs. (18) and (23), respectively. From Eq. (12), the equivalent Neumann boundary condition on S_R is obtained using Eq. (24) as

$$S_1 p = \frac{\partial p}{\partial n} = -C_1 \frac{\partial p}{\partial t} - C_2 p. \tag{25}$$

Using Eqs. (25) and (9), the last term in Eq. (8) is written as

$$-\int_{S_R} S_1 p\,\delta p\,ds = \int_{S_R} C_1 \dot{p}\,\delta p\,ds + \int_{S_R} C_2 p\,\delta p\,ds$$

$$= \sum_e \delta p_e^T \left(\int_{S_R} C_1 N_F^{eT} N_F^e ds \dot{p}_e + \int_{S_R} C_2 N_F^{eT} N_F^e ds p_e \right). \quad (26)$$

Substituting Eqs. (10) and (26) in Eq. (8), we obtain

$$\sum_e \left[\left\{ \int_{\Omega_F^e} \nabla N_F^{eT} \nabla N_F^e d\Omega + \int_{S_R} C_2 N_F^{eT} N_F^e ds \right\} p_e + \left\{ \int_{S_R} C_1 N_F^{eT} N_F^e ds \right\} \dot{p}_e + \left\{ \frac{1}{c^2} \int_{\Omega_F^e} N_F^{eT} N_F^e d\Omega \right\} \ddot{p}_e \right] = F(t), \quad (27)$$

which conforms to the form

$$[K]\{p\} + [C]\{\dot{p}\} + [M]\{\ddot{p}\} = \{F\}. \quad (28)$$

$$[K] = \sum_e \int_{\Omega_F^e} \nabla N_F^{eT} \nabla N_F^e d\Omega + \sum_e \int_{S_R^e} C_2 N_F^{eT} N_F^e ds,$$

$$[C] = \sum_e \int_{S_R^e} C_1 N_F^{eT} N_F^e ds \quad (29)$$

$$[M] = \sum_e \frac{1}{c^2} \int_{\Omega_F^e} N_F^{eT} N_F^e d\Omega,$$

$$\{F\} = \sum_e -\rho_F \int_{S_{B1}^e} N_F^{eT} \overline{\ddot{u}_n}\, ds,$$

where $[C]$ and the second term of $[K]$ are global radiation damping and stiffness matrices, respectively, arising from S_R, first terms of $[K]$ and $[M]$ are global fluid stiffness and mass matrices, respectively, and $\{F\}$ is global load vector. It should be noted that, as in the first-order BGT damper given by Eq. (11), which does not have any term involving the circumferential derivative, the v derivative in Eq. (19) is ignored in the present formulation. Its consideration would have led to asymmetric FE matrices destroying the symmetry of Eqs. (28) and (29), making it unattractive as a numerical approximation.

VI. THREE-DIMENSIONAL SPHEROIDAL DAMPER

For $m=1$, the condition given by Eq. (5), using Eq. (4), can be written as

$$B_1 p = \left(\partial_r + \frac{1}{c}\partial_t + \frac{1}{r} \right) p = 0, \quad (30)$$

and the corresponding Neumann condition is given by Eq. (12). Comparing Eqs. (30) and Eq. (12), we obtain

$$S_1 p = -\frac{p}{r} - \frac{\dot{p}}{c}. \quad (31)$$

Equations (30) and (31) are valid for a spherical absorbing boundary (S_R). If this boundary is a prolate spheroid, then we work in the prolate spheroidal coordinate system (u,v,ϕ) related to the Cartesian coordinate system as

$$x = h\cosh u \cos v; \quad y = h \sinh u \sin v \cos\phi;$$
$$z = h \sinh u \sin v \sin\phi \quad (0 \leqslant u < \infty, 0 \leqslant v \leqslant \pi, 0 \leqslant \phi \leqslant 2\pi), \quad (32)$$

where a constant u surface is an ellipsoid with major axis length $2h\cosh u$, a constant v surface is a two-sheeted hyperboloid of revolution, and a constant ϕ surface is a plane containing the major axis which coincides with x axis. It may be noted that a prolate spheroid is obtained by rotating an ellipse about its major axis. Thus, Eqs. (16a) and (16b) are the equations for the family of confocal ellipses and hyperbolas, respectively, in this system also. The radial distance r from the origin to the prolate spheroidal surface ($r = \sqrt{x^2+y^2+z^2}$) is given by Eq. (18) and hence Eq. (19) also holds.

Substituting Eq. (19) in Eq. (30), the spheroidal damper condition is obtained as

$$B_1 p = \frac{\sqrt{2(\cosh 2u + \cos 2v)}}{h \sinh 2u} \frac{\partial p}{\partial u} + \frac{1}{c}\frac{\partial p}{\partial t} + \frac{\sqrt{2}}{h} \frac{p}{\sqrt{\cosh 2u + \cos 2v}} = 0. \quad (33)$$

Following the developments of the elliptical boundary damper, Eq. (21) becomes

$$dn = \sqrt{\left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2 + \left(\frac{\partial z}{\partial u}\right)^2}\, du = l_1\, du, \quad (34)$$

and also from Eq. (32)

$$\frac{\partial x}{\partial u} = h \sinh u \cos v, \quad \frac{\partial y}{\partial u} = h \cosh u \sin v \cos\phi,$$

$$\frac{\partial z}{\partial u} = h \cosh u \sin v \sin\phi.$$

Hence, Eqs. (22) and (23) hold for this case also. Using Eq. (22) in Eq. (33), we get

$$\frac{\partial p}{\partial n} = -\frac{h^2 \sinh 2u}{2rcl_1}\frac{\partial p}{\partial t} - \frac{h^2 \sinh 2u}{2r^2 l_1} p = -C_1 \frac{\partial p}{\partial t} - C_2 p, \quad (35)$$

where r and l_1 are given by Eqs. (18) and (23), respectively. For this case also, the equivalent Neumann boundary condition given by Eq. (25) holds with definitions of C_1 and C_2 as in Eq. (35), and finally the FE equations given by Eqs. (27) to (29) also hold.

VII. NUMERICAL EXAMPLES OF EXTERIOR ACOUSTICS

A. Harmonically loaded sphere and prolate spheroid

Consider a sphere (radius $R=1$ m) and a 1:8 prolate spheroid ($a=1$ m, $b=0.125$ m) in an infinite fluid medium [c(m/s)$=\rho_F$ (kg/m^3)$=1$], on the entire wet surface of which a sinusoidal pressure $p(t)=p_0 \sin\omega t (\omega=1$ rad/s) is applied. Pinsky and Abboud (1991) studied FE solution of the sphere problem whose analytical solution is

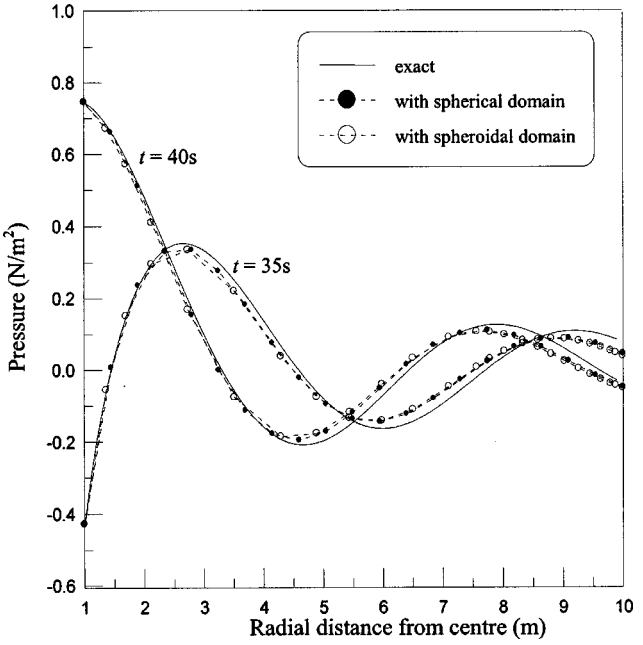


FIG. 2. Pressure along the radial line at $t=35$ and 40 s for spherical radiator.

$$p(r,t) = -p_0 \frac{1}{r} \sin k(r-R-ct).$$

No analytical solution of the prolate spheroid problem is found in the literature. The sphere problem is solved using one of two FE fluid domains: (i) spherical domain [S_R: $r=10R$ with 5 and 1 quadratic elements in r and meridional ($10°\times 10°$ sector) directions, respectively] and (ii) 1:5 spheroidal domain (one quadrant, S_R: $a=10$ m and $b=2$ m; using automesh technique with 2 quadratic elements along minor axis). The prolate spheroid problem is solved using 3 and 8 quadratic elements in r and meridional directions, respectively, in one quadrant of (i) circular domain (S_R: $r=2$ m) and (ii) spheroidal domain (S_R: $a=2$ m and $b=1.125$ m). The transient FE solution of both problems has been carried out with a time step of 0.5 s such that the system starts from $t=0$ and allowed to reach steady-state solution with time.

The pressure from the sphere boundary to S_R boundary is plotted in Fig. 2 for $t=35$ and 40 s. It is clear that both spherical (10 elements in a sector due to spherical symmetry) and spheroidal (4000 elements in a quadrant) boundary dampers yield practically identical results with excellent comparison with the analytical solution, thus proving correct formulation and implementation of the spheroidal damper.

The pressure along the major and minor axes from the spheroid boundary to S_R boundary is plotted in Fig. 3 for $t=35$ s. The time history of pressure at the point on S_R boundary lying on the major axis is shown in Fig. 4. From these plots, it is clear that both spherical and spheroidal boundary dampers yield practically identical results, thus demonstrating the use of the spheroidal damper.

This prolate spheroid example also brings out the economy of computation achievable using the spheroidal damper. There are 2000 fluid elements with a half-bandwidth of 242 for the spherical domain mesh and 1600 fluid elements with a half-bandwidth of 206 for the spheroidal domain mesh. In transient calculations, where the system of equations is solved many times (100 time steps in the present example), the computational economy can be quite large (119 s vs 242 s for spheroidal and spherical domain meshes, respectively). It may be noted that meshing a spherical domain with smaller number of elements (say, about 1600, as in the case of spheroidal domain mesh) leads to poor element aspect ratio of some of the elements, which yields results with significant errors.

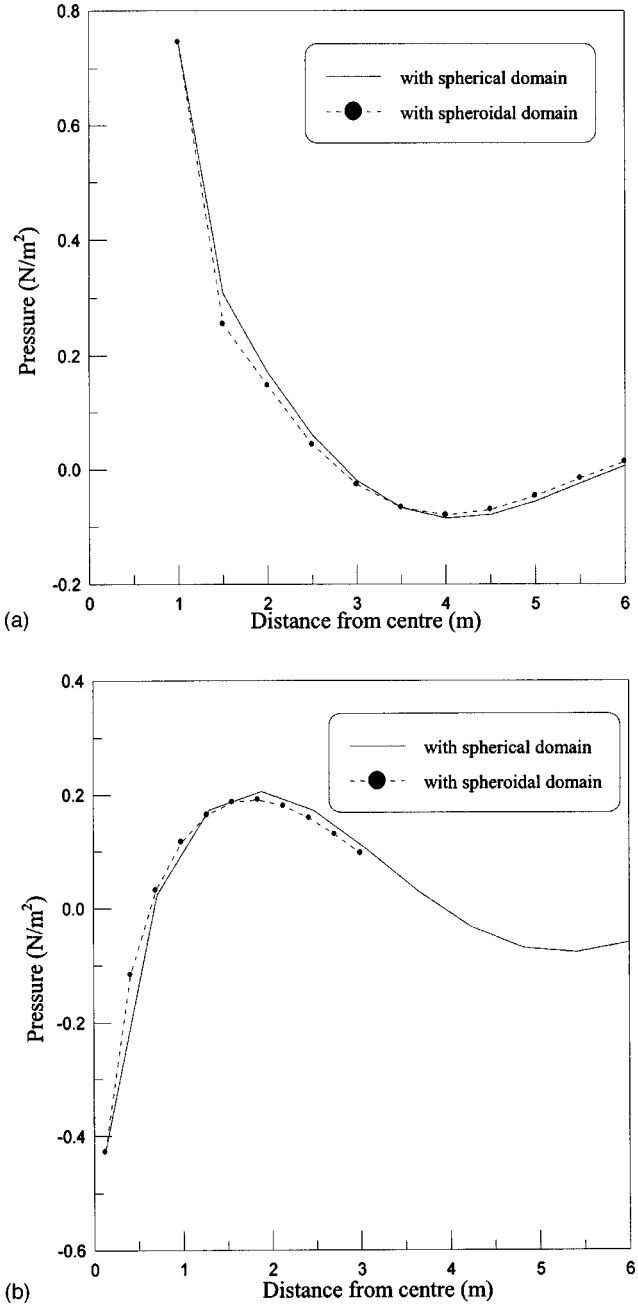


FIG. 3. (a) Pressure along the major axis at $t=35$ s for spheroidal radiator. (b) Pressure along the minor axis at $t=35$ s for spheroidal radiator.

VIII. GOVERNING EQUATIONS OF FLUID-STRUCTURE INTERACTION

In order to formulate transient FSI, consider the coupled system in Fig. 5 consisting of the computational fluid domain

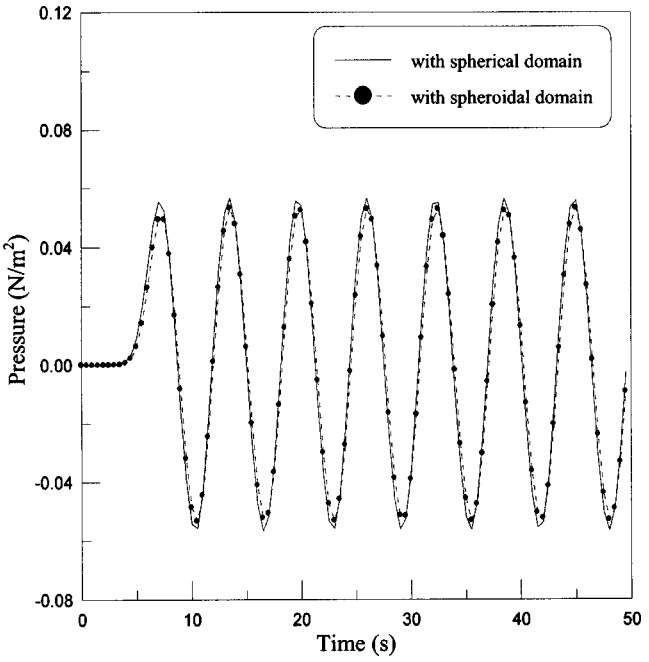


FIG. 4. Time history of pressure at the point on major axis of the radiation boundary for spheroidal radiator.

Ω_F and structural domain Ω_S. The boundary of the fluid domain consists of the fluid–structure interface S_I and radiation boundary S_R. The structural domain boundary consists of S_I and the dry body boundary S_B so that S_B is that part of the structure boundary that is not part of S_I. The outward normal to Ω_S (i.e., inward normal to Ω_F) is denoted **n**. The governing equations of the FSI problem are

$$\nabla \cdot \sigma - \rho_S \ddot{u} = 0 \quad \text{in } \Omega_S, \tag{36}$$

$$\sigma = D\varepsilon \quad \text{in } \Omega_S, \tag{37}$$

$$\nabla^2 p = \frac{\ddot{p}}{c^2} \quad \text{in } \Omega_F, \tag{38}$$

with boundary conditions

$$\dot{u} \cdot \mathbf{n} = \mathbf{v} \cdot \mathbf{n} = \nabla \phi \cdot \mathbf{n} \tag{39a}$$

or

$$p_{,n} = -\rho_F \ddot{u}_n \quad \text{on } S_I, \tag{39b}$$

$$\sigma_n = -p \cdot \mathbf{n}; \quad \tau_n = 0 \quad \text{on } S_I, \tag{40}$$

$$\sigma_n = \overline{\sigma_n}; \quad \tau_n = \overline{\tau_n} \quad \text{on } S_B, \tag{41}$$

$$r^{(h-1)/2}\left(\frac{\partial p}{\partial r} + \frac{\dot{p}}{c}\right) = \mathrm{O}\left(\frac{1}{r}\right) \quad \text{on } S_R, \tag{42}$$

where $\sigma = \langle \sigma_x, \sigma_y, \sigma_z, \tau_{xy}, \tau_{xz}, \tau_{yz} \rangle$ is the stress vector; $\varepsilon = \langle \varepsilon_x, \varepsilon_y, \varepsilon_z, \gamma_{xy}, \gamma_{xz}, \gamma_{yz} \rangle$ is the strain vector, D is the constitutive matrix, ρ_S is the density of structural material, ϕ is the velocity potential of the fluid, **v** is the fluid velocity vector, $u, \dot{u}$, and $\ddot{u}$ are the displacement, velocity, and acceleration of the structure, respectively, σ_n and τ_n are the normal and tangential (or traction) loading on the structural boundary, respectively, $\overline{\sigma_n}$ and $\overline{\tau_n}$ are the prescribed value of σ_n and τ_n, respectively.

Equation (36) is the linear momentum balance equation of the structure, Eq. (37) is the linear constitutive relation, and Eq. (38) is the linear acoustic wave equation [Eq. (1) restated]. Since the acoustic fluid is coupled to the structure and $p-u$ formulation is used, only one interface condition is imposed explicitly on the structure and one on fluid. The condition on the structure is given by Eq. (40) and it essentially is a force (dynamic) boundary condition on the structure. The condition on the fluid is given by Eq. (39a) and it essentially is a kinematic boundary condition involving velocity. When this condition is recast in the form of Eq. (39b), it directly relates the equilibrium of the fluid elements at the interface. Equation (41) is the prescribed loading applied to the structure and Eq. (42) is the radiation condition [Eq. (3) restated].

A four-noded bilinear degenerated shell (BDS) element (Nukulchai, 1979), evolving from an eight-noded, isoparametric brick element with four straight sides forming a hyperbolic paraboloid enclosing its midsurface, has been used to model the structure. The assumptions made are that the normal to midsurface remains straight after deformation and stresses normal to midsurface are zeros. The shell formula-

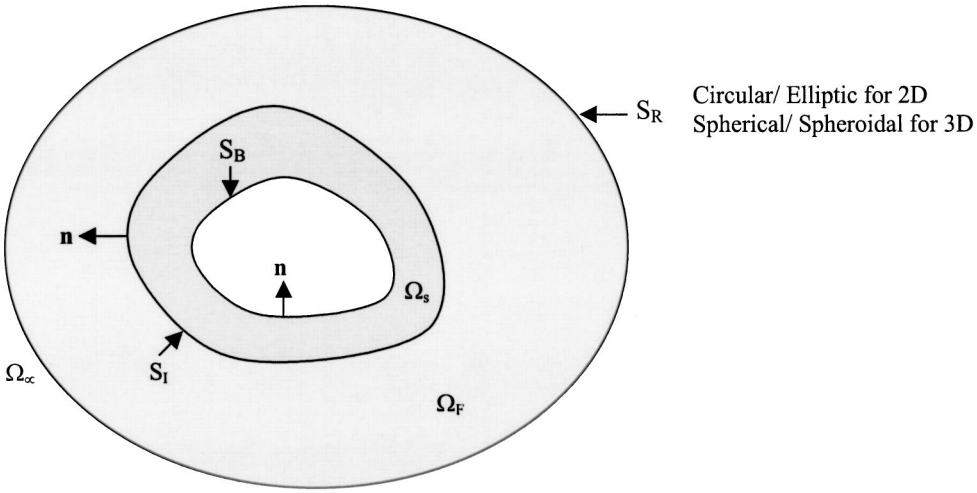


FIG. 5. Domain and boundaries of FSI problem.

tion includes transverse shear deformation and Kirchhoff–Love hypothesis is not assumed. The element thickness in the direction normal to the midsurface at each node is specified as input. In BDS elements, convergence is spoiled by the weakly restrained torsional mode after the mesh reaches some state of refinement. The element employs six degrees of freedom (dof) per node; however, no stiffness corresponding to the torsional dof exists locally in the formulation. When FE mesh is refined, the angles between elements may become close to 2π and this results in a very small amount of stiffness for torsional rotation. Therefore, any slight disturbance in the load corresponding to this dof amplifies the torsional rotation to an unrealistic value, affecting global solution. Thus, the element stiffness matrix is divided into three parts, namely those corresponding to the bending and membrane effects, transverse shear effects, and torsional contribution. However, the rotation of the normal and midsurface displacement field is independent. Hence, an additional constraint between the torsional rotation of the normal and the rotation of the midsurface is required to overcome the problem of nonconvergence due to torsional mode.

IX. FINITE-ELEMENT FORMULATION OF FLUID–STRUCTURE INTERACTION

Let the vector u containing translational and rotational displacements over the structure and p over the fluid be approximated in a FE sense as [also see Eq. (9)]

$$u=\sum_e N_s^e u_e; \quad p=\sum_e N_F^e p_e, \tag{43}$$

where N_s^e is the shape function matrix corresponding to a shell element (e denotes element) and u_e is its nodal displacement array. The structure FE equations can be derived from the Lagrange's equation of motion where the fluid dynamic force is treated as loading. This yields the familiar system of equations

$$[M_S]\{\ddot{u}\}+[C_S]\{\dot{u}\}+[K_S]\{u\}=\{f_s\}+\{f_p\}, \tag{44}$$

where $[K_S]$ is the global stiffness matrix, $[M_S]$ is the global mass matrix, and $[C_S]$ is the global viscous damping matrix for the structure. Also, f_p denotes the global load vector due to pressure acting at the fluid–structure interface and f_S denotes applied loading on the structure, and usually one assumes

$$[C_S]=\alpha_c[M_S]+\beta_c[K_S], \tag{45}$$

where α_c and β_c are the parameters for viscous structural damping.

The fluid FE equations are derived from the governing equations using the Rayleigh–Ritz technique employing the variational formulation as given in Sec. IV, and can be written as

$$[M_F+M_R]\{\ddot{p}\}+[C_R]\{\dot{p}\}+[K_F+K_R]\{p\}=\{f_F\}+\{f_u\}, \tag{46}$$

where $[M_F]$ and $[K_F]$ are the global fluid mass and stiffness matrices, $[M_R]$, $[C_R]$, and $[K_R]$ are, respectively, the radiation mass, damping, and stiffness matrices which depend on the type and order of the damper adopted. For the damper presented in Secs. V and VI, $[M_R]=0$. The loading term $\{f_u\}$ originates from the interface accelerations and $\{f_F\}$ from the other boundary conditions (either Neumann or Dirichlet). It is evident that $\{f_p\}$ in Eq. (44) and $\{f_u\}$ in Eq. (46) denote the mutual influences or interaction between the structure and fluid and provide the coupling between the two systems. Considering the pressure loading on the structure, $\{f_p\}$ in Eq. (44) can be written as

$$\{f_p\}=\sum_e \int_{S_I} N_s^{eT}\mathbf{n}p\, ds. \tag{47}$$

In the above, $\mathbf{n}$ represents the outward normal to Ω_s (i.e., inward normal to Ω_F). Using Eq. (43), Eq. (47) may be rewritten as

$$\{f_p\}=\sum_e \left(\int_{S_I} N_s^{eT}\mathbf{n}N_F^e\, ds\right)p_e. \tag{48}$$

Considering the normal acceleration $\ddot{u}_n$ of the fluid at the interface S_I, $\{f_u\}$ in Eq. (46) can be written in view of Eq. (39) as

$$\{f_u\}=\sum_e \int_{S_I} N_F^{eT}\mathbf{n}p\, ds=\sum_e \left(-\rho_F\int_{S_I} N_F^{eT}\mathbf{n}N_S^e\, ds\right)\ddot{u}_e, \tag{49}$$

where the fact that the shape functions of the interface element are the same as that of the shell element has been used. Using Eqs. (48) and (49) in Eqs. (44) and (46), we get

$$[M_S]\{\ddot{u}\}+[C_S]\{\dot{u}\}+[K_S]\{u\}=\left(\sum \int_{S_I} N_S^{eT}\mathbf{n}N_F^e\, ds\right)\{p\}+\{f_s\}, \tag{50}$$

$$[M_F+M_R]\{\ddot{p}\}+[C_R]\{\dot{p}\}+[K_F+K_R]\{p\}=\left(\sum -\int_{S_I}\rho_F N_F^{eT}\mathbf{n}\overline{N_S^e}\, ds\right)\{\ddot{u}\}+\{f_F\}. \tag{51}$$

Combining Eqs. (50) and (51), the coupled equations of motion can be written as

$$\begin{bmatrix} M_S & 0 \\ M_{FS} & M_F+M_R\end{bmatrix}\begin{Bmatrix}\ddot{u}\\ \ddot{p}\end{Bmatrix}+\begin{bmatrix} C_S & 0 \\ 0 & C_R\end{bmatrix}\begin{Bmatrix}\dot{u}\\ \dot{p}\end{Bmatrix}+\begin{bmatrix} K_S & K_{SF} \\ 0 & K_F+K_R\end{bmatrix}\begin{Bmatrix}u\\ p\end{Bmatrix}=\begin{Bmatrix}f_S\\ f_F\end{Bmatrix}, \tag{52}$$

where $[M_{FS}]$ and $[K_{SF}]$, the mass and stiffness coupling matrices, respectively, are

$$[M_{FS}]=\sum_e -\rho_F\int_{S_I} N_F^{eT}\mathbf{n}N_S^e\, ds, \qquad [K_{SF}]=\sum_e \int_{S_I} N_s^{eT}\mathbf{n}N_F^e\, ds. \tag{53}$$

X. COUPLING MATRIX COMPUTATION

In Eq. (52), special attention is required to compute the coupling matrices in Eq. (53) at the fluid–structure interface. The computation of structural and fluid FE matrices is standard and does not need elucidation. The coupling matrices

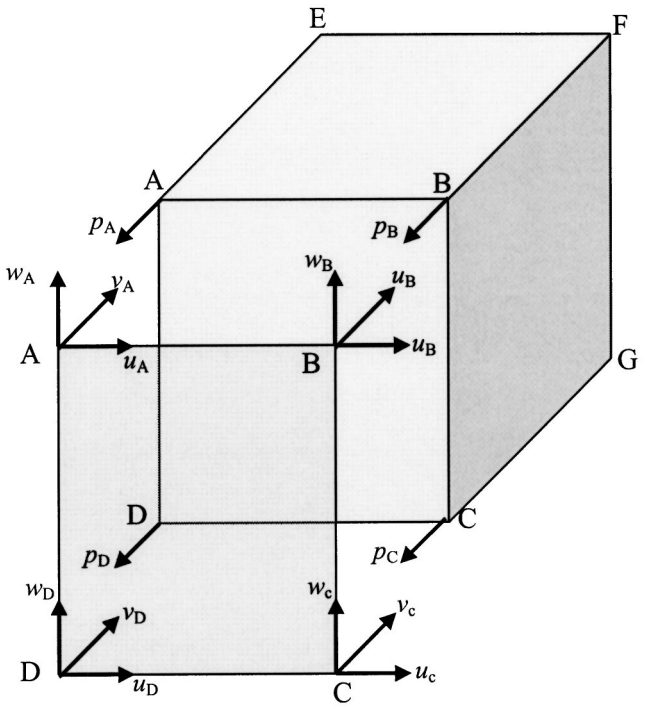


FIG. 6. Shell–fluid interface. (ABCD: Shell element, ABCDEFGH: Fluid element. Interface: ABCD. Coupled Shell dof's: u_i, v_i, w_i (j=A,B,C,D). Coupled fluid dof's: p_j (j=A,B,C,D).

are envisaged to belong to ''coupling elements,'' serving as transition between the structure and fluid, and therefore have both structural and fluid nodes (Fig. 6). The four-noded shell element and an eight-noded hexahedral fluid element are coupled by a four-noded interface element. From Eq. (52), it is clear that coupling is through both the mass and stiffness matrices, which is the essence of the compact $p-u$ formulation. From Eq. (53), the mass coupling matrix can be written, after ignoring the shell rotational dof as

$$[M_{FS}]=\sum_e M^e_{FS};$$

$$[M^e_{FS}]=[M^e_{FS_{jk}}]=-\rho_F\int_{S_I} N^{eT}_{Fj}\mathbf{n}N^e_{sk}\,ds \qquad (j=1 \text{ to } 4 \text{ and } k=1 \text{ to } 12), \tag{54}$$

where index j denotes the number of fluid dof's in the interface element and its maximum value is 4 because the eight-noded fluid element has four nodes in a face. The index k denotes the number of coupled shell dof's in the interface element, and its maximum value is 12 because the three shell displacements are assumed to be coupled to the corresponding fluid dof at each of the four nodes, leading to 4×3 or 12 dof's. Hence, the size of the element mass coupling matrix is 4×12. The interface element matrix being rectangular, unlike the usual square form of FE matrices, requires special attention during the assembly of element matrices. The shape function matrix of the interface element consists of four fluid shape functions depending on which face (out of six faces) of the fluid element is coupled with the shell element and 12 shape functions corresponding to 12 coupled shell dof's. Since we assume that only shell displacements (not the rotations) and fluid pressure are coupled at a node of the interface element, one can write the displacement vector of the shell by considering only the nodal translations as

$$\mathbf{u}(\xi,\eta,\zeta)=\sum_{i=1}^{4} N^e_{Si}\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}d_i, \tag{55}$$

at a point with isoparametric FE natural coordinate (ξ,η,ζ), where $d_i=\langle u_i v_i w_i\rangle$ is the nodal translational displacement vector of the shell, u_i, v_i, and w_i being displacements in global x, y, and z directions. Now, using Eq. (56), the product of the last two terms in the integrand appearing in Eq. (54) can be written as

$$\langle \mathbf{n}N^e_{sk}\rangle=\langle l_{3i}m_{3i}n_{3i}\rangle\times\begin{bmatrix}N_{si}&0&0\\0&N_{si}&0\\0&0&N_{si}\end{bmatrix} \quad (i=1 \text{ to } 4)$$
$$=\langle N_{s1}l_{31},N_{s1}m_{31},N_{s1}n_{31},\ldots,N_{s4}l_{34},$$
$$N_{s4}m_{34},N_{s4}n_{34}\rangle_{(12\times1)}=Q^e_{sk}, \tag{56}$$

where $\langle l_3 m_3 n_3\rangle$ is the direction cosine vector which is normal to the shell midsurface. Thus, Q^e_{sk} is a vector containing 12 shape functions corresponding to the coupled structural dof's of the interface coupling element and N^e_{Fj} is the vector containing four fluid shape functions corresponding to the face of the fluid element which is in contact with shell surface. By using Eq. (56), Eq. (54) can be modified as

$$[M^e_{FS_{jk}}]=-\rho_F\int_{S_I} N^{eT}_{Fj}Q^e_{sk}\,ds, \tag{57}$$

where index j takes values from 1 to 4 and k takes values from 1 to 12. Following a similar procedure, the second of Eq. (53) yields the stiffness-coupling matrix of size 4×12 as

$$[K^e_{SF_{jk}}]=\int_{S_I} Q^{eT}_{sk}N^e_{Fj}ds. \tag{58}$$

XI. COMPUTATIONAL ASPECTS

A FORTRAN code has been implemented using isoparametric C^0-type 20-noded quadratic and 8-noded linear hexahedron (or brick) acoustic finite elements, shell element described in Sec. VIII and the fluid–solid interface coupling element described in Sec. X. The details of 3D fluid elements are standard (Bathe, 1996) and hence not reproduced here. For time integration, the damped Newmark's method with $\beta=0.3025$ and $\gamma=0.6$ is adopted in the calculations (Hilber and Hughes, 1978; Manoj and Bhattacharyya, 2000a, b). However, the system of FE equations, Eq. (52), being unsymmetric but banded with a symmetric profile, the unsymmetric, out-of-core block solver developed by Manoj and Bhattacharyya (1997) has been used for solution.

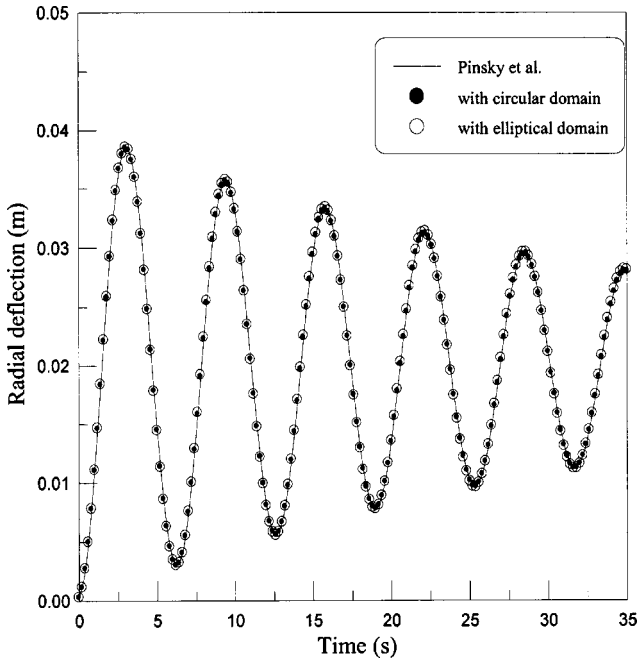


FIG. 7. Time history of radial deflection of the circular cylindrical shell.

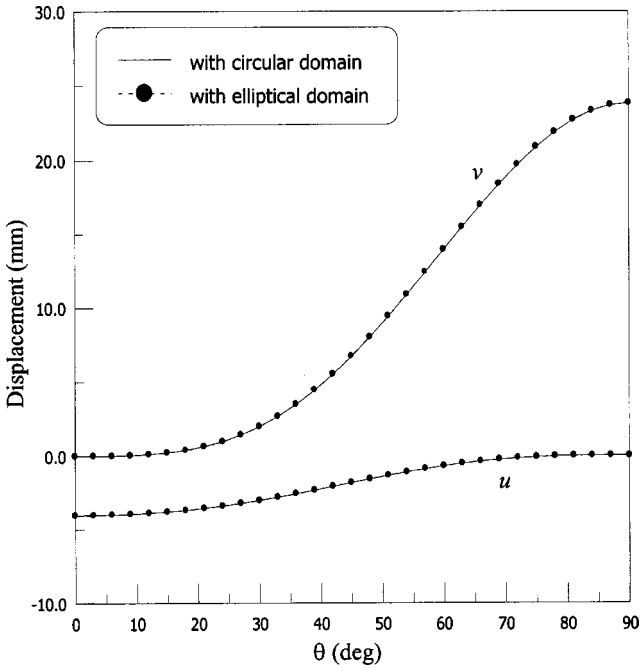


FIG. 8. Circumferential variation of u- and v-displacements at $t=0.06$ s for elliptical cylindrical shell.

XII. NUMERICAL EXAMPLES OF FLUID-STRUCTURE INTERACTION

A. Exterior circular and elliptical cylindrical shell–fluid interaction

Consider infinite circular (radius $R=1$ m, thickness $h=0.05$ m, Young's modulus $E=910$ N/m^2, Poisson's ratio $v=0.3$, material density $\rho_s=1000$ kg/m^3) and 1:5 elliptical ($a=1$ m, $b=0.2$ m, $h=0.05$ m, $E=2.1\times10^{11}$ N/m^2, $v=0.3$, $\rho_S=7850$ kg/m^3, material: steel) cylindrical shells in an infinite fluid medium [$\rho_F=2$ kg/m^3 and $c=1$ m/s for circular shell and $\rho_F=1000$ kg/m^3 and $c=1500$ m/s (water) for elliptical shell], on the entire wet surface of which a step pressure loading $p(t)=p_0$ for $t>0$ [p_0(N/m^2)$=1$ for circular and 1×10^6 for elliptical shell] is applied. Pinsky *et al.* (1992) and Manoj (1998) studied FE solution of the circular shell problem for which properties are such that the shell breathing mode characteristic frequency in vacuum is 1 rad/s and this choice accentuates the transient character of the solution.

The circular shell problem is solved using one quadrant of two FE fluid domains: (i) circular domain (S_R: $r=1.5R$) and (ii) 1:2 elliptical domain (S_R: $a=6$ m and $b=3$ m) using 5 and 15 linear elements in the r and θ directions, respectively. The elliptical shell problem is solved using one quadrant of two FE fluid domains: (i) circular domain (S_R: $r=6$ m) and (ii) 1:2 elliptical domain (S_R: $a=6$ m and $b=3$ m) using 10 and 15 linear elements in the r and θ directions, respectively. The transient FE solution of both problems has been carried out with a time step of 0.2 s for circular shell and 0.001 s for elliptical shell such that the system starts from $t=0$ and the response is tracked up to $t=35$ s for circular shell and $t=0.25$ s for elliptical shell.

The time history of radial deflection of the circular shell is compared with the published results in Fig. 7. It is clear that both circular and elliptical boundary dampers yield practically identical results with excellent comparison with the reference results, thus proving correct formulation and implementation of the elliptical damper in transient FSI problems.

For the elliptical shell with circular and elliptical radiation boundaries, the deflections u (x displacement) and v (y displacement) along circumference at $t=0.06$ s are shown in Fig. 8 and the time histories of u at point A (on the major axis) and of v at point B (on the minor axis) are shown in Fig. 9. The "shell-alone" (i.e., in *vacuo* vibration) case is also presented in Fig. 9, which compares almost exactly with the results using a commercial FE code, to bring out the fluid-coupling effect vividly. From these plots, it is clear that both circular and elliptical boundary dampers yield practi-

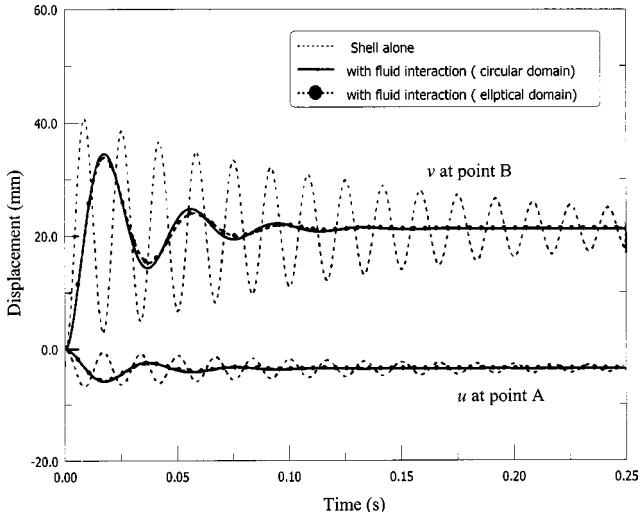
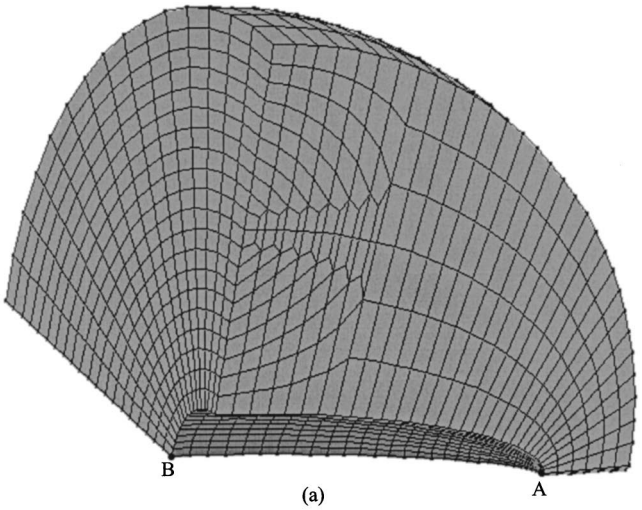


FIG. 9. Time histories of u and v displacements for elliptical cylindrical shell.



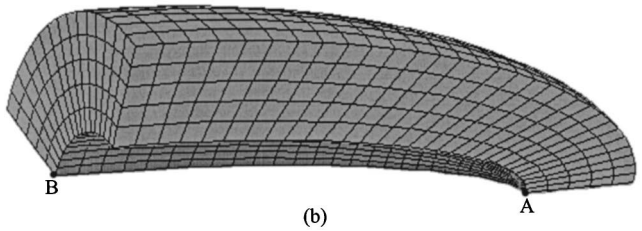


FIG. 10. Finite-element mesh for spheroidal shell fluid interaction problems. (a) with spherical radiation boundary; (b) with spheroidal radiation boundary.

cally identical results, thus demonstrating the use of the elliptical damper.

B. Exterior spherical and spheroidal steel shell–water interaction

Consider a spherical ($R=1.905$ m) and a 1:5 prolate spheroidal ($a=1$ m, $b=0.2$ m) steel shell in an infinite water medium, on the entire wet surface of which a step pressure loading $p(t)=p_0$ for $t>0$ is applied. The data of both shells are: $h=0.0254$ m, $E=2\times10^{11}$ N/m^2, $v=0.3$, $\rho_S=7850$ kg/m^3, $\rho_F=1020$ kg/m^3, $c=1499$ m/s, and p_0(N/m^2)$=7.02\times10^3$ (spherical shell), 1×10^6 (spheroidal shell). Akkas *et al.* (1979) and Bhattacharyya *et al.* (2001b) studied FE solution of the spherical shell problem, for which analytical solution may be found in Bedrosian and DiMaggio (1972). No analytical solution of the spheroidal shell problem is found in the literature. The spherical shell problem is solved using one quadrant of a spherical domain (S_R: $r=1.955$ m) modeled by 3° by 3° shell elements and only one stack of fluid elements of thickness 0.05 m over the entire shell surface. The spheroidal shell problem is solved using one quadrant of two FE fluid domains: (i) spherical domain (S_R: $r=1.25$ m) and (ii) spheroidal domain (S_R: $a=1.25$ m and $b=0.45$ m). Both domains have 30 and 15 shell elements along the major and minor axes, respectively.

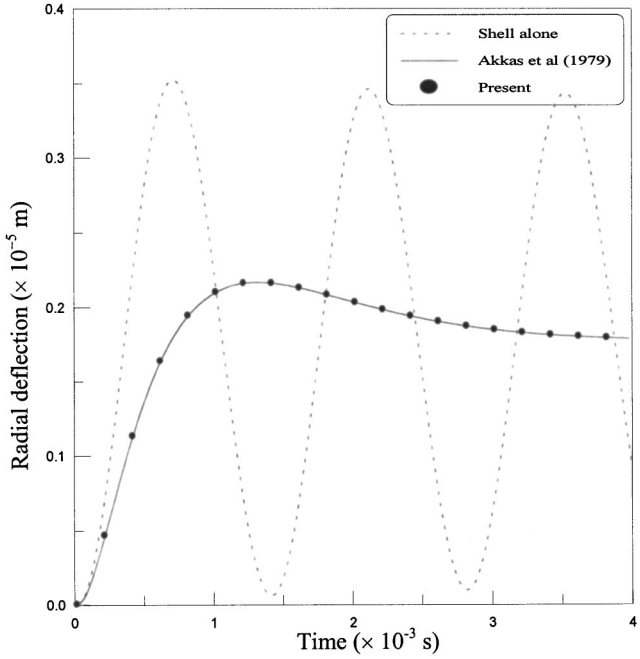


FIG. 11. Time history of radial deflection of spherical shell.

The fluid elements of the spherical domain are generated using automesh technique with 6 and 20 linear elements along the major and minor axes, respectively, and the spheroidal domain has five stacks of fluid elements over the shell surface (see Fig. 10). The transient FE solution of both problems has been carried out with a time step of 0.000 02 s such that the system starts from $t=0$ and the response is tracked up to $t=0.004$ s for spherical shell and $t=0.02$ s for spheroidal shell.

The time history of radial deflection of the spherical shell is compared with analytical result in Fig. 11. The shell-

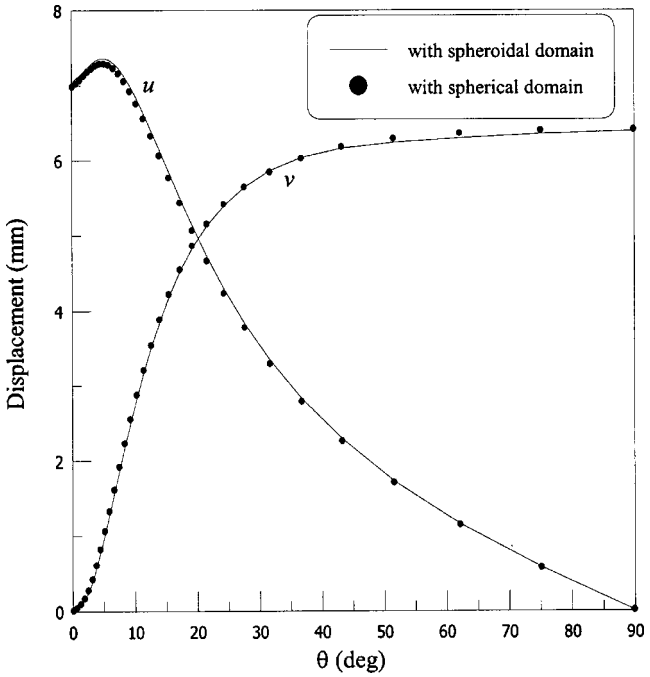


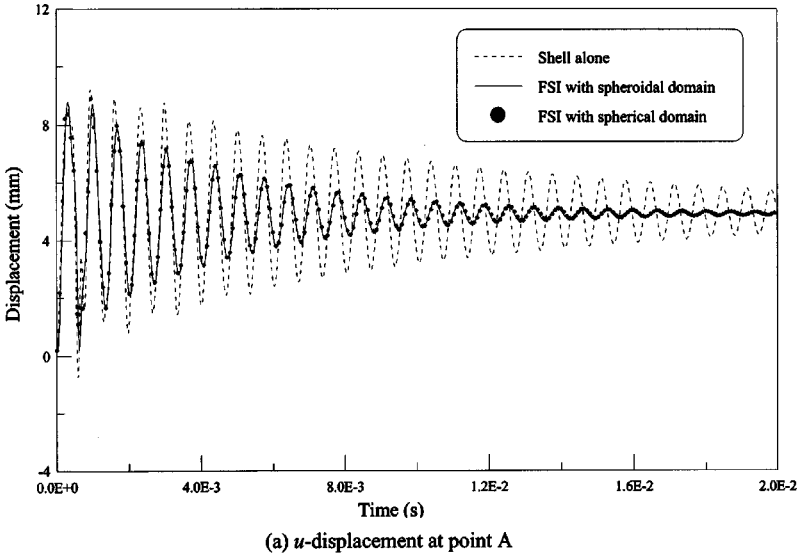
FIG. 12. Circumferential variation of u- and v-displacements at $t=0.003$ s for spheroidal shell.

alone case is also presented in this figure, which compares almost exactly with the results using a commercial FE code, to bring out the fluid coupling effect vividly. Akkas *et al.* (1979) obtained similar quality of comparison for this problem using displacement $(u-v-w)$ formulation for the fluid, but using eight stacks of fluid finite elements. In the present work, only one stack of elements yields virtually the same exact results. This demonstrates the efficacy of the FE procedures implemented in this work.

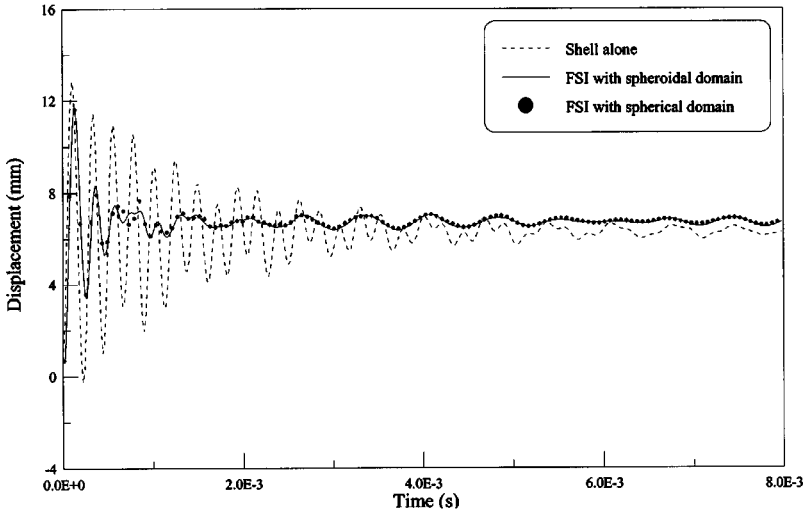
For the spheroidal shell with spherical and spheroidal radiation boundaries, the deflections u (x displacement) and v (y displacement) along circumference at $t=0.06$ s are shown in Fig. 12 and the time histories of u at point A (on the major axis) and of v at point B (on the minor axis) are shown in Fig. 13. The shell-alone case is also presented in Fig. 13, which compares almost exactly with the results using a commercial FE code, to bring out the fluid-coupling effect vividly. From these plots, it is clear that both spherical and spheroidal boundary dampers yield practically identical results, thus demonstrating the use of the spheroidal damper. It should be noted that a spherical domain with the same number of elements (five stacks) as the spheroidal domain mesh would yield grossly inaccurate results due to poor element sizing.

The spheroidal shell example also brings out the economy of computation achievable using the spheroidal damper. There are 3750 fluid elements with a half-bandwidth of 1423 for the spherical domain mesh and 2250 fluid elements with a half-bandwidth of 854 for the spheroidal domain mesh. In transient calculations, where the system of equations is solved many times (1000 time steps in the present example), the computational economy can be quite large (4.9 h vs 33.2 h for the spheroidal and spherical domain meshes, respectively). It may be noted that meshing the spherical domain with smaller number of elements (say, about 2250, as in the case of spheroidal domain mesh) leads to poor element aspect ratio of some of the elements, which yields results with significant errors.

All the examples treated in this section and in Sec. VIII have been solved using a variety of meshes, including the ones reported, in order to obtain converged results. The use of elliptical and spheroidal radiation boundaries, apart from computational economy, results in better element shapes when used for elongated bodies. Whereas this economy in a



(a) u-displacement at point A



(b) v-displacement at point A

FIG. 13. Time histories of deflection for spheroidal shell (a) u displacement at point A; (b) v displacement at point A.

2D problem is at best moderate, it can be substantial in 3D problems.

XIII. CONCLUSIONS

Transient nonreflecting boundary conditions applicable on elliptical and spheroidal boundaries have been developed and implemented in the context of finite-element analysis of transient fluid–structure interaction in exterior acoustics. The FE capability has been validated with numerical examples with excellent comparison with either analytical or published numerical results. The potential of computational economy, especially for 3D problems, is shown to be significant.

Akkas, N., Akay, H. U., and Yilmaz, C. (**1979**). ''Applicability of general-purpose finite element program in solid–fluid interaction problems,'' Comput. Struct. **10**, 773–783.

Antoine, X., and Barucq, H. (**1999**). ''Bayliss–Turkel-like radiation conditions on surfaces of arbitrary shape,'' J. Math. Anal. Appl. **229**, 184–211.

Astley, R. J. (**1998a**). ''Transient spheroidal elements for unbounded wave problems,'' Comput. Methods Appl. Mech. Eng. **164**, 3–15.

Astley, R. J. (**1998b**). ''Mapped spheroidal wave envelope elements for unbounded wave problems,'' Int. J. Numer. Methods Eng. **41**, 1235–1254.

Bathe, K. J. (**1996**). *Finite Element Procedures* (Prentice-Hall, Englewood Cliffs, NJ).

Bayliss, A., Gunzburger, M., and Turkel, E. (**1982**). ''Boundary conditions for the numerical solution of elliptic equations in exterior regions,'' SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **42**, 430–451.

Bayliss, A., and Turkel, E. (**1980**). ''Radiation boundary condition for wave like equations,'' Commun. Pure Appl. Math. **33**, 707–725.

Bedrosian, B., and DiMaggio, F. L. (**1972**). ''Transient response of submerged spheroidal shells,'' Int. J. Solids Struct. **8**, 111–129.

Bhattacharyya, S. K., Premkumar, R., and Manoj, K. G. (**2001a**). ''Boundary damper approximations for finite element analysis of transient fluid structure interaction,'' *Computational Fluid and Solid Mechanics*, edited by K. J. Bathe, in Proceedings of 1st MIT Conference on Computational Fluid and Solid Mechanics (Elsevier, New York), pp. 1081–1083.

Bhattacharyya, S. K., Sathyapal, S., and Vendhan, C. P. (**2001b**). ''Absorbing boundary conditions on elliptic boundary for finite element analysis of wave diffraction by large elongated bodies,'' Int. J. Numer. Methods Fluids **37**, 249–277.

Burnett, D. S., and Holford, R. L. (**1998a**). ''Prolate and oblate spheroidal acoustic infinite elements,'' Comput. Methods Appl. Mech. Eng. **158**, 117–141.

Burnett, D. S., and Holford, R. L. (**1998b**). ''Ellipsoidal acoustic infinite element,'' Comput. Methods Appl. Mech. Eng. **164**, 49–76.

Cipolla, J. L., and Butler, M. J. (**1998**). ''Infinite elements in the time domain using a prolate spheroidal multipole expansion,'' Int. J. Numer. Methods Eng. **43**, 889–908.

Djellouli, R., Farhat, C., Macedo, A., and Tezaur, R. (**2000**). ''Finite element solution of two-dimensional acoustic scattering problems using arbitrarily shaped convex artificial boundaries,'' J. Comput. Acoust. **8**, 81–99.

Grote, M. J., and Keller, J. B. (**1995**). ''On reflecting boundary conditions,'' J. Comput. Phys. **122**, 231–243.

Hilber, H. M., and Hughes, T. J. R. (**1978**). ''Collocation, dissipation, and overshoot for time integration schemes in structural dynamics,'' J. Earthquake Eng. Struct. Dyn. **6**, 99–118.

Manoj, K. G. (**1998**). ''Finite element procedures for transient exterior acoustics including fluid structure interaction,'' Ph.D. thesis, IIT Madras.

Manoj, K. G., and Bhattacharyya, S. K. (**1997**). ''A block solver for large unsymmetric, sparse, banded matrices with symmetric profile,'' Int. J. Numer. Methods Eng. **40**, 3279–3295.

Manoj, K. G., and Bhattacharyya, S. K. (**2000a**). ''A family of second-order boundary dampers for finite element analysis of two and three-dimensional problems in transient exterior acoustics,'' Int. J. Numer. Methods Eng. **48**, 925–948.

Manoj, K. G., and Bhattacharyya, S. K. (**2000b**). ''Transient acoustic radiation from impulsively accelerated bodies by the finite element method,'' J. Acoust. Soc. Am. **107**(3), 1179–1188.

Meade, D. B., Slade, G. W., and Peterson, A. F. (**1995**). ''Comparison of local radiation boundary conditions for the scalar Helmholtz equation with general boundary shapes,'' IEEE Trans. Antennas Propag. **43**(1), 6–10.

Nukulchai, W. K. (**1979**). ''A simple and efficient finite element for general shell analysis,'' Int. J. Numer. Methods Eng. **14**, 179–200.

Pinsky, P. M., and Abboud, N. N. (**1991**). ''Finite element solution of the transient exterior structural acoustic problem based on the use of radially asymptotic boundary operators,'' Comput. Methods Appl. Mech. Eng. **85**, 311–348.

Pinsky, P. M., Thompson, L. L., and Abboud, N. N. (**1992**). ''Local high-order radiation boundary conditions for the two-dimensional time-dependent structural acoustics problem,'' J. Acoust. Soc. Am. **91**, 1320–1335.

Tezaur, R., Macedo, A., Farhat, C., and Djellouli, R. (**2002**). ''Three-dimensional finite element calculations in acoustic scattering using arbitrarily shaped convex artificial boundaries,'' Int. J. Numer. Methods Eng. **53**, 1461–1476.

Acoustic field of a wedge-shaped section of a spherical cap transducer

Jeffrey A. Ketterling[a)]
Riverside Research Institute, 156 William Street, New York, New York 10038

(Received 12 January 2003; revised 9 August 2003; accepted 25 August 2003)

The acoustic pressure field at an arbitrary point in space is derived for a wedge-shaped section of a spherical cap transducer using the spatial impulse response (SIR) method. For a spherical surface centered at the origin, a wedge shape is created by taking cuts in the $X-Y$ and $X-Z$ planes and removing the smallest surface component. Analytic expressions are derived for the SIR based on spatial location. The expressions utilize the SIR solutions for a spherical cap transducer [Arditi *et al.*, Ultrason. Imaging **3**, 37–61 (1981)] with additional terms added to account for the reduced surface area of the wedge. Results from the numerical model are compared to experimental measurements from a wedge transducer with an 8-cm outer diameter and 9-cm geometric focus. The experimental and theoretical -3-dB beamwidths agreed to within $10\% \pm 5\%$. The SIR model for a wedge-shaped transducer is easily extended to other spherically curved transducer geometries that consist of combinations of wedge sections and spherical caps. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1618753]

PACS numbers: 43.20.Rz, 43.40.Rj, 43.35.Wa [LLT] Pages: 3065–3075

I. INTRODUCTION

Spherically curved cap transducers are used extensively in medical applications for both imaging and high-intensity focused ultrasound (HIFU) tissue exposures. The spherical curvature is ideal for achieving high lateral resolution for imaging and localizing energy deposition for HIFU exposures. It is becoming more common to modify the surface electrode pattern of spherical caps by cutting out holes (circles or rectangles) or etching novel conductive patterns onto the surface (annular rings or strips). The modeling of the sound fields from these new transducer designs is important for both interpreting backscattered ultrasound signals and for predicting *in vivo* thermal effects.

The spatial impulse response (SIR) method provides an efficient means of calculating the sound fields of these new electrode patterns.[1,2] To calculate the time-domain sound pressure at some point $\mathbf{P}(\mathbf{r})$ in space, the SIR method requires two time-domain functions: the transducer's normal surface velocity $v(t)$ and the impulse response $\mathbf{h}(\mathbf{r},t)$. Typically, the excitation velocity is considered uniform across the transducer surface but the excitation may be nonuniform.[3,4] The pressure is then found by convolving the time derivative of $v(t)$ with $\mathbf{h}(\mathbf{r},t)$. Because the waveform $v(t)$ is arbitrary, both pulsed and continuous excitations can be modeled. The method may also be applied to either transmit or transmit/receive systems.[5]

Analytic SIR expressions at an arbitrary point in space have been derived for only a few transducer geometries including circular pistons,[1] rectangular pistons,[6,7] triangular pistons,[8] planar polygons,[9] and spherical caps.[2,10] For more complicated geometries, methods exist for numerically approximating the SIR.[11,12] For some geometries, such as a cylinder, a combination of analytic and numerical solutions has been used.[13] Other methods have been presented to approximate a transducer with a mesh of triangular[8] or rectangular subelements.[4,14] The total SIR at **P** is then the summation of the subelement SIRs. Numerical methods to directly calculate the sound field such as the line integral method[15,16] and direct numerical integration of the Rayleigh integral may also be implemented.

The summation property of the SIR method makes the method ideal for studying phased arrays.[4,14,17–19] To apply the method to phased arrays, the SIR $\mathbf{h}(\mathbf{r},t)$ of each array element is first computed and then convolved with a time-delayed drive signal $\partial v(t-t_n)/\partial t$. The resulting signals are summed to give the pressure waveform $p(t)$ at $\mathbf{r}$. In a similar fashion, the function for $\mathbf{h}(\mathbf{r},t)$ from subelements that completely overlap can be subtracted to give $\mathbf{h}(\mathbf{r},t)$ for the nonoverlapping area. For instance, $\mathbf{h}(\mathbf{r},t)$ for an annular ring on a spherical cap is the difference between $\mathbf{h}(\mathbf{r},t)$ from two full spherical caps.[10]

Spherically curved transducers lend themselves particularly well to the SIR method due to the solution for $\mathbf{h}(\mathbf{r},t)$ reducing to a line integral in a plane normal to $\mathbf{r}$. However, full analytic expressions have only been worked out for spherical caps and related shapes such as annular rings.[10] Partial solutions for specific planes in space have also been derived for a curved strip shape.[20] Additional SIR solutions for spherical cap-based geometries are therefore useful.

One such new geometry is a wedge-shaped section of a spherically curved transducer (Fig. 2). For a spherical cap transducer with radius a and geometric focus R, a wedge is formed by making a cut normal to a at $z=h$ where $h<a$. If $h=0$, the wedge would be 1/2 of a spherical cap. The wedge shape is a useful geometry to model because its solution lends itself to solving numerous other geometries including a spherical cap with a rectangular cutout, a curved strip,[20] and a "strip electrode."[21,22]

We derive the analytic expressions necessary to describe

a)Electronic mail: ketterling@rrinyc.org

the SIR at some arbitrary point in the sound field of a wedge section of a spherically curved transducer. The derivation relies heavily on the SIR solution for a spherical cap transducer developed by Arditi *et al.*[10] Examples of the SIR for a wedge are compared to a full spherical cap at various points in space. We then compare hydrophone measurements of a wedge-shaped transducer to the theoretical predictions. Finally, we show how the wedge SIR solution can be extended to other spherically curved geometries.

II. THEORY SUMMARY

A. SIR method

The general formulation of the SIR at **P** for an arbitrarily shaped source is briefly summarized. If we take a piston in an infinite baffle with surface **S** excited with a uniform normal surface velocity $v(t)$, then the pressure at $\mathbf{P}=(x,y,z)$ is $p(t)=-\rho\,\partial\phi(\mathbf{r},t)/\partial t$. Here, ρ is the medium density and the velocity potential $\phi(\mathbf{r},t)$ is described by the Rayleigh integral[23]

$$\phi(\mathbf{r},t)=\frac{1}{2\pi}\int_S\frac{v(t-r'/c)}{r'}d\mathbf{S}, \tag{1}$$

where c is the speed of sound and r' is the distance between **P** and a surface element $d\mathbf{S}$.

This description of $\phi(\mathbf{r},t)$ holds equally well for unbaffled planar and weakly focused sources.[2,10,24] For a spherical cap transducer, this condition implies that $a\gg\lambda$, where a is the transducer's radius and λ is the wavelength of sound.

The description of $\phi(\mathbf{r},t)$ is simplified by writing it in terms of a convolution[1,6] $\phi(\mathbf{r},t)=v(t)*\mathbf{h}(\mathbf{r},t)$, where

$$\mathbf{h}(\mathbf{r},t)=\frac{1}{2\pi}\int_S\frac{\delta(t-r'/c)}{r'}d\mathbf{S} \tag{2}$$

is the SIR and $\delta(t)$ is the Dirac delta function. Equation (2) states that at a time r'/c, the value of $\mathbf{h}(\mathbf{r},r'/c)$ is the sum of the transducer area that is a distance r' from **P** divided by $2\pi r'$. If $r'_{\min}$ and $r'_{\max}$ represent the nearest and farthest points on **S** from **P**, then $\mathbf{h}(\mathbf{r},r'/c)$ is nonzero on the time interval $[r'_{\min}/c,\, r'_{\max}/c]$ or $[t_{\min},\, t_{\max}]$.

The pressure at **P** in terms of a convolution is then

$$p(\mathbf{r},t)=-\rho\frac{\partial v(t)}{\partial t}*\mathbf{h}(\mathbf{r},t). \tag{3}$$

B. Spherical cap SIR

The analytic SIR expressions for the wedge transducer will utilize the spherical cap results of Arditi *et al.*,[10] which we summarize and expand upon. For a spherical cap with geometric focus R and radius a, we first define several key parameters (Figs. 1 and 2).

The depth of the spherical cap is

$$d=R-\sqrt{R^2-a^2}. \tag{4}$$

The point **P** utilizes the coordinate system shown in Fig. 1(a). To obtain the equivalent coordinate system used by Arditi *et al.*, we need to rotate about the Y axis by an angle

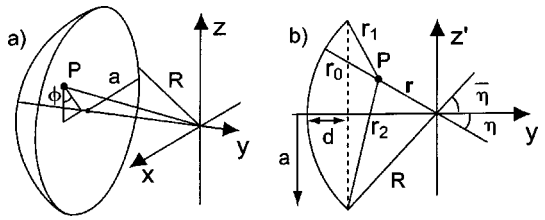


FIG. 1. Geometry of the spherical cap. (a) The basic geometric parameters. (b) When **P** is rotated to the $x=0$ plane, several key distances are defined.

$\phi=\tan^{-1}(|x/z|)$. After the rotation, the point **P** is in the $Y-Z'$ reference frame with $x'=0$ [Fig. 1(b)].

The angles $\eta=\sin^{-1}(\sqrt{x^2+z^2}/|\mathbf{r}|)$ and $\bar{\eta}=\sin^{-1}(a/R)$ are used to determine whether **P** lies within the cone formed by a line drawn between the edge of the spherical cap and the origin. If $\eta<\bar{\eta}$ then **P** lies within this geometric shadow, and the value

$$r_0=\begin{cases}R-|\mathbf{r}|; & y<0\\ R+|\mathbf{r}|; & y>0\end{cases} \tag{5}$$

is valid. If $y<0$ then r_0 is the shortest distance between **S** to **P**; if $y>0$ then r_0 is the farthest.

Next, the distances from **P** to the top and bottom edges of the spherical cap in the $Y-Z'$ plane are

$$r_1=\sqrt{(a-p_1)^2+(R-d+y)^2}, \tag{6}$$

and

$$r_2=\sqrt{(a+p_1)^2+(R-d+y)^2}, \tag{7}$$

where $p_1=z\sqrt{(x/z)^2+1}$ is defined in order to maintain the sign of z. If $z=0$ then $p_1=x$.

With r_0, r_1, and r_2, the expressions from Table I of Ref. 10 are used to compute $\mathbf{h}(\mathbf{r},t)$ for any point **P**

$$\frac{\mathbf{h}(\mathbf{r},t)}{\frac{Rc}{2\pi|\mathbf{r}|}}=\begin{cases}
\begin{array}{c|cc|c}
 & \eta<\bar{\eta} & & \eta>\bar{\eta}\\
 & y<0 & y>0 & \\
\hline
0 & ct<r_0 & r_0<ct & ct<r_1\\
2\pi & r_0<ct<r_1 & r_2<ct<r_0 & -\\
\beta(t) & r_1<ct<r_2 & r_1<ct<r_2 & r_1<ct<r_2\\
0 & r_2<ct & ct<r_1 & r_2<ct
\end{array}
\end{cases}. \tag{8}$$

The value $\beta(t)$ is defined by Eq. (A10) in Ref. 10 and is summarized in the Appendix.

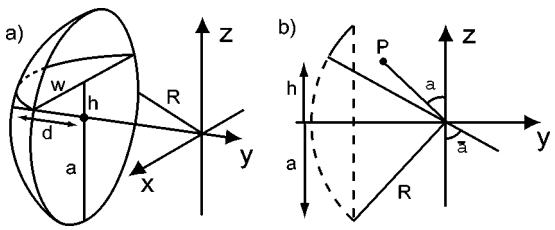


FIG. 2. Geometry of wedge section of a spherical cap transducer. (a) The geometric focus of the spherical cap is R and the radius is a. A wedge is formed by making a cut at $z=h$. (b) Projection of **P** onto $Y-Z$ plane.

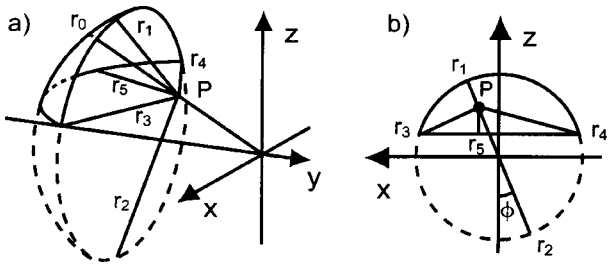


FIG. 3. (a) Key distances between **P** and the wedge used to calculate the SIR. (b) Projections of key distances r_n onto the $X-Z$ plane. The distance r_1 is always in the $z>0$ space and r_2 in the $z<0$ space. The distances r_3 and r_4 are to the wedge corners. The distance r_5 is the shortest path between the cut at $z=h$ and **P**. Some distances are not valid under certain conditions.

To find the SIR for a spherical cap, Arditi *et al.* made a coordinate substitution to reduce Eq. (2) to

$$\mathbf{h}(\mathbf{r},t)=\frac{cR}{2\pi|\mathbf{r}|}\int d\beta\int_{t_{\min}}^{t_{\max}}\delta(t-t_i)dt_i\,. \tag{9}$$

In this form, the value of $\mathbf{h}(\mathbf{r},t)$ at time t_i reduces to solving the line integral $\int d\beta$. The angle β lies in a plane normal to $\mathbf{r}$ and represents the angle swept out by the intersection of the plane and the spherical cap at t_i. The radius of the arc segment $\beta(t)$ in this reference frame is constant and equal to

$$\sigma(t)=\sqrt{(ct)^2-\left(\frac{(ct)^2+|\mathbf{r}|^2-R^2}{2|\mathbf{r}|}\right)^2}\,. \tag{10}$$

For a spherical cap, the solutions for $\beta(t)$ on the interval $[t_{\min},t_{\max}]$ are either 2π or some fraction thereof [Eq. (8)]. For a wedge, the value of $\beta(t)$ is further reduced because of the reduced surface area compared to a spherical cap.

III. SIR OF A WEDGE

A. General formulation

The geometry of a wedge is shown in Fig. 2(a). The wedge is formed by making a cut at $z=h$, where $h\geqslant 0$. For $h=0$ the wedge is half a spherical cap and for $h=a$ the wedge has zero surface area. At the cut of height h, w represents one half of the maximum width of the wedge and $a=\sqrt{w^2+h^2}$. The value d represents the depth of the full spherical cap, as was defined earlier [Eq. (4)].

A projection of **P** onto the $Y-Z$ plane is shown in Fig. 2(b). Two new angles are defined analogous to the η angles that determine if **P** is in the geometric shadow of the spherical cap. They are defined as

$$\alpha=\begin{cases}\cos^{-1}(z/|\mathbf{r}|); & y<0\\ \cos^{-1}(-z/|\mathbf{r}|); & y>0,\end{cases} \tag{11}$$

and $\overline{\alpha}=\cos^{-1}(h/R)$. For a given **P**, if $\alpha<\overline{\alpha}$ and $\eta<\overline{\eta}$, then **P** lies within the shadow of the wedge.

Due to the added complexity of the wedge geometry, several new distances between **P** and **S** need to be defined [Fig. 3(a)]. These include

$$r_3=\sqrt{(x-w)^2+(z-h)^2+(y+\sqrt{R^2-a^2})^2} \tag{12}$$

for the distance between $(w,d-R,h)$ and **P**, and

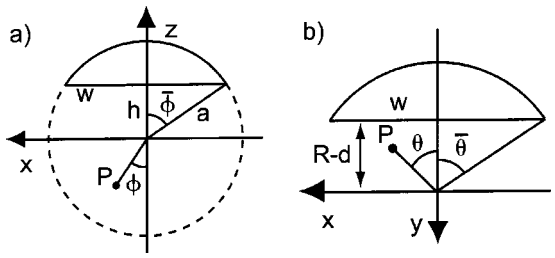


FIG. 4. Angles used to test for validity of r_1 and r_5. (a) If $\phi<\overline{\phi}$ then r_1 lies on wedge. (b) If $\theta<\overline{\theta}$ then r_5 lies on wedge.

$$r_4=\sqrt{(x+w)^2+(z-h)^2+(y+\sqrt{R^2-a^2})^2} \tag{13}$$

for the distance between $(-w,d-R,h)$ and **P**. Finally, the shortest distance between **P** and the wedge edge at $z=h$ is

$$r_5=\sqrt{(\sqrt{R^2-h^2}+k_1)^2+(z-h)^2}, \tag{14}$$

where $k_1=y\sqrt{(x/y)^2+1}$. If $y=0$ then $k_1=|x|$. The values r_0, r_1, and r_2 are still valid with r_1 representing a distance between **P** and a point in the $z>0$ space and r_2 a point in the $z<0$ space.

Figure 3(b) shows a projection of **P** and the r_n values onto the $X-Z$ plane. Due to the definitions for r_n, not all the values will always lie on the wedge: r_0 lies on the wedge when $\alpha<\overline{\alpha}$ and $\eta<\overline{\eta}$; r_2 never lies on the wedge because it is in the $z<0$ space and the wedge is in the $z>0$ space; r_1 and r_5 lie on the wedge only under certain conditions requiring the definition of additional angles.

Figure 4(a) shows a projection onto the $X-Z$ plane and has two angles of interest: the angle $\overline{\phi}=\cos^{-1}(h/a)$ and the angle $\phi=\tan^{-1}(|x/z|)$ defined earlier. When $\phi<\overline{\phi}$, r_1 lies on the wedge. Figure 4(b) shows two additional angles found from a projection onto the $X-Y$ plane: the angle $\theta=\tan^{-1}(|x/y|)$ and the angle $\overline{\theta}=\tan^{-1}(w/(R-d))$. If $\theta<\overline{\theta}$ then r_5 lies on the wedge. If $\theta>\overline{\theta}$ then r_5 is not used.

With the above parameters defined, we now outline the method to derive the SIR at **P**. The SIR at a time t_i on $[t_{\min},t_{\max}]$ represents a path integral defined by the intersection of a plane normal to $\mathbf{r}$ and the wedge [Eq. (9)]. An example of the intersection between a sphere $\mathbf{S}_2$ of radius ct_i centered at **P** and a wedge cut from a spherical cap $\mathbf{S}_1$ is shown in Fig. 5. The full path of intersection between $\mathbf{S}_1$ and $\mathbf{S}_2$ is $\Gamma=\beta(t)$ and the portion of the intersection within the wedge is Υ. The key to finding the SIR for the wedge at a time t_i is to modify $\beta(t)$ based on the intersection of Γ and $z=h$.

The sphere of radius ct centered at **P** is

$$(x-P_x)^2+(y-P_y)^2+(z-P_x)^2=(ct)^2. \tag{15}$$

This sphere intersects the spherical cap $x^2+y^2+z^2=R^2$ along the path Γ. The equation describing Γ is

$$\frac{R^2+|\mathbf{r}|^2-(ct)^2}{2}=xP_x+yP_y+zP_z\,, \tag{16}$$

where $P_x^2+P_y^2+P_z^2=\mathbf{r}^2$. To find the intersection points of Γ and $z=h$, spherical coordinates $(R,\hat{\theta},\hat{\phi})$ are used and Eq. (16) reduces to

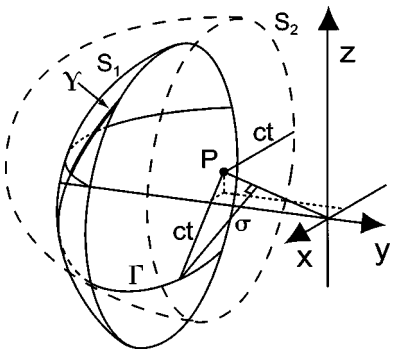


FIG. 5. Intersection of wedge and sphere of radius ct centered at **P**. The intersection occurs in a plane normal to **r** and is described by the arc Γ. The arc Υ is the segment of Γ in the wedge.

$$D = A\cos(\hat{\theta}) + B\sin(\hat{\theta}), \tag{17a}$$

where

$$A = P_x \sin(\hat{\phi}); B = P_y \sin(\hat{\phi}), \tag{17b}$$

and

$$D = \frac{R^2 + |\mathbf{r}|^2 - (ct)^2}{2R} - P_z \cos(\hat{\phi}). \tag{17c}$$

The intersection will occur at $\hat{\phi} = \overline{\alpha} = \cos^{-1}(h/R)$. The solutions to Eq. (17a) are

$$\hat{\theta}(t) = -\cos^{-1}\left[\frac{AD \pm B\sqrt{(A^2+B^2-D^2)}}{A^2+B^2}\right], \tag{18}$$

where the conditions $\cos^{-1}[\ldots] \leqslant 1$ and $\hat{\theta} = [3\pi/2 - \overline{\theta}, 3\pi/2 + \overline{\theta}]$ [Fig. 4(b)] must both be valid. If a solution to $\hat{\theta}$ is valid, then a conversion back to rectangular coordinates is performed. Three cases may arise on the time interval $[t_{\min}, t_{\max}]$: no valid values of $\hat{\theta}$, a single valid point $\mathbf{P}^1$ (Fig. 5), or two valid points $\mathbf{P}^1$ and $\mathbf{P}^2$.

For the case with no valid $\hat{\theta}$, then the path Υ is either $\beta(t)$ or 2π. A case with a single valid point $\mathbf{P}^1$ is shown in Fig. 6. The figure represents a projection of **P** and the arc Γ onto the $X-Z$ plane. The radius of the arc is $\sigma(t) = |\overline{\mathbf{P} - \mathbf{P}^1}|$ and the arc of the wedge segment is $\Upsilon = \beta(t)/2 - \beta_1$. The value of β_1 is

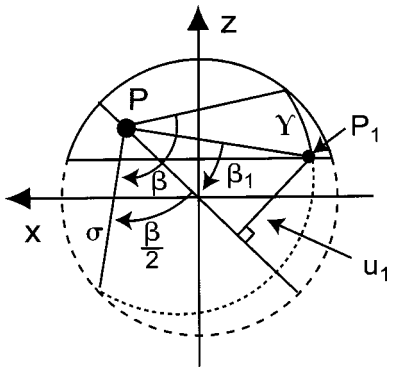


FIG. 6. Projection of **P** onto $X-Z$ plane showing how $\beta_1(t)$ is calculated. The arc Υ is found from $\beta(t)/2 - \beta_1(t)$.

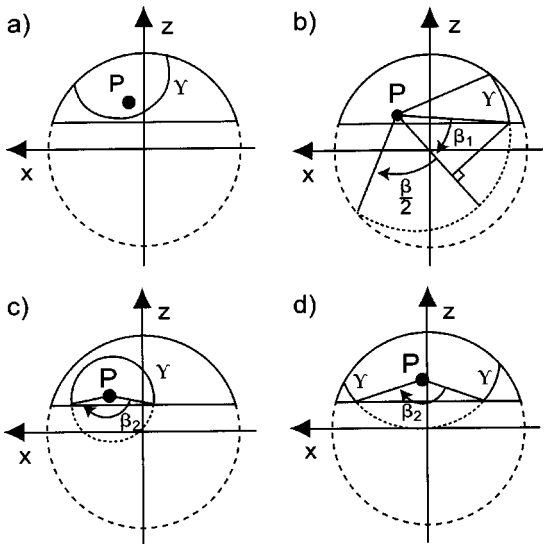


FIG. 7. Four cases of a projection of **P** and Γ onto the $X-Z$ plane showing the path Υ in the wedge. (a) No solution for $\hat{\theta}$; (b) a valid solution for $\mathbf{P}^1$; and (c), (d) valid solutions for $\mathbf{P}^1$ and $\mathbf{P}^2$.

$$\beta_1(t) = \begin{cases} \sin^{-1}(u_1(t)/\sigma(t)); & \mathbf{P''}^1_z < 0 \\ \pi - \sin^{-1}(u_1(t)/\sigma(t)); & \mathbf{P''}^1_z > 0, \end{cases} \tag{19}$$

where

$$u_1(t) = \frac{P^1_z P_x - P^1_x P_z}{\sqrt{P_x^2 + P_z^2}} \tag{20}$$

represents the minimum distance between $\mathbf{P}^1$, and the line through **P** and the origin after these objects have been projected onto the $X-Z$ plane. If **P** were rotated about the Y axis by ϕ (Fig. 4), then u_1 would represent the distance between the new $\mathbf{P'}^1$ and $x' = 0$ plane. An additional rotation about the new X' axis would place $\mathbf{P''}$ along the $\pm Y''$ axis. The value test $\mathbf{P''}^1_z$ is required because the arcsine alone does not work for cases when $|\beta_1| > \pi/2$.

For the final case with two valid values of $\hat{\theta}$, $\mathbf{P}^1$ and $\mathbf{P}^2$, the technique to modify Γ is more direct. For this case, the law of cosines is used to calculate

$$\beta_2(t) = \cos^{-1}\left(\frac{2\sigma(t)^2 - |\overline{P^1 - P^2}|^2}{2\sigma(t)^2}\right). \tag{21}$$

As in the case with a single point $\mathbf{P}^1$, β_2 is then used to modify $\beta(t)$ to find the arc segment Υ on the wedge.

Figure 7 graphically displays the three cases discussed above. For the case with no valid values of $\hat{\theta}$, $\beta(t)$ represents Υ [Fig. 7(a)]. For the case with $\mathbf{P}^1$ [Fig. 7(b)], the arc segment is $\Upsilon = \beta(t)/2 - \beta_1$. For cases with values for $\mathbf{P}^1$ and $\mathbf{P}^2$, arcs $\Upsilon = 2\pi - \beta_2$ [Fig. 7(c)] or $\Upsilon = \beta(t) - \beta_2$ [Fig. 7(d)] are two of the possible solutions. Additional cases may also arise but the approach remains the same: Eq. (17a) is solved, and then if necessary, either β_1 or β_2 is calculated. These values along with $\beta(t)$ are then used to determine the path Υ in the wedge.

TABLE I. Location tests of **P** used to determine appropriate solution for $\mathbf{h}(\mathbf{r},t)$.

Test	Description
$\eta<\overline{\eta}$	If **P** in spherical cap shadow
$\alpha<\overline{\alpha}$	If **P** in wedge shadow
$\phi<\overline{\phi}$	If r_1 lies on wedge
$\theta<\overline{\theta}$	If r_5 valid value
$r_1<r_5$	Which distance shorter
$r_2<r_5$	Which distance shorter
$y\leqslant 0$	Reverses sequence of r_n
$z<0$	Which r_n shortest

B. Analytic SIR expressions

For a spherical cap, the SIR expressions were grouped into three subsets based on the location of **P**: at the origin, in the cap shadow, and out of the cap shadow. The wedge shares a similar grouping but now with four subsets: at the origin, in the wedge shadow, in the cap shadow but out of the wedge shadow, and out of the cap shadow. These four subsets are then broken down into additional cases. The increased number of unique solutions for the wedge arises because, unlike for the spherical cap, the acoustic field of the wedge does not have radial symmetry. The tests used to further subdivide the wedge cases relate to angle (α, η, etc.) and r_n relations and are summarized in Table I.

In total, **P** may lie in one of 23 regions of space, and ten unique analytic expressions are required to describe $\mathbf{h}(\mathbf{r},t)$ at **P**. Nine of these expressions have a solution for both $y\leqslant 0$ and $y>0$; the tenth is at the origin. In the following four sections, the specific expressions to calculate $\mathbf{h}(\mathbf{r},t)$ are summarized, relying on the expressions for β, β_1, and β_2 derived earlier. Because of the symmetry about the $x=0$ plane, the solutions will be restricted to the space $x>0$ for $y\leqslant 0$ and $x<0$ for $y>0$.

1. P at origin

The solution at the origin is a unique case and is a function of the surface area of the wedge. The case of the wedge is similar to that of the spherical cap, where Arditi *et al.* showed that the SIR was $h(0,t)=\hat{\mathbf{S}}/(2\pi R)\,\delta(t-R/c)$, where $\hat{\mathbf{S}}=2\pi Rd$ for a spherical cap. For the case of a wedge described by the function $F(x,y,z)=x^2+y^2+z^2=R^2$, the surface area is found from

$$S=\int\int\frac{|\nabla F|}{|\nabla F\cdot\mathbf{n}|}dA, \tag{22}$$

where dA represents the projection of F onto the $z=0$ plane and $\mathbf{n}$ is the surface normal to dA. After converting to polar coordinates $(\hat{r},\hat{\theta},z)$, Eq. (22) can be expressed as

$$S=2\int_0^{\overline{\theta}}\int_{\hat{r}_1}^{\hat{r}_2}\frac{\hat{r}\,d\hat{r}}{\sqrt{R^2-\hat{r}^2}}d\hat{\theta}, \tag{23}$$

where $\hat{r}_1=\sqrt{R^2-a^2}/\cos(\hat{\theta})$, $\hat{r}_2=\sqrt{R^2-h^2}$, and $\overline{\theta}=\tan^{-1}(w/(R-d))$. The analytic solution[25] to Eq. (23) is

$$S=2R\mathbf{A}\sqrt{a^2-R^2}+2R^2\sin^{-1}\left(\frac{R\sin(\overline{\theta})}{a}\right)-2Rh\overline{\theta}, \tag{24a}$$

where

$$\mathbf{A}=\tanh^{-1}\left(\sin(\overline{\theta})\sqrt{\frac{2a^2-2R^2}{2a^2-R^2+R^2\cos(2\overline{\theta})}}\right). \tag{24b}$$

The SIR at the origin for a wedge is then

$$h(0,t)=S/(2\pi R)\,\delta(t-R/c). \tag{25}$$

2. P in wedge shadow

When **P** is in the wedge shadow, then the conditions $\alpha<\overline{\alpha}$ and $\eta<\overline{\eta}$ are true. For these cases, the value r_0 falls on the wedge surface and either r_5 or r_1 is the next closest r_n value. The analytic expressions are derived by examining the intervals between the increasing values of r_n and determining how Γ intersects the wedge at $z=h$. For the case $r_5<r_1$ and $y\leqslant 0$, the expressions are

$$\frac{\mathbf{h}(\mathbf{r},t)}{\frac{Rc}{2\pi|\mathbf{r}|}}=\begin{cases}\begin{array}{c|c|c}\Upsilon & y\leqslant 0,\ r_5<r_1 & y>0,\ r_1<r_5\\ \hline 0 & ct<r_0 & r_0<ct\\ 2\pi & r_0<ct<r_5 & r_5<ct<r_0\\ 2\pi-\beta_2 & r_5<ct<r_1 & r_1<ct<r_5\\ \beta-\beta_2 & r_1<ct<r_3 & r_3<ct<r_1\\ \frac{1}{2}\beta\mp\beta_1 & r_3<ct<r_4 & r_4<ct<r_3\\ 0 & r_4<ct & ct<r_4\end{array}\end{cases}. \tag{26}$$

In order of increasing r_n, the solution is graphically described by a full circle $\Upsilon=2\pi$ on $r_0<ct<r_5$, then Fig. 7(c) on $r_5<ct<r_1$, then Fig. 7(d) on $r_1<ct<r_3$, and finally Fig. 7(b) on $r_3<ct<r_4$. The expressions for $r_1<r_5$ and $y>0$ are identical but in the reverse sequence, with r_4 the smallest and r_0 the largest values for r_n. The $\mp$ sign used before β_1 indicates that for the first column of expressions $\beta/2-\beta_1$ is used and for the second $\beta/2+\beta_1$ is used. The change is required because the sign of β_1 is different for $y>0$ and $y\leqslant 0$ [Eq. (20)].

The equations for the case with $r_5>r_1$ and $y\leqslant 0$ are similar to Eq. (26) with

$$\frac{\mathbf{h}(\mathbf{r},t)}{\frac{Rc}{2\pi|\mathbf{r}|}}=\begin{cases}\begin{array}{c|c|c}\Upsilon & y\leqslant 0,\ r_1<r_5 & y>0,\ r_5<r_1\\ \hline 0 & ct<r_0 & r_0<ct\\ 2\pi & r_0<ct<r_1 & r_1<ct<r_0\\ \beta & r_1<ct<r_5 & r_5<ct<r_1\\ \beta-\beta_2 & r_5<ct<r_3 & r_3<ct<r_5\\ \frac{1}{2}\beta\mp\beta_1 & r_3<ct<r_4 & r_4<ct<r_3\\ 0 & r_4<ct & ct<r_4\end{array}\end{cases}. \tag{27}$$

Again, the two cases displayed in the equation represent a reversal in the sequence of events when $y>0$ versus when $y\leqslant 0$.

3. P inside spherical cap but outside wedge shadow

When **P** is outside the wedge shadow but inside the spherical cap shadow, the conditions $\alpha>\overline{\alpha}$ and $\eta<\overline{\eta}$ hold true. Unlike the case when **P** was inside the shadow of the wedge, the angles θ and ϕ are now needed to distinguish the various cases. For these cases, there is no valid value of r_0 on the wedge and it will not appear in any of the solutions.

The first case is when $\theta>\overline{\theta}$, which implies that $\phi>\overline{\phi}$ is also true. As noted in Table I, these conditions imply that r_5 is not a valid value and r_1 does not lie on the edge. The analytic expression therefore reduces to just one interval

$$\frac{\mathbf{h}(\mathbf{r},t)}{\frac{Rc}{2\pi|\mathbf{r}|}}=\begin{cases} & \theta>\bar{\theta} & \\ \Upsilon & y\leqslant 0 & y>0 \\ \hline 0 & ct<r_3 & r_3<ct \\ \frac{1}{2}\beta\mp\beta_1 & r_3<ct<r_4 & r_4<ct<r_3 \\ 0 & r_4<ct & ct<r_4 \end{cases}. \tag{28}$$

The remaining cases all have $\theta<\overline{\theta}$. When $\theta<\overline{\theta}$ and $\phi>\overline{\phi}$ the expressions for $\mathbf{h}(\mathbf{r},t)$ are

$$\frac{\mathbf{h}(\mathbf{r},t)}{\frac{Rc}{2\pi|\mathbf{r}|}}=\begin{cases} & \theta<\bar{\theta}\ \text{and}\ \phi>\bar{\phi} & \\ \Upsilon & y\leqslant 0 & y>0 \\ \hline 0 & ct<r_5 & r_5<ct \\ \beta_2 & r_5<ct<r_3 & r_3<ct<r_5 \\ \frac{1}{2}\beta\mp\beta_1 & r_3<ct<r_4 & r_4<ct<r_3 \\ 0 & r_4<ct & ct<r_4 \end{cases}. \tag{29}$$

Here, r_1 is still not on the wedge and its value does not appear in any of the intervals for r_n. For cases when $\phi<\overline{\phi}$, r_1 does lie on the wedge and the analytic expressions are

$$\frac{\mathbf{h}(\mathbf{r},t)}{\frac{Rc}{2\pi|\mathbf{r}|}}=\begin{cases} & \theta<\bar{\theta}\ \text{and}\ \phi<\bar{\phi} & \\ & y\leqslant 0 & y>0 \\ \Upsilon & z>0 & z<0 \\ \hline 0 & ct<r_5 & r_5<ct \\ \beta_2 & r_5<ct<r_1 & r_1<ct<r_5 \\ \beta_2-(2\pi-\beta) & r_1<ct<r_3 & r_3<ct<r_1 \\ \frac{1}{2}\beta\mp\beta_1 & r_3<ct<r_4 & r_4<ct<r_3 \\ 0 & r_4<ct & ct<r_4 \end{cases}, \tag{30}$$

and

TABLE II. Summary of expressions for $\mathbf{h}(\mathbf{r},t)$ based on the location of **P** for $y\leqslant 0$.

P location	$\mathbf{h}(\mathbf{r},t)$
Origin	Eq. (25)
$\eta<\overline{\eta},\ \alpha<\overline{\alpha},\ r_5<r_1$	Eq. (26)
$\eta<\overline{\eta},\ \alpha<\overline{\alpha},\ r_5>r_1$	Eq. (27)
$\eta<\overline{\eta},\ \alpha>\overline{\alpha},\ \theta>\overline{\theta}$	Eq. (28)
$\eta<\overline{\eta},\ \alpha>\overline{\alpha},\ \theta<\overline{\theta},\ \phi>\overline{\phi}$	Eq. (29)
$\eta<\overline{\eta},\ \alpha>\overline{\alpha},\ \theta<\overline{\theta},\ \phi<\overline{\phi},\ z>0$	Eq. (30)
$\eta<\overline{\eta},\ \alpha>\overline{\alpha},\ \theta<\overline{\theta},\ \phi<\overline{\phi},\ z<0$	Eq. (31)
$\eta>\overline{\eta},\ \theta<\overline{\theta},\ z>0$	Eq. (32)
$\eta>\overline{\eta},\ \theta<\overline{\theta},\ z<0$	Eq. (31)
$\eta>\overline{\eta},\ \theta>\overline{\theta},\ \phi<\overline{\phi},\ z>0$	Eq. (33)
$\eta>\overline{\eta},\ \theta>\overline{\theta},\ \phi<\overline{\phi},\ z<0$	Eq. (34)
$\eta>\overline{\eta},\ \theta>\overline{\theta},\ \phi>\overline{\phi}$	Eq. (28)

$$\frac{\mathbf{h}(\mathbf{r},t)}{\frac{Rc}{2\pi|\mathbf{r}|}}=\begin{cases} & \theta<\bar{\theta}\ \text{and}\ \phi<\bar{\phi} & \\ & y\leqslant 0 & y>0 \\ \Upsilon & z<0 & z>0 \\ \hline 0 & ct<r_5 & r_5<ct \\ \beta_2 & r_5<ct<r_3 & r_3<ct<r_5 \\ \frac{1}{2}\beta\mp\beta_1 & r_3<ct<r_4 & r_4<ct<r_3 \\ \beta & r_4<ct<r_1 & r_1<ct<r_4 \\ 0 & r_1<ct & ct<r_1 \end{cases}. \tag{31}$$

In these cases, the condition $z<0$ provides an additional test to further subdivide the solutions.

4. Outside spherical cap shadow

Outside of the spherical cap shadow, the cases again divide down based on the tests in Table I, with the condition $\eta>\overline{\eta}$ true for all cases. For the first case, $\theta<\overline{\theta}$, which also implies $\phi<\overline{\phi}$. The expression for $\mathbf{h}(\mathbf{r},t)$ when $y\leqslant 0$ and $z>0$ is

$$\frac{\mathbf{h}(\mathbf{r},t)}{\frac{Rc}{2\pi|\mathbf{r}|}}=\begin{cases} & \theta<\bar{\theta} & \\ & y\leqslant 0 & y>0 \\ \Upsilon & z>0 & z<0 \\ \hline 0 & ct<r_1 & r_1<ct \\ \beta & r_1<ct<r_5 & r_5<ct<r_1 \\ \beta-\beta_2 & r_5<ct<r_3 & r_3<ct<r_5 \\ \frac{1}{2}\beta\mp\beta_1 & r_3<ct<r_4 & r_4<ct<r_3 \\ 0 & r_4<ct & ct<r_4 \end{cases}. \tag{32}$$

When $y\leqslant 0$ with $z<0$ and $y>0$ with $z>0$, the expressions for $\mathbf{h}(\mathbf{r},t)$ are identical to those in Eq. (31).

When $\theta>\overline{\theta}$ and $\phi<\overline{\phi}$ the expressions for $\mathbf{h}(\mathbf{r},t)$ are

$$\frac{\mathbf{h}(\mathbf{r},t)}{\frac{Rc}{2\pi|\mathbf{r}|}}=\left\{\begin{array}{c|cc} & \theta>\bar{\theta}\ \text{and}\ \phi<\bar{\phi} & \\ & y\leqslant 0 & y>0 \\ \Upsilon & z>0 & z<0 \\ \hline 0 & ct<r_1 & r_1<ct \\ \beta & r_1<ct<r_3 & r_3<ct<r_1 \\ \frac{1}{2}\beta\mp\beta_1 & r_3<ct<r_4 & r_4<ct<r_3 \\ 0 & r_4<ct & ct<r_4 \end{array}\right. \tag{33}$$

and

$$\frac{\mathbf{h}(\mathbf{r},t)}{\frac{Rc}{2\pi|\mathbf{r}|}}=\left\{\begin{array}{c|cc} & \theta>\bar{\theta}\ \text{and}\ \phi<\bar{\phi} & \\ & y\leqslant 0 & y>0 \\ \Upsilon & z<0 & z>0 \\ \hline 0 & ct<r_3 & r_3<ct \\ \frac{1}{2}\beta\mp\beta_1 & r_3<ct<r_4 & r_4<ct<r_3 \\ \beta & r_4<ct<r_1 & r_1<ct<r_4 \\ 0 & r_1<ct & ct<r_1 \end{array}\right. . \tag{34}$$

Finally, when $\theta>\overline{\theta}$ and $\phi>\overline{\phi}$ the form of $\mathbf{h}(\mathbf{r},t)$ is equivalent to Eq. (28). A summary of all possible solutions for $y\leqslant 0$ is shown in Table II. A table for the cases when $y>0$ looks nearly identical and the two tables combined would yield the 23 possible expressions for $\mathbf{h}(\mathbf{r},t)$ based on the location of **P**.

IV. RESULTS

The analytic expressions listed above permit the calculation of the SIR at any **P**. With the SIR and a drive signal $v(t)$, the pressure is found from Eq. (3). In the following sections, examples are given of $\mathbf{h}(\mathbf{r},t)$ for different locations of **P**. Then, hydrophone measurements for a wedge transducer are compared to theoretical predictions. Finally, experimental results for a more complex transducer with two symmetrically opposed wedges in a ''strip electrode'' pattern[21,22] are compared to theoretical predictions.

A. Numerical results

The numerical examples all utilize a transducer having dimensions $R=9$ cm and $a=4$ cm that is driven with a continuous wave (cw) of 4.7 MHz. The numerical sampling rate is $\Delta t=10^{-8}$ s except when noted otherwise. Figure 8 shows $\mathbf{h}(\mathbf{r},t)$ at $\mathbf{P}=(0.2,-4,1.4)$ cm for several values of h. The solid line represents $\mathbf{h}(\mathbf{r},t)$ for a full spherical cap and all the curves are normalized to this case.

The curve for $h=0$ cm is found from Eq. (27). The initial segment at amplitude 1 occurs when $\Upsilon=2\pi$. The curve follows that of the spherical cap until $ct=r_5$. With further increase in time, the curve decreases until Υ goes to zero. As h increases, the break point from the spherical cap case occurs at earlier and earlier times. At $h=2$ cm, the curve is found from Eq. (26). For the curves when $h=3$ and 3.5 cm, r_0 no longer lies on the wedge and $\mathbf{h}(\mathbf{r},t)$ is found from Eq. (30). The peak amplitude is less than that of the spherical cap because Υ never equals 2π. Note that as h increases, the duration of $\mathbf{h}(\mathbf{r},t)$ decreases, consistent with $r_{\max}-r_{\min}$ also decreasing. Because r_0 no longer lies on the wedge, the curve has a delay relative to the spherical cap before the first nonzero point. The delay is minimal for $h=3$ cm but visible for $h=3.5$ cm.

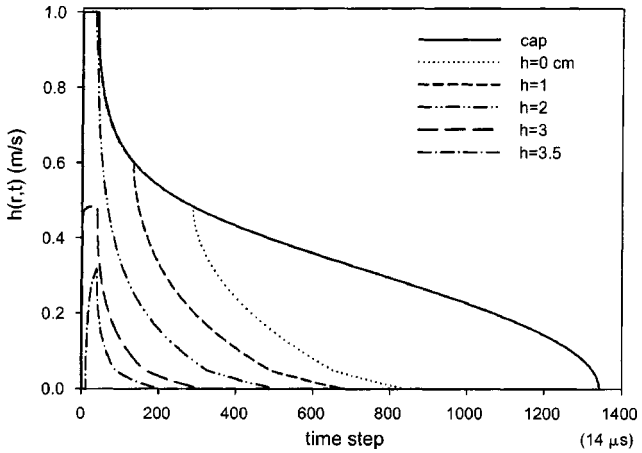


FIG. 8. Examples of $\mathbf{h}(\mathbf{r},t)$ at $\mathbf{P}=(0.2,-4,1.4)$ cm for a full spherical cap and a wedge with $h=0$, 1, 2, 3, and 3.5 cm.

Figure 9 shows another example of a collection of $\mathbf{h}(\mathbf{r},t)$ for a fixed wedge height of $h=2$ cm. In this plot, the curves represent the SIRs for $\mathbf{P}=(0.2,-1,z)$ cm as z ranges on $[-1,1]$ cm in 1-mm intervals. The amplitudes are normalized to 1 relative to the peak value of $\mathbf{h}(\mathbf{r},t)$. Starting at $z=-1$, **P** is outside of the spherical cap shadow. As z increases, **P** moves into the spherical cap shadow, then into the wedge shadow, and then once again out of the spherical cap shadow. In all, seven analytic expressions were needed to generate the full set of curves. The delays of the curves are all relative to $\mathbf{h}(\mathbf{r},t)$ at $z=1$, indicating $\mathbf{P}=(0.2,-1,1)$ cm is closest to the wedge and $\mathbf{P}=(0.2,-1,-1)$ cm farthest. The peak amplitudes occur for the cases when r_0 lies on the wedge and $\Upsilon=2\pi$.

With the SIR results over a two-dimensional grid of spatial locations, the peak pressure profile in a single plane for a cw drive can be generated using Eq. (3).[26] Figure 10 shows a plot of peak pressure from a 4.7-MHz drive in the $x=0$ plane. Unlike the case of a spherical cap, the beam profile of a wedge is not radially symmetric. The figure shows a series of isolines decreasing in steps of -3 dB from the peak pressure at the geometric focus $\mathbf{P}=(0,0,0)$. The transducer is located to the upper left in the figure and the beam propagates from there to the lower right corner. The -3-dB widths of the beam are 1.3 mm in the Z direction and 3.8 mm in the Y (axial) direction compared to 0.38 and 2.7 mm, respectively, for a spherical cap.

A view in a plane of constant Y aids in visualizing the beam profile. Figure 11 shows the peak pressure in the $y=0$ plane. The -3-dB beamwidths are 0.46 mm in the X direction and 1.4 mm in Z. The primary lobe has symmetry about the $x=0$ and $z=0$ planes. The symmetry about $z=0$

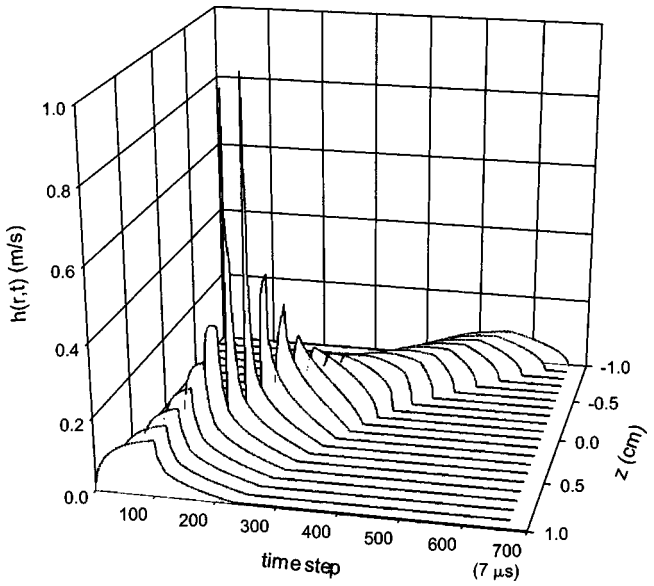


FIG. 9. Examples of $\mathbf{h}(\mathbf{r},t)$ for $\mathbf{P}=(0.2,-1,z)$ cm where z is in the range $[-1,1]$ cm. The curves for $\mathbf{h}(\mathbf{r},t)$ were generated using 7 of the 23 analytic SIR expressions.

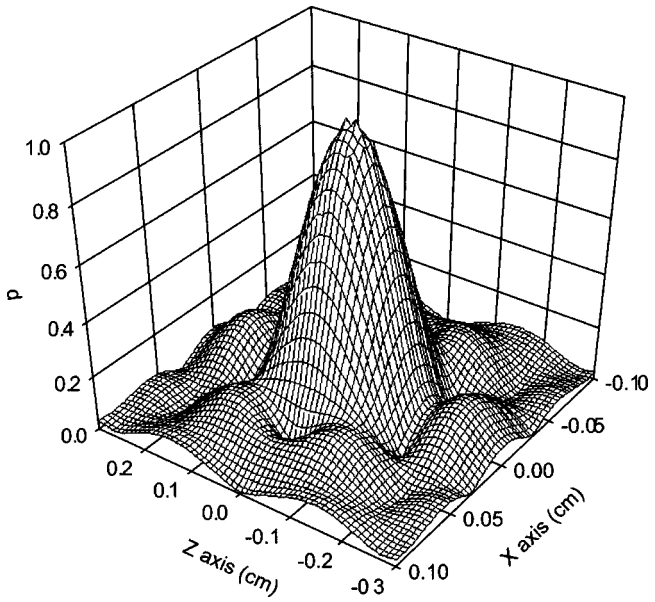


FIG. 11. Pressure plot in the $X-Z$ plane with $y=0$ cm. The numerical sampling rate was $\Delta t=10^{-9}$ s.

disappears when the pressure field is viewed in the $y=-0.5$-cm plane (Fig. 12). For this case, the -3-dB beamwidths are 2.2 mm in the X direction and 1.2 mm in the Z direction. Here, the main lobe has broadened and shifted upward in z, consistent with a view prior to the geometric focus (Fig. 10).

B. Experimental results

To validate the numerical simulations, a hydrophone was used to acquire acoustic data from a wedge transducer. The transducer was made from a spherical cap piezoelectric ceramic and had dimensions of $R=9$ cm and $a=4$ cm.[21,22] Wedges were created by etching the electrode on one side of the transducer at $h=-2$, -1, 1, and 2 cm. The etching pattern created five electrode zones that could be excited in any combination.

Experiments were conducted with degassed water using a stationary 0.2-mm hydrophone (Precision Acoustics Ltd., Dorchester, UK). The transducer was attached to motorized $X-Y-Z$ motion stages under control of a customized LABVIEW program (National Instruments, Austin, TX) that permitted single-axis and two-dimensional scanning of the sound field. The drive signal originated in an Agilent 33120 function generator (Palo Alto, CA) and was amplified by an ENI 1040L (Rochester, NY) power amplifier connected to a matching network. The hydrophone signal was sampled with an Acqiris DP 110 digitizer (Monroe, NY) also under LABVIEW control. The motion and data acquisition components were fully integrated, permitting real-time scanning and data visualization. An experiment consisted of first aligning the transducer's $y=0$ plane normal to and centered on the hydrophone. This location became the origin and all further scans were relative to this zero position.

Figure 13 shows a comparison of curves generated from experimental data (dashed) and theoretical predictions (solid) for a wedge with $h=2$ cm. The figure shows results in the

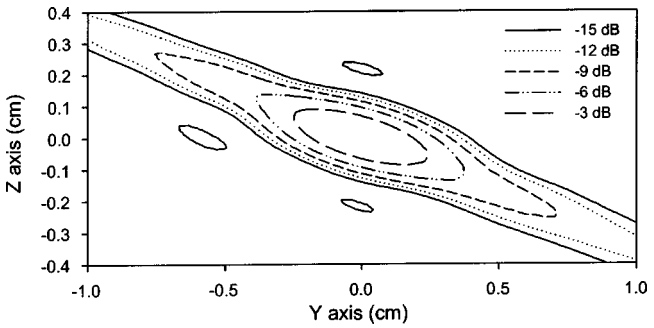


FIG. 10. Pressure plot in $Y-Z$ plane with $x=0$ cm. The wedge is located to the upper left. The isolines represent decreasing steps of -3 dB from the peak pressure at the origin. The numerical sampling rate was $\Delta t=10^{-9}$.

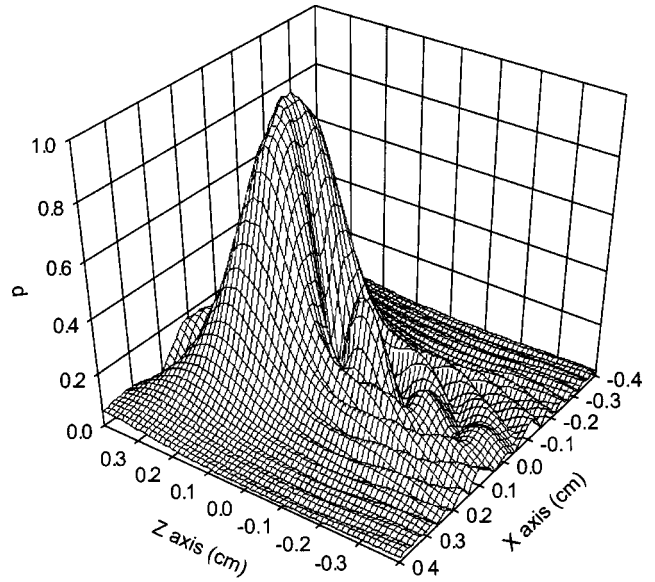


FIG. 12. Pressure plot in the $X-Z$ plane with $y=0.5$ cm. The numerical sampling rate was $\Delta t=10^{-9}$ s.

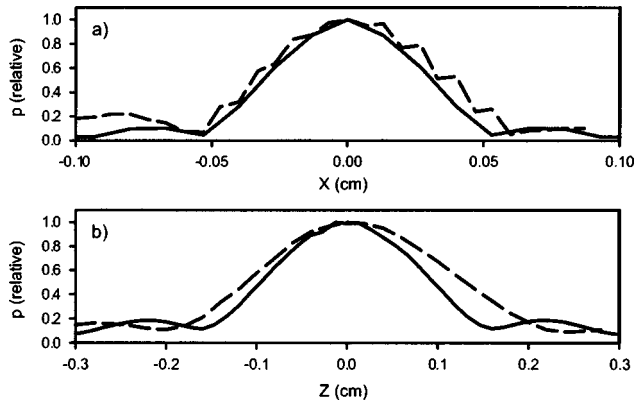


FIG. 13. Comparison of experimental data (dashed) and theoretical predictions (solid) in the $y=0$-cm plane for a wedge with $h=2$ cm. Both curves are normalized to 1.

$y=0$ plane along the X- [Fig. 13(a)] and Z-axes [Fig. 13(b)]. The peak pressure values in both curves are normalized to 1. The experimental -3-dB beamwidths are 10.55 mm on the X axis and 1.7 mm on the Z axis, in close agreement with the theoretical predictions of 0.46 and 1.4 mm, respectively. A similar result can be found in the $y=-0.5$ cm plane and is shown in Fig. 14. The beamwidths for this case are 2.2 mm on the X axis and 1.5 mm on the Z axis, again in close agreement with the theoretical predictions of 2.2 and 1.2 mm.

The results from Figs. 13 and 14 show that the analytic SIR wedge model is able to capture the beam profile of the transducer. Features observed in the theoretical model are clearly seen in the experimental results: the beamwidths broaden as the plane of observation moves from the $y=0$ to $y=-0.5$-cm planes and the location of the peak pressure moves to a more positive z value. To further validate the numerical model, a more complicated strip electrode[21,22] transducer geometry consisting of two symmetrically located wedges was examined.

The theoretical beam pattern in the $y=0$ plane for a strip electrode consisting of two wedges at $h=2$ cm and -2 cm is shown in Fig. 15. The SIR at a point **P** is found by computing the SIR for each wedge and then summing the results to find a single SIR for the combined surface area of

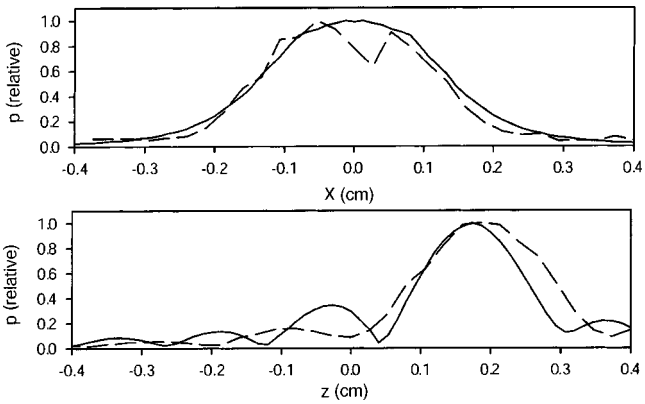


FIG. 14. Comparison of experimental data (dashed) and theoretical predictions (solid) in the $Y=-0.5$-cm plane for a wedge with $h=2$ cm. Both curves are normalized to 1.

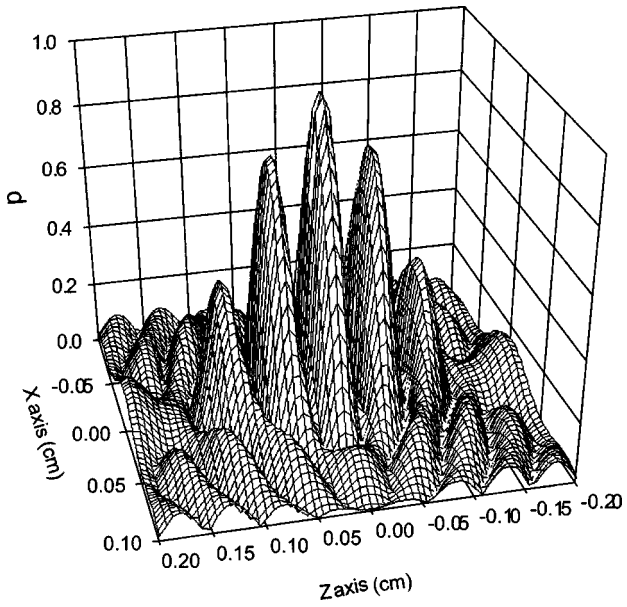


FIG. 15. Beam profile in the $y=0$ plane for a strip electrode. The transducer consists of two wedges at $h=2$ and -2 cm excited simultaneously with the same phase.

both wedges. The result is then convolved with $\partial v(t)/\partial t$ to find the pressure. For cases when each wedge is excited with its own drive signal, then each wedge is convolved with its respective $\partial v(t)/\partial t$ prior to summing to find the pressure. As noted earlier, the derivation for the SIR of a wedge was restricted to cases with $h \geqslant 0$. Therefore, to find the SIR at $\mathbf{P}=(x,y,z)$ for a wedge with $h=-2$ cm, the wedge needs to be rotated about the Y axis until it is at $h=2$ cm. For the case when the rotation is $\phi=\pi$, the SIR for $h=-2$ cm is found by using the point $\mathbf{P}=(x,y,-z)$ for $h=2$ cm.

Figure 15 shows that the beam pattern for a strip electrode is symmetric about the $x=0$ and $z=0$ planes, and that it consists of a main lobe and a series of sidelobes that are aligned in the Z direction. The width of the center lobe is greater in the X direction than in the Z direction. For the case shown here, the -3-dB beamwidths are 0.26 mm in the Z direction, and 0.46 mm in the X direction, and the lobes are spaced every 0.5 mm.

Comparisons of experimental results to theoretical predictions are shown in Fig. 16 for the Z and X axis. The experimental beamwidths are 0.26 and 0.46 mm, respectively, and the lobe spacing is ≈ 0.5 mm.

V. CONCLUSION

The analytic expressions for the SIR of a wedge-shaped section of a spherical cap transducer were derived for any point P in space. The SIR expressions for a spherical cap derived by Arditi *et al.*[10] form the core of the wedge derivation with additional terms included to account for the reduced surface area of the wedge. Because the solution of the wedge is symmetric about the $x=0$ plane, the solutions were restricted to $x>0$ when $y \leqslant 0$ and $x<0$ when $y>0$. Based on the location of **P**, one of 23 analytic expressions was used to calculate the SIR. The expressions were divided into four classes: at the origin, in the wedge shadow, out of the spheri-

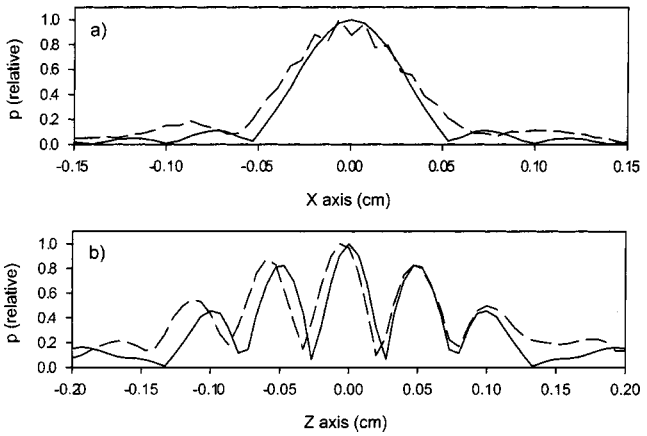


FIG. 16. Comparison of experimental data (dashed) and theoretical predictions (solid) in the $y=0$ plane for a pair of wedges at $h=2$ and -2 cm. Both curves are normalized to 1.

cal cap shadow, and in the spherical cap shadow but not in the wedge shadow. Additional tests were used to further subdivide the cases as summarized in Table I.

Numerical calculations of the wedge pressure field were compared to the experimental results acquired with a hydrophone. Measurements in the $y=0$- and -0.5-cm planes were taken for a wedge with $R=9$ cm, $a=4$ cm, and $h=2$ cm. Measurements were also taken for a strip electrode transducer geometry consisting of two wedges at $h=2$ and -2 cm. The experimental measurements and theoretical -3-dB beamwidths agreed to within $10\%\pm5\%$ for both the wedge and strip electrode transducers.

One utility of the SIR method is that the SIR at **P** can be assembled by summing or subtracting the SIRs from smaller subareas. In the case of the strip electrode the SIR is a sum of two wedges [Fig. 17(a)]. For an annular ring, the SIR is the difference between two spherical caps. More complex transducer geometries can be modeled by combining the SIR of a wedge and a spherical cap. Strips in the interior of the spherical cap can be found by taking the difference between two wedges [Fig. 17(b)]. If one of the interior strip wedges has $h=0$, then a single strip down the middle of the spherical cap can be modeled [Fig. 17(c)]. Cases with a polygon cut out of the interior of a spherical cap are modeled as a collection of wedges [Fig. 17(d)] with the restriction that the polygon's vertices all lie on a common circle. If the polygon is in the interior of the spherical cap, then the SIR is modeled as a collection of wedges and an annular ring [Fig. 17(e)]. In the reverse situation, a polygon-shaped transducer cut out of a spherical cap, such as a curved strip[20] [Fig. 17(f)], is found by subtracting the wedge SIRs from a spherical cap.

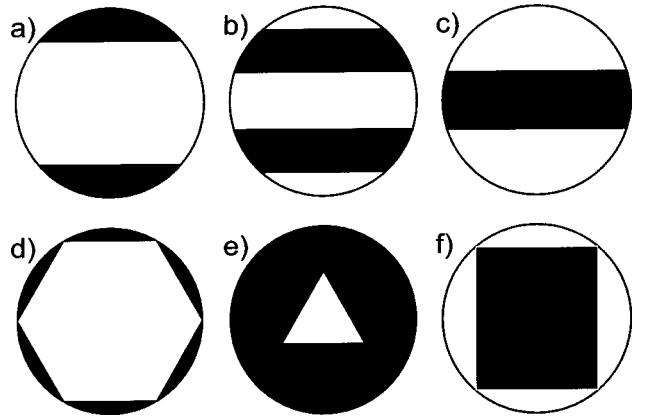


FIG. 17. Spherically curved transducer electrode patterns that can be modeled with a combination of wedges and spherical caps.

ACKNOWLEDGMENTS

The author wishes to thank Fred Lizzi, Cheri Deng, and Dan Reuster for many useful discussions. Portions of this work were supported by NIH Grants EY01212 and CA84588.

APPENDIX: EXPRESSION FOR $\beta(t)$

The expression for $\beta(t)$ derived by Arditi *et al.*[10] is

$$\beta(t)=2\cos^{-1}(\xi(t)/\sigma(t)), \tag{A1}$$

where $\sigma(t)$ was defined in Eq. (10)

$$\xi(t)=\frac{R-d}{\sin(\varphi)}+\frac{R^2+r^2-(ct)^2}{2|\mathbf{r}|\tan(\varphi)}, \tag{A2}$$

and

$$\varphi=\begin{cases}\eta; & y>0\\ \pi-\eta; & y\leqslant 0.\end{cases} \tag{A3}$$

[1] P. Stepanishen, "Transient radiation from pistons in an infinite planar baffle," J. Acoust. Soc. Am. **49**, 1629–1683 (1971).
[2] P. Penttinen and M. Luukkala, "The impulse response and pressure nearfield of a curved ultrasonic radiator," J. Phys. D **9**, 1547–1557 (1976).
[3] W. A. Verhoef, M. Cloostermans, and J. M. Thijssen, "The impulse-response of a focused source with an arbitrary axisymmetric surface velocity distribution," J. Acoust. Soc. Am. **75**, 1716–1721 (1984).
[4] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 262–267 (1992).
[5] D. Dietz, S. J. Norton, and M. Linzer, "Wideband annular array response," IEEE Ultrason. Symp. Proc. (1978), pp. 206–211.
[6] J. Lockwood and J. G. Willette, "High-speed method for computing the exact solution for the pressure variations in the nearfield of a baffled piston," J. Acoust. Soc. Am. **53**, 735–741 (1973).
[7] J. L. San Emeterio and L. G. Ullate, "Diffraction impulse response of rectangular transducers," J. Acoust. Soc. Am. **92**, 651–662 (1992).
[8] J. A. Jensen, "Ultrasound fields from triangular apertures," J. Acoust. Soc. Am. **100**, 2049–2056 (1996).
[9] J. A. Jensen, "A new calculation procedure for spatial impulse responses in ultrasound," J. Acoust. Soc. Am. **105**, 3266–3274 (1999).
[10] M. Arditi, F. S. Foster, and J. W. Hunt, "Transient fields of concave annular arrays," Ultrason. Imaging **3**, 37–61 (1981).
[11] B. Piwakowski and B. Delannoy, "Method for computing spatial pulse response: Time-domain approach," J. Acoust. Soc. Am. **86**, 2422–2432 (1989).
[12] B. Piwakowski and K. Sbai, "A new approach to calculate the field radiated from arbitrarily structured transducer arrays," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 422–440 (1999).
[13] J. F. Theumann, M. Arditi, J. J. Meister, and E. Jaques, "Acoustic fields of concave cylindrical transducers," J. Acoust. Soc. Am. **88**, 1160–1169 (1990).
[14] P. Wu and T. Stepinski, "Spatial impulse response method for predicting pulse-echo fields from a linear array with cylindrically concave surface," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 1283–1297 (1999).
[15] D. Cathignol, P. Faure, and F. Chavrier, "Acoustic field of plane or spherical transducers," Acustica **83**, 410–418 (1997).
[16] A. T. de Hoop, S. Zeroug, and S. Kostek, "Transient analysis of the transmitting properties of a focused acoustic transducer with an arbitrary rim," J. Acoust. Soc. Am. **98**, 1767–1777 (1995).

[17] D. H. Turnbull and F. S. Foster, ''Beam steering with pulsed two-dimensional transducer arrays,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **38**, 320–333 (1991).
[18] P. Crombie, P. A. J. Bascom, and R. S. C. Cobbold, ''Calculating the pulsed response of linear arrays: Accuracy versus computational efficiency,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 997–1009 (1997).
[19] M. Arditi, W. B. Taylor, F. S. Foster, and J. W. Hunt, ''An annular array system for high-resolution breast echography,'' Ultrason. Imaging **4**, 1–31 (1982).
[20] P. Faure, D. Cathignol, J. Y. Chapelon, and V. L. Newhouse, ''On the pressure field of a transducer in the form of a curved strip,'' J. Acoust. Soc. Am. **95**, 628–637 (1994).
[21] F. Lizzi, M. Astor, C. Deng, A. Rosado, R. H. Silverman, M. Rondeau, and D. J. Coleman, ''Control of lesion geometry using asymmetric beams for ultrasonic tumor therapy,'' in Proceedings of SPIE Conference on Ultrasonic Transducer Engineering (1998), pp. 99–106.
[22] F. L. Lizzi, C. X. Deng, P. Lee, A. Rosado, R. H. Silverman, and D. J. Coleman, ''A comparison of ultrasonic beams for thermal treatment of ocular tumors,'' Eur. J. Ultrasound **9**, 71–78 (1999).
[23] G. R. Harris, ''Review of transient field theory for a baffled planar piston,'' J. Acoust. Soc. Am. **70**, 10–20 (1981).
[24] H. T. O'Neil, ''Theory of focusing radiators,'' J. Acoust. Soc. Am. **21**, 516–526 (1949).
[25] I. S. Gradshteyn, I. M. Ryzhik, and A. Jeffrey, *Table of Integrals, Series, and Products*, 6th ed. (Academic, San Diego, 2000).
[26] J. A. Ketterling and F. L. Lizzi, ''Time-domain pressure response of arrays with periodic excitation,'' J. Acoust. Soc. Am. **114**, 48–51 (2003).

The acoustic impedance of a circular orifice in grazing mean flow: Comparison with theory

Keith S. Peat[a)]
Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough, Leicestershire LE11 3TU, United Kingdom

Jeong-Guon Ih and Seong-Hyun Lee
Center for Noise and Vibration Control, Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology, Science Town, Taejon 305-701, Korea

(Received 25 October 2002; revised 20 June 2003; accepted 8 September 2003)

It is well known that the presence of a grazing mean flow affects the acoustic impedance of an aperture, but the detailed nature and understanding of the influence is still unknown. In this paper, results from a recent theoretical analysis of the problem are compared with a new set of experimental results. The purpose is twofold. First, the experimental results are used to validate the theory. It is found that the theory predicts the resistance quite well, but not the reactance. Second, the theory is used to try and give some physical understanding to the experimental results. In particular, some scaling laws are confirmed, and it is also shown that measured negative resistance values are to be expected. They are not erroneous, as previously thought. Former sets of experimental data for this problem are notable for the amount of variation that they display. Thus, both the theory and the new experimental results are also compared with those earlier detailed results that most closely conform to the conditions assumed here, namely fully developed turbulent pipe flow of low Mach number past circular orifices. The main field of application is in flow ducts, in particular, flow through perforated tubes in exhaust mufflers. [DOI: 10.1121/1.1621865]

PACS numbers: 43.20.Rz, 43.20.Tb [LLT] Pages: 3076–3086

I. INTRODUCTION

Perforations in the walls of flow ducts are common sound-attenuating features in many applications. The mean flow may be forced through the perforations, or be tangential to the plane of the perforations, or be a mixture of both. In any event, it is well known that the flow affects the acoustic impedance of the orifices in the perforate. Attention is focused here on tangential flow over a circular orifice where the flow is turbulent, fully developed, and of low Mach number. Such conditions are typically found in exhaust mufflers of internal combustion engines. A typical orifice for this application would have a diameter of 4 mm and a thickness of 1 mm. Particular attention is given to those sizes of orifice that may be found in exhaust mufflers.

Although it has been known for many years that mean flow affects the acoustic impedance of an orifice, the detailed mechanism and nature of that influence is still unknown. Consequently the grazing flow effect on orifice impedance is generally accounted for by some empirical equation derived from one or more of the many experimental investigations[1–13] into the effect. There is much variation between experimental results, although general trends are observed through them all, such as an increase of orifice resistance and a decrease of orifice reactance with increasing mean flow. The variation can in part be explained by differences in the form of mean flow, and in particular the boundary layer development, between the different sets of results. Theoretical models[5,14–18] of various complexity for the grazing flow effect on orifice impedance have at best predicted the general trends, but a detailed validation against relevant experimental results has been poor. Furthermore, they have generally employed some form of empirical parameter.

More recently, Howe *et al.*[19] have presented a purely theoretical analysis of the influence of grazing flow upon the conductivity of a circular aperture when the flow is of low Mach number but high Reynolds number, such that viscous effects are unimportant, except in the generation of vorticity. The aperture is then assumed to be spanned by a vortex sheet of zero thickness. Thus, the assumptions in the theory basically coincide with conditions for exhaust muffler applications, except that the analysis is for an orifice of zero thickness. Howe[20] gives a general method by which this form of an analysis can be extended to an orifice of finite thickness. However, the extension is only valid for an orifice of small thickness to radius ratio, and also when the wavelength of disturbances on the vortex sheet is much greater than the orifice thickness. In a further paper, Howe[21] has given an approximate method by which the general analysis can be extended to rectangular apertures of large thickness to width ratio. In this paper, the first form of extension[20] is applied to the theory for grazing flow past a circular orifice to generate results for the conductivity and impedance of an orifice of finite thickness. The trends for the theoretical change of conductivity and impedance with increasing orifice thickness are

[a)]Author to whom correspondence should be addressed. Keith S. Peat, Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough, Leicestershire LE11 3TU, United Kingdom. Telephone: +44.1509.227232; fax: +44.1509.227275; electronic mail: K.S.Peat@lboro.ac.uk

thereby obtained, in addition to specific results for a given size of orifice.

The theoretical results are then compared against a new set of experimental data.[22] The comparison is best if one makes a minor revision in the interpretation of the theory. The variable that had been introduced as the mean flow velocity is best regarded as the velocity of the convection of vorticity. This change results in predictions of orifice resistance that not only give the correct trend with an increasing Strouhal number, but that are also fairly accurate in value. The theoretical results for reactance, by contrast, bear little relation to the experimental results in either general trend with an increasing Strouhal number or value.

Jing *et al.*[23] adapted the theory of Howe *et al.*[19] to impose the continuity of normal velocity across the shear layer rather than continuity of displacement. They argued that the shear layer for fully developed turbulent flow is thick relative to the transverse dimension of the orifice and hence the continuity of velocity is more appropriate than the continuity of displacement. Furthermore, theoretical and experimental work by Nelson *et al.*[24,25] lends some weight to their argument. Jing *et al.*[23] introduced the shorthand PDM (particle displacement match) and PVM (particle velocity match) to distinguish the two cases, and the same terminology will be used here. They compared results for orifice impedance as a function of Mach number from both the PDM and PVM with measured values and concluded that the PVM values were more accurate for both resistance and reactance. Here we conduct a more exacting test by comparing results for orifice impedance as a function of Strouhal number from both the PDM and PVM with the recent measured values.[22]

The theoretical results are also used to help to interpret the experimental data. In particular, the theory suggests that the resistance and reactance should vary linearly with Mach number, and thereafter be functions only of Strouhal number and the thickness to the radius ratio of the orifice. This form of functional dependence is confirmed in full for the orifice resistance. The linear dependence on Mach number is essentially confirmed for reactance also, but thereafter results for reactance are disappointing.

The theoretical and experimental results are then compared against earlier empirical formulas,[12,13] which were developed from experiments of fully developed turbulent flow over circular orifices. These formulas are in terms of a Strouhal number based on friction velocity, following the practice of previous work[11] that considered young boundary layers. The Strouhal number in this paper is based on mean flow velocity, since the detailed boundary layer structure forms no part of the theoretical model. It would appear that all results for orifice impedance, where the grazing flow is fully developed and turbulent, scale just as well using mean flow velocity as using friction velocity.

In previous experimental tests,[12,13] negative values of resistance had been measured but then discarded, since they were thought to be erroneous. Likewise in the new experimental results, negative resistance values have been measured over some range of high Strouhal numbers. However, the data is not erroneous, but is correctly predicted by the theory.[19] In essence, the values are negative since in this Strouhal number range the oscillation of the shear layer over the orifice generates sound. For most Strouhal numbers, the incident sound field forces the oscillation of the shear layer, which thereby absorbs sound energy and leads to positive orifice resistance.

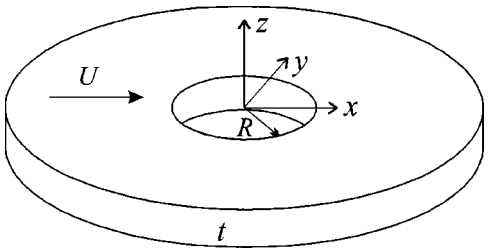


FIG. 1. Circular orifice in an infinite plate of thickness t with a grazing flow of velocity U above the plate.

II. GOVERNING EQUATIONS

Howe *et al.*[19] have shown that grazing flow of a low Mach number past an acoustically compact circular orifice in a plate of infinitesimal thickness in a harmonic sound field of frequency ω causes a perpendicular displacement $\zeta(x,y)$ of the shear layer in the plane of the aperture, Ω, given by

$$\left[\left(1+\frac{i}{S}\frac{\partial}{\partial x}\right)^2+1\right]\frac{1}{2}\int\int_{\Omega}\frac{\zeta(\xi,\eta)}{\sqrt{(x-\xi)^2+(y-\eta)^2}}\,d\xi\,d\eta=1; \tag{1}$$

see Fig. 1. Here $S=\omega R/U$ is the Strouhal number, where U is the mean flow velocity above the plate. All lengths have been nondimensionalized with respect to the orifice radius R and the nondimensional displacement $\zeta(x,y)$ also incorporates the dimensionless applied acoustic pressure differential across the orifice, such that

$$\zeta=\frac{\rho\omega^2 R\tilde{\zeta}}{\pi(p_+-p_-)}, \tag{2}$$

where $\tilde{\zeta}$ is the actual vertical displacement, ρ is the mean fluid density, and p_+, p_- are the farfield pressures above and below the plate, respectively. $\tilde{\zeta}$ and the vertical coordinate z are measured positive upward; see Fig. 1.

Howe[20] also gives a general approximate method for the inclusion of finite plate thickness in such problems, for orifices of small thickness (t) to radius ($R=d/2$) ratio $h\ll 1$ and also given $h\ll\lambda/R=2\pi/S$, where λ is the wavelength of disturbances on the vortex sheet. In essence, the approximation requires the vertical displacement of fluid particles within the thickness of the orifice to be independent of z and Eq. (1) is modified to become

$$\left[\left(1+\frac{i}{S}\frac{\partial}{\partial x}\right)^2+1\right]\frac{1}{2}\int\int_{\Omega}\frac{\zeta(\xi,\eta)}{\sqrt{(x-\xi)^2+(y-\eta)^2}}\,d\xi\,d\eta + \pi h\zeta(x,y)=1. \tag{3}$$

This can be rewritten as

$$\left[\left(1+\frac{i}{S}\frac{\partial}{\partial x}\right)^2+1\right]\left\{\frac{1}{2}\int\int_{\Omega}\frac{\zeta(\xi,\eta)}{\sqrt{(x-\xi)^2+(y-\eta)^2}}d\xi\,d\eta + \int_{\Omega_y}\pi h G(x,\xi)\zeta(\xi,y)d\xi\right\}=1, \tag{4}$$

where

$$G(x,\xi)=\frac{S}{2}[H(x-\xi)e^{i\sigma_1(x-\xi)}+H(\xi-x)e^{i\sigma_2(x-\xi)}], \tag{5}$$

with

$$\sigma_1=S(i+1), \quad \sigma_2=S(i-1), \tag{6}$$

since the Green's function $G(x,\xi)$ then satisfies

$$\left[\left(1+\frac{i}{S}\frac{\partial}{\partial x}\right)^2+1\right]G(x,\xi)=\delta(x-\xi). \tag{7}$$

The region of integration Ω_y is that portion of x within Ω for given y. It may be noted that σ_1 and σ_2 are the dimensionless Kelvin–Helmholtz wave numbers of the instability waves of the vortex sheet.

Equation (4) can now be integrated with respect to the second-order differential operator to give

$$\int\int_{\Omega}\frac{\zeta(\xi,\eta)}{\sqrt{(x-\xi)^2+(y-\eta)^2}}d\xi\,d\eta + 2\pi h\int_{\Omega_y}G(x,\xi)\zeta(\xi,y)d\xi+A(y)e^{i\sigma_1 x}+B(y)e^{i\sigma_2 x}=1. \tag{8}$$

The amplitudes $A(y)$ and $B(y)$ are to be determined so as to satisfy the Kutta condition, namely, that the vortex sheet leaves the upstream semicircular arc of the aperture edge tangentially.

This equation can be solved numerically for ζ, A, and B using the same technique as for an orifice of zero thickness.[19] Values of ζ are found for each element in a set of square elements overlying the circular orifice. For a given streamwise strip of elements, ζ is set to be zero on the first two upstream elements to enforce the Kutta condition such that, with A and B for the strip unknown, one has the same number of equations as unknowns. The Rayleigh conductivity of the orifice, K, can then be evaluated by numerical integration, since[20]

$$K/(2R)=\Gamma-i\Delta=\frac{\pi}{2}\int\int_{S}\zeta(x,y,S)dx\,dy. \tag{9}$$

PDM results for the real and imaginary parts of the conductivity as a function of Strouhal number are shown in Fig. 2, for varying hole thickness to radius ratios, namely $h=0$, 0.25, and 0.5. The results for $h=0$ correspond precisely with those of Howe *et al.*,[19] confirming the validity of the computations, especially with regard to the fineness of the mesh. The general effect of increasing the orifice thickness is seen to be non-negligible. The general form remains unaltered, but it is skewed to higher Strouhal numbers for increasing h. At similar parts of the general pattern, the magnitude of the real part of the conductivity decreases with increasing h, while the imaginary part of the conductivity increases with increasing h. These results should be treated with a degree of caution, however, as the assumption that the displacement is constant throughout the thickness of the orifice is only assuredly valid when $h \ll 1$ and therefore is already suspect even at $h=0.25$.

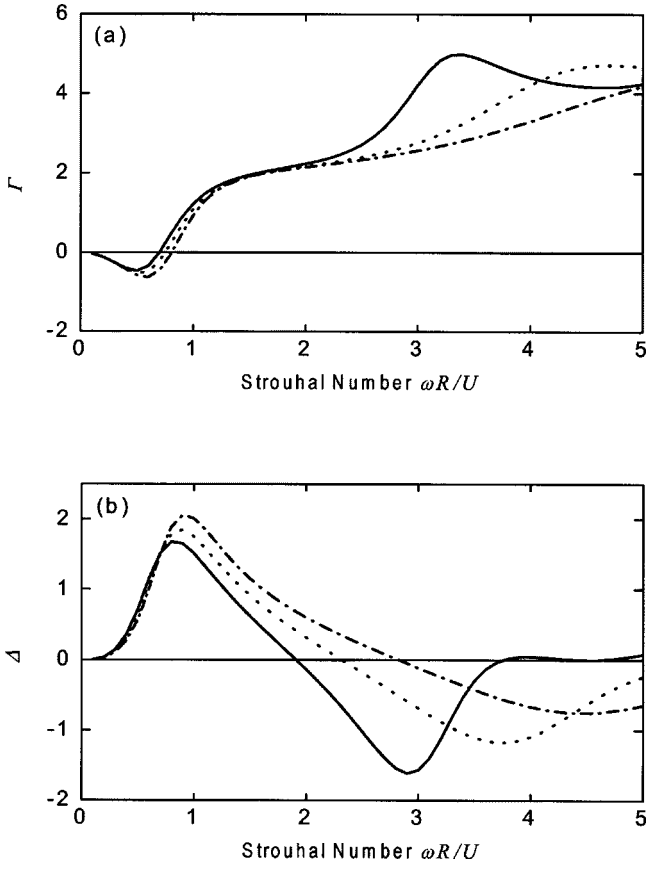


FIG. 2. The influence of plate thickness on the conductivity of a circular aperture (PDM theory). (a) Real part; (b) imaginary part. —, $h=0$; ---, $h=0.25$; –·–, $h=0.5$.

In the high Strouhal number limit, which corresponds to the zero mean flow case, the imaginary part of the conductivity tends to zero for all h, as indeed it should, since viscous effects are ignored other than in generation of the shear layer. For nonzero hole thickness, this happens at Strouhal numbers above those shown in Fig. 2. The real part of the conductivity should become equal to $\pi/2(h+\delta_0/R)$ in the no-flow limit, where[26] the end correction $\delta_0 \approx 2\times 0.85R$. It is seen from Fig. 2 that Howe's theory gives an incorrect value, even in the case of $h=0$, and indeed the same is found to be true for orifices of nonzero thickness. These observations cause no great surprise or concern since the theory becomes invalid in the limit of a large Strouhal number. The theory does, however, predict the correct trend of decreasing conductivity with an increasing orifice thickness in this limit.

The continuity of normal particle velocity W across the shear layer instead of displacement results in Eq. (3) being replaced by[23]

$$\left(2iS-\frac{\partial}{\partial x}\right)\frac{1}{2\pi}\int\int_{\Omega}\frac{w(\xi,\eta)}{\sqrt{(x-\xi)^2+(y-\eta)^2}}d\xi\,d\eta$$
$$+iShw(x,y)=1, \tag{10}$$

where

$$w=\frac{iS\rho U\omega W}{\pi(p_+-p_-)}. \tag{11}$$

Once again the treatment of finite thickness is approximate, as this time w is assumed to be constant with respect to z within the thickness of the orifice. This can be rewritten as

$$\left(1+\frac{i}{2S}\frac{\partial}{\partial x}\right)\left\{\int\int_{\Omega}\frac{w(\xi,\eta)}{\sqrt{(x-\xi)^2+(y-\eta)^2}}d\xi\,d\eta\right.$$
$$\left.+\int_{\Omega_y}\pi hG(x,\xi)w(\xi,y)d\xi\right\}=1, \tag{12}$$

where

$$G(x,\xi)=-2iSH(x-\xi)e^{2iS(x-\xi)}. \tag{13}$$

Equation (13) can now be integrated with respect to the first-order differential operator to give

$$\int\int_{\Omega}\frac{w(\xi,\eta)}{\sqrt{(x-\xi)^2+(y-\eta)^2}}d\xi\,d\eta+\pi\int_{\Omega_y}G(x,\xi)w(\xi,y)d\xi$$
$$+A(y)e^{2iSx}=1. \tag{14}$$

This equation can be solved in a similar manner to Eq. (8), although now the Kutta condition is applied by setting $w=0$ at only the first upstream element of a given streamwise strip.

PVM results for the real and imaginary parts of the conductivity as a function of Strouhal number are different in value from those shown in Fig. 2 but have a similarly marked variation with h.

III. COMPARISON WITH EXPERIMENTS

A detailed set of experiments[22] has been conducted to determine the specific impedance of an orifice as a function of various parameters. The specific impedance is related to the conductivity as

$$\varphi=\theta-i\chi=\frac{\pi kR(\Gamma-i\Delta)}{\Gamma^2+\Delta^2}=\frac{\pi MS(\Gamma-i\Delta)}{\Gamma^2+\Delta^2}, \tag{15}$$

where k is the wave number and M is the Mach number of the mean flow. Since, according to the theory of Sec. II, Γ and Δ are functions only of S and h, it follows that for orifices of constant h, θ/M and χ/M should be functions only of S. Figures 3 and 4 show experimental values of θ_f/M and χ/M, respectively, plotted against S to test this deduction from the theory. θ_f is the orifice resistance due to flow, in accordance with the theoretical value of θ for which viscous effects were ignored. θ_f is evaluated by subtracting experimental values for the no-flow resistance from the measured values with flow. In any event the no-flow resistance is very small. Results shown are for varying flow Mach number,

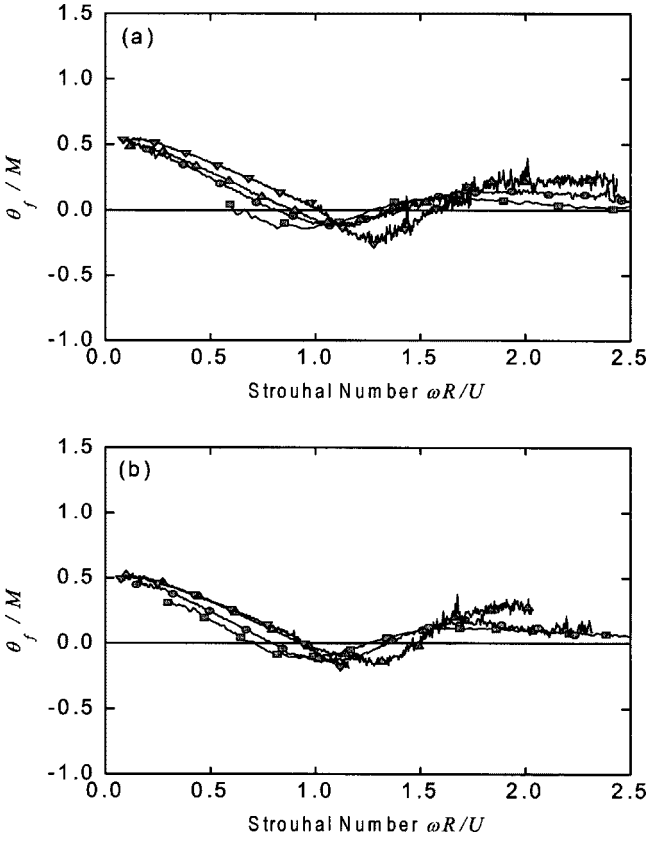


FIG. 3. Experimental resistance for varying grazing flow Mach number of a circular aperture ($d=8$ mm, $t=1$ mm). (a) -□-, $M=0.021$; -○-, $M=0.064$; -△-, $M=0.106$; -▽-, $M=0.149$; (b) -□-, $M=0.043$; -○-, $M=0.085$; -△-, $M=0.128$; -▽-, $M=0.170$.

$0.02<M<0.2$, but for orifices of a single size $R=4$ mm and thickness $t=1$ mm, hence constant $h=0.25$. The resistance results in Fig. 3 do show some slight variation with Mach number. However, the difference between the curves is not consistent in trend with increasing Mach number, and thus may be due to experimental error. An exception to the last statement occurs for the curves for the two lowest Mach numbers. These two curves are the most distinct from the rest, and a clear trend with increasing Mach number is observed. The Reynolds number of the mean flow, based on the duct width for these two lowest flow speeds are 1.5×10^4 and 3×10^4, respectively. The theory assumes high Reynolds number flow, such that the only effect of viscosity is in the generation of an infinitely thin shear layer. Thus, the failure of the experimental results for lowest Mach numbers to collapse onto a single curve is probably due to the Reynolds number being too low.

It is seen from Fig. 4 that the reactance data does effectively collapse onto a single curve for both low and high Strouhal numbers. However, there is an intermediate range of Strouhal number, $0.75<S<1.5$, where a small but clear distinction between the results for a different Mach number is observed. The trend is for χ/M to decrease with increasing Mach number at a given Strouhal number.

Figures 5 and 6 show the theoretical PDM and PVM results for θ/M and χ/M, respectively, for orifices with $h=0$, 0.25, and 0.5. For nonzero h, the PVM analysis assumes

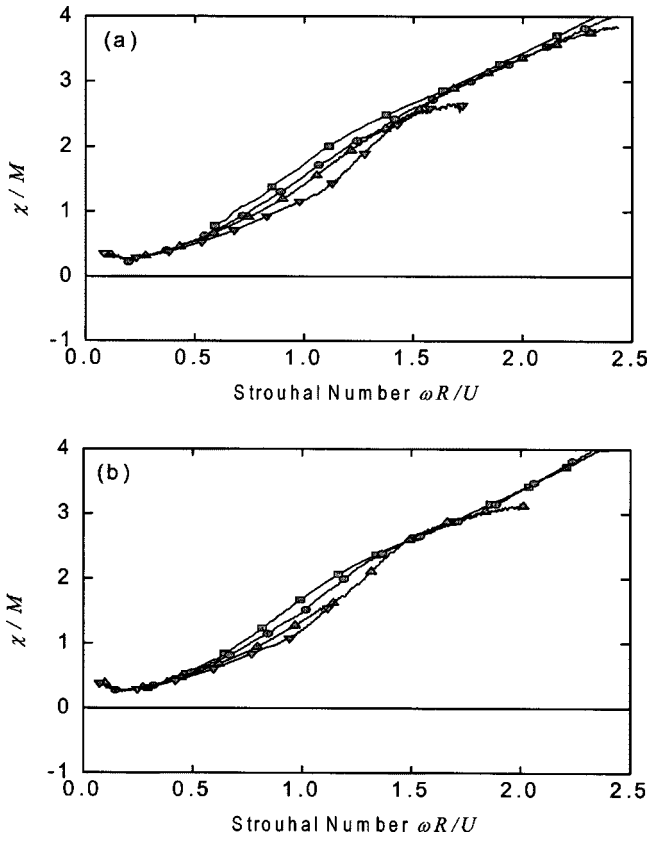


FIG. 4. Experimental reactance for varying grazing flow Mach number of a circular aperture ($d=8$ mm, $t=1$ mm). (a) -□-, $M=0.021$; -○-, $M=0.064$; -△-, $M=0.106$; -▽-, $M=0.149$; (b) -□-, $M=0.043$; -○-, $M=0.085$; -△-, $M=0.128$; -▽-, $M=0.170$.

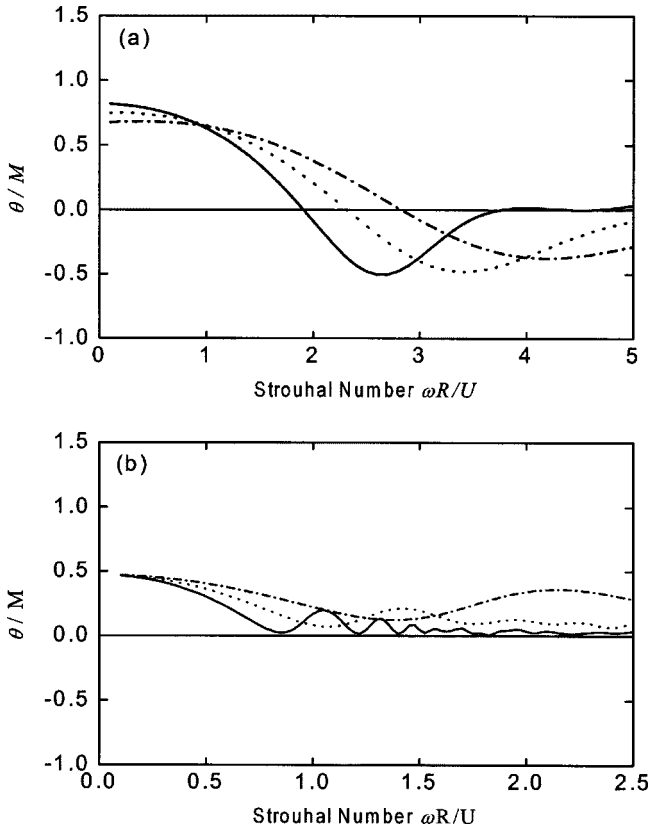


FIG. 5. The influence of plate thickness on the resistance of a circular aperture. —, $h=0$; ---, $h=0.25$; –·–, $h=0.5$. (a) PDM; (b) PVM.

that the vertical velocity is constant throughout the thickness of the hole, in the same way that the PDM assumed constant displacement. Thus, as mentioned earlier, only results for $h \ll 1$ can be expected to meet this requirement, and the results for the higher values of h are circumspect. However, the general trend is seen to be the same for all three values of h and thus some general comparison between the theoretical values for $h=0$ and 0.25 and the preceding experimental results can be made. We concentrate first on PDM results. It is seen from Fig. 5(a) that the general trend of θ/M in the theoretical results is indeed the same as for the corresponding experimental results of Fig. 3. However, in detail it is seen that the Strouhal numbers in the theoretical results are much higher than for the corresponding part of the curve from the experimental results. In addition, the absolute values of θ/M at corresponding points on the curves are greater in the theoretical than in the experimental results. Now in the theoretical development, the mean flow velocity U is assumed to affect only the development of the shear layer and the convection of vorticity along it. Thus it may be argued that the pertinent velocity is not that of the mean flow, but that of the convection of vorticity, say U_c, in the same manner as that adopted for the study of bias flow on the conductivity of apertures[20] and flow-induced resonances.[27] Now[20,24,27] $U_c \approx 0.5U$ and assuming that in the theoretical results S is actually $S_c=\omega R/U_c$, then it follows that $S_c=2\omega R/U$. Thus, if Figs. 5(a) and 6(a) are interpreted as a plot of impedance against S_c, the scale of the horizontal axis

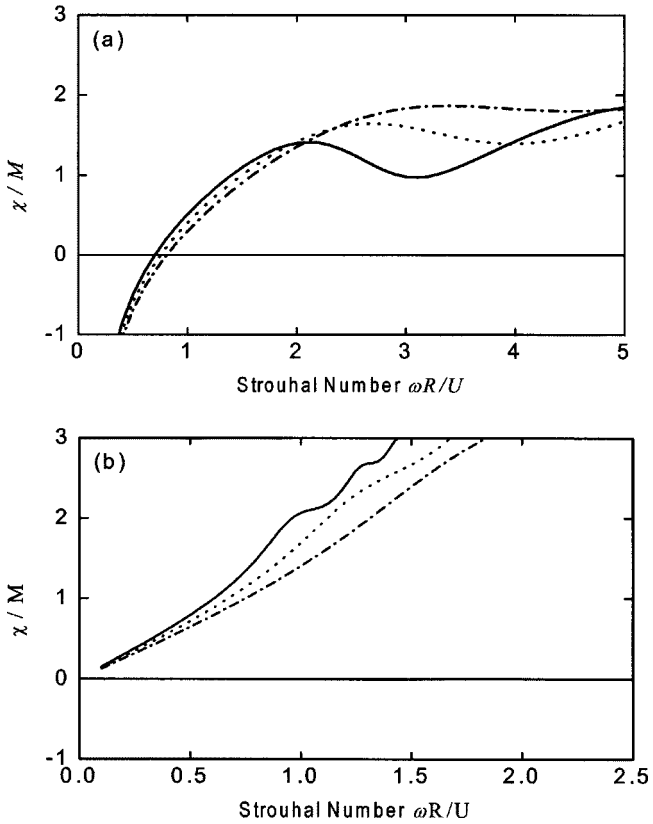


FIG. 6. The influence of plate thickness on the reactance of a circular aperture. —, $h=0$; ---, $h=0.25$; –·–, $h=0.5$. (a) PDM; (b) PVM.

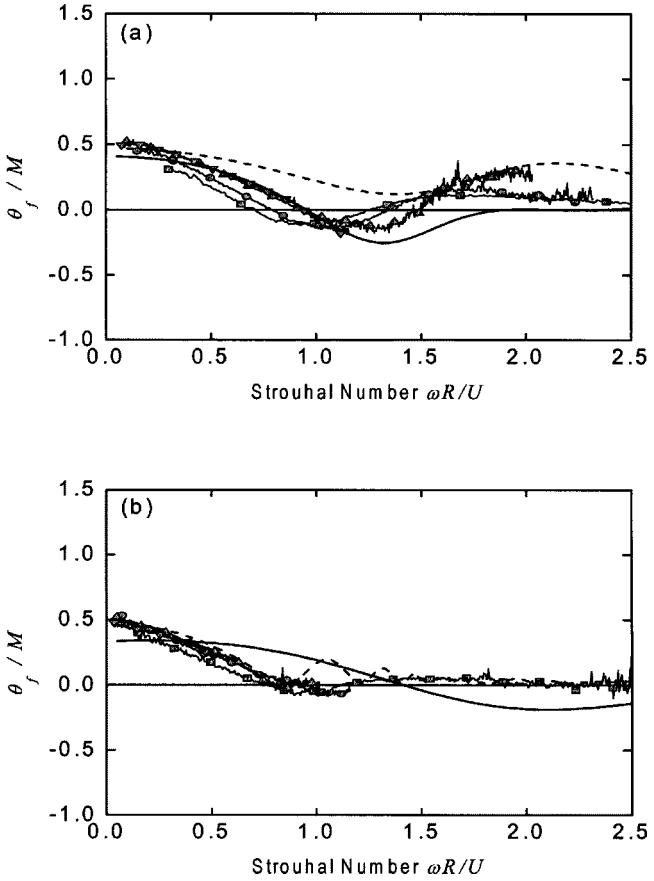


FIG. 7. Experimental resistance for two different sizes of orifice with the same value $h=0.5$ with varying grazing flow Mach number. (a) ($d=8$ mm, $t=2$ mm) -■-, $M=0.043$; -●-, $M=0.085$; -▲-, $M=0.128$; -▼-, $M=0.170$; —, PDM*, $h=0$; ---, PVM, $h=0$. (b) ($d=4$ mm, $t=1$ mm) -■-, $M=0.043$; -●-, $M=0.085$; -▲-, $M=0.128$; -▼-, $M=0.170$; —, PDM*, $h=0.5$; ---, PVM, $h=0.5$.

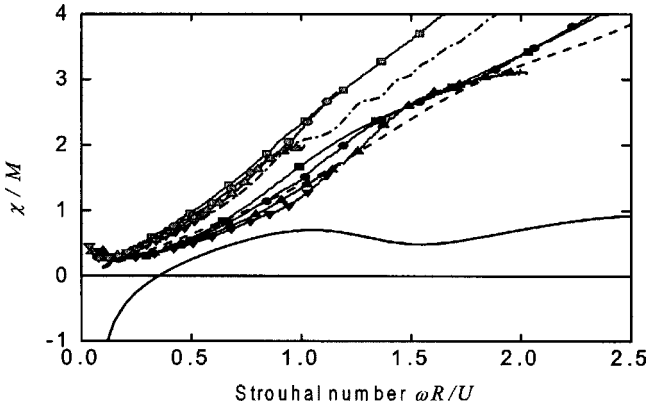


FIG. 8. Experimental reactance for two different sizes of orifice with the same value $h=0.5$ with varying grazing flow Mach number. ($d=8$ mm, $t=2$ mm) -■-, $M=0.043$; -●-, $M=0.085$; -▲-, $M=0.128$; -▼-, $M=0.170$; ($d=4$ mm, $t=1$ mm) -■-, $M=0.043$; -●-, $M=0.085$; -▲-, $M=0.128$; -▼-, $M=0.170$; —, PDM*, $h=0$; ---, PVM, $h=0$; –·–, PVM, $h=0.5$.

will be from 0 to 2.5 in terms of $S=\omega R/U$. The results then give a much closer correlation with experimental results, and indeed a secondary benefit is found because the amplitude of the theoretical results is also reduced. From this point onward it will be assumed that in the PDM results U should actually be U_c, and they are referred to as PDM* results to remind us of this modification.

A comparison of Figs. 4 and 6(a) shows that there is less similarity in general form between experimental and PDM results for the case of reactance χ/M. However, a change from mean flow velocity to the convection velocity of vorticity in the theoretical results, as noted above, would certainly be beneficial.

Turning now to the PVM results in Figs. 5(b) and 6(b), a comparison with the experimental results of Figs. 3 and 4, respectively, indicates that no such change in the definition of a Strouhal number is required here, and this time the correct trend is predicted for both the resistance and the reactance.

Figures 7 and 8 show experimental results for θ/M and χ/M, respectively, for two different sizes of orifice with the same value $h=0.5$. One form of orifice has a radius of 4 mm and a thickness of 2 mm, while the other has a radius of 2 mm and a thickness of 1 mm. According to the theory, results for both holes and for all Mach numbers should collapse onto a single curve. It is seen from Fig. 7 that this is approximately true for θ/M. The small discrepancies that are observed can probably be attributed to experimental error, since there is no clear trend of distinction by either orifice type or Mach number, at least at low Strouhal number. At Strouhal numbers above 1.5, there is some evidence that the larger diameter orifice gives the greater resistance. The PDM* and PVM curves for $h=0$ and $h=0.5$ are also shown for comparison in Figs. 7(a) and (b), respectively. The comparison is now reasonable in terms of values as well as the general trend. It may be noted that for PDM* results the curve for $h=0$ gives the best comparison, both in terms of the location of the minimum point of the curve with Strouhal number and the amplitude of the results. In contrast, for PVM results the $h=0.5$ curve clearly gives the best fit to experimental results. It is in noticeable error only at Strouhal numbers close to one, where the experimental results for resistance are negative. It seems to be a feature of the PVM results that resistance values are always positive, in contrast to the PDM* results that correctly predict negative values at some range of Strouhal number.

In contrast to Fig. 7, the results for χ/M in Fig. 8 very clearly do not collapse onto a single curve. The χ/M results for different Mach numbers for each individual size of orifice do nearly collapse onto single curves, although as noted earlier there is a small intermediate range of Strouhal numbers for which even this is not true. It can be seen from Fig. 8 that the magnitude of such variations is less for the larger orifice than for the smaller one. However, the key feature from Fig. 8 is that the results for the two different orifice sizes are very distinctly different, the larger and thicker orifice giving the higher values of χ/M. In the high Strouhal number, no-flow limit, the reactance should be $kR(h+\delta_0/R)$, or $\chi/M=S(h+\delta_0/R)$, where δ_0 is the no-flow end correction. Thus the gradients of the basic curves in Fig. 8 should be the same at a high Strouhal number and are clearly not. In the case of the theoretical curves, this is merely a consequence of the limit value of the conductivity always being incorrect, as noted earlier. The PVM results for $h=0$ and $h=0.5$ show the cor-

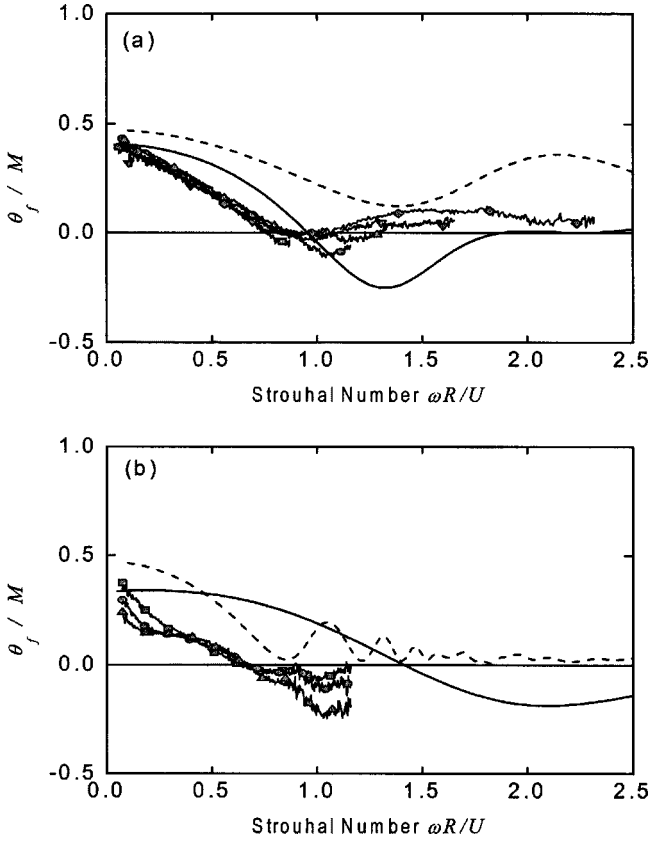


FIG. 9. The influence of variation of h upon experimental resistance with grazing flow Mach number of 0.085. (a) ($t=1$ mm) -■-, $d=3$ mm; -●-, $d=4$ mm; -▲-, $d=4.6$ mm; -▼-, $d=5.7$ mm; -◆-, $d=8$ mm; —, PDM*, $h=0$; ---, PVM, $h=0$. (b) ($d=4$ mm) -■-, $t=2$ mm; -●-, $t=3$ mm; -▲-, $t=5$ mm; —, PDM*, $h=0.5$; ---, PVM, $h=0.5$.

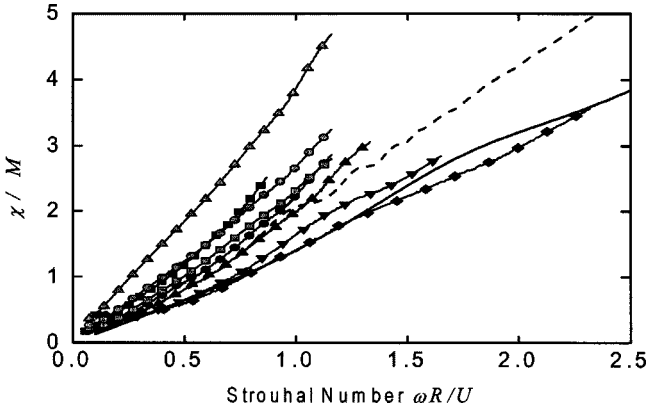


FIG. 10. The influence of variation of h upon experimental reactance with grazing flow Mach number of 0.085. ($t=1$ mm) -■-, $d=3$ mm; -●-, $d=4$ mm; -▲-, $d=4.6$ mm; -▼-, $d=5.7$ mm; -◆-, $d=8$ mm; ($d=4$ mm) -■-, $t=2$ mm; -●-, $t=3$ mm; -▲-, $t=5$ mm; —, PVM, $h=0$; ---, PVM, $h=0.5$.

rect variation with Strouhal number. The $h=0$ curve gives a very good fit to the experimental data for the larger diameter orifice and the $h=0.5$ curve lies within the data for the two sizes of orifice. The PVM results for reactance are very much better than the PDM* results.

No detailed measurements of the boundary layer structure were conducted during the experiments.[22] However, the range of Reynolds number that the tests cover implies that a one-seventh power law should be appropriate to describe the velocity profile, and this in turn would indicate a boundary layer thickness of the order of 2–3 mm, i.e., of comparative order to the orifice radius and thickness. Thus the shear layer in the experiments is not thin and clearly defined, as assumed in Howe's theory and in such a case continuity of particle velocity[23,28] is probably a more appropriate boundary condition than the continuity of displacement.

The effects of the variation of h upon θ/M and χ/M, for a range of Strouhal numbers, are shown in Figs. 9 and 10, respectively. In both cases, the variation with h is obtained in two different ways, first by varying the thickness of a 4 mm diameter orifice to give $1 \leqslant h \leqslant 2.5$, and, second, by varying the diameter of a 1 mm thick orifice to give $0.25 \leqslant h \leqslant 0.6$. It is clear from Fig. 9 that there is no noticeable change of resistance with h, except maybe for large values of h that are unrepresentative of exhaust silencer applications. Even then the variation does not occur for all Strouhal numbers and may well be attributable to experimental error. The observed variation is certainly much less than that predicted by the theory, shown in Fig. 5 for a smaller range of h. The theory is inexact for nonzero h and, since it would appear that it overpredicts the dependence on h, there is some justification to compare the experimental results against the $h=0$ theoretical curve. Accordingly, in later figures only $h=0$ results are shown.

In contrast to the results for resistance, it is seen from Fig. 10 that the reactance is strongly affected by h. This is scarcely surprising, since the zero flow reactance increases with both orifice thickness and end correction, and hence orifice diameter. The PDM* results for reactance, given in Fig. 6, bear little resemblance to the measured values. In contrast the PVM results give the correct trend and sensible values. PVM results for $h=0$ and $h=0.5$ are shown in Fig. 10 for comparative purposes. The PVM result for $h=0$ is very similar to the experimental result of least h, namely $h=0.25$. The PVM result for $h=0.5$ is most similar to an experimental result of $h=0.435$ for a thin aperture, even though there is an experimental result for $h=0.5$. Once again, though, it is noticeable that the experimental results for apertures of different size but the same value of h do not collapse onto a single curve.

IV. COMPARISON WITH PREVIOUS RESULTS

From the many previous works on the measurement of orifice impedance mentioned earlier, that of Cummings[12] and Kirby and Cummings[13] are singled out for special attention here since they address the impedance of an orifice measured in a flow duct containing fully developed turbulent flow. This is not only the situation prevalent in exhaust silencer applications, but is also that which is closest to the assumed conditions in the theory of Howe *et al.*[19] Cummings[12] followed previous work by Kooi and Sarin,[11] in that he sought to collapse data by plotting $\theta c/fd$ and δ/δ_0 against u_*/fd and u_*/ft, respectively, where u_* is the friction velocity and f is the cyclic frequency. The work of Kooi and Sarin was done for mean flow with "young" boundary layers and low turbulence levels, conditions in which the precise form of the

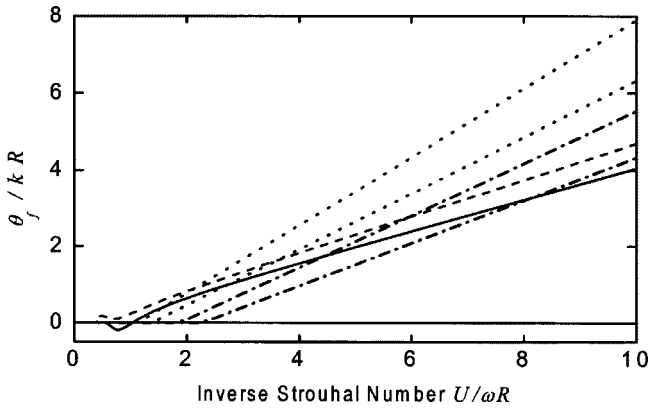


FIG. 11. Theoretical resistance compared to empirical formulas as a function of the inverse Strouhal number. —, PDM*, $h=0$; – – –, PVM, $h=0$; ---, Cummings (Ref. 12); – –, Kirby and Cummings (Ref. 13).

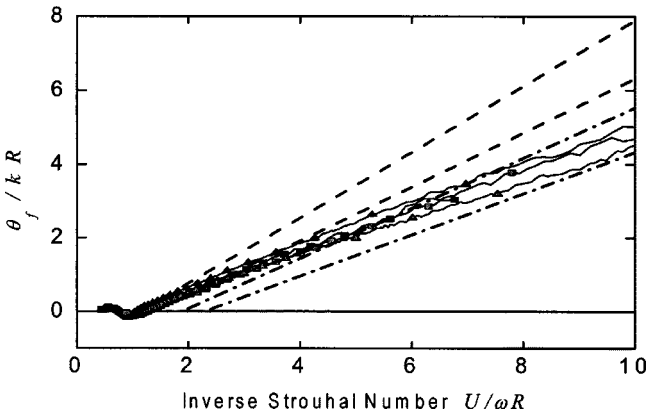


FIG. 12. Experimental resistance compared to empirical formulas as a function of inverse Strouhal number. ($d=8$ mm, $t=2$ mm) -■-, $M=0.085$; -▲-, $M=0.170$; ($d=4$ mm, $t=1$ mm) -■-, $M=0.085$; -▲-, $M=0.170$; ---, Cummings (Ref. 12); – –, Kirby and Cummings (Ref. 13).

boundary layer as expressed through the friction velocity would be expected to vary greatly and hence to be of major significance on the measured impedance. It is by no means as obvious that the friction velocity will play such an important role if the boundary layer is always turbulent and fully developed. The experiments of Cummings were conducted over a Reynolds number range of $3.8\times10^4<\mathrm{Re}<1.8\times10^5$, for which[12]

$$\frac{u_*}{fd}=\frac{\pi}{S}\sqrt{\frac{0.306\,\mathrm{Re}^{-0.25}}{8}}. \tag{16}$$

Thus, it is not possible to relate u_*/fd and u_*/ft directly to the Strouhal number, but since the Reynolds number range is quite small, a given Strouhal number corresponds to a fairly narrow range of possible u_*/fd or u_*/ft values. Cummings gave empirical formulas for $\theta c/fd$ as a function of u_*/fd and t/d, and δ/δ_0 as a function of u_*/ft and t/d, obtained from curve fitting to his experimental results. Kirby and Cummings[13] later gave different empirical formulas following a further set of experiments on orifices with reduced thickness. No explanation had been found for the differences between the two sets of experimental results, which are quite marked.

Figure 11 shows PDM* and PVM results in comparison with both aforementioned sets of empirical formulas in the form of θ/kR against an *inverse* Strouhal number, for an orifice with $h=0.5$. This is the same format in which the experimental results were plotted, from which the empirical formulas were obtained. The range of data given in Fig. 11 is within that of the experimental results, which were strongly clustered around a low inverse Strouhal number. The experimental results were once again for the flow-dependent resistance θ_f only, but this time the theoretical no-flow resistance values were subtracted from the measured data with flow. Once again, the theoretical results are seen to have the correct general trend, but do not give an accurate prediction. Figure 12 shows the new experimental results in comparison to the earlier empirical formulas. The new results lie perfectly within the band of the Cummings formula at high Strouhal number and the band of the Kirby and Cummings formula at low Strouhal number, which suggests that possibly the Kirby and Cummings formula has put a greater emphasis on results for low Strouhal number.

In Figs. 11 and 12, the empirical curves are truncated at zero resistance, as only positive resistance results were retained in the experimental results from which they were derived. Negative resistance values were measured,[12] but assumed to be caused by experimental error. The trend is clearly for the results to become negative at a high Strouhal number, both as measured here and as predicted by the theory. The difference between the two sets of empirical formulas is seen to be greater than the maximum difference in either set, which could be caused by a change from friction velocity to mean flow velocity. Since the new experimental results also collapse very well using a Strouhal number based on mean flow velocity, it would appear that detailed knowledge of the boundary layer structure is not required provided that the boundary layer is always turbulent and fully developed.

A presentation of the results in terms of the inverse Strouhal number compresses the high Strouhal number results into a very small region. Furthermore, the ordinate θ/kR contains a division by the Strouhal number, causing the high Strouhal number results to become of insignificant am-

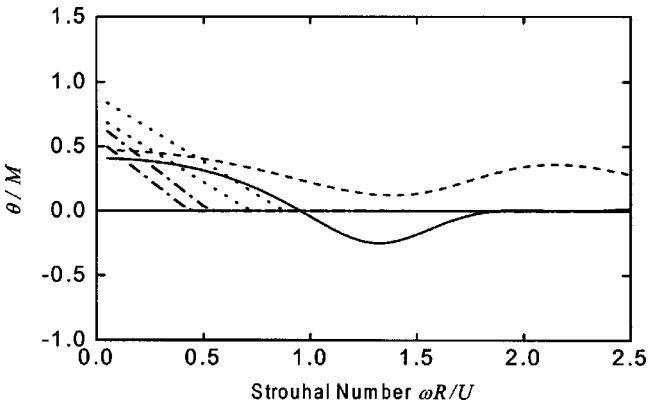


FIG. 13. Theoretical resistance compared to empirical formulas as a function of Strouhal number. —, PDM*, $h=0$; – – –, PVM, $h=0$; ---, Cummings (Ref. 12); – –, Kirby and Cummings (Ref. 13).

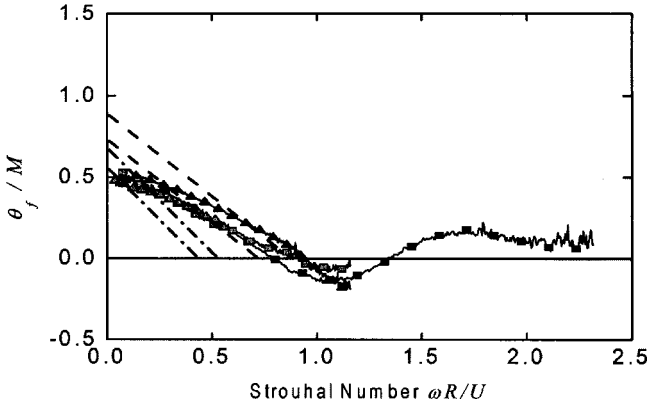


FIG. 14. Experimental resistance compared to empirical formulas as a function of the Strouhal number. ($d=8$ mm, $t=2$ mm) -■-, $M=0.085$; -●-, $M=0.170$; ($d=4$ mm, $t=1$ mm) -■-, $M=0.085$; -●-, $M=0.170$; ---, Cummings (Ref. 12); –––, Kirby and Cummings (Ref. 13).

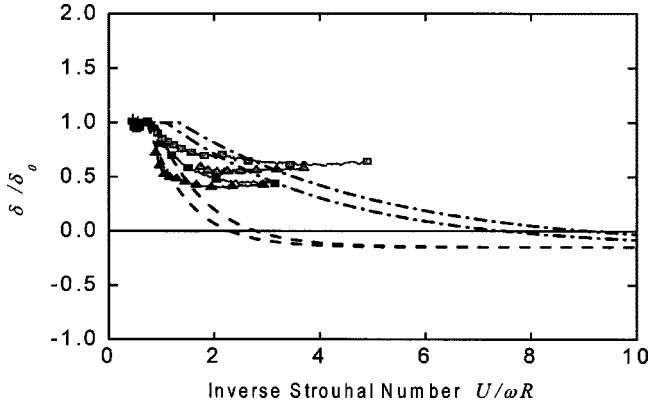


FIG. 16. Experimental end correction compared to empirical formulas as a function of the inverse Strouhal number. ($d=8$ mm, $t=2$ mm) -■-, $M=0.085$; -▲-, $M=0.170$; ($d=4$ mm, $t=1$ mm) -■-, $M=0.085$; -▲-, $M=0.170$; ---, Cummings (Ref. 12); –––, Kirby and Cummings (Ref. 13).

plitude. A more accurate impression of the effect of Strouhal number upon the resistance is therefore given by plots of θ/M against the Strouhal number, as used earlier. Figures 13 and 14 give repetitions of Figs. 11 and 12 in this format. In both figures, the comparison between the empirical formulas and current results does not look as good as in the previous format, but this is only to be expected. It would be preferential to have empirical formulas obtained from curve fitting to experimental results plotted in this format. However it is seen from Figs. 5 and 13 that, once again, for PDM* results the $h=0$ curve gives a better prediction than would the $h=0.5$ curve, and that the change of Strouhal number is again beneficial. For PVM results, the $h=0.5$ curve would actually give a better than fit than the $h=0$ curve shown here.

Reactance data is given by Cummings[12] and Kirby and Cummings[13] in the form of the proportional change in the end correction due to flow, δ/δ_0, where $\chi=k(t+\delta)$. Once again they plotted their experimental results against the inverse of the Strouhal number and then obtained empirical formulas by curve fitting to this data. Consequently the theoretical and experimental results are initially compared to the empirical formulas in this format, in Figs. 15 and 16, respectively. It is seen that the two sets of empirical formulas give very different results, but they do at least have the same general trend of a decreasing end correction as the Strouhal number is decreased. It is also worth noting that there was large scatter in the experimental results in both cases, so to some extent this similarity in trend might be due to the fact that the nature of the underlying curve fit was the same in both cases. In any event, the theoretical results and the new experimental results are not only different in magnitude, but also different in basic form. The theoretical result, given by $\delta/\delta_0=(\chi/kR-t/R)/(\delta_0/R)$, bears no resemblance to any set of experimental data. The new experimental results show the same trend as the empirical formulas at high Strouhal numbers, but are very different at low Strouhal numbers. The same end-correction data is shown again in Figs. 17 and 18, but this time plotted against the Strouhal number, to reduce the bias toward the low Strouhal number results. It is then clear that there is an agreement in trend between all sets of results at high Strouhal numbers, and even in value between all the experimental and PVM results. The PDM* result by contrast is very poor.

V. CONCLUSIONS

The theory of Howe *et al.*[19] gives a reasonable estimate of the effect of flow upon the resistance of a circular orifice.

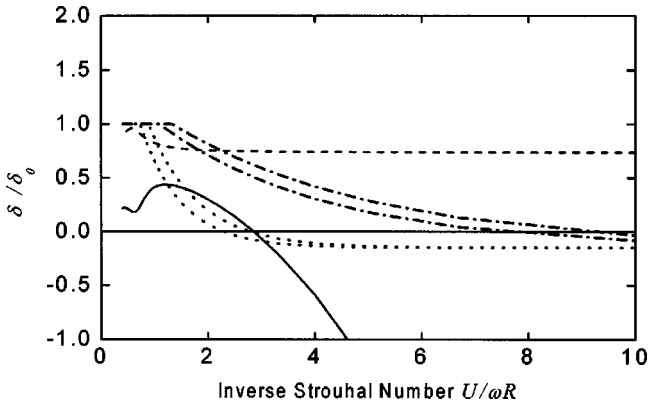


FIG. 15. Theoretical end correction compared to empirical formulas as a function of the inverse Strouhal number. —, PDM*, $h=0$; –––, PVM, $h=0$; ---, Cummings (Ref. 12); –––, Kirby and Cummings (Ref. 13).

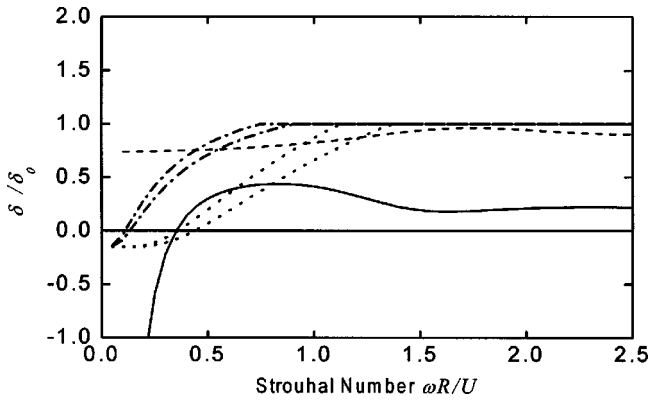


FIG. 17. Theoretical end correction compared to empirical formulas as a function of the Strouhal number. —, PDM*, $h=0$; –––, PVM, $h=0$; ---, Cummings (Ref. 12); –––, Kirby and Cummings (Ref. 13).

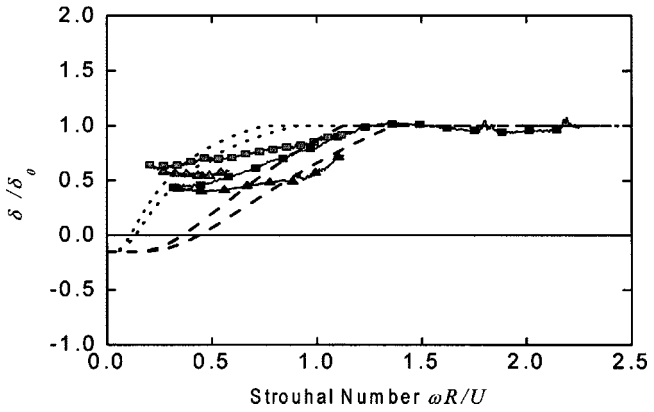


FIG. 18. Experimental end correction compared to empirical formulas as a function of Strouhal number. ($d=8$ mm, $t=2$ mm) -■-, $M=0.085$; -▲-, $M=0.170$; ($d=4$ mm, $t=1$ mm) -■-, $M=0.085$; -▲-, $M=0.170$; ---, Cummings (Ref. 12); –·–, Kirby and Cummings (Ref. 13).

The quality of prediction is improved greatly if the variable U in the theory, introduced as the velocity of the mean flow, is interpreted as the velocity of the convection of vorticity. In contrast, the results for the reactance or end correction given by the theory are poor in terms of both the general trend as well as value, although once again the change of the interpretation of U is definitely beneficial.

A modification to the theory by Jing *et al.*[23] results in far better predictions of reactance, without any recourse to the change in interpretation of U noted above. The trend is predicted perfectly and even the values are reasonable. In contrast, the benefit for the predictions of resistance is less clearly defined. Generally the resistance values predicted are at least as good as from Howe's theory, especially for the finite thickness of aperture, but the trend is worse in so much as the results from Jing's theory fail to predict negative resistance over a limited range of Strouhal number.

The theory predicts that both resistance and reactance should scale linearly with respect to Mach number. This is generally found to be the case in the experimental results, provided that the Reynolds number is high enough, although there is an intermediate band of Strouhal number for which the reactance varies slightly with Mach number. Thereafter, the theory predicts that the resistance and reactance vary only with h and the Strouhal number. For resistance this is found to be the case. The form of the variation of resistance with the Strouhal number is correctly predicted, and the absolute values are reasonable. The degree of the variation of resistance with h is overestimated by the theory. The experimental results for reactance show some variation for different sizes of orifice, for constant h and S. Thus there appears to be some separate dependence upon the orifice radius and thickness. The cause is unknown, but could account for much of the variation between earlier empirical formulas for reactance, which were based on sets of experiments with a very dissimilar orifice thickness. It implies some significant physical mechanism present that is not accounted for within the theory.

The analysis is exact only for an orifice of zero thickness. The extension to finite thickness is an approximation that is valid only for an orifice of small thickness to radius ratio, h. The values of h used for a comparison with experimental results violate this condition. Experimental results indicate very little change of resistance, but a very pronounced change of reactance, with orifice thickness. Indeed, for resistance, the PDM* results for $h=0$ give a noticeably better prediction than those for $h=0.5$, compared to experimental results from an orifice with $h=0.5$. In contrast, the PVM results for resistance and reactance with $h=0.5$ are at least as good as those as for $h=0$, when compared against experimental values for $h=0.5$. In both cases, it would certainly be a valuable exercise to extend the theory to analyze an orifice of any thickness to the radius ratio without recourse to the assumption that the displacement or velocity are invariant through the thickness of the orifice.

It is particularly noteworthy that the PDM* theory predicts negative resistance values over some range of high Strouhal numbers, in accordance with the measured data. Previously it had been thought that such measurements were erroneous, but, in fact, they simply represent cases where the oscillation of the shear layer is generating sound. Silencers that include perforated tubes are generally found to whistle at certain flow speeds due to this effect.

Earlier empirical formulas were based on the use of friction velocity rather than mean flow velocity. Previous experimental results for which the boundary layers were young had been successfully collapsed using the friction velocity, indicating that the boundary layer structure was of importance. However, if the boundary layer is turbulent and fully developed, it is by no means obvious that the friction velocity will play any significant role. The results presented here suggest that a Strouhal number based on mean flow velocity is indeed adequate in this case.

ACKNOWLEDGMENTS

Dr. Peat is grateful to the Korea Science and Engineering Foundation (KOSEF) for providing a Brainpool Fellowship for financial support. The work was completed while he was a Visiting Professor at KAIST, Korea. This work was partially supported by BK21 and NRL.

[1] G. D. Garrison, "Suppression of combustion oscillations with mechanical damping devices," Pratt and Whitney Aircraft Rep. PWA FR3299, 1969.
[2] D. Ronneberger, "The acoustic impedance of holes in the wall of flow ducts," J. Sound Vib. **24**, 133–150 (1972).
[3] P. D. Dean, "An *in situ* method of wall acoustic impedance measurements in flow duct," J. Sound Vib. **34**, 97–130 (1974).
[4] A. B. Bauer, "Impedance theory and measurements on porous acoustic liners," J. Aircr. **14**, 720–728 (1977).
[5] A. S. Hersh, B. Walker, and M. Bucka, "Effect of the grazing flow on the acoustic impedance of Helmholtz resonators consisting of single and clustered orifices," AIAA Paper 78-1124, 1978.
[6] J. W. Sullivan, "A method for modeling perforated tube muffler components. II. Application," J. Acoust. Soc. Am. **66**, 779–788 (1979).
[7] K. Jayaraman and K. Yam, "Decoupling approach to modeling perforated tube muffler components," J. Acoust. Soc. Am. **69**, 390–396 (1981).
[8] A. Goldman and R. L. Panton, "Measurement of the acoustic impedance of an orifice under a turbulent boundary layer," J. Acoust. Soc. Am. **60**, 1397–1404 (1976).
[9] A. Goldman and C. H. Chung, "Impedance of an orifice under a turbulent boundary layer with pressure gradient," J. Acoust. Soc. Am. **71**, 573–579 (1982).
[10] K. N. Rao and M. L. Munjal, "Experimental evaluation of impedance of perforate with grazing flow," J. Sound Vib. **108**, 283–295 (1986).

[11] J. W. Kooi and S. L. Sarin, ''An experimental study of the acoustic impedance of Helmholtz resonator arrays under a turbulent boundary layer,'' AIAA Paper 81-1998, 1981.
[12] A. Cummings, ''The effects of grazing turbulent pipe-flow on the impedance of an orifice,'' Acustica **61**, 233–242 (1986).
[13] R. Kirby and A. Cummings, ''The impedance of perforated plates subjected to grazing gas flow and backed by porous media,'' J. Sound Vib. **217**, 619–636 (1998); **217**, 619–636 (1998).
[14] M. S. Howe, ''The influence of grazing flow on the acoustic impedance of a cylindrical wall cavity,'' J. Sound Vib. **67**, 533–544 (1979).
[15] B. E. Walker and A. F. Charwat, ''Correlation of the effects of grazing flow on the impedance of Helmholtz resonators,'' J. Acoust. Soc. Am. **72**, 550–555 (1982).
[16] D. Ronneberger, ''The dynamics of shearing flow over a cavity—a visual study related to the acoustic impedance of small orifices,'' J. Sound Vib. **71**, 565–581 (1980).
[17] S. Kaji, M. Hiramoto, and T. Okazaki, ''Acoustic characterization of holes exposed to grazing flow,'' Bull. JSME **27**, 2388–2396 (1984).
[18] E. J. Rice, ''A theoretical study of the acoustic impedance of orifices in the presence of a steady grazing flow,'' NASA TM X-71903, 1976.
[19] M. S. Howe, M. I. Scott, and S. R. Sipcic, ''The influence of tangential mean flow on the Rayleigh conductivity of an aperture,'' Proc. R. Soc. London, Ser. A **452**, 2302–2317 (1996).
[20] M. S. Howe, *Acoustics of Fluid–Structure Interactions* (Cambridge University Press, Cambridge, 1998).
[21] M. S. Howe, ''Low Strouhal number instabilities of flow over apertures and wall cavities,'' J. Acoust. Soc. Am. **102**, 772–780 (1997).
[22] S.-H. Lee and J.-G. Ih, ''Empirical model of the acoustic impedance of a circular orifice in grazing mean flow,'' J. Acoust. Soc. Am. **114**, 98–113.
[23] X. Jing, X. Sun, J. Wu, and K. Meng, ''Effect of grazing flow on the acoustic impedance of an orifice,'' AIAA J. **39**, 1478–1484 (2001).
[24] P. A. Nelson, N. A. Halliwell, and P. E. Doak, ''Fluid dynamics of a flowexcited resonance, Part I: Experiment,'' J. Sound Vib. **78**, 15–38 (1981).
[25] P. A. Nelson, N. A. Halliwell, and P. E. Doak, ''Fluid dynamics of a flow excited resonance, Part II: Flow acoustic interaction,'' J. Sound Vib. **91**, 375–402 (1983).
[26] A. W. Guess, ''Calculation of perforated plate liner parameters from specified acoustic resistance and reactance,'' J. Sound Vib. **40**, 119–137 (1975).
[27] T. D. Mast and A. D. Pierce, ''Describing function theory for flow excitation resonators,'' J. Acoust. Soc. Am. **97**, 163–172 (1995).
[28] S. D. Savkar, ''Radiation of cylindrical duct acoustic modes with flow mismatch,'' J. Sound Vib. **42**, 363–386 (1975).

Measurement of elastic nonlinearity of soft solid with transient elastography

S. Catheline,[a] J.-L. Gennisson, and M. Fink
Laboratoire Ondes et Acoustique, ESPCI, Université Paris VII, U.M.R. C.N.R.S. 7587, 10 rue Vauquelin, 75231 Paris cedex 05, France

(Received 23 September 2002; revised 11 July 2003; accepted 21 July 2003)

Transient elastography is a powerful tool to measure the speed of low-frequency shear waves in soft tissues and thus to determine the second-order elastic modulus μ (or the Young's modulus E). In this paper, it is shown how transient elastography can also achieve the measurement of the nonlinear third-order elastic moduli of an Agar-gelatin-based phantom. This method requires speed measurements of polarized elastic waves measured in a statically stressed isotropic medium. A static uniaxial stress induces a hexagonal anisotropy (transverse isotropy) in solids. In the special case of uniaxially stressed isotropic media, the anisotropy is not caused by linear elastic coefficients but by the third-order nonlinear elastic constants, and the medium recovers its isotropic properties as soon as the uniaxial stress disappears. It has already been shown how transient elastography can measure the elastic (second-order) moduli in a media with transverse isotropy such as muscles. Consequently this method, based on the measurement of the speed variations of a low-frequency (50-Hz) polarized shear strain waves as a function of the applied stress, allows one to measure the Landau moduli *A, B, C* that completely describe the third-order nonlinearity. The several orders of magnitude found among these three constants can be justified from the theoretical expression of the internal energy.
 [DOI: 10.1121/1.1610457]

PACS numbers: 43.25.Dc, 43.25.Ed [MFH] Pages: 3087–3091

I. INTRODUCTION

Acoustoelasticity is a well-established technique[1] to experimentally measure nonlinear third-order elastic constants in solids such as metals,[2] crystals,[3] or rocks.[4] It consists of measuring the speed of ultrasonic waves in stressed solids. More precisely, the third-order moduli are deduced from the slope of the speed as a function of the uniaxial stress. So far no such measurements have been made in soft tissue since it has long been considered a liquid-like medium from an ultrasonic point of view. However, like in all solids, shear waves do propagate in soft tissues at low frequency (50 Hz typically).[5,6] As shown in this paper, the speed of these shear waves is modified if the medium is submitted to a uniaxial stress (which is the evidence of a deviation from the Hooke's law). In such a medium, the uniaxial stress induces anisotropy, the transverse isotropy,[7] which remains as long as the uniaxial stress is applied. Thus, a quantitative evaluation of this temporary transverse isotropy with the technique of transient elastography leads to the measurement of the third-order constants. In the literature, the quantitative measurement of soft solids nonlinearity lies on the measurement of ultrasound speed[8] which cannot lead to the measurement of the Landau moduli *A, B, C* (as shown in Sec. II) without the complementary measurement of shear wave speeds shown in the present paper. From a more qualitative point of view, it is shown that nonlinear effects can alter the contrast in the elastography images of breast tissues[9] and are of fundamental importance in the touch–contact interaction with soft tissues.[10]

In Sec. II, a brief theoretical recall introduces the set of three equations that gives the speed of the two polarized shear waves (with a polarization parallel or perpendicular to the stress axis) and of the compressional wave (ultrasonic wave) as a function of the applied stress. Section III describes the experiment and explains how polarized shear strain waves can be generated by a rod source. Then, with the help of theoretical Green's functions, a simulation shows the ability of the transient elastography technique to measure the small speed variations of polarized shear waves induced by nonlinear phenomena. Experimental values of the third-order Landau moduli *A, B, C* are discussed in Sec. III.

II. THEORY

Hugues and Kelly[2] have established expressions of the speed of elastic waves in a uniaxially stressed solid as function of the second-order Lamé coefficients (λ, μ) and the third-order Landau coefficients (A, B, C). The detailed development of the elastic nonlinear theory can be found in numerous textbooks.[11,12] Here we summarize the general equations. The equation of motion is

$$\rho \ddot{u}_i = \frac{\partial \sigma_{ij}}{\partial x_j}, \tag{1}$$

where ρ, σ, and $\ddot{u}$ designate the density, the stress tensor, and the particle acceleration, respectively. The stress tensor is given by

$$\sigma_{ij} = \frac{\partial e}{\partial \left(\frac{\partial u_i}{\partial x_j} \right)}, \tag{2}$$

[a]Electronic mail: stefan.catheline@espci.fr

where e designates the elastic strain energy which, developed to the third order in strain ε in a general elastic medium, is

$$e=\tfrac{1}{2}C_{ijkl}\varepsilon_{ij}\varepsilon_{kl}+\tfrac{1}{6}C_{ijklmn}\varepsilon_{ij}\varepsilon_{kl}\varepsilon_{mn}, \tag{3}$$

where c_{ijkl} and c_{ijklmn} designate the component of the second-order elastic tensor and the third-order elastic tensor, respectively.

In Eq. (3), the components of the Lagrangian strain tensor ε are given by

$$2\varepsilon_{ij}=\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}+\frac{\partial u_k}{\partial x_i}\frac{\partial u_k}{\partial x_j}. \tag{4}$$

Thus, the general nonlinear elastodynamic equation is

$$\rho\ddot{u}_i=\frac{\partial^2}{\partial x_j\partial x_l}\Bigg[c_{ijkl}+(c_{ijklmn}+c_{ijln}\delta_{km}+c_{jlmn}\delta_{ik}+c_{jnkl}\delta_{im})\frac{\partial u_m}{\partial x_n}\Bigg]u_k. \tag{5}$$

The speeds V are given by the solutions of

$$\mathrm{Det}\left|\sum_{jl}\mathbf{B}_{ijkl}N_jN_l-V^2\delta_{jk}\right|=0, \tag{6}$$

where δ is the Kronecker symbol, and N_j, N_l are the components of a unit vector that represents a plane-wave direction of propagation. It should be noted that $\mathbf{B}_{rstu}$ designates the elastic tensor as defined by Hugues and Kelly [deduced from Eq. (5)] in isotropic solid and thus contains both second- and third-order elastic constants and the stress along one axis σ. The stress σ comes when replacing the strain in brackets in (5) by means of the formulas of the linear theory (Hooke's law). Now, in order to use the Landau coefficients, we must recall how they were introduced. A third-order development of the elastic internal energy in an isotropic solid, Eq. (3), is also expressed as

$$e=\mu\varepsilon_{ik}^2+\frac{\lambda}{2}\varepsilon_{ll}^2+\frac{A}{3}\varepsilon_{ik}\varepsilon_{il}\varepsilon_{kl}+B\varepsilon_{ik}^2\varepsilon_{ll}+\frac{C}{3}\varepsilon_{ll}^3. \tag{7}$$

The Lamé and the Landau coefficients as a function of the elastic moduli in Voigt's notation are, respectively, $\lambda=c_{12}$, $\mu=c_{66}$, $A=4c_{456}$, $B=c_{144}$, $C=c_{123}/2$. Thus, the three plane-wave solutions that propagate perpendicularly to the static uniaxial stress are associated with the three following eigenvalues:

$$\rho(V_p)^2=\lambda+2\mu-\frac{\sigma}{3\lambda+2\mu}\Bigg[-\frac{\lambda}{\mu}A+2B\left(1-\frac{\lambda}{\mu}\right)+2C-4\lambda-2\frac{\lambda^2}{\mu}\Bigg], \tag{8}$$

$$\rho(V_s^{\parallel})^2=\mu-\frac{\sigma}{3\lambda+2\mu}\left[\frac{A}{2}\left(1+\frac{\lambda}{2\mu}\right)+B+\lambda+2\mu\right], \tag{9}$$

$$\rho(V_s^{\perp})^2=\mu-\frac{\sigma}{3\lambda+2\mu}\left(-\frac{A\lambda}{2\mu}+B-2\lambda\right). \tag{10}$$

In Eqs. (8), (9), and (10), V_P stands for the speed of the compressional wave, $V_s^{\parallel}$ for the speed of the shear wave with a polarization parallel to the stress axis, and $V_s^{\perp}$ for the speed of the shear wave with a polarization perpendicular to the stress axis. Let's compare Eqs. (8), (9), and (10) to the set of equations obtained from a transverse isotropic medium for a propagation direction perpendicular to the plan of isotropy

$$\rho(V_p)^2=c_{11}, \tag{11}$$

$$\rho(V_s^{\parallel})^2=c_{44}, \tag{12}$$

$$\rho(V_s^{\perp})^2=c_{66}. \tag{13}$$

c_{11}, c_{44}, and c_{66} stand for the elastic constants of the Christoffel's tensor using Voigt's notation. As in the acoustoelasticity experiment, three plane-wave solutions with the same polarization are found. This is the expression of the formal equivalence between transverse isotropic media and stressed media. One difference is that, in the latter configuration, the anisotropy is stress dependent; in an unstressed medium ($\sigma=0$), one can easily verify from Eqs. (8), (9), and (10) that the speeds correspond to an isotropic solid. But, at a given stress, one can define a unique set of elastic constants c_{11}, c_{44}, and c_{66} in Eqs. (11), (12), and (13) that perfectly accounts for the temporary nonlinear anisotropy. Moreover, changing the stress in Eqs. (8), (9), and (10) is equivalent to modifying c_{11}, c_{44}, and c_{66} according to a linear law with a slope defined by the second- and third-order elastic constants. In the following part, we are taking advantage of this equivalence to predict the behavior of the transient elastography technique used in an acoustoelasticity experiment.

III. EXPERIMENT AND SIMULATION

The experiments presented in the following section are conducted in a model of soft tissues: an Agar-gelatin based phantom (5% gelatin, 3% Agar). Its dimensions are $80\times80\times80$ mm. The mechanical characteristics are highly dependent on temperature and maturation. A 5-MHz transducer is mounted in the middle of a rod (80 mm long) fixed on a vibrator (Brüel & Kjaer, type 4810). The whole system (the shear elasticity probe) is applied at the surface of the phantom so that the ultrasonic beam is horizontal (see Fig. 1). A rigid Plexiglas plate is placed on the top, and loads (a container filled with water) can be added to control the uniaxial stress in the sample. The transducer works as a pulse–echo system and backscattered signals sampled at 50 MHz are stored in a 2-MB memory with a repetition frequency of typically 3000 Hz. The low-frequency pulse (50-Hz central frequency) propagates in the medium mainly as a shear wave, and its longitudinal displacements along the ultrasonic beam are measured with a cross-correlation algorithm between successive A lines.[13] This technique is known as transient elastography.[14]

On the seismic-like representation of the displacements (Fig. 2) obtained in an unstressed medium, the maximum amplitude of the 50-Hz pulse is 120 μm. A small-amplitude compressional wave (P) first propagates almost instantaneously within the medium (1500 m s^{-1}). The shear wave (S) longitudinal component appears at each depth with a phase delay inversely proportional to its speed. Actually, the speed (2.52 ± 0.02 m s^{-1}) is extracted with a simple phase analysis at the central frequency. The standard deviation of the phase

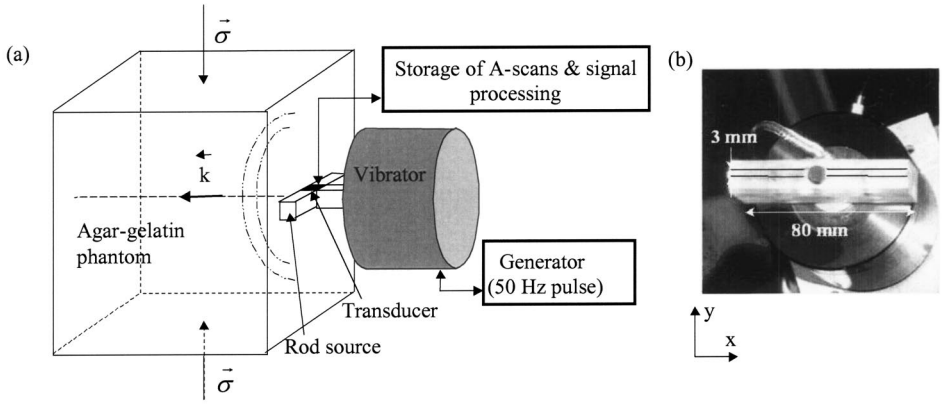


FIG. 1. (a) Experimental setup. A transducer is set in the middle of a rod mounted on a vibrator. A low-frequency pulse propagates in the medium and the displacements are computed from the A-scan stored in a memory. (b) Picture of the shear elasticity probe. The transducer (the black disk) is in the middle of a Plexiglas rod and the black vibrator is visible in the back ground.

from a linear fit gives the error. We have shown in a recent paper[15] that a rod source allows one to generate shear waves with a strain field polarized perpendicularly to the rod in the first few centimeters. Since the speed of shear waves in a uniaxially stressed medium depends on its polarization in regard to the stress, the speed of the shear wave is measured for each amount of stress with the rod in the horizontal and vertical position.

Contrary to the theoretical analysis, we are not dealing with plane waves in the experiments, so how close to plane polarized shear waves are the speed measurements obtained with the rod system? Is it accurate enough to detect small variations due to nonlinear effects? The theoretical Green's functions in transverse isotropic elastic media computed by Vavryčuk[16] from a higher-order ray theory are used to simulate the displacement field induced by the rod system in an acoustoelasticity experiment. The transverse isotropy is controlled by the following two parameters: the parallel and perpendicular elastic moduli, c_{44} and c_{66}, respectively (the compression modulus is kept constant, $c_{11}=2.25$ GPa). The simulation (introduced in Ref. 13) is run as many time as the number of desired stressed states of the medium. The parameters are adjusted so as to have a comparable value with the experiments (a typical density for soft tissues is 1000 kg m^{-3}): the parallel and perpendicular elastic moduli range from 6250 to 8000 Pa and from 6250 to 6500 Pa, respectively. The impulse response of the rod is computed from the sum of the Green's function of point sources uniformly distributed on the contact surface. A 1-mm square grid is sufficient for a 5-cm typical wavelength. This summation is the expression of the Rayleigh–Sommerfeld integral. Finally, a simple convolution of this impulse response with the source excitation (a one-cycle sinusoid with a 50-Hz central frequency) gives the theoretical displacement field sampled at 3000 Hz. As in the experiment, a phase analysis at 50 Hz from the longitudinal displacements in the axis of the rod (3×80 mm^2) is used to compute the speed in Fig. 3. The error bars are deduced from the standard deviation of this phase measurement from a linear fit. It clearly appears that the speeds measured with the two perpendicular positions of the rod relative to the main axis of the transverse isotropy properly reveal the two shear moduli introduced in the simulation. In other words the rod source generates a shear wave which has the speed of a plane shear wave polarized perpendicularly to its bigger dimension. Consequently, any changes on one of the shear moduli modify the speed of either polar-

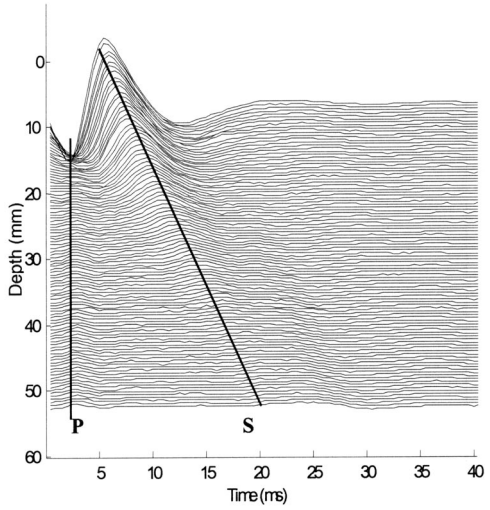


FIG. 2. Typical experimental displacements in an Agar-gelatin-based phantom. The maximum amplitude of the 50-Hz pulse is 120 μm. A phase analysis gives the speed of the shear wave: 2.52 ± 0.02 m s^{-1}.

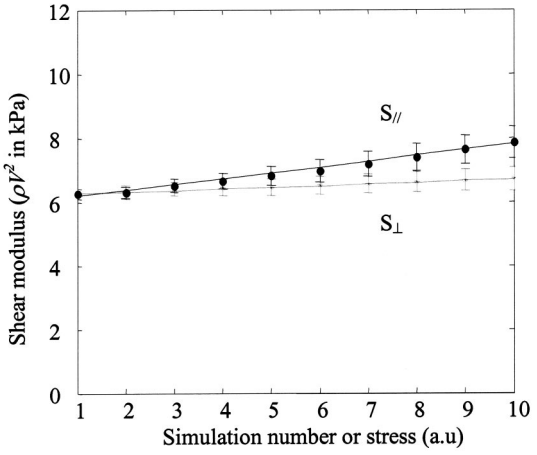


FIG. 3. In order to simulate an uniaxial stress ranging from 0 to several hundred Pa, we modify the shear elasticities in a transverse isotropic medium. In this model, the increasing stress corresponds to a parallel and perpendicular elastic moduli ranging from 6250 to 8000 Pa and from 6250 to 6500 Pa, respectively. The simulation is run twice for each stress state (two positions of the rod). The rod system shows its ability to recover the right values of elasticity used in the simulation.

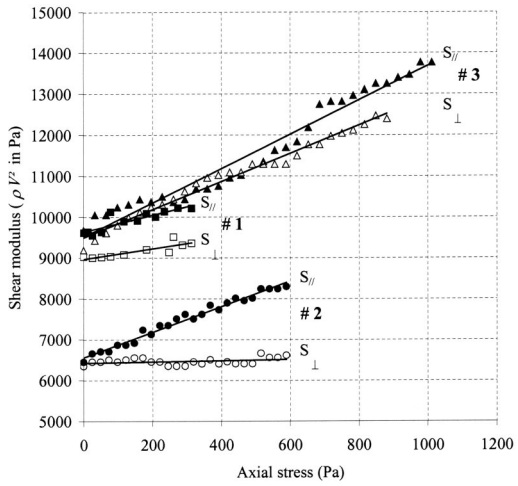


FIG. 4. Experimental shear and compression moduli as function of the uniaxial stress for three Agar-gelatin-based phantom. The nonlinear Landau coefficients A and B are computed from each pair of slopes.

ized strain shear wave. The expected slopes of the moduli (simulation parameters) that characterize their evolution are 8% and 33%. The measured slopes are, respectively, 7% and 20%. The difference is attributed to the fact that the measurements are not performed on purely polarized shear waves and the speed may be influenced by diffraction effects: the phase as a function of depth is not perfectly linear (error bars). However, as a conclusion to this paragraph, one can say that the rod system gives a rather accurate description of the anisotropy of a medium. Thus, it is a reasonable tool to study a nonlinear transverse isotropic medium.

IV. RESULTS AND DISCUSSION

Three phantoms are tested. As observed in Fig. 4, all pairs of the experiment, numbered from 1 to 3, show that both elastic moduli (perpendicular and parallel) increase with static axial stress. This is a clear evidence of a nonlinear behavior of the Agar-gelatin-based phantom. Compared to a typical acoustoelasticity experiment in metals where the shear perpendicular modulus decreases, this result constitutes a first particularity of the gelatin-based phantom. Now, from a quantitative point of view, the shear elasticity is extracted from the first point (null stress) of each pair of experiments: $\mu = 9.00$, 6.35, and 9.67 kPa. The experiment #1 corresponds to our first results. As the experimental setup has then been improved, the stress range increases with experiments #2 and #3. The anisotropy visible on the experimental data #1 at zero applied stress is thought to be a consequence of the phantom's own weight. It was found experimentally that the speed measurements at 4 cm beneath the top surface of the phantom were not subjected to the nonlinear effect of a stress induced by its weight. Indeed, in experiments #2 and #3, the same speed is found for the two polarized strain shear waves in the unstressed medium.

When the uniaxial stress is linearly increased within the phantoms #1, #2, #3, the perpendicular and parallel elastic moduli increase with slopes of 1.27, 0.13, 3.41, and 2.05, 3.12, 4.17, respectively. From each pairs of slopes and using the set of equations (9) and (10), we have found the following values: -64, -101, -68 kPa and -12, -14, -26 GPa for the Landau coefficients A and B, respectively. The huge difference between these third-order moduli is striking since in more conventional media such as metal, rocks, or crystals they are of the same order. Now, the last Landau coefficient can be deduced from the results found in the literature. Indeed, in Ref. 8, with a thermodynamic experimental setup (the pressure is changed while the temperature is kept constant), Everbach measured the nonlinear parameter $\beta = 3.64$ in a gelatin-based phantom from speed measurements of ultrasound. This value of the nonlinear parameter does not significantly change from one sample to another and is in fact close to the parameter found in the water, $\beta = 3.50$. Since β is expressed as function of the Landau coefficients[14] as $\beta = -(3/2) - [(A+3B+C)/\rho_o c_l^2] = 3.64$, and using the experimental values of A and B, one finally obtains $C = 24$, 31, and 67 GPa.

The experimental errors in these measurements are difficult to quantify. They may emerge from the diffraction biases cited earlier, although the error bar on each individual speed measurement (less than 2%) is too small to be represented in Fig. 4. An unperfected experimental uniaxial stress may cause wrong estimations of nonlinear moduli. Finally, the linear relationship between elastic modulus and stress may be a hazardous approximation if higher order nonlinear moduli are involved. Thus, one can take as an error estimate the standard deviation of the experimental measurements from a linear fit. The results are summarized in Table I. The gap between the first (A) and the two last Landau coefficient (B and C) is huge for the three samples. This result can be justified by the following considerations: for soft solids, in Eq. (7), the first coefficient μ in front of a shear strain is 10^6 smaller than the coefficient λ in front of the compression strain (one has to keep in mind the order of magnitude of the second-order moduli in the Agar-gelatin-based phantom; typically λ is in GPa, whereas μ is in kPa). Thus, it is not surprising for the third-order coefficient A in front of shear strain terms to present a very small value compared to the coefficients B and C in front of terms that contain compres-

TABLE I. Elastic moduli measured in three Agar-gelatin-based phantoms.

	Linear second-order elastic moduli (Lamé coefficients)		Nonlinear third order elastic moduli (Landau coefficients)		
Phantom #	λ (GPa)	μ (kPa)	A (kPa)	B (GPa)	C (GPa)
1	2.25	9.0 ± 0.2	-64 ± 13	-12 ± 3	24 ± 6
2	2.25	6.35 ± 0.04	-101 ± 7	-14 ± 2	31 ± 3
3	2.25	9.67 ± 0.06	-68 ± 3	-26 ± 2	67 ± 4

sion strain. This intuitive justification does not hold for a theoretical explanation; some efforts still have to be done to completely understand the results of Table I.

V. CONCLUSION

In this paper, we have used the transient elastography technique in an acoustoelasticity experiment in order to characterize the nonlinear behavior of a model of soft solid: the gelatin-based phantom. The results clearly show the particularity of soft solids. Indeed, the shear Lamé coefficient μ is roughly 10^6 times smaller than the compression Lamé coefficient λ. The contribution of this paper is to show that the Landau coefficient A is 10^6 times smaller than the two Landau coefficients B and C. These first quantitative results on nonlinearity of soft solids encourage us to pursue the experiments on *in vitro* and *in vivo* biological tissues. Our point of view is that the nonlinear behavior of living tissues may be as important as its linear behavior.

[1] F. D. Murnaghan, ''Finite deformation of an elastic solid,'' Am. J. Math. **49**, 235 (1937).

[2] D. S. Hugues and J. L. Kelly, ''Second-order elastic deformation of solids,'' Phys. Rev. **92**, 1145–1149 (1953).

[3] T. Bateman, W. P. Mason, and H. J. McSkimin, ''Third-order elastic moduli of germanium,'' J. Appl. Phys. **32**, 928–936 (1961).

[4] F. Birch, ''Compressibility; elastic constants,'' in *Handbook of Physical Constants*, edited by S. P. Clark, Jr., Mem. Geol. Soc. Am. **97**, 97–174 (1966).

[5] K. J. Parker and R. M. Lerner, ''Sonoelasticity of organs: Shear waves ring a bell,'' J. Ultrasound Med. **11**, 387–392 (1992).

[6] Y. Yamakoshi, J. Sato, and T. Sato, ''Ultrasonic imaging of internal vibration of soft tissue under forced vibration,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**, 45–53 (1990).

[7] P. A. Johnson and P. N. J. Rasolofosaon, ''Nonlinear elasticity and stress-induced anisotropy in rock,'' J. Geophys. Res. **101**, 3113–3124 (1996).

[8] E. C. Everbach and R. E. Apfel, ''An interferometric technique for B/A measurement,'' J. Acoust. Soc. Am. **98**, 3428–3436 (1995).

[9] T. Varghese, J. Ophir, and T. A. Krouskop, ''Nonlinear stress-strain relationships in tissue and their effect on the contrast-to-noise ratio in elastograms,'' Ultrasound Med. Biol. **26**, 839–851 (2000).

[10] D. T. Pawluk, W. J. Peine, P. S. Wellman, and R. D. Howe, ''Simulating Soft Tissue with a Tactile Shape Display,'' IMECE Haptics Symposium (1997).

[11] M. F. Hamilton and D. T. Blackstock, *Nonlinear Acoustics* (Academic, San Diego, 1998).

[12] W. P. Mason, *Physical Acoustics* (Academic, New York and London, 1964).

[13] J. Ophir, I. Céspedes, H. Ponnekanti, Y. Yasdi, and X. Li, ''Elastography: A quantitative method for imaging the elasticity of biological tissues,'' Ultrason. Imaging **13**, 111–134 (1991).

[14] L. Sandrin, M. Tanter, J.-L. Gennisson, S. Catheline, and M. Fink, ''Shear modulus imaging with 2D transient elastography,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 426–435 (2002).

[15] J. L. Gennisson, S. Catheline, S. Chaffaï, and M. Fink, ''Transient elastography in anisotropic medium: Application to the measurement of slow and fast shear waves velocities in muscles,'' J. Acoust. Soc. Am. **114**, 536–541 (2003).

[16] V. Vavryčuk, ''Exact elastodynamic Green functions for simple types of anisotropy derived from higher order ray theory,'' Stud. Geophys. Geod. **45**, 67–84 (2001).

Thermal effects on acoustic streaming in standing waves

Mark F. Hamilton, Yurii A. Ilinskii, and Evgenia A. Zabolotskaya
Department of Mechanical Engineering, The University of Texas at Austin, Austin, Texas 78712-1063
and Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 78713-8029

(Received 14 March 2003; revised 2 August 2003; accepted 25 August 2003)

Acoustic streaming generated by standing waves in channels of arbitrary width is investigated analytically. In a previous paper by the authors [J. Acoust. Soc. Am. **113**, 153–160 (2003)], a purely viscous fluid in a two-dimensional channel was considered. That analysis is extended here to a gas in which heat conduction and dependence of the viscosity on temperature are taken into account. Calculations are presented for typical working gases used in thermoacoustic engines at standard temperature and pressure. In channels that are very wide in comparison with the viscous penetration depth, which is the Rayleigh streaming regime, the influence of the two thermal effects is comparable but small. The same is true in very narrow channels, having widths on the order of the viscous penetration depth. In channels having intermediate widths, 10–20 times the viscous penetration depth, the effect of heat conduction can be substantial. The analysis is performed for cylindrical tubes as well as two-dimensional channels, and the results are found to be qualitatively the same. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1618752]

PACS numbers: 43.25.Nm [RR] Pages: 3092–3101

I. INTRODUCTION

Acoustic streaming generated by standing waves in channels is due primarily to shear forces within the viscous boundary layer along the walls. Solutions for streaming generated in this way have been derived for channels that are very wide in comparison with the viscous penetration depth, by Rayleigh[1] for two-dimensional channels and by Schuster and Matz[2] for cylindrical tubes. Thermal effects on streaming were first considered by Rott.[3] His result, also restricted to wide channels, includes the effects of heat conduction and dependence of the viscosity on temperature in a gas, as well as the effect of a mean temperature gradient imposed along the channel walls. Inclusion of a temperature gradient was motivated by interest in processes that occur in thermoacoustic engines. His result reveals that in the absence of an imposed temperature gradient, thermal effects alter the streaming velocity in wide channels by only a few percent.

Apparently unaware of the article by Rott, Qi[4] derived a solution for streaming in a wide two-dimensional channel containing a thermoviscous gas. Qi's expression for the streaming velocity in the center of the channel matches Rott's result when mean temperature gradient and viscosity variations are ignored in the latter. However, the physical interpretation proposed by Qi is misleading and deserves comment. Qi asserted that his result differs from those of Rayleigh and others because he took compressibility into account. This is incorrect, because Qi's solution reduces to Rayleigh's solution in the absence of heat conduction. Several methods of accounting for compressibility in classical analyses of streaming are discussed in Sec. III D of Ref. 5.

The purpose of the present paper is to obtain solutions for streaming in channels of arbitrary width, taking into account heat conduction and dependence of the viscosity on temperature in an ideal gas. Solutions are obtained for two-dimensional channels and cylindrical tubes. The solution for two-dimensional channels is in closed form, whereas the solution for cylindrical tubes contains integrals over Bessel functions. It is revealed that for channel widths 10–20 times the viscous penetration depth, thermal effects may alter the streaming velocity substantially. For significantly wider or narrower channels, thermal effects influence the streaming velocity by only a few percent.

Thermal effects were considered in two previous investigations of acoustic streaming in channels of arbitrary width. Waxler[6] included heat conduction in a study of streaming in a gas between parallel plates. Bailliet *et al.*[7] included temperature dependence of the viscosity in addition to heat conduction in analyses of streaming in both two-dimensional channels and cylindrical tubes.[8] In both cases, the mean temperature was assumed to vary along the channel walls to account for such variations within the stacks of thermoacoustic engines. We do not account for this temperature variation, and because of this the solution for streaming in a two-dimensional channel can be obtained analytically.

The present analysis follows closely an earlier paper on acoustic streaming in a two-dimensional channel containing a purely viscous fluid.[5] In Sec. II of the present paper, on the sound field and streaming in a two-dimensional channel, the reader is referred frequently to the earlier paper for calculations that are common to both analyses. The derivations in Sec. III, for the sound field and streaming in a cylindrical tube, are straightforward modifications of the calculations described in Sec. II. The derivations in Sec. III are therefore kept relatively short in the interest of brevity. Figures illustrating thermal effects on streaming in cylindrical tubes were presented previously by the authors in Ref. 9.

II. TWO-DIMENSIONAL CHANNELS

The resonator geometry shown in Fig. 1, formed by rigid walls at $x=\pm x_0$ and $y=\pm y_0$, is the same as in Ref. 5. The gas inside is assumed to be viscous and heat conducting. A standing wave is produced in the gas by vibration of the

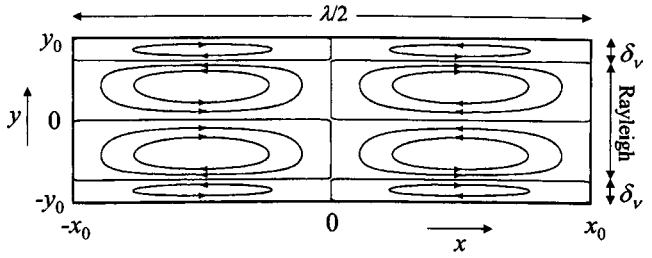


FIG. 1. Acoustic streaming patterns in a resonator with rigid ends, excited in its lowest mode. Coordinates for a two-dimensional channel are shown. For the cylindrical tube considered in Sec. III, replace x by z and $\pm y_0$ by r_0.

resonator along the x axis with velocity $v(t)$. An Eulerian coordinate system attached to the resonator is used, in terms of which the fluid motion is assumed to vanish at the walls. The wall temperature is assumed constant and equal to the equilibrium temperature of the gas in the absence of sound. Considerations of symmetry dictate that the pressure, density, temperature, and x component of the particle velocity are even functions of y, and that the y component of the particle velocity is an odd function of y. As before,[5] we use the approximate forms of the equations of motion for channels in which the viscous penetration depth is a small fraction of the acoustic wavelength.[10] Within this framework the pressure is independent of y.

The analysis in the present section is a direct extension of that in Ref. 5, here taking heat conduction into account. Included is the dependence of viscosity on temperature. Not included is the dependence of thermal conductivity on temperature, because that effect is of higher order. Also not included is variation in mean temperature of the gas, again because it is higher order.

A. Linear theory

In the linear approximation the continuity, momentum, and entropy equations are

$$\frac{1}{\rho_0}\frac{\partial \rho'}{\partial t}+\frac{\partial u_x}{\partial x}+\frac{\partial u_y}{\partial y}=0, \tag{1}$$

$$\rho_0\frac{\partial u_x}{\partial t}+\frac{\partial p}{\partial x}=\mu_0\frac{\partial^2 u_x}{\partial y^2}-\rho_0\dot{v}(t), \tag{2}$$

$$\frac{\partial p}{\partial y}=0, \tag{3}$$

$$\rho_0 T_0\frac{\partial s'}{\partial t}=\kappa_0\frac{\partial^2 T'}{\partial y^2}, \tag{4}$$

where p is the sound pressure, ρ_0 and ρ' are the equilibrium and excess densities, respectively, T_0 and T' are the equilibrium and excess temperatures, u_x and u_y are the components of the particle velocity vector, s' is the excess entropy per unit mass, μ_0 is the equilibrium shear viscosity, κ_0 is the equilibrium thermal conductivity, and $-\rho_0\dot{v}(t)$ is an effective body force per unit volume due to vibration of the resonator. The state equations for the gas are

$$P=\rho RT, \tag{5}$$

$$T=T(P,s), \tag{6}$$

where P, T, and s are the total pressure, temperature, and entropy per unit mass, respectively, and R is the gas constant. Equations (1)–(6) may be combined to solve for the six unknowns u_x, u_y, p, ρ', T', and s'.

The procedure begins by eliminating density from Eq. (1). Linearizing Eq. (5) yields

$$\frac{\rho'}{\rho_0}=\frac{p}{P_0}-\frac{T'}{T_0}, \tag{7}$$

where P_0 is the equilibrium pressure, and Eq. (1) becomes

$$\frac{1}{P_0}\frac{\partial p}{\partial t}-\frac{1}{T_0}\frac{\partial T'}{\partial t}+\frac{\partial u_x}{\partial x}+\frac{\partial u_y}{\partial y}=0. \tag{8}$$

To eliminate entropy from Eq. (4) we use the following differential form of Eq. (6):

$$\frac{\partial T'}{\partial t}=\left(\frac{\partial T}{\partial P}\right)_s\frac{\partial p}{\partial t}+\left(\frac{\partial T}{\partial s}\right)_P\frac{\partial s'}{\partial t} \tag{9}$$

$$=\frac{(\gamma-1)T_0}{\gamma P_0}\frac{\partial p}{\partial t}+\frac{T_0}{c_p}\frac{\partial s'}{\partial t}, \tag{10}$$

where the derivatives are evaluated at equilibrium, c_p is the specific heat at constant pressure, and γ is the ratio of specific heats. The first coefficient in Eq. (10) follows from $(T/T_0)_s=(P/P_0)^{(\gamma-1)/\gamma}$, the second from definition. Equation (10) is used to eliminate entropy from Eq. (4):

$$\frac{\partial T'}{\partial t}=\chi\frac{\partial^2 T'}{\partial y^2}+\frac{(\gamma-1)T_0}{\gamma P_0}\frac{\partial p}{\partial t}, \tag{11}$$

where $\chi=\kappa_0/\rho_0 c_p$ is the equilibrium thermal diffusivity. The four equations (2), (3), (8), and (11) are now solved for the four unknowns u_x, u_y, p, and T'.

Harmonic excitation of the resonator is assumed,

$$v(t)=\mathrm{Re}\{v_0 e^{i\omega t}\}=\tfrac{1}{2}v_0 e^{i\omega t}+\text{c.c.}, \tag{12}$$

in which case

$$u_{x,y}=\mathrm{Re}\{\tilde{u}_{x,y}(x,y)e^{i\omega t}\},\quad p=\mathrm{Re}\{\tilde{p}(x)e^{i\omega t}\},$$
$$T'=\mathrm{Re}\{\tilde{T}(x,y)e^{i\omega t}\}, \tag{13}$$

where the functional form of p follows from Eq. (3). Equations (8), (2), and (11) thus become, respectively,

$$\frac{i\omega}{P_0}\tilde{p}-\frac{i\omega}{T_0}\tilde{T}+\frac{\partial \tilde{u}_x}{\partial x}+\frac{\partial \tilde{u}_y}{\partial y}=0, \tag{14}$$

$$\tilde{u}_x+\frac{1}{i\omega\rho_0}\frac{d\tilde{p}}{dx}=\frac{\nu}{i\omega}\frac{\partial^2 \tilde{u}_x}{\partial y^2}-v_0, \tag{15}$$

$$\tilde{T}=\frac{\chi}{i\omega}\frac{\partial^2 \tilde{T}}{\partial y^2}+\frac{(\gamma-1)T_0}{\gamma P_0}\tilde{p}, \tag{16}$$

where $\nu=\mu_0/\rho_0$ is the equilibrium kinematic viscosity.

As previously,[5] the solution of Eq. (15) for $\tilde{u}_x$ may be written

$$\tilde{u}_x=\tilde{u}_{x0}(x)\left(1-\frac{\cosh\beta_\nu y}{\cosh\beta_\nu y_0}\right), \tag{17}$$

where the x dependence is to be determined, and

$$\beta_\nu=\frac{1+i}{\delta_\nu},\quad \delta_\nu=\sqrt{\frac{2\nu}{\omega}}, \tag{18}$$

where δ_ν is the viscous penetration depth. It is helpful to introduce the average axial velocity

$$U_x(x)=\frac{1}{2y_0}\int_{-y_0}^{y_0}\tilde{u}_x(x,y)dy, \tag{19}$$

which may be interpreted as a normalized volume velocity. We thus have

$$U_x=(1-f_\nu)\tilde{u}_{x0}(x) \tag{20}$$

following substitution of Eq. (17), where

$$f_\nu=\frac{\tanh\beta_\nu y_0}{\beta_\nu y_0}. \tag{21}$$

Equation (17) thus becomes

$$\tilde{u}_x=\frac{U_x(x)}{1-f_\nu}\left(1-\frac{\cosh\beta_\nu y}{\cosh\beta_\nu y_0}\right), \tag{22}$$

where $U_x(x)$ is determined below. A relation for the y component of the particle velocity is obtained by integrating Eq. (14) from 0 to y:

$$\tilde{u}_y=-\frac{i\omega}{P_0}\tilde{p}(x)y+\frac{i\omega}{T_0}\int_0^y\tilde{T}(x,y')dy'-\int_0^y\frac{\partial\tilde{u}_x}{\partial x}dy', \tag{23}$$

where the boundary condition $\tilde{u}_y=0$ at $y=0$ was taken into account.

Similarly, the solution of Eq. (16) for $\tilde{T}$ may be written

$$\tilde{T}(x,y)=\tilde{T}_{x0}(x)\left(1-\frac{\cosh\beta_\chi y}{\cosh\beta_\chi y_0}\right), \tag{24}$$

where

$$\beta_\chi=\frac{1+i}{\delta_\chi},\quad \delta_\chi=\sqrt{\frac{2\chi}{\omega}}, \tag{25}$$

δ_χ is the thermal penetration depth, and

$$\tilde{T}_{x0}(x)=\frac{(\gamma-1)T_0}{\gamma P_0}\tilde{p}(x). \tag{26}$$

Average Eq. (24) across the channel to obtain

$$\frac{1}{2y_0}\int_{-y_0}^{y_0}\tilde{T}(x,y)dy=\frac{(\gamma-1)T_0}{\gamma P_0}(1-f_\chi)\tilde{p}(x), \tag{27}$$

where

$$f_\chi=\frac{\tanh\beta_\chi y_0}{\beta_\chi y_0}. \tag{28}$$

Taking the same average of Eq. (14) yields

$$\frac{dU_x}{dx}=-\frac{i\omega}{P_0}\tilde{p}(x)+\frac{i\omega}{T_0}\frac{1}{2y_0}\int_{-y_0}^{y_0}\tilde{T}(x,y)dy. \tag{29}$$

Now, use Eq. (27) to eliminate $\tilde{T}$ in Eq. (29) to obtain

$$\frac{dU_x}{dx}=-\frac{i\omega}{\rho_0 c_0^2}[1+(\gamma-1)f_\chi]\tilde{p}(x), \tag{30}$$

where we have introduced the adiabatic sound speed c_0:

$$c_0^2=\gamma P_0/\rho_0. \tag{31}$$

A second equation relating U_x and $\tilde{p}$ is obtained by averaging Eq. (15) across the channel to obtain[5]

$$\frac{1}{i\omega\rho_0}\frac{d\tilde{p}}{dx}+\frac{U_x(x)}{1-f_\nu}=-v_0. \tag{32}$$

Combining Eqs. (30) and (32) yields a single equation for U_x:

$$\frac{d^2U_x}{dx^2}-\alpha^2U_x=(1-f_\nu)\alpha^2v_0, \tag{33}$$

where

$$\alpha^2=-\left(\frac{1+(\gamma-1)f_\chi}{1-f_\nu}\right)\frac{\omega^2}{c_0^2}. \tag{34}$$

As before,[5] we obtain the solution of Eq. (33),

$$U_x(x)=(1-f_\nu)v_0\left(\frac{\cosh\alpha x}{\cosh\alpha x_0}-1\right), \tag{35}$$

by requiring $U_x=0$ at $x=\pm x_0$, and from Eq. (20) we have

$$\tilde{u}_{x0}(x)=v_0\left(\frac{\cosh\alpha x}{\cosh\alpha x_0}-1\right). \tag{36}$$

The acoustic pressure can now be obtained from Eqs. (30) and (35)

$$\tilde{p}(x)=-i\omega\rho_0v_0\frac{\sinh\alpha x}{\alpha\cosh\alpha x_0}, \tag{37}$$

which in turn yields the excess temperature from Eqs. (24) and (26):

$$\tilde{T}(x,y)=(\gamma-1)T_0\frac{\tilde{p}(x)}{\rho_0c_0^2}\left(1-\frac{\cosh\beta_\chi y}{\cosh\beta_\chi y_0}\right). \tag{38}$$

To calculate streaming, we also require the solution for the excess density $\rho'=\mathrm{Re}\{\tilde{\rho}(x,y)e^{i\omega t}\}$, obtained from Eq. (7):

$$\tilde{\rho}(x,y)=\frac{\tilde{p}(x)}{c_0^2}\left(1+(\gamma-1)\frac{\cosh\beta_\chi y}{\cosh\beta_\chi y_0}\right). \tag{39}$$

The x component of the particle velocity follows now from Eqs. (17) and (36):

$$\tilde{u}_x(x,y)=v_0\left(\frac{\cosh\alpha x}{\cosh\alpha x_0}-1\right)Y_x(y), \tag{40}$$

where

$$Y_x(y)=1-\frac{\cosh\beta_\nu y}{\cosh\beta_\nu y_0}. \tag{41}$$

The y component follows from Eq. (23):

$$\tilde{u}_y(x,y)=-v_0y_0f_\nu\frac{\alpha\sinh\alpha x}{\cosh\alpha x_0}Y_y(y), \tag{42}$$

where

$$Y_y(y)=\frac{y}{y_0}-\frac{\sinh\beta_\nu y}{\sinh\beta_\nu y_0}+\frac{(\gamma-1)(1-f_\nu)f_\chi}{[1+(\gamma-1)f_\chi]f_\nu}\times\left(\frac{y}{y_0}-\frac{\sinh\beta_\chi y}{\sinh\beta_\chi y_0}\right). \tag{43}$$

The above linear solutions are free of restrictions on channel width (provided boundary layer thickness is small compared with wavelength). Whereas effects of both viscosity and heat conduction are included in these solutions, the dependence of viscosity on temperature does not appear at the linear order of approximation.

In the absence of heat condition, $\kappa_0=0$, the solutions reduce to those derived previously for a purely viscous fluid.[5] In this limit the y dependence vanishes in Eqs. (38) and (39). Furthermore, $f_\chi=0$ and the earlier solutions for the particle velocity are recovered; see Eqs. (28) and (29) of Ref. 5. In the previous work, limiting forms of the solutions for a purely viscous fluid were presented for wide (large y_0/δ_ν) and narrow (small y_0/δ_ν) channels. Although similar asymptotic forms can be obtained in the present case of a thermoviscous gas, they shall not be considered here.

As noted previously,[5] although the solution for the particle velocity vanishes at the side walls, it does not at the ends. Specifically, although $\tilde{u}_x$ vanishes at $x=\pm x_0$, $\tilde{u}_y$ does not. Since the ends have negligible influence on the field beyond the boundary layer, and given the assumption that the boundary layer thickness is a very small fraction of channel length, failure to satisfy completely the end conditions has negligible effect on the streaming patterns.

The transverse spatial dependence of the acoustic field, when plotted versus y/y_0, is determined by the three dimensionless parameters y_0/δ_ν, γ, and the Prandtl number

$$\mathrm{Pr}=\nu/\chi. \tag{44}$$

Shown in Fig. 2 are the y dependencies of the particle velocity components, Eqs. (41) and (43), for several channel widths and Prandtl numbers. The smallest channel width considered, $y_0/\delta_\nu=2$ in Fig. 2(d), corresponds to a nominal channel width for thermoacoustic engine stacks.[11] Values of γ and Pr were taken for working gases used in thermoacoustic engines at standard temperature and pressure.[12] We set $\gamma=5/3$ for all cases, representative of monatomic gases. Solid lines in Fig. 2 are for $\mathrm{Pr}=\infty$ (heat conduction ignored), dashed are for $\mathrm{Pr}=0.67$ (for a single monatomic gas), and dot-dash for $\mathrm{Pr}=0.20$ (for a mixture of 60% helium and 40% xenon). Note that $Y_x(y)$ is independent of γ and Pr. The main feature revealed by $Y_y(y)$ is that the boundary layer thickness connected with $\tilde{u}_y$ increases slightly with thermal conductivity.

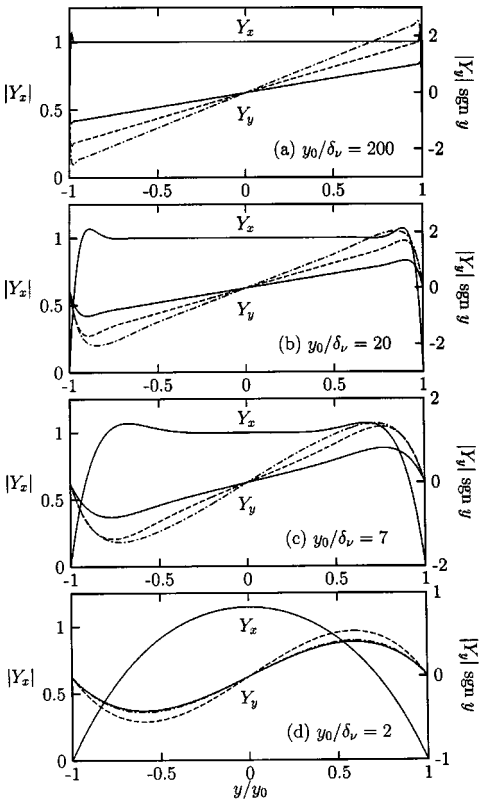


FIG. 2. Profiles of the x and y components of the acoustic particle velocity for standing waves in two-dimensional channels of different widths, with $\mathrm{Pr}=\infty$ (solid lines), $\mathrm{Pr}=0.67$ (dashed lines), and $\mathrm{Pr}=0.20$ (dot-dash lines).

B. Acoustic streaming

1. Governing equations

The analysis of acoustic streaming begins with the x component of the nonlinear momentum equation[5]

$$\frac{\partial(\rho u_x)}{\partial t}+\frac{\partial(\rho u_x^2)}{\partial x}+\frac{\partial(\rho u_x u_y)}{\partial y}=-\frac{\partial p}{\partial x}+\mu\frac{\partial^2 u_x}{\partial y^2}-\rho\dot{v}(t), \tag{45}$$

where we allow for dependence of the viscosity of the gas on temperature. The functional relation is assumed to be $\mu=\mu_0(T/T_0)^b$, expansion of which through first order in T' yields

$$\mu=\mu_0+\frac{\mu_0 b}{T_0}T'. \tag{46}$$

Higher-order terms are unnecessary for describing acoustic streaming at quadratic order in the acoustic variables. The thermal conductivity of a gas is usually assumed to vary in a similar way, as $\kappa=\kappa_0(T/T_0)^c$, where the exponent c is comparable to b.[11] However, this effect does not enter into calculations of acoustic streaming at our order of approximation.

Substitution of Eq. (46) in (45), expansion of the latter through quadratic order, and taking the time average yields

$$\mu_0\frac{\partial^2\overline{u_x}}{\partial y^2}=K(x)-F(x,y), \tag{47}$$

where

$$K(x)=\frac{d\overline{p}}{dx}+\frac{\gamma}{c_0^2}\langle\dot{v}p\rangle, \tag{48}$$

$$F(x,y)=-\rho_0\frac{\partial\langle u_x^2\rangle}{\partial x}-\rho_0\frac{\partial\langle u_x u_y\rangle}{\partial y}+\frac{\mu_0 b}{T_0}\frac{\partial}{\partial y}\left\langle T'\frac{\partial u_x}{\partial y}\right\rangle+\frac{\rho_0}{T_0}\langle\dot{v}T'\rangle, \tag{49}$$

and where we have used the linear relation $\rho'/\rho_0 = \gamma p/\rho_0 c_0^2 - T'/T_0$ obtained from Eqs. (7) and (31). The angular brackets $\langle\cdot\rangle$ designate the time average of the quantity inside, $\overline{u}_{x,y} = \langle u_{x,y}\rangle$ is the x component of the Eulerian streaming velocity, and $\overline{p} = \langle p\rangle$ is the dc pressure. For harmonic excitation of the resonator given by Eq. (12) we have

$$K(x) = \frac{d\overline{p}}{dx} + \frac{\gamma\omega}{2c_0^2}\,\mathrm{Im}\{v_0^*\tilde{p}\}, \tag{50}$$

$$F(x,y) = -\frac{\rho_0}{2}\frac{\partial|\tilde{u}_x|^2}{\partial x} - \mathrm{Re}\left\{\frac{\rho_0}{2}\frac{\partial(\tilde{u}_x\tilde{u}_y^*)}{\partial y} - \frac{\mu_0 b}{2T_0}\frac{\partial}{\partial y}\left(\tilde{T}\frac{\partial\tilde{u}_x^*}{\partial y}\right)\right\} + \frac{\rho_0\omega}{2T_0}\,\mathrm{Im}\{v_0^*\tilde{T}\}. \tag{51}$$

2. Eulerian streaming velocity

Equation (47) is integrated twice with respect to y to obtain the x component of the Eulerian streaming velocity:

$$\overline{u}_x = V(x,y) + C_2(x)y^2 + C_0(x), \tag{52}$$

where the coefficients C_0 and C_2, which follow from the integrations of $K(x)$, will be determined below by boundary conditions on the mass transport velocity. There is no term that is linear in y because $\overline{u}_x$ must be an even function of y. In what follows it is convenient to recognize that the x dependence of $\tilde{u}_x$ is proportional to $\tilde{u}_{x0}(x)$, and that for each of the quantities $\tilde{u}_y$, $\tilde{p}$, $\tilde{\rho}$, and $\tilde{T}$ it is proportional to $d\tilde{u}_{x0}/dx$. From the integrations of $F(x,y)$ one obtains

$$V(x,y) = \frac{\delta_\nu^2}{4\nu}\mathrm{Re}\left\{\tilde{u}_{x0}\frac{d\tilde{u}_{x0}^*}{dx}[H_1(y)+H_2(y)+H_3(y)] + \tilde{u}_{x0}^*\frac{d\tilde{u}_{x0}}{dx}H_4(y) - v_0^*\frac{d\tilde{u}_{x0}}{dx}H_5(y)\right\}, \tag{53}$$

where

$$H_1(y) = \frac{\cosh(2y/\delta_\nu)-\cos(2y/\delta_\nu)}{2|\cosh\beta_\nu y_0|^2} - 4\,\mathrm{Im}\frac{\cosh\beta_\nu y}{\cosh\beta_\nu y_0}, \tag{54}$$

$$H_2(y) = i\left[f_2^*\frac{\cosh\beta_\nu y - \beta_\nu y\sinh\beta_\nu y}{\cosh\beta_\nu y_0} + \frac{\cosh\beta_\nu^* y}{\cosh\beta_\nu^* y_0} - \frac{\cosh(2y/\delta_\nu)+i\cos(2y/\delta_\nu)}{2\beta_\nu\delta_\nu|\cosh\beta_\nu y|^2}\right], \tag{55}$$

$$H_3(y) = i(\gamma-1)\frac{f_1^*}{\mathrm{Pr}}\left[\frac{\cosh\beta_\chi^* y}{\cosh\beta_\chi^* y_0} - \frac{\mathrm{Pr}}{\beta_\chi\delta_\nu^2\cosh\beta_\nu y_0\cosh\beta_\chi^* y_0}\times\left(\frac{\cosh(\beta_\nu+\beta_\chi^*)y}{\beta_\nu+\beta_\chi^*} - \frac{\cosh(\beta_\nu-\beta_\chi^*)y}{\beta_\nu-\beta_\chi^*}\right)\right], \tag{56}$$

$$H_4(y) = (\gamma-1)f_1 b\beta_\nu\left[\frac{\cosh\beta_\nu^* y}{\beta_\nu^*\cosh\beta_\nu^* y_0} - \frac{1}{2\cosh\beta_\chi y_0\cosh\beta_\nu^* y_0}\left(\frac{\cosh(\beta_\chi+\beta_\nu^*)y}{\beta_\chi+\beta_\nu^*} - \frac{\cosh(\beta_\chi-\beta_\nu^*)y}{\beta_\chi-\beta_\nu^*}\right)\right], \tag{57}$$

$$H_5(y) = i(\gamma-1)\frac{f_1}{\mathrm{Pr}}\frac{\cosh\beta_\chi y}{\cosh\beta_\chi y_0}, \tag{58}$$

and we have introduced the notation

$$f_1 = \frac{1-f_\nu}{1+(\gamma-1)f_\chi}, \quad f_2 = \frac{f_\nu+(\gamma-1)f_\chi}{1+(\gamma-1)f_\chi}. \tag{59}$$

The y component of the Eulerian streaming velocity, should it be needed, can be obtained after calculation of the mass transport velocity, which is considered next.

3. Mass transport velocity

Our interest is in the mass transport velocity $\overline{\mathbf{u}}^M$, which is related to the Eulerian streaming velocity by[13]

$$\overline{\mathbf{u}}^M = \overline{\mathbf{u}} + \langle\rho'\mathbf{u}\rangle/\rho_0. \tag{60}$$

The x component is therefore given by

$$\overline{u}_x^M = \overline{u}_x + \frac{1}{2\rho_0}\mathrm{Re}\{\tilde{\rho}\tilde{u}_x^*\} \tag{61}$$

$$= \overline{u}_x + \frac{\delta_\nu^2}{4\nu}\mathrm{Re}\left\{\tilde{u}_{x0}^*\frac{d\tilde{u}_{x0}}{dx}H_6(y)\right\}, \tag{62}$$

where

$$H_6(y) = if_1\left(1+(\gamma-1)\frac{\cosh\beta_\chi y}{\cosh\beta_\chi y_0}\right)\left(1-\frac{\cosh\beta_\nu^* y}{\cosh\beta_\nu^* y_0}\right). \tag{63}$$

The y component of the mass transport velocity is determined by recognizing that $\overline{\mathbf{u}}^M$ is divergence-free,

$$\boldsymbol{\nabla}\cdot\overline{\mathbf{u}}^M = 0, \tag{64}$$

which permits definition of a stream function ψ:

$$\overline{u}_x^M = \frac{\partial\psi}{\partial y}, \quad \overline{u}_y^M = -\frac{\partial\psi}{\partial x}. \tag{65}$$

Integration of the first relation with respect to y, using Eq. (62), yields

$$\psi(x,y) = \Psi(x,y) + A_3(x)\frac{y^3}{y_0^3} + A_1(x)\frac{y}{y_0}, \tag{66}$$

in which even powers of y were set to zero because ψ must be an odd function of y, and where

$$\Psi(x,y)=\frac{\delta_\nu^2}{4\nu}\mathrm{Re}\Bigg\{\tilde{u}_{x0}\frac{d\tilde{u}_{x0}^*}{dx}[H_7(y)+H_8(y)+H_9(y)]+\tilde{u}_{x0}^*\frac{d\tilde{u}_{x0}}{dx}[H_{10}(y)+H_{11}(y)]-v_0^*\frac{d\tilde{u}_{x0}}{dx}H_{12}(y)\Bigg\}, \tag{67}$$

in which

$$H_7(y)=\frac{\sinh(2y/\delta_\nu)-\sin(2y/\delta_\nu)}{4|\cosh\beta_\nu y_0|^2}-4\,\mathrm{Im}\frac{\sinh\beta_\nu y}{\beta_\nu\delta_\nu\cosh\beta_\nu y_0}, \tag{68}$$

$$H_8(y)=i\Bigg[f_2^*\frac{2\sinh\beta_\nu y-\beta_\nu y\cosh\beta_\nu y}{\beta_\nu\delta_\nu\cosh\beta_\nu y_0}+\frac{i\sinh\beta_\nu^* y}{\beta_\nu\delta_\nu\cosh\beta_\nu^* y_0}-\frac{\sinh(2y/\delta_\nu)+i\sin(2y/\delta_\nu)}{4\beta_\nu\delta_\nu|\cosh\beta_\nu y|^2}\Bigg], \tag{69}$$

$$H_9(y)=i(\gamma-1)\frac{f_1^*}{\mathrm{Pr}}\Bigg[\frac{i\sinh\beta_\chi^* y}{\beta_\chi\delta_\nu\cosh\beta_\chi^* y_0}-\frac{\mathrm{Pr}}{\beta_\chi\delta_\nu^3\cosh\beta_\nu y_0\cosh\beta_\chi^* y_0}\times\Bigg(\frac{\sinh(\beta_\nu+\beta_\chi^*)y}{(\beta_\nu+\beta_\chi^*)^2}-\frac{\sinh(\beta_\nu-\beta_\chi^*)y}{(\beta_\nu-\beta_\chi^*)^2}\Bigg)\Bigg], \tag{70}$$

$$H_{10}(y)=\frac{1}{2}(\gamma-1)f_1 b\beta_\nu\delta_\nu\Bigg[\frac{i\sinh\beta_\nu^* y}{\cosh\beta_\nu^* y_0}-\frac{1}{\delta_\nu^2\cosh\beta_\chi y_0\cosh\beta_\nu^* y_0}\Bigg(\frac{\sinh(\beta_\chi+\beta_\nu^*)y}{(\beta_\chi+\beta_\nu^*)^2}-\frac{\sinh(\beta_\chi-\beta_\nu^*)y}{(\beta_\chi-\beta_\nu^*)^2}\Bigg)\Bigg], \tag{71}$$

$$H_{11}(y)=if_1\Bigg[\frac{y}{\delta_\nu}+\frac{(\gamma-1)\sinh\beta_\chi y}{\beta_\chi\delta_\nu\cosh\beta_\chi y_0}-\frac{i\sinh\beta_\nu^* y}{\beta_\nu\delta_\nu\cosh\beta_\nu^* y_0}-\frac{\gamma-1}{2\delta_\nu\cosh\beta_\chi y_0\cosh\beta_\nu^* y_0}\Bigg(\frac{\sinh(\beta_\chi+\beta_\nu^*)y}{\beta_\chi+\beta_\nu^*}+\frac{\sinh(\beta_\chi-\beta_\nu^*)y}{\beta_\chi-\beta_\nu^*}\Bigg)\Bigg], \tag{72}$$

$$H_{12}(y)=\frac{i(\gamma-1)f_1}{\beta_\chi\delta_\nu\mathrm{Pr}}\frac{\sinh\beta_\chi y}{\cosh\beta_\chi y_0}. \tag{73}$$

Instead of the two unknown functions C_0 and C_2 in Eq. (52), we now have A_1 and A_3 in Eq. (66). We thus write

$$\overline{u}_x^M=\overline{U}_x^M(x,y)+\frac{1}{y_0}\Bigg[3A_3(x)\frac{y^2}{y_0^2}+A_1(x)\Bigg], \tag{74}$$

where

$$\overline{U}_x^M=\frac{\delta_\nu^2}{4\nu}\mathrm{Re}\Bigg\{\tilde{u}_{x0}\frac{d\tilde{u}_{x0}^*}{dx}[H_1(y)+H_2(y)+H_3(y)]+\tilde{u}_{x0}^*\frac{d\tilde{u}_{x0}}{dx}[H_4(y)+H_6(y)]-v_0^*\frac{d\tilde{u}_{x0}}{dx}H_5(y)\Bigg\}. \tag{75}$$

The unknowns A_1 and A_3 are determined by requiring that the x and y components of $\overline{\mathbf{u}}^M$ vanish along the wall at $y=y_0$, which yields, respectively,

$$\overline{u}_x^M(x,y_0)=\overline{U}_x^M(x,y_0)+\frac{3}{y_0}A_3(x)+\frac{1}{y_0}A_1(x)=0, \tag{76}$$

$$\psi(x,y_0)=\Psi(x,y_0)+A_3(x)+A_1(x)=0, \tag{77}$$

the solutions of which are

$$A_1(x)=-\frac{3}{2}\Psi(x,y_0)+\frac{y_0}{2}V(x,y_0), \tag{78}$$

$$A_3(x)=\frac{1}{2}\Psi(x,y_0)-\frac{y_0}{2}V(x,y_0), \tag{79}$$

where it was taken into account that $\overline{U}_x^M(x,y_0)=V(x,y_0)$.

Equation (74) is the complete solution for $\overline{u}_x^M$. The solution for $\overline{u}_y^M$ follows from substituting Eq. (66) into the second of Eqs. (65) and taking the indicated derivatives of $\tilde{u}_{x0}(x)$. The resulting closed-form solution for the vector field $\overline{\mathbf{u}}^M$ is free of restrictions on channel width.

4. Wide channels

In the limit of wide channels ($y_0\gg\delta_\nu$) and outside the boundary layer, it can be shown that the x component of the Eulerian streaming velocity reduces to

$$\overline{u}_x=-\Gamma u_R\Bigg(1-3\frac{y^2}{y_0^2}\Bigg)\sin 2kx,\quad u_R=\frac{3}{16}\frac{u_0^2}{c_0}, \tag{80}$$

where

$$\Gamma=1+\frac{2}{3}(\gamma-1)(1-b)\frac{\sqrt{\mathrm{Pr}}}{1+\mathrm{Pr}}. \tag{81}$$

The acoustic field is described in this limit by $\tilde{u}_x=u_0\cos kx$, where $k=\omega/c_0$. In the absence of thermal effects $\Gamma=1$ and Eq. (80) is the solution obtained by Rayleigh.[1] The extension of Rayleigh's analysis to include thermal effects, taken into account by Γ, was provided by Rott.[3] Values of γ, b, and Pr at standard temperature and pressure[11] yield $\Gamma=1.03$ for air and $\Gamma=1.07$ for helium. Thermal effects thus have little influence on streaming in wide channels. Rott called attention to the curious result that thermal effects on streaming in wide channels disappear when viscosity is proportional to absolute temperature ($b=1$).

We note that Eq. (80) applies also to the mass transport velocity, because $\overline{u}_x\simeq\overline{u}_x^M$ in wide channels outside the boundary layer.[5]

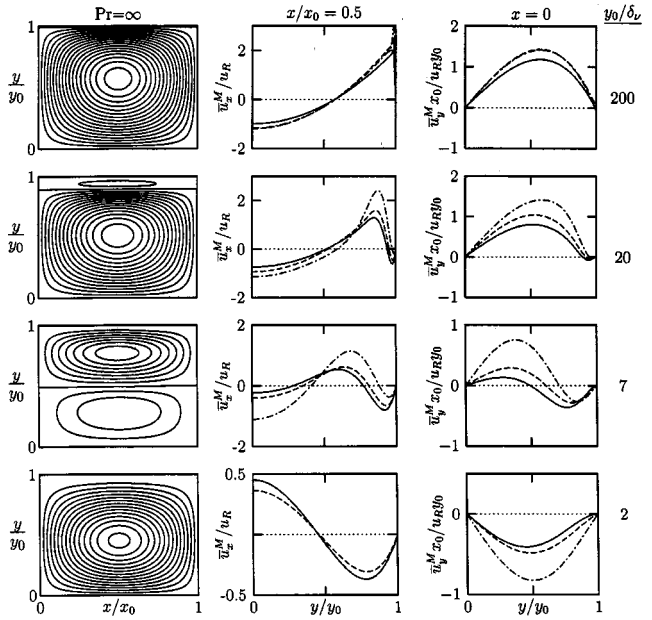


FIG. 3. Acoustic streaming patterns for the average mass transport velocity in a two-dimensional resonator with rigid ends, excited in its lowest mode, as a function of channel width and Prandtl number. Left column: streamlines for Pr$=\infty$. Middle column: transverse distribution of x component at $x/x_0=0.5$, with Pr$=\infty$ (solid lines), Pr$=0.67$ (dashed lines), and Pr$=0.20$ (dot-dash lines). Right column: transverse distribution of y component at $x/x_0=0$, with same Prandtl numbers and corresponding line types. In all cases, $\gamma=5/3$ and $b=0$.

5. Results

As in the case of non-heat-conducting fluids,[5] it is convenient to consider the following dimensionless forms of the streaming quantities:

$$\frac{\psi(x,y)}{y_0 u_R}, \quad \frac{\overline{u}_x^M(x,y)}{u_R}, \quad \frac{x_0 \overline{u}_y^M(x,y)}{y_0 u_R}, \tag{82}$$

where $u_R=\frac{3}{16}u_0^2/c_0$ is Rayleigh's result for the maximum streaming velocity in the center of a wide channel containing a purely viscous fluid [see Eq. (80)], and u_0 is the axial value of $|\tilde{u}_x|$ at a velocity antinode. When expressed in terms of x/x_0 and y/y_0, the quantities in Eq. (82) are determined completely by the five dimensionless parameters ω/ω_0, y_0/δ_ν, Pr, γ, and b. Here, $\omega_0=\pi c_0/2x_0$ is the fundamental natural frequency of the resonator in the absence of viscosity and heat conduction. The drive frequency ω is chosen to be the lowest resonance frequency in the presence of viscosity and heat conduction. We identify this frequency as the one that maximizes $u_0=|\tilde{u}_x(0,0)|$. Equation (40) reveals that the resonance frequency defined in this way is the one that maximizes the quantity $|1-\operatorname{sech}\alpha x_0|$. As in Fig. 2 we set $\gamma=5/3$, which leaves y_0/δ_ν, Pr, and b as the remaining free parameters.

In Fig. 3 we examine the effect of heat conduction in channels of different widths, with $b=0$ for all cases. The left column displays the streamlines ($\psi=$const) for the upper-right quadrant of Fig. 1 when heat conduction is ignored (Pr$=\infty$). Different rows are associated with the different channel widths listed in the right of the figure. The dependence of these streaming patterns on channel width has been discussed previously.[5] The main point is that for wide channels the streaming pattern is dominated by the outer, Rayleigh streaming vortex. For sufficiently narrow channels the Rayleigh streaming disappears, leaving only the inner streaming vortices within the boundary layer, rotating in directions opposite those of the Rayleigh streaming.

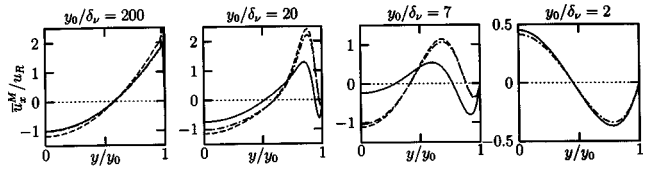


FIG. 4. Acoustic streaming patterns corresponding to Fig. 3, but illustrating influence of viscosity dependence on temperature in channels of different widths. Solid lines are for Pr$=\infty$ and $b=0$, dashed lines are for Pr$=0.2$ and $b=0$, and dot-dash lines are for Pr$=0.2$ and $b=1$.

In the second and third columns of Fig. 3 are plotted, respectively, $\overline{u}_x^M(y)$ at $x/x_0=0.5$ and $\overline{u}_y^M(y)$ at $x/x_0=0$. Solid lines are for Pr$=\infty$, dashed lines are for Pr$=0.67$, and dot-dash lines are for Pr$=0.20$. In the wide-channel limit, $y_0/\delta_\nu=200$, heat conduction is seen to increase the streaming velocity by about 20%, as predicted by Eq. (80) with $b=0$. For the intermediate channel widths, represented by the second and third rows, it is seen that heat conduction can double the streaming velocity. For $y_0/\delta_\nu=7$, heat conduction is also seen to alter the spatial distribution. In this case the shift in the zero crossing of $\overline{u}_y^M$ indicates that the size of the inner streaming vortex is reduced. In the narrow channel, $y_0/\delta_\nu=2$, the effect of heat conduction is relatively small at $x/x_0=0.5$, but for Pr$=0.20$ it is relatively large at $x=0$. This particular variation is connected with the source term $\langle\rho\dot{v}\rangle$ following from Eq. (45).

Figure 4 illustrates the effect of viscosity dependence on temperature, which is determined by the exponent b [see Eq. (46)]. The graphs show $\overline{u}_x^M(y)$ at $x/x_0=0.5$. Solid lines are for Pr$=\infty$ and $b=0$, dashed lines are for Pr$=0.2$ and $b=0$, and dot-dash lines are for Pr$=0.2$ and $b=1$. In the wide channel, $y_0/\delta_\nu=200$, the dot-dash line is nearly identical to the solid line. As predicted by Eq. (80), linear dependence of μ on T ($b=1$) compensates for the effect of heat conduction in wide channels. For the intermediate channel widths, $y_0/\delta_\nu=20$ and $y_0/\delta_\nu=7$, the temperature dependence of viscosity has a negligible effect on the streaming.

III. CYLINDRICAL TUBES

The geometry of the cylindrical resonator corresponds formally to the upper half of Fig. 1, with the rectangular coordinates (x, y) replaced by the cylindrical coordinates (z, r). Specifically, the tube is formed by rigid walls at $z=\pm z_0$ and $r=r_0$. The boundary conditions are again that fluid motion vanishes at the walls, and that wall temperature equals the equilibrium temperature of the gas. Axial symmetry is assumed, which requires the following quantities to vanish at $r=0$: the r component of the particle velocity, and the derivative with respect to r of the pressure, density, temperature, and z component of the particle velocity.

A. Linear theory

Solutions for the acoustic variables in the case of cylindrical channels are derived following essentially the same steps as in Sec. II A. For the acoustical variables we now write

$$u_{z,r}=\mathrm{Re}\{\tilde{u}_{z,r}(z,r)e^{i\omega t}\},\quad p=\mathrm{Re}\{\tilde{p}(z)e^{i\omega t}\},$$
$$T'=\mathrm{Re}\{\tilde{T}(z,r)e^{i\omega t}\},\quad \rho'=\mathrm{Re}\{\tilde{\rho}(z,r)e^{i\omega t}\}. \tag{83}$$

Equations (14)–(16) are thus replaced by

$$\frac{i\omega}{P_0}\tilde{p}-\frac{i\omega}{T_0}\tilde{T}+\frac{\partial\tilde{u}_z}{\partial z}+\frac{1}{r}\frac{\partial(r\tilde{u}_r)}{\partial r}=0, \tag{84}$$

$$\tilde{u}_z+\frac{1}{i\omega\rho_0}\frac{d\tilde{p}}{dz}=\frac{\nu}{i\omega r}\frac{\partial}{\partial r}\left(r\frac{\partial\tilde{u}_z}{\partial r}\right)-v_0, \tag{85}$$

$$\tilde{T}=\frac{\chi}{i\omega r}\frac{\partial}{\partial r}\left(r\frac{\partial\tilde{T}}{\partial r}\right)+\frac{(\gamma-1)T_0}{\gamma P_0}\tilde{p}. \tag{86}$$

Instead of Eq. (19), we use here for the average axial particle velocity

$$U_z(z)=\frac{2}{r_0^2}\int_0^{r_0}\tilde{u}_z(z,r)r\ dr. \tag{87}$$

The same integral, but over $\tilde{T}$ rather than $\tilde{u}_z$, is used in place of Eq. (27) to calculate the average excess temperature.

The solutions for the particle velocity components are

$$\tilde{u}_z(z,r)=\tilde{u}_{z0}(z)R_z(r), \tag{88}$$

$$\tilde{u}_r(z,r)=-\frac{1}{2}f_\nu r_0\frac{d\tilde{u}_{z0}}{dz}R_r(r), \tag{89}$$

where

$$\tilde{u}_{z0}(z)=v_0\left(\frac{\cosh\alpha z}{\cosh\alpha z_0}-1\right), \tag{90}$$

$$R_z(r)=1-\frac{J_0(\beta_\nu r)}{J_0(\beta_\nu r_0)}, \tag{91}$$

$$R_r(r)=\frac{r}{r_0}-\frac{J_1(\beta_\nu r)}{J_1(\beta_\nu r_0)}+\frac{(\gamma-1)(1-f_\nu)f_\chi}{[1+(\gamma-1)f_\chi]f_\nu}\times\left(\frac{r}{r_0}-\frac{J_1(\beta_\chi r)}{J_1(\beta_\chi r_0)}\right), \tag{92}$$

where J_0 and J_1 are Bessel functions. Here, the coefficients have the following definitions:

$$\alpha^2=-\left(\frac{1+(\gamma-1)f_\chi}{1-f_\nu}\right)\frac{\omega^2}{c_0^2},$$
$$f_{\nu,\chi}=\frac{2J_1(\beta_{\nu,\chi}r_0)}{\beta_{\nu,\chi}r_0J_0(\beta_{\nu,\chi}r_0)},\quad \beta_{\nu,\chi}=\frac{i-1}{\delta_{\nu,\chi}}. \tag{93}$$

Note that the definitions of $f_{\nu,\chi}$ and $\beta_{\nu,\chi}$ used here differ from those for two-dimensional channels, while the definitions of $\delta_{\nu,\chi}$ are the same as before, $\delta_\nu=(2\nu/\omega)^{1/2}$ and $\delta_\chi=(2\chi/\omega)^{1/2}$. The solutions for the remaining field variables are

$$\tilde{p}(z)=-\frac{i\omega\rho_0}{\alpha^2}\frac{d\tilde{u}_{z0}}{dz}, \tag{94}$$

$$\tilde{T}(z,r)=(\gamma-1)T_0\frac{\tilde{p}(z)}{\rho_0c_0^2}R_T(r), \tag{95}$$

$$\tilde{\rho}(z,r)=\frac{\tilde{p}(z)}{c_0^2}R_\rho(r), \tag{96}$$

where

$$R_T(r)=1-\frac{J_0(\beta_\chi r)}{J_0(\beta_\chi r_0)}, \tag{97}$$

$$R_\rho(r)=1+(\gamma-1)\frac{J_0(\beta_\chi r)}{J_0(\beta_\chi r_0)}. \tag{98}$$

The solutions for the acoustic field variables are qualitatively the same as for a two-dimensional channel (see Fig. 2).

B. Acoustic streaming

1. Eulerian streaming velocity

In place of Eqs. (47), (50), and (51) we have the following differential equation for the z component of the Eulerian streaming velocity, $\overline{u_z}$:

$$\frac{\mu_0}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\overline{u_z}}{\partial r}\right)=K(z)-F(z,r), \tag{99}$$

where

$$K(z)=\frac{d\overline{p}}{dz}+\frac{\gamma\omega}{2c_0^2}\mathrm{Im}\{v_0^*\tilde{p}\}, \tag{100}$$

$$F(z,r)=-\frac{\rho_0}{2}\frac{\partial|\tilde{u}_z|^2}{\partial z}-\mathrm{Re}\left\{\frac{\rho_0}{2r}\frac{\partial(r\tilde{u}_r\tilde{u}_z^*)}{\partial r}-\frac{\mu_0 b}{2T_0r}\frac{\partial}{\partial r}\left(r\tilde{T}\frac{\partial\tilde{u}_z^*}{\partial r}\right)\right\}+\frac{\rho_0\omega}{2T_0}\mathrm{Im}\{v_0^*\tilde{T}\}. \tag{101}$$

The first integration of Eq. (99) with respect to r gives

$$\frac{\partial\overline{u_z}}{\partial r}=\frac{1}{\mu_0 r}\int_0^r[K(z)-F(z,r')]r'dr'$$
$$=S_1(z,r)+S_2(z,r)+S_3(z,r)+S_4(z,r)+B_2(z)r, \tag{102}$$

where

$$S_1(z,r)=\frac{1}{2\nu r}\int_0^r\frac{\partial|\tilde{u}_z|^2}{\partial z}r'dr'=\mathrm{Re}\left\{\frac{d\tilde{u}_{z0}}{dz}\tilde{u}_{z0}^*T_1(r)\right\}, \tag{103}$$

$$S_2(z,r)=\frac{1}{2\nu}\mathrm{Re}\{u_ru_z^*\}=\mathrm{Re}\left\{\frac{d\tilde{u}_{z0}}{dz}\tilde{u}_{z0}^*T_2(r)\right\}, \tag{104}$$

$$S_3(z,r)=-\frac{b}{2T_0}\mathrm{Re}\left\{\tilde{T}\frac{\partial\tilde{u}_z^*}{\partial r}\right\}=\mathrm{Re}\left\{\frac{d\tilde{u}_{z0}}{dz}\tilde{u}_{z0}^*T_3(r)\right\}, \tag{105}$$

$$S_4(z,r) = -\frac{\omega}{2\nu T_0 r}\int_0^r \mathrm{Im}\{v_0^* \tilde{T}\} r'\,dr'$$

$$= \mathrm{Re}\left\{\frac{d\tilde{u}_{z0}}{dz} v_0^* T_4(r)\right\}, \tag{106}$$

and

$$T_1(r) = \frac{1}{\nu r}\int_0^r |R_z(r')|^2 r'\,dr', \tag{107}$$

$$T_2(r) = -\frac{r_0 f_\nu}{4\nu} R_r(r) R_z^*(r), \tag{108}$$

$$T_3(r) = (\gamma - 1)\frac{b r_0 f_1 f_\nu^*}{4\nu} R_T(r)\frac{J_1^*(\beta_\nu r)}{J_1^*(\beta_\nu r_0)}, \tag{109}$$

$$T_4(r) = -(\gamma - 1)\frac{f_1}{2\nu r}\int_0^r R_T(r') r'\,dr', \tag{110}$$

with

$$f_1 = \frac{1 - f_\nu}{1 + (\gamma - 1) f_\chi}. \tag{111}$$

We also have $B_2 = K(z)/2\mu_0$, while the other constant of integration in Eq. (102) was set to zero because $\partial \overline{u_z}/\partial r$ must vanish at $r = 0$.

Integration of Eq. (102) with respect to r yields

$$\overline{u_z}(z,r) = V(z,r) + \tfrac{1}{2} B_2(z) r^2 + B_0(z), \tag{112}$$

where

$$V(z,r) = \mathrm{Re}\left\{\frac{d\tilde{u}_{z0}}{dz}\tilde{u}_{z0}^*[Q_1(r) + Q_2(r) + Q_3(r)] + \frac{d\tilde{u}_{z0}}{dz} v_0^* Q_4(r)\right\}, \tag{113}$$

and

$$Q_i(r) = \int_0^r T_i(r')\,dr', \quad 1 \leqslant i \leqslant 4. \tag{114}$$

The functions $B_2(z)$ and $B_0(z)$ are determined by boundary conditions on the mass transport velocity.

2. Mass transport velocity

Equation (62) for the x component of the mass transport velocity becomes

$$\overline{u}_z^M = \overline{u_z} + \mathrm{Re}\left\{\frac{d\tilde{u}_{z0}}{dz}\tilde{u}_{z0}^* Q_5(r)\right\}, \tag{115}$$

where

$$Q_5(r) = -\frac{f_1}{2i\omega} R_\rho(r) R_z^*(r). \tag{116}$$

For the stream function we have, in place of Eqs. (65),

$$\overline{u}_z^M = \frac{1}{r}\frac{\partial(r\psi)}{\partial r}, \quad \overline{u}_r^M = -\frac{\partial \psi}{\partial z}. \tag{117}$$

Therefore

$$\psi = \frac{1}{r}\int_0^r \overline{u}_z^M(z,r') r'\,dr', \tag{118}$$

which after substitution of Eq. (115) can be expressed as

$$\psi(z,r) = \Psi(z,r) + A_3(z)\frac{r^3}{r_0^3} + A_1(z)\frac{r}{r_0}, \tag{119}$$

where

$$\Psi(z,r) = \mathrm{Re}\left\{\frac{d\tilde{u}_{z0}}{dz}\tilde{u}_{z0}^*[P_1(r) + P_2(r) + P_3(r) + P_5(r)] + \frac{d\tilde{u}_{z0}}{dz} v_0^* P_4(r)\right\}, \tag{120}$$

and

$$P_i(r) = \frac{1}{r}\int_0^r Q_i(r') r'\,dr', \quad 1 \leqslant i \leqslant 5. \tag{121}$$

The expression for mass transport velocity is now rewritten as

$$\overline{u}_z^M = \overline{U}_z^M(z,r) + \frac{2}{r_0}\left[2A_3(z)\frac{r^2}{r_0^2} + A_1(z)\right], \tag{122}$$

where

$$\overline{U}_z^M = \mathrm{Re}\left\{\frac{d\tilde{u}_{z0}}{dz}\tilde{u}_{z0}^*[Q_1(r) + Q_2(r) + Q_3(r) + Q_5(r)] + \frac{d\tilde{u}_{z0}}{dz} v_0^* Q_4(r)\right\}. \tag{123}$$

The unknown functions $A_1(z)$ and $A_3(z)$ are determined by the condition that the motion must vanish at the walls, such that $\overline{u}_z^M(z,r_0) = 0$ and $\psi(z,r_0) = 0$. Solving these equations for A_1 and A_3 yields

$$A_1(z) = -2\Psi(z,r_0) + \frac{r_0}{2}\overline{U}_z^M(z,r_0), \tag{124}$$

$$A_3(z) = \Psi(z,r_0) - \frac{r_0}{2}\overline{U}_z^M(z,r_0). \tag{125}$$

The solution is now complete. The z component of the mass transport velocity is given by Eq. (122). The r component is obtained by substituting Eq. (119) in the second of Eqs. (117) and taking derivatives of $\tilde{u}_{z0}(z)$, the functional forms of which are identical to those for two-dimensional channels.

3. Wide tubes

Here, we have in the wide-tube limit ($r_0 \gg \delta_\nu$) and outside the boundary layer

$$\overline{u_z} = -\Gamma u_R\left(1 - 2\frac{r^2}{r_0^2}\right)\sin 2kz, \quad u_R = \frac{3}{8}\frac{u_0^2}{c_0}, \tag{126}$$

where Γ is given again by Eq. (81), and $\tilde{u}_z = u_0 \cos kz$, where $k = \omega/c_0$. Without thermal effects, in which case $\Gamma = 1$, Eq. (126) is the solution derived by Schuster and Matz.[2] Note that for the same acoustic particle velocity amplitude, the axial streaming velocity in the cylindrical tube is twice as

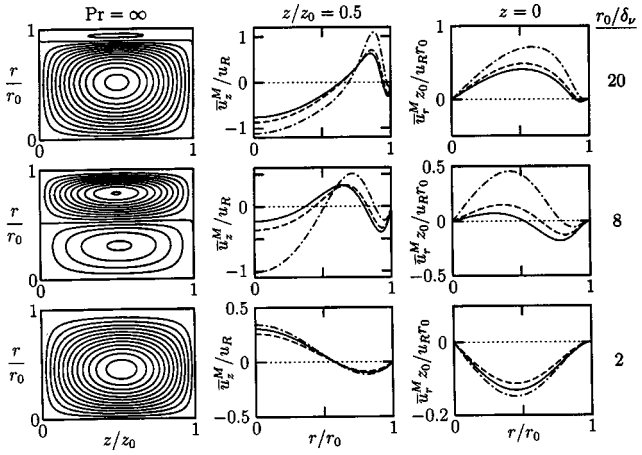


FIG. 5. Acoustic streaming patterns for the average mass transport velocity in a cylindrical tube resonator with rigid ends, excited in its lowest mode, as a function of tube radius and Prandtl number. Left column: streamlines for $\mathrm{Pr}=\infty$. Middle column: transverse distribution of z component at $z/z_0=0.5$, with $\mathrm{Pr}=\infty$ (solid lines), $\mathrm{Pr}=0.67$ (dashed lines), and $\mathrm{Pr}=0.20$ (dot-dash lines). Right column: transverse distribution of r component at $z/z_0=0$, with same Prandtl numbers and corresponding line types. In all cases, $\gamma=5/3$ and $b=0$.

large as in the two-dimensional channel. Also, in a two-dimensional channel the maximum velocity varies (at $\sin 2kz=1$) from $-\Gamma u_R$ in the center to $2\Gamma u_R$ near the wall, whereas in a cylindrical tube it varies from $-\Gamma u_R$ in the center to Γu_R near the wall.

4. Results

The dimensionless quantities in Eq. (82) become ψ/r_0u_R, $\overline{u}_z^M/u_R$, and $z_0\overline{u}_r^M/r_0u_R$, where $u_R=\frac{3}{8}u_0^2/c_0$ from Eq. (126). It is again assumed that the tube is excited at its lowest resonance frequency, calculated as before and with $u_0=|\tilde{u}_z(0,0)|$. Figure 5 is comparable to Fig. 3, illustrating the influence of heat conduction in channels of different widths. Figure 6 is comparable to Fig. 4, illustrating the influence of viscosity dependence on temperature. The behavior in 2D channels and cylindrical tubes is seen to be qualitatively the same.

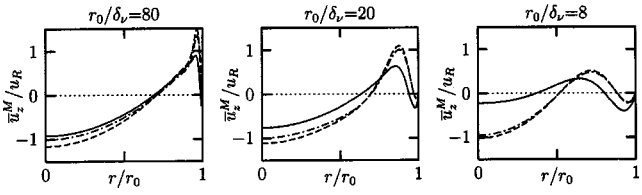


FIG. 6. Acoustic streaming patterns corresponding to Fig. 5, but illustrating influence of viscosity dependence on temperature in cylindrical tubes of different radii. Solid lines are for $\mathrm{Pr}=\infty$ and $b=0$, dashed lines are for $\mathrm{Pr}=0.2$ and $b=0$, and dot-dash lines are for $\mathrm{Pr}=0.2$ and $b=1$.

IV. CONCLUSION

An earlier analysis of acoustic streaming in a viscous fluid between parallel plates is generalized here to include two thermal effects, heat conduction and dependence of the viscosity on temperature in a gas. Cylindrical tubes as well as two-dimensional channels are considered. Calculations are presented for typical working gases used in thermoacoustic engines. In the Rayleigh streaming limit (channels that are orders of magnitude wider than the viscous penetration depth) and in very narrow channels (widths on the order of the viscous penetration depth), thermal effects influence the acoustic streaming by only a few percent. In the intermediate range, for widths on the order of 10–20 times the viscous penetration depth, thermal conductivity can have a substantial influence on the streaming. The streaming is qualitatively the same in two-dimensional channels and cylindrical tubes.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research, and by the Internal Research and Development Program at Applied Research Laboratories. Figures 5 and 6, which appeared previously in Ref. 9, are reprinted with permission.

[1] Lord Rayleigh, "On the circulation of air observed in Kundt's tubes, and on some allied acoustical problems," Philos. Trans. R. Soc. London **175**, 1–21 (1884), Sec. 3. See also Lord Rayleigh (J. W. Strutt), *The Theory of Sound*, 2nd ed. (Dover, New York, 1945), Vol. 2, Sec. 352.

[2] K. Schuster and W. Matz, "Über stationäre Strömungen im Kundtschen Rohr" ("On stationary streaming in a Kundt's tube"), Akust. Z. **5**, 349–352 (1940).

[3] N. Rott, "The influence of heat conduction on acoustic streaming," Z. Angew. Math. Phys. **25**, 417–421 (1974).

[4] Q. Qi, "The effect of compressibility on acoustic streaming near a rigid boundary for a plane traveling wave," J. Acoust. Soc. Am. **94**, 1090–1098 (1993).

[5] M. F. Hamilton, Yu. A. Ilinskii, and E. A. Zabolotskaya, "Acoustic streaming generated by standing waves in two-dimensional channels of arbitrary width," J. Acoust. Soc. Am. **113**, 153–160 (2003).

[6] R. Waxler, "Stationary velocity and pressure gradients in a thermoacoustic stack," J. Acoust. Soc. Am. **109**, 2739–2750 (2001).

[7] H. Bailliet, V. Gusev, R. Raspet, and R. A. Hiller, "Acoustic streaming in closed thermoacoustic devices," J. Acoust. Soc. Am. **110**, 1808–1821 (2001).

[8] In Ref. 5 it was stated incorrectly that mass transport velocity was not considered in Ref. 7.

[9] M. F. Hamilton, Yu. A. Ilinskii, and E. A. Zabolotskaya, "Acoustic streaming in a cylindrical tube," in *Nonlinear Acoustics at the Beginning of the 21st Century*, edited by O. V. Rudenko and O. A. Sapozhnikov (Moscow State University Press, Moscow, 2002), Vol. 1, pp. 147–150.

[10] M. F. Hamilton, Yu. A. Ilinskii, and E. A. Zabolotskaya, "Nonlinear two-dimensional model for thermoacoustic engines," J. Acoust. Soc. Am. **111**, 2076–2086 (2002).

[11] G. W. Swift, *Thermoacoustics* (Acoustical Society of America, New York, 2002).

[12] J. R. Belcher, W. V. Slaton, R. Raspet, H. E. Bass, and J. Lightfoot, "Working gases in thermoacoustic engines," J. Acoust. Soc. Am. **105**, 2677–2684 (1999).

[13] W. L. Nyborg, "Acoustic streaming," in *Physical Acoustics*, edited by W. P. Mason (Academic, New York, 1965), Vol. 2B, Chap. 11.

Excess attenuation of an acoustic beam by turbulence

Naixian Pan
Department of Atmospheric Science, Peking University, Beijing 100871, China

(Received 18 December 2000; revised 18 September 2003; accepted 22 September 2003)

A theory based on the concept of a spatial sinusoidal diffraction grating is presented for the estimation of the excess attenuation in an acoustic beam. The equation of the excess attenuation coefficient shows that the excess attenuation of acoustic beam not only depends on the turbulence but also depends on the application parameters such as the beam width, the beam orientation and whether for forward propagation or back scatter propagation. Analysis shows that the excess attenuation appears to have a frequency dependence of cube-root. The expression for the excess attenuation coefficient has been used in the estimations of the temperature structure coefficient, C_T^2, in sodar sounding. The correction of C_T^2 values for excess attenuation reduces their errors greatly. Published profiles of temperature structure coefficient and the velocity structure coefficient in convective conditions are used to test our theory, which is compared with the theory by Brown and Clifford. The excess attenuation due to scattering from turbulence and atmospheric absorption are both taken into account in sodar data processing for deducing the contribution of the lower atmosphere to seeing, which is the sharpness of a telescope image determined by the degree of turbulence in the Earth's atmosphere. The comparison between the contributions of the lowest 300-m layer to seeing with that of the whole atmosphere supports the reasonableness of our estimation of excess attenuation.
[DOI: 10.1121/1.1628679]

PACS numbers: 43.28.Fp, 43.20.Fn, 43.20.Hg [LCS] Pages: 3102–3111

I. INTRODUCTION

Tyndall[1] discovered the phenomenon of sound scattering by turbulence in 1874. After a long silence, acoustic scattering theory started to build up in the 1940s.[2,3] In the 1950s, Kraichnan[4] and Lighthill[5] proposed the correct form of the velocity-dependent part of the scattering cross section. Lighthill estimated the excess attenuation by integrating the energy flux due to the scattering cross section, including the velocity fluctuations only, over all angles on a spherical surface. That theory also predicted a quadratic frequency dependence.

In the 1960s, Tatarskii[6] and Monin[7] introduced the effective scattering cross section under the condition of homogeneous and isotropic random fields. They showed that the acoustic wave scattering resulted from the inhomogeneity of the temperature and velocity fields and that scattering at a given scattering angle is determined by a single spectral component of the inhomogeneities, corresponding to Bragg's condition of a spatial sinusoidal diffraction grating. Under the condition of the locally homogeneous and isotropic random fields with Kolmogorov spectra, the effective scattering cross section[8] is related to two measurable parameters, i.e., the temperature structure coefficient C_T^2 and velocity structure coefficient C_V^2, and has been widely used in sodar sounding.

Since the mid 1970s experiments on measuring C_T^2 or C_T^2 plus C_V^2 were carried out by many scientists.[9–15] Haugen *et al.*[11] found that the difference between the sodar and the tower instrumentation measurements was large, and attributed it to attenuation from scattering by turbulence and beam bending. Neff[9] and Asimakopoulos[10,13] have similar opinions. Brown and Clifford[16,17] presented a theory of attenuation by turbulence. Since then no further experimental verification of their theory has been published to the author's knowledge. The utilization of echo intensity seems to have slowed down.

Measurement of C_T^2 with sodar is very useful in line-of-sight communication links and turbulence studies of the atmospheric boundary layer. Moreover, C_T^2 and C_V^2 are important parameters in deduction of sensible heat flux[18,19] in the study of the boundary layer meteorology. Knowing the attenuation along the path of a sodar beam has become important to the quantitative measurements of the intensity of echo signals and retrieval of turbulent parameters. In the present paper, we are only concerned with the excess attenuation and derive an equation for this excess attenuation of an acoustic beam based on the classic scattering theory. This equation has been used in sodar data processing to correct the value of C_T^2. A comparison between our theory and another theory by Brown and Clifford[16] is presented.

II. THE EXCESS ATTENUATION OF A PLANE WAVE

Using the Kolmogorov spectrum of the locally homogeneous and isotropic turbulence, the effective scattering cross section of the scattered wave is expressed as follows:[8]

$$\sigma(\theta)=0.033k^{1/3}\cos^2\theta\left[\frac{C_V^2}{c^2}\cos^2\frac{\theta}{2}+0.136\frac{C_T^2}{T^2}\right]\times\left(\sin\frac{\theta}{2}\right)^{-11/3}, \tag{1}$$

where $\sigma(\theta)$, in m^2/m^3 or m^{-1}, is the power of the scattered wave per unit flux density of the incident wave per unit scattering volume per unit solid angle, θ is the scattering

angle, k is the wave number, c is the sound speed, and T is the air temperature. C_T^2 in $\mathrm{K}^2\,\mathrm{m}^{-2/3}$ and C_V^2 in $\mathrm{m}^{4/3}\,\mathrm{s}^{-2}$ are defined as

$$C_T^2=\frac{\langle[T(x)-T(x+r)]^2\rangle}{r^{2/3}} \tag{2}$$

and

$$C_V^2=\frac{\langle[V(x)-V(x+r)]^2\rangle}{r^{2/3}}, \tag{3}$$

respectively, where r is the separation distance between two simultaneous measurements of temperature or velocity and $\langle\ \rangle$ indicates an ensemble average, instead of which a time average is used in practice. The scale of effective eddies which form a spatial diffraction grating with the fixed spacing, $l(\theta)$, and scattering angle of scattered wave, θ, satisfy the Bragg condition:

$$l(\theta)=\frac{\lambda}{2\sin\theta/2}, \tag{4}$$

where λ is the wave length. Equation (4) means the smaller the scattering angle θ, the larger the effective eddies, relative to the wavelength.

For back scattered waves, with $\theta=\pi$, Eq. (1) yields

$$\sigma(\pi)=0.0045k^{1/3}\frac{C_T^2}{T^2}. \tag{5}$$

At this point we need to mention that in the deduction of Eqs. (1) and (5), the acoustic wave is assumed to be a plane wave and the mean velocity of the media is zero. Equations (1) and (5) hold true only under the condition of $l_0 \ll l(\theta) \ll L_0$, where l_0 is called the inner scale of turbulence and L_0 is the outer scale of turbulence. In other words, Eqs. (1) and (5) are suitable only for the inertial subrange of turbulence.

To calculate the excess attenuation coefficient of sound intensity, which is the total scattering power removed from the incident plane wave per unit distance, we simply integrate $\sigma(\theta)$ over all angles on a spherical surface as Lighthill[5] did, that is,

$$a_e=\int_0^{2\pi}\int_0^{\pi}\sigma(\theta)\sin\theta\cdot d\theta\cdot d\varphi. \tag{6}$$

However, Eqs. (6) and (1) can not be used to estimate the excess attenuation because the integral in Eq. (6) is divergent. Considering the outer scale of turbulence, L_0 as a limit value to define $\theta_{\min}$, we have

$$\theta_{\min}=2\arcsin\left(\frac{\lambda}{2L_0}\right). \tag{7}$$

Equation (7) shows that the acoustic wave will not be scattered at a scattering angle $\theta<\theta_{\min}$. Equation (6) becomes

$$a_e=\int_0^{2\pi}\int_{\theta_{\min}}^{\pi}\sigma(\theta)\sin\theta\cdot d\theta\cdot d\varphi. \tag{8}$$

Using Eq. (7) and integrating Eq. (8) yields

$$a_e=0.828k^{1/3}\frac{C_V^2}{c^2}(15l_*^{1/3}+0.6l_*^{-5/3}-13.09)$$

$$+0.113k^{1/3}\frac{C_T^2}{T^2}(12l_*^{1/3}+0.6l_*^{-5/3}-10.89), \tag{9}$$

where $l_*=\lambda/2L_0$. Considering $l_*<1$, the smaller terms for this integration, $-\frac{24}{7}l_*^{7/3}+\frac{12}{13}l_*^{13/3}$ and $-\frac{12}{7}l_*^{7/3}$ in brackets, have been neglected. Weber *et al.*[20] suggested the outer scale of turbulence, L_0, is

$$L_0=2\pi U\int_0^{\infty}R(\tau)d\tau, \tag{10}$$

where U is the average wind speed and $R(\tau)$ is the autocorrelation coefficient of wind speed. Theoretically we have $R(\tau)\to 0$, when $\tau\to\infty$. But, in reality, when τ is large enough, $R(\tau)$ will not be zero and takes different values around zero. Whether we take the average of L_0 or not, the L_0 may not have a clear relevance. Another problem is that when $l(\theta)\to L_0$, the scales of turbulence are out of the inertial subrange. Then Eq. (1) does not hold true. In this case using Eqs. (1), (7) and (9) to calculate the excess attenuation will cause a large overestimation.

III. THE EXCESS ATTENUATION OF AN ACOUSTIC BEAM

In acoustic remote sensing and for most acoustic measurements, the transmitted energy is concentrated into a beam. The acoustic beam is then quite different from an ideal plane wave. According to the concept of turbulence theory, the field of random fluctuations of any parameter is composed of numerous sinusoid components which sometime are called eddies. We can infer that for a certain scale of component, the parameter is relatively uniform in an area with a scale much smaller than the scale of the component. The scattering of sound is caused by the inhomogeneity of refractivity. Eddies, which are larger than the cross section of acoustic beam, must produce a very small contribution to a scattered wave and the eddies which are smaller than the cross section play the most important role in acoustic scattering. So only the eddies with effective scales, which are smaller than the beam diameter, scatter the sound in an acoustic beam. We will follow this concept to solve the scattering of an acoustic beam. Another problem to be considered is the frequency dependence of the excess attenuation.

A. Three modes of the acoustic beam

For the simplest mode, we assume that the acoustic beam is a cylinder and the distribution of energy inside the beam cross section is uniform. We assume this mode suits the estimation of excess attenuation of the scattered wave at a certain direction and the receiver is far away from the scattering source. We name it as mode A. According to the theory of the sinusoidal diffraction grating, the direction of diffraction depends on the wave length and the distance of two adjacent slits. We assume that the minimum number of slits for a grating is two. Therefore the eddies, whose scale is larger than half the diameter D of the cylinder beam, can not

form a spatial diffraction grating and contribute no scattered energy. Pan[21] defined an effective turbulent outer scale $L_e = D/2$ as the largest scale of eddies which scatter the acoustic energy. If L_e is not too large, Eqs. (1) and (8) with $\theta_{\min} = 2\arcsin(1/2)$ meet the condition of $l_0 \ll l(\theta) \ll L_0$ and Eq. (9), with $l_* = \lambda/2L_e$, can be used in the estimation of excess attenuation. If $l_* < 0.05$ and using $l_* = \lambda/2L_e$, Eq. (9) reduces to

$$a_e = \frac{L_e^{5/3}}{\lambda^2}\left(2.9\frac{C_V^2}{c^2} + 0.40\frac{C_T^2}{T^2}\right). \tag{11}$$

Considering that the values of C_V^2 are generally larger than that of C_T^2 by about one to two orders in the atmosphere, the first item in Eq. (11) is larger than the second item by about one to three orders. Even if C_T^2 equals to C_V^2, the former is still larger than the latter about 4.5 to 5.5 times. So the inhomogeneity of velocity is the main factor in excess attenuation. Equation (11) shows that the excess attenuation coefficient is sensitive to L_e. The larger the cross section of the cylinder beam, the larger is the excess attenuation.

The diameter of beam cross section increases with its distance to the source for a conical beam. The effective turbulent outer scale is then written as

$$L_e = R\tan\left(\frac{\gamma}{2}\right), \tag{12}$$

where R is the distance from the source to scattering volume and γ is the angular beam width. Thus, the excess attenuation coefficient α_e increases with the distance even in a homogeneous turbulent field. The effective turbulent outer scale L_e in Eq. (12) has a greater weighting for the turbulence farther away from the transmitter. The larger the beam width, the larger is the weighting. The excess attenuation of a beam type acoustic wave is not reversible during forward propagating in the turbulent atmosphere. In other words, the reversibility of the excess attenuation does not exist between the transmitter and the receiver at the two ends of a path. This mode suits a directional source and a receiver at the far field. We name it as mode B.

In sodar sounding, the scattering volume receives the transmitting energy and scatters, as a second source, a part of the incident energy. For the transmitting wave, the excess attenuation from antenna to the scattering volume can be estimated by using mode B with a small modification. This modification is that the $\theta_{\min}$ in Eqs. (7) and (8) is taken to be half of the angular beam width. This means that the part of the scattered wave, which has a scattering angle less than half of the beam width, joins with the transmitting waves and these become the incident energy at the next scattering volume. So in the sodar mode, L_e, instead of L_0 in Eq. (7), is always smaller than the value in Eq. (12) and becomes a constant which depends on the antenna and sound frequency and loses its weighting effect on turbulence. For the scattered wave, mode A is assumed to be used in both monostatic and bistatic sodars. The weighting effect of L_e on turbulence is not on the path but on the scale of the scattering volume. Here we name this mixed mode as mode C.

B. The frequency dependence of the excess attenuation

Lighthill[5] presented the frequency dependence of excess attenuation as $a_e \propto \lambda^{-2}$. But this frequency dependence has never been verified by experiments. What kind frequency dependence of excess attenuation could it be? Let us define a nondimensional outer turbulence scale, L_e^*:

$$L_e^* = \frac{L_e}{R} = \tan\left(\frac{\gamma}{2}\right). \tag{13}$$

Equation (11) then can be rewritten as

$$a_e = L_e^{*5/3} f^2 \frac{R^{5/3}}{c^2}\left(2.9\frac{C_V^2}{c^2} + 0.40\frac{C_T^2}{T^2}\right). \tag{14}$$

The L_e^* only depends on the acoustic beam width, which is related to the sound frequency and the antenna configuration. Generally speaking, the higher the frequency, the narrower is the beam. So the decreasing of L_e^*, as the frequency gets higher, causes less excess attenuation. Finally, the frequency dependence should be smaller than quadratic.

If the Fraunhofer diffraction is the only factor that causes the beam spreading, then the relation[16]

$$D_f^2 = D_0^2 + 16\frac{R^2}{k^2 D_0^2} \tag{15}$$

yields the half beam width, expressed as follows,

$$\tan\frac{\gamma}{2} = \frac{2}{kD_0}\sqrt{\frac{D_f - D_0}{D_f + D_0}}, \tag{16}$$

where D_0 is the diameter of the source and D_f is the diameter that the far field beam would have at the propagation distance R. If $D_f \gg D_0$, Eq. (16) is simplified to

$$\tan\frac{\gamma}{2} = \frac{2}{kD_0}. \tag{17}$$

Applying Eq. (13) and inserting Eq. (17) into Eq. (14) yields

$$a_e = f^{1/3}c^{-1/3}\left(\frac{R}{\pi D_0}\right)^{5/3}\left(2.9\frac{C_V^2}{c^2} + 0.40\frac{C_T^2}{T^2}\right), \tag{18}$$

where D_0 is also the diameter of the source antenna. Equation (18) shows a cube-root frequency dependence. Neff[9] measured the antenna-gain patterns at 1000, 2000, and 3000 Hz; the beam widths at -3 dB can be read from his antenna-gain patterns to be about 16°, 8°, and 4°, respectively. Fitting these data, we have

$$L_e^{*5/3} f^2 = A \times f^{-0.066}, \tag{19}$$

where A is a constant. The beam width at 3000 Hz may have an error. If we take off this point, then the result of fitting becomes

$$L_e^{*5/3} f^2 = A \times f^{0.322}, \tag{20}$$

which is consistent with Eq. (18). This cube-root frequency dependence of excess attenuation needs a verification by experiment.

IV. THE EVALUATION OF THE EXCESS ATTENUATION OF AN ACOUSTIC BEAM IN THE ATMOSPHERE

DeLoach[22] concluded in his survey that the instantaneous value of the excess-attenuation coefficient undergoes rapid fluctuations over a wide range of magnitude. During the observation period of the atmospheric refractive index,[23] which was carried out at Xinglong Observatory during October–December 1994, we did frequently find that the values of C_T^2 measured over a 3.5-min average by a thermometer with a pair of W–Re filament sensors separated by a distance of 40 cm, varied by over one to three orders of magnitude during a whole night. We also found that the instantaneous C_T^2, measured using a pair of W filament sensors separated by 1 m, varied in the range of two orders of magnitude during minutes at night under a quiescent wind condition. The fluctuation of vertical wind speed was often within 1.5–3 m/s over 3-min periods under convective conditions. Considering the rapid fluctuation of turbulent parameters, taking an average over 10–20 min for the excess attenuation coefficient may be suitable.

Direct measurements of C_V^2 and C_T^2 can hardly be considered a routine task. Sodar, as remote sensing equipment, can measure the wind profile, turbulent parameters and the echo intensity. Using the data for C_V^2 and C_T^2 collected by sodar, the excess attenuation coefficient can be evaluated. There are several approaches to determine C_V^2 and C_T^2. The first one is based on Eqs. (1) and (5). The echo intensities in the monostatic mode and bistatic mode of sodar can be measured at the same time. Joining these two modes we can separate C_T^2 and C_V^2 from each other.[12] The second approach is to determine C_T^2 with the monostatic mode and C_V^2 from the vertical velocity spectra. Under the condition of a Kolmogorov local homogeneous isotropic random field, the one-dimensional spectrum of energy can be deduced from the structure function:[24]

$$E(\kappa)=\frac{1}{2\pi}\Gamma\left(\frac{5}{3}\right)\sin\left(\frac{\pi}{3}\right)C_V^2\kappa^{-5/3}, \tag{21}$$

where κ is the independent variable for the spectral function of the velocity field. But this requires a fast sampling rate to cover enough length of inertial subrange in the velocity spectrum. For most sodars it is impossible to do so. The third method is presented by Pan.[21] The variance of vertical velocity spectrum of sodar echo, σ_V^2, is related to the turbulent energy in the scattering volume:[21]

$$\langle V'^2\rangle=\sigma_V^2, \tag{22}$$

where $V'=V-\langle V\rangle$ and the $\langle\ \rangle$ means averaging the velocity fluctuation in the scattering volume. The one-dimensional spectrum can also be expressed by a von Karman spectrum as[24]

$$E(\kappa)=\frac{\Gamma(\frac{5}{6})L_0\overline{V'^2}}{3\pi\Gamma(\frac{1}{3})(1+\kappa^2L_0^2)^{5/6}}, \tag{23}$$

where $\overline{V'^2}$ is the mean square of the velocity fluctuation in the locally homogeneous and isotropic random field. L_0 and $\overline{V'^2}$ are effective parameters for the locally homogeneous and isotropic turbulent field but not suitable for a sodar beam. To match the sodar situation, we take $\langle V'^2\rangle$ and L_e instead of the $\overline{V'^2}$ and L_0 in Eq. (23), to yield an approximation equation,

$$E(\kappa)=\frac{\Gamma(\frac{5}{6})L_e\langle V'^2\rangle}{3\pi\Gamma(\frac{1}{3})(1+\kappa^2L_e^2)^{5/6}}. \tag{24}$$

The combination of Eq. (21) with Eq. (24) with $\kappa=1$ yields

$$C_V^2=0.636L_e^{-2/3}\langle V'^2\rangle. \tag{25}$$

C_V^2 can be thus obtained from Eqs. (22) and (25).

V. CASE STUDIES AND INDIRECT VERIFICATION OF THEORIES

It is hard to find the profiles of C_T^2 and C_V^2 measured by sodar in published papers. The data from Moulsley *et al.*[25] is taken to be a case study here. Moulsley *et al.* got the C_V^2 profiles by using a bistatic sodar and the C_T^2 profiles by a monostatic sodar. Moulsley calculated the atmospheric absorption attenuation but did not consider the excess attenuation in data processing. We also use our data from an experiment for astronomical site selection to test the theories.

A. The sodar equation

The sodar equation of monostatic mode can be written as follows,

$$\begin{aligned}P_r&=P_t\frac{c\tau AG}{2R^2}\sigma(\pi)T_aT_e,\\&=c_s\frac{1}{R^2}C_T^2T_aT_e,\end{aligned} \tag{26}$$

where P_r is the received sound power in $\mathrm{W\,m^{-2}}$, P_t is the transmitting sound power, τ is the transmitting pulse length in seconds, G is the antenna gain, A is the antenna aperture in $\mathrm{m^2}$, R is the range from antenna to scattering volume, c_s is a constant with the units $\mathrm{W\,m^{2/3}\,K^{-2}}$ composed of parameters for the sodar system, T_a is the transmissivity of sound intensity for the atmospheric absorption defined by

$$T_a=\exp\left(-2\int_0^R\alpha_I dr\right), \tag{27}$$

where α_I is the atmospheric absorption coefficient of sound intensity in $\mathrm{m^{-1}}$ and has a relation to atmospheric absorption coefficient of sound pressure, α in Nepers $\mathrm{m^{-1}}$, as follows $\alpha_I=2\alpha$, and T_e is the transmissivity for the excess attenuation defined by

$$T_e=\exp\left(-\int_0^R a_e dr-\int_R^0 a_e' dr\right), \tag{28}$$

where a_e is the excess attenuation coefficient for the transmitting wave in $\mathrm{m^{-1}}$ and a_e' is the excess attenuation coefficient for the back scattered wave. For both Eqs. (27) and (28), the integrals allow for variation of the attenuation coefficients along the dual propagation paths, 0-R and R-0.

If attenuation factors have not been considered in data processing, the first retrieved value of C_T^2 can be written as

TABLE I. Transmisitivity of excess attenuation calculated with different methods. Frequency=2048 Hz, beam width=9° and $T=12\,^{\circ}\mathrm{C}$. The antenna is a 1 m×1 m horn array (from Ref. 25). a_e is the excess attenuation coefficient of forward propagation, a_e' represents the excess attenuation coefficient of back scattered wave from 242 m to the antenna. T_e is the transmissivity of excess attenuation from source to target level and T_e' is that from scattering volume back to the antenna. (P), (B&C) and (L_0) represent the theory in this paper, the Brown and Clifford theory, and the plane wave method, respectively.

Height (m)	$a_e(\mathrm{P})(\mathrm{m}^{-1})$	$a_e'(\mathrm{P})(\mathrm{m}^{-1})$	$T_e(\mathrm{P})$	$T_e'(\mathrm{P})$	$T_e(\mathrm{B\&C})$	$T_e'(\mathrm{B\&C})$	$T_e(L_0)$
46	1.00×10^{-4}	4.11×10^{-3}	9.954×10^{-1}	9.86×10^{-1}	9.9994×10^{-1}	9.9796×10^{-1}	6.10×10^{-1}
60	8.08×10^{-5}	3.31×10^{-3}	9.943×10^{-1}	9.74×10^{-1}	9.9992×10^{-1}	9.9653×10^{-1}	3.74×10^{-1}
67	7.11×10^{-5}	2.92×10^{-3}	9.938×10^{-1}	9.67×10^{-1}	9.9991×10^{-1}	9.9576×10^{-1}	2.80×10^{-1}
81	6.78×10^{-5}	2.78×10^{-3}	9.928×10^{-1}	9.50×10^{-1}	9.9989×10^{-1}	9.9390×10^{-1}	1.37×10^{-1}
95	6.45×10^{-5}	2.65×10^{-3}	9.919×10^{-1}	9.29×10^{-1}	9.9987×10^{-1}	9.9154×10^{-1}	5.50×10^{-2}
109	6.77×10^{-5}	2.78×10^{-3}	9.910×10^{-1}	9.02×10^{-1}	9.9985×10^{-1}	9.8803×10^{-1}	1.62×10^{-2}
137	7.09×10^{-5}	2.91×10^{-3}	9.890×10^{-1}	8.35×10^{-1}	9.9982×10^{-1}	9.7674×10^{-1}	4.94×10^{-4}
193	6.12×10^{-5}	2.51×10^{-3}	9.857×10^{-1}	6.63×10^{-1}	9.9976×10^{-1}	9.4745×10^{-1}	2.65×10^{-8}
242	4.83×10^{-5}	1.98×10^{-3}	9.833×10^{-1}	5.02×10^{-1}	9.9971×10^{-1}	9.2173×10^{-1}	5.59×10^{-13}

$$C_{Ts}^2=\frac{P_r R^2}{c_s}. \tag{29}$$

C_{Ts}^2 can be used to estimate the excess attenuation.

B. Estimation of excess attenuation with sodar data

The C_T^2 and C_V^2 profiles taken from Figs. 7 and 8 of Ref. 25, which were measured under convective condition at run 2a and 2b, are used to estimate the excess attenuation. The results of calculation for these two runs are similar to each other, so only the results of run 2a are shown in tables. For comparison, three methods are used in estimation.

The first method is based on the theory presented in this paper and mode C is used in calculation.

The second method is based on Brown and Clifford's theory.[16] According to their theory, the transmissivity of the excess attenuation is expressed as follows:

$$T_e=\left\{1+1.56k^{5/12}D_0^2\left[\int_0^R\left(\frac{R-r}{R}\right)C_n^2(r)dr\right]^{6/5}\right\}^{-1}, \tag{30}$$

where R is the distance from the source, D_0 is the diameter of the source, and C_n^2 is the refractivity structure coefficient. C_n^2 depends on C_T^2 and C_V^2 according to the relation[24]

$$C_n^2=\frac{C_T^2}{4T^2}+\frac{C_V^2}{C^2}. \tag{31}$$

If $C_n^2(r)$ is constant, then Eq. (30) reduces to[16]

$$T_e=(1+0.48k^{12/5}D_0^2R^{6/5}C_n^{12/5})^{-1}. \tag{32}$$

When calculating the transmissivity with sodar data, D_0 is taken to be the diameter of antenna or the diameter of the cross section of the scatter source. We assume C_n^2 to be constant within the gate spacing of sodar. According to their theory, the transmissivity does not directly relate to the beam width of antenna which is measured in calibration of sodar equipment, but to that implied by the Fraunhofer diffraction, which is expressed by Eq. (16) or (17). Thus, the diameter of the cross section of the scatter source is $D_f=4R/kD_0$.

The third method is the plane wave method with $L_0=z$ and Eqs. (7) and (8) are used in the calculation.

Table I shows the results of the calculation using these three methods. The raw data, C_{Ts}^2 and C_V^2, are listed in the left two columns in Table II. It is interesting to note that the values of excess attenuation coefficient for back scattered waves are much larger than that for the transmitting wave at the same path by about one to two orders of magnitude depending on the path length. The transmissivities calculated with the methods in this paper and Brown and Clifford's methods, $T(\mathrm{P})$ and $T(\mathrm{B\&C})$, respectively, have a similar variation with height but a large difference for the back scattered wave. The ratio of $[T_e(\mathrm{B\&C})-T_e(\mathrm{P})]/T_e(\mathrm{P})$ (forward-propagation) varies from 0.5% to 1.7% with the increase of height and the ratio of $[T_e'(\mathrm{B\&C})-T_e'(\mathrm{P})]/T_e'(\mathrm{P})$ (back scattering propagation) varies from 1.2% to 83.6%. This variation reflects the difference of these two theories. In the former, the factor $L_e=R\tan(\gamma/2)$ in Eq. (11) weights the turbulence farther away from the transmitter and, in the latter, the factor $((R-r)/R)$ in Eq. (30) always weight the turbulence near the transmitter most heavily.[16] It

TABLE II. The correction of C_T^2 profiles with the excess attenuation factor. D_f is the diameter of scattering volume deduced from Eq. (15). Frequency=2048 Hz, beam width=9° and $T=12\,^{\circ}\mathrm{C}$. Time: 1544–1640 (GMT), 21 June 1977 (from Ref. 25).

Height (m)	C_V^2	C_{Ts}^2	$C_T^2(\mathrm{P})$	L_e (m)	$C_T^2(\mathrm{B\&C})$	D_f (m)	$C_T^2(L_0)$
46	3.1×10^{-2}	6.31×10^{-4}	6.43×10^{-4}	4.1	6.32×10^{-4}	5.8	1.03×10^{-3}
60	2.5×10^{-2}	6.0×10^{-4}	6.19×10^{-4}	5.2	6.02×10^{-4}	7.3	1.60×10^{-3}
67	2.2×10^{-2}	5.0×10^{-4}	5.20×10^{-4}	5.8	5.02×10^{-4}	8.1	1.79×10^{-3}
81	2.1×10^{-2}	4.0×10^{-4}	4.24×10^{-4}	6.9	4.03×10^{-4}	9.5	2.93×10^{-3}
95	2.0×10^{-2}	3.0×10^{-4}	3.26×10^{-4}	8.0	3.03×10^{-4}	11.0	5.45×10^{-3}
109	2.1×10^{-2}	2.21×10^{-4}	2.47×10^{-4}	9.1	2.24×10^{-4}	12.5	1.36×10^{-2}
137	2.2×10^{-2}	1.51×10^{-4}	1.83×10^{-4}	11.3	1.55×10^{-4}	15.4	3.05×10^{-1}
193	1.9×10^{-2}	8.0×10^{-5}	1.22×10^{-4}	15.7	8.5×10^{-5}	21.3	3.02×10^{-7}
242	1.5×10^{-2}	5.0×10^{-5}	1.01×10^{-4}	19.5	5.4×10^{-5}	26.5	8.94×10^{11}

TABLE III. The comparison between the beam widths predicted by Fraunhofer diffraction and the measured values.

Authors	f (Hz)	D_0 (m)	$\gamma_{\text{predicted}}$	$\gamma_{\text{measured}}(^\circ)$
Moulsley[a]	2048	1	6	9
Neff[b]	1000	1.5	8.3	16
Neff	2000	1.5	4.1	8
Neff	3000	1.5	2.8	4
Ito[c]	2400	~1	5	10
Fan[d]	1615	1.7	4.5	9

[a]See Ref. 25.
[b]See Ref. 9.
[c]See Ref. 31.
[d]See Ref. 32.
Note: The temperature is assumed to be 15 °C except that of Moulsley, which uses 12 °C.

is apparent that the longer the propagation distance, the larger is the difference between these two estimates. Brown and Clifford's theory seems to underestimate the excess attenuation. We attribute this partly to the underestimation of beam width by Fraunhofer diffraction [see Eq. (17)]. Table III lists the measured beam widths by different authors and the predicted widths by Eq. (17). The predicted beam width is about 50% to 70% of the measured beam width. The essential difference between the scattering theory of optics and that of acoustics which is deduced from the equations of fluid dynamics is outside of the coverage of this paper.

The plane wave method seriously overestimates the excess attenuation.

According to the value of $T_e(\mathrm{P})$ in Table I we can see that the excess attenuation for the first 100-m path with modified mode B is about 1% of the incident energy. According to the value of $T'_e(\mathrm{P})$, the excess attenuation for mode A is about 10% of the incident energy. For comparison with other attenuations, to define the excess attenuation per 100 m may be a practical measure of an excess attenuation coefficient.

C. The indirect verification of the theories

To verify the theories from field experiments directly is not easy, especially to verify the weighting effect. Also it is difficult to find useful published data. Here we present some of indirect verification. The C_T^2 profile in the well-mixed layer follows the relation[26]

$$C_T^2 = 2.67 H_0^{4/3} (T/g)^{2/3} z^{-4/3}, \tag{33}$$

where H_0 is the surface sensible heat flux in K m s^{-1}, T is the temperature in K, g is the acceleration of gravity, and z is the height above the ground. This relationship was observed at first in the surface layer but also found to hold well beyond the surface layer under convective conditions.[27] The height dependence of C_T^2 in Eq. (33) is often referred to as the "$z^{-4/3}$" law. Plotting log (C_T^2) as a function of log (z), the profiles will show a slope of $-\frac{4}{3}$. The surface sensible heat flux, H_0, can then be deduced from the nodal increment of the log(C_T^2) curve, which equals the value of log(C_T^2) when log(z) equals zero. Even though Neff[9] observed with sodar that the C_T^2 profiles in a mixed layer agree with Eq. (33), Haugen and Kaimal[11] claimed that the excess attenuation can

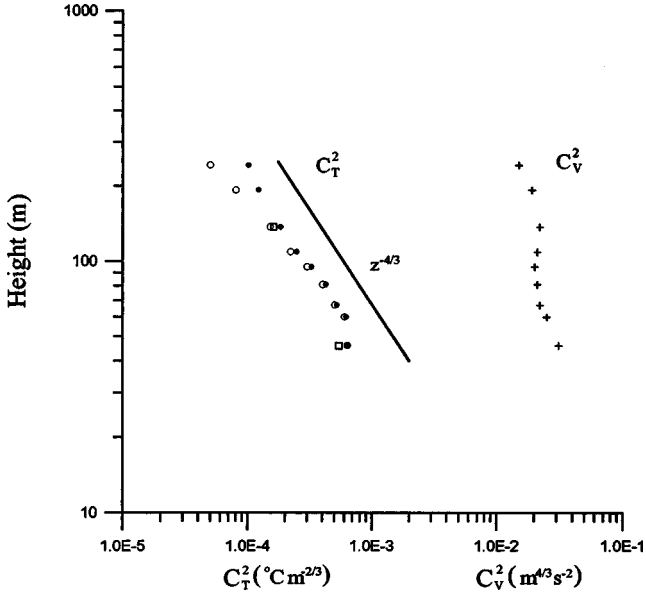


FIG. 1. Comparison between corrected C_T^2 profile and uncorrected profile: ● the corrected value, ○ the uncorrected value, □ the direct measured value, and + the C_V^2 profile.

lead to large errors in C_T^2 estimates and should be included in the sodar equation. Our observations show the C_T^2 profiles in the mixed layer do not often meet the "$z^{-4/3}$" law. The main factors, which cause the slope of the curve log(z) vs log(C_T^2) to be smaller than $-\frac{4}{3}$, are both the atmospheric absorption attenuation of sound and excess attenuation. If the theories are correct, by using sodar equations (26) and (29), the retrieval of C_T^2 profiles makes the profile slopes close to $-\frac{4}{3}$. Here we will ignore the effects of beam bending. Considering the atmospheric absorption has been calculated by Moulsley[25] for C_{Ts}^2, let $T_a = 1$ in Eq. (26). Then combining Eqs. (26) and (29) yields

$$C_T^2 = \frac{C_{Ts}^2}{T_e}. \tag{34}$$

Using Eq. (34) and the transmissivities in Table I, the C_T^2 profiles are retrieved from the raw data. The values of the profiles including the raw profiles and the corrected profiles with different methods are given in Table II and illustrated in Fig. 1. From Table II we can find that the modification magnitude of excess attenuation to C_{Ts}^2, that is $C_T^2(\mathrm{P}) - C_{Ts}^2$, increases with height very rapidly. Usually the excess attenuation can only be ignored in the first 100-m layer above the ground. In this case study, the value of C_{Ts}^2 at 242 m is about half of the value of corrected C_T^2. The C_{Ts}^2 equals one-seventh of the corrected C_T^2 at 210 m above ground in our observation.[21] From Table II we can find that the plane wave method overestimates the excess attenuation too much and cannot be used for an acoustic beam. Brown and Clifford's theory offers only a very small part of the correction. Its modification magnitude to C_{Ts}^2 is less than 13% of the modification magnitude by the theory in this paper.

Moulsley used the signal level corresponding to heights greater than 300 m as a estimate of the noise level. To avoid the low signal-to-noise ratio, the highest point at 242 m is omitted in curve fitting. Also to avoid the influence on C_T^2

TABLE IV. Slopes of profiles and surface heat fluxes deduced from C_T^2 profile.

Profile	Theoretical	C_T^2(P)	C_T^2(B&C)	C_{Ts}^2	$C_T^2(L_0)$
Slope	-1.333[a]	-1.420	-1.695	-1.735	11.43
H_0 (K m s^{-1})	0.028[b]	0.027	0.063	0.072	···

[a]Taking the logarithm of Eq. (33), the theoretical slope of the line $\log(C_T^2)$ $-\log(z)$ in convective condition is $-\frac{4}{3}$.
[b]Directly measured at 46 m. See Table 2 of Ref. 25.

from the local surface, the lowest point at 46 m is omitted in curve fitting. Table IV lists the slopes of the profiles and surface heat fluxes which are deduced from the nodal increment of the C_T^2 profiles in log-C_T^2 and log-height coordinate. The C_T^2 profile corrected with our theory has a slope of -1.420, which is close to -1.333 and indicates a value of surface heat flux close to the direct measured value at 46 m. But the C_T^2 profile corrected with Brown and Clifford's theory has a slope of -1.695, which nearly equals to the slope of C_{Ts}^2 profile and the deviation of a deduced value of surface heat flux from the direct measured value seems too large.

There are no direct measured surface heat flux values in Ref. 25 for comparison with our deduced values. A formula of sensible heat flux in Ref. 18 is used to provide a reference value that is

$$\overline{w'T'} = \beta\sqrt{C_T^2 C_V^2}\, z^{2/3}\frac{z_0 - z}{z_i}, \tag{35}$$

where $\overline{w'T'}$ is the sensible heat flux in K m/s, β is a dimensionless empirical constant, z_0 is the height of the zero-heat-flux layer, z_i is the height of the mixed layer, and z is the height above the surface, both in meters. Here we assume $\beta=1$. Moulsley *et al.*[25] offer us directly measured values of sensible heat flux at two levels that are 0.00054, 0.00016 for C_T^2 and 0.028, 0.019 for C_V^2 at 46 and 137 m, respectively. Using these two sets of values, we can deduce the values of z_0 and z_i in Eq. (35). They are $z_0=168$ m and $z_i=217$ m. The values of z_0 and z_i seem smaller than usual but are reasonable if we consider the echo pattern in Fig. 5 of Ref. 25 and notice that the amount of low clouds is 87% from 1448 to 1640 on that day (Table II of Ref. 25). It was obviously a cloudy day. We can deduce a sensible heat flux profile with Eq. (35). Figure 2 shows the deduced sensible heat flux profile and the surface sensible heat fluxes deduced from the C_T^2 profiles. Comparing with the surface sensible heat flux deduced from the C_{Ts}^2 profile, our estimation of excess attenuation reduces the error of estimate of the surface sensible heat flux because Eq. (33) suits the well-mixed layer but run 2a in Ref. 25 is in a weak convective condition. If the measurements are made under a typical convective condition, the estimation may be better than that in this case.

Another indirect verification is from our "seeing" monitoring experiment. The seeing is a term used in astronomy, which means the sharpness of a telescope image determined by the degree of turbulence in the Earth's atmosphere.[28] Temperature alters the refractive index of air. Physically, the effect of an inhomogeneous refractive index corrugates the incoming planar wave front from a star, and its image does

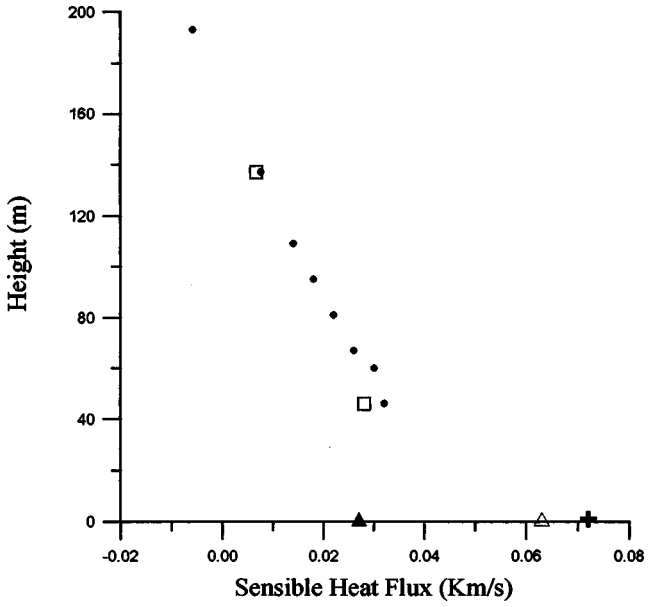


FIG. 2. The profile of sensible heat flux retrieved from C_T^2 and C_V^2 profiles: ● The sensible heat flux profile deduced from Eq. (35), □ the directly measured value of sensible heat flux, ▲ the calculated surface sensible heat flux from the nodal increment of the C_T^2(P) profile in Fig. 1, △ the calculated surface sensible heat flux from the nodal increment of the C_T^2(B&C) profile, and + the calculated surface sensible heat flux from the nodal increment of the C_{Ts}^2 profile.

not focus to a point. The image quality of a star is greatly affected by the integral value of the atmospheric refractive index structure coefficient C_n^2 in m$^{-2/3}$ on the optical path. In fact, the instantaneous images have a random deviation from their average center and the deviation has a Gaussian distribution. To define this, Fried[29] introduced a parameter r_0,

$$r_0 = 0.1847\lambda^{1.2}\left[\sec\gamma_z \int_0^\infty C_n^2(z)\,dz\right]^{-3/5}, \tag{36}$$

which is a measure of the transverse span of an optically coherent image called the transverse coherence. The larger the value of r_0, the sharper the image. The angle γ_z is the zenith distance, which is the angle between the line-of-sight of a star and the line to the point in the heavens directly above an observer, and z is the altitude along the optical path. The transverse coherence is one of the important factors determining the width of the Gaussian distribution of the optical image. Equation (36) shows that the larger the integral of C_n^2, the smaller the r_0 and the worse the seeing. For the optical wave, the water vapor contribution to the atmospheric refractive index structure coefficient can be ignored except on a maritime site or in humid weather, then C_n^2 only depends on C_T^2:

$$C_n^2 = \left(\frac{Ap}{T^2}\right)^2 C_T^2, \tag{37}$$

where $A=78.7\times10^{-6}$ K hPa^{-1}, 1 hPa=100 Pa, and p is the atmospheric pressure in hPa. Then the contribution to the seeing from the lower part of the atmosphere can be written as

$$r_{0p}=15\,534\lambda^{1.2}\left[\int_{z_{\min}}^{z_{\max}}\left(\frac{p}{T^2}\right)^2 C_T^2 dz\right]^{-3/5}, \tag{38}$$

where r_{0p} is named as transverse quasi-coherence and the constant 15 534 has the units $\mathrm{hPa}^{6/5}\,\mathrm{K}^{-6/5}$. Sodar measures the C_T^2 profiles of the lower part of the atmosphere and can thus help to monitor the variation of seeing.

Our experiment was conducted from 29 November to 10 December 1997 at Gaomeigu, 3200 m above sea level, in the Yunnan province of China. The instruments related to this paper include the differential image motion monitor with three subapertures (DIMM for short) which is the essential instrument for measuring seeing, a multi-parameter monitoring system developed by the author, and conventional surface-based observation instruments. The multi-parameter monitoring system includes a sodar, Gill-type anemometer, and W–Re filament thermometer which is composed of two sensors, separated by a distance of 40 cm. The latter two instruments were installed on a 30-m meteorological tower. The sodar antenna is a 1 m×1 m array and the range gate is 10 m. The transmitted acoustic frequency is 2000 Hz. The tower is at the eastside of the DIMM at a distance of only 12 m. The sodar is at the northwest side of the DIMM about 50 m away. The surface around the tower is not flat. Therefore the sodar antenna is 1.3 m below the tower bottom and the objective of the DIMM is about 6.5 m above the base of the tower. The DIMM samples r_0 with a sampling frequency of 50 Hz and derives a 8-s average value of r_0 about every 4 min. The sampling frequencies of the Gill-type anemometer and W–Re filament thermometer both are 5 Hz, measured for 10 min in every 15 min. The multi-parameter monitoring system derived the parameters of wind speed U and temperature T at 25 m, C_T^2 profiles, and r_{0p} over 10-min averages. We had no C_V^2 profile data. Considering that the site, which has an area about 60×60 m^2, is located at the top of a hill and measurements were running at night, a local empirical formula $C_V^2=0.005+0.001U$ was accepted. No variation with height in C_V^2 is assumed in the 0–300-m layer. The atmospheric absorption[30] and excess attenuation both are taken into account in the sodar data processing. The temperature, humidity and pressure, which are necessary for calculating the atmospheric absorption coefficient, are taken from the records of conventional surface-based observation.

Here a case study of one night's data, 0015–0615 8 December 1997, is presented. We take three steps to finish the retrieval of r_{0p} from the C_{Ts}^2 profile. The first step is to let parameters C_T^2, p and T equal C_{Ts}^2 measured by sodar, and the average pressure and temperature at the site, respectively, and use Eq. (38) to calculate the approximate value of r_{0p}, written as $r_{0p}(0)$. The second step is using the corrected values of C_{Ts}^2 for excess attenuation to get the corrected values of $r_{0p}(0)$, written as $r_{0p}(1)$. The third step uses both the corrected values of C_{Ts}^2 for excess attenuation and atmospheric absorption to get the final r_{0p}, written as $r_{0p}(2)$. A comparison between time variations of r_0, which is measured by DIMM, and the $r_{0p}(0)$, $r_{0p}(1)$ and $r_{0p}(2)$ is shown in Fig. 3. The curve $r_{0p}(0)$, which has no correction for excess attenuation and atmospheric absorption, does not appear in Fig. 3 because it is too close to the curve $r_{0p}(1)$, which does have a correction for excess attenuation. Because we do not know the percent contribution to r_0 from the lower layer of the atmosphere, the accuracy of estimation of excess attenuation is difficult to determine. Generally the refractive index structure coefficient is relatively uniform on horizontal surfaces in the upper part of the atmosphere than in the boundary layer, so the time variation of seeing is mainly caused by the boundary layer. The attenuation of sound distorts the C_T^2 profiles measured by sodar and reduces the correlation of time variations between r_0 and r_{0p}. The average values of r_0, r_{0p} and the correlation coefficients between curve r_0 and curves $r_{0p}(0)$, $r_{0p}(1)$, $r_{0p}(2)$ on that night are shown in Table V. The portions of the curves in Fig. 3 for local times from 0:15 to 1:45 and 4:15 to 6:15 have a good correlation between $r_{0p}(0)$, $r_{0p}(1)$, $r_{0p}(2)$ and r_0. The correlation coefficients in the period from 0:15 to 1:45 are 0.688, 0.767 and 0.938, respectively. The correlation coefficient increases after the correction, obviously. The integral value of C_n^2 for the whole atmosphere can be calculated from r_0 by using Eq. (36). An approximate integral value of C_n^2 for the 0–300-m layer can be calculated by using Eqs. (38) and (37). The ratios of integral values of C_n^2 in the 0–300-m layer to that for the whole optical path, $r_{C_n^2}=\int_0^{300\,\mathrm{m}} C_n^2 dz/\int_0^{\infty} C_n^2 dz$, are also listed in Table V. From Table V we can see that the modification magnitude of $r_{C_n^2}$ for excess attenuation is only 1%, but is 8% for atmospheric absorption. Because C_T^2 and C_V^2 decrease rapidly with height, atmospheric absorption attenuation usually plays a more important role than the excess attenuation does in short distance upward propagation. The underestimation of excess attenuation of Brown and Clifford's theory also appears in Table V.

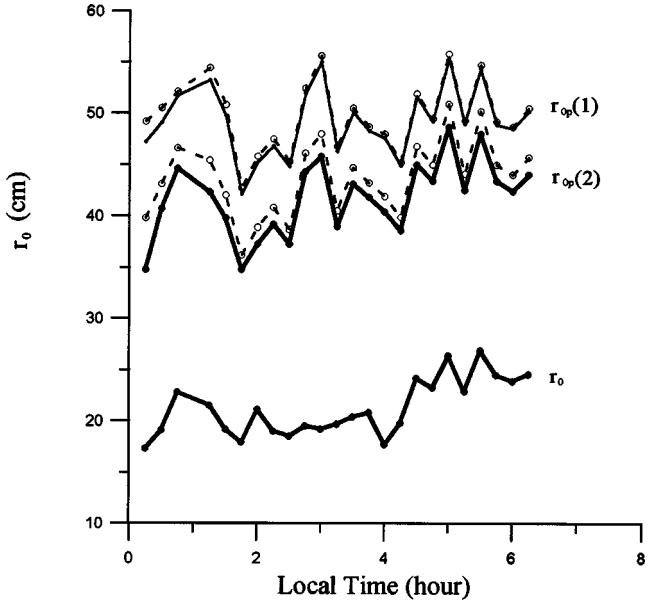


FIG. 3. Comparisons between r_0 and r_{0p} at different steps of retrieval. r_0 measured by DIMM. $r_{0p}(1)$ after the correction for excess attenuation. $r_{0p}(2)$ after the correction for excess attenuation and atmospheric absorption both. Solid lines of r_{0p} are retrieved according to the theory in this paper. Dashed lines are retrieved according to Brown and Clifford's theory, 8 December 1997, Gaomeigu, Yunnan province of China.

TABLE V. Comparison between the average transverse coherence r_0 and the average transverse quasi-coherence r_{0p} measured in one night.[a]

Theory	r_0 (cm)	$r_{0p}(0)$	$r_{0p}(1)$	$r_{0p}(2)$	$r_{C_n^2}(0)$[b]	$r_{C_n^2}(1)$	$r_{C_n^2}(2)$	$R(0)$[c]	$R(1)$	$R(2)$
Pan	21.3	49.9	49.2	41.7	24%	25%	33%	0.495	0.552	0.763
B&C	21.3	49.9	49.8	43.6	24%	24%	30%	0.495	0.494	0.734

[a]For time period of measurement, see Fig. 3. r_0 is measured by DIMM. r_{0p} is measured by sodar and represents the contribution to r_0 from the lowest 300-m layer. (0) represents no correction for attenuations. (1) represents a correction for excess attenuation. (2) represents a correction for both excess and absorption attenuation.
[b]$r_{C_n^2}=\int_0^{300\text{ m}} C_n^2\,dz/\int_0^{\infty} C_n^2\,dz$ calculated by Eqs. (36) and (38).
[c]R is the correlation coefficient between r_0 and r_{0p}.

VI. CONCLUSIONS

Linking the theory for scattering by homogeneous and isotropic turbulence with the propagation of an acoustic beam, we have presented a theory for estimating the excess attenuation of different beam patterns due to scattering by turbulence. The case studies show that the estimates of excess attenuation and its utilization to correct the temperature structure coefficient measured by sodar give reasonable results and reduces the errors of sodar measured C_T^2 substantially.

The excess attenuation is not an intrinsic property of just the atmosphere. It depends not only on the inhomogeneity of the refractive index in the atmosphere but also on the sound beam pattern. Basically, the effect of the beam pattern on the excess attenuation is that the larger scale of the beam cross section covers more larger scattering eddies. For a conical beam, mode B, which is a common mode for acoustic measurements in the atmosphere, the effective turbulent outer scale weights more for the turbulence farther away from the transmitter and causes the nonreversibility of excess attenuation on the propagation path.

Some of the characteristics of the excess attenuation can be summarized as follows.

(1) The excess attenuation coefficient averaged in 5–10 min varies in space with a large range up to two orders of magnitude in the lower atmosphere. It decreases with height rapidly and varies relatively slowly horizontally. Its time variation could sometimes reach two orders of magnitude in a few hours.
(2) The total excess attenuation can be neglected for path lengths less than 100 m but increases rapidly with the propagation distance.
(3) The larger the beam width or cross section, the stronger the excess attenuation in the same turbulent field.
(4) The excess attenuation of acoustic beam has a weak frequency dependence, theoretically a cubic-root frequency dependence.
(5) The turbulence of the velocity field contributes to the excess attenuation much more than that in the temperature field. In most cases, using only C_V^2 to estimate the excess attenuation will not cause a large error.

The estimation of the excess attenuation can be used to reduce the error of retrieved C_T^2 and some parameters, which are deduced from sodar data, such as the sensible heat flux and transverse quasi-coherence in astronomy observations.

ACKNOWLEDGMENTS

The author, as a cooperator, thanks Professor Zhang Bairong and Tan Huisong, Yunnan Observatory China, for their support in the field experiment at Gaomeigu in 1997 which is a part of the work under the guidance of the Expert Group for the Astronomical Site Resources Investigation. The author also thanks Dr. Liu Zizhong, Dr. Qian Tongling and Ma Kaiquan for their cooperation in this experiment. The author is grateful to Dr. H. Y. Weng for her help on my English writing of this paper.

[1] J. Tyndall, "On the atmosphere as a vehicle of sound," Philos. Trans. R. Soc. London, Ser. A **164**, 183–244 (1874).
[2] A. M. Obukhov, "Scattering of sound in turbulent flow," Dokl. Akad. Nauk SSSR **30**, 611 (1941).
[3] D. I. Blokhintsev, "The propagation of sound in an inhomogeneous and moving medium," J. Acoust. Soc. Am. **18**, 322–334 (1946).
[4] R. H. Kraichnan, "The scattering of sound in a turbulent medium," J. Acoust. Soc. Am. **25**, 1096–1109 (1953).
[5] M. J. Lighthill, "On the energy scattered from the interaction of turbulence with sound or shock waves," Proc. Cambridge Philos. Soc. **49**, 531 (1953).
[6] V. I. Tatarskii, *The Effects of the Turbulent Atmosphere on Wave Propagation* (translated from Russian) (Israel Program for Scientific Translations, Chap. 2, 1971).
[7] A. S. Monin, "Characteristics of the scattering of sound in a turbulent atmosphere," Sov. Phys. Acoust. **7**, 370–373 (1962).
[8] C. G. Little, "Acoustic methods for the remote probing of the lower atmosphere," Proc. IEEE **57**, 571–578 (1969).
[9] W. D. Neff, "Quantitative evaluation of acoustic echoes from the planetaty boundary layer," NOAA Technical Report ERL 322-WPL 38 (1975).
[10] D. N. Asimakopoulos, R. S. Cole, S. J. Caughey, and B. A. Crease, "A quantitative comparison between acoustic sounder returns and the direct measurement of atmospheric temperature fluctuations," Boundary-Layer Meteorol. **10**, 137–147 (1976).
[11] D. A. Haugen and J. C. Kaimal, "Measuring temperature structure parameter profiles with an acoustic sounder," J. Appl. Meteorol. **17**, 895–899 (1978).
[12] S. J. Caughey, B. A. Crease, D. N. Asimakopoulos, and R. S. Cole, "Quantitative bistatic acoustic sounding of the atmospheric boundary layer," Q. J. R. Meteorol. Soc. **104**, 147–161 (1978).
[13] D. N. Asimakopoulos, T. J. Moulsley, C. G. Helmis, D. P. Lalas, and J. Gaynor, "Quantitative low-level acoustic sounding and comparison with direct measurements," Boundary-Layer Meteorol. **27**, 1–26 (1983).
[14] M. Y. Zhou, Y. J. Chen, N. P. Lu, and S. M. Li, "A comparative study between the temperature structure coefficients detected by sodar and direct methods," Acta Geophys. Sinica **25**, 492–499 (1982).
[15] T. J. Moulsley, D. N. Asimakopoulos, C. G. Helmis, D. P. Lalas, and J. Gaynor, "A quantitative comparison of horizontal and vertical acoustic sounding with *in-situ* measurements," Boundary-Layer Meteorol. **33**, 85–100 (1985).
[16] E. H. Brown and S. F. Clifford, "On the attenuation of sound by turbulence," J. Acoust. Soc. Am. **60**, 788–794 (1976).
[17] S. F. Clifford and E. H. Brown, "Excess attenuation in echosonde signals," J. Acoust. Soc. Am. **67**, 1967–1973 (1980).

[18] N. X. Pan, ''Expressions of sensible heat flux based on a dimensional analysis,'' JTECH **19**, 1163–1169 (2002).
[19] R. L. Coulter and M. L. Wesely, ''Estimates of surface heat flux from sodar and laser scintillation measurements in the unstable boundary layer,'' J. Appl. Meteorol. **19**, 1209–1222 (1980).
[20] A. H. Weber, J. S. Irwin, J. J. Mathis, and J. R. Kahler, ''Spectral scales in the atmospheric boundary layer,'' J. Appl. Meteorol. **21**, 1622–1632 (1982).
[21] N. X. Pan, ''Determination of the turbulent structure parameter,'' in *Acoustic Remote Sensing Applications*, edited by S. P. Singal (Narosa, 1997), pp. 179–190.
[22] R. DeLoach, ''On the excess attenuation of sound in the atmosphere,'' NASA TN D-7823 (1975).
[23] The co-operators in this observation program are Anhui Institute of Optics and Fine Mechanics, Beijing Astronomical Observatory and Peking University.
[24] V. I. Tatarskii, *Wave Propagation in a Turbulent Medium* (translated from Russia) (Dover, New York, 1961).
[25] T. J. Moulsley, D. N. Asimakopoulos, R. S. Cole, B. A. Crease, and S. J. Caughey, ''Measurement of boundary layer structure parameter profiles by acoustic sounding and comparison with direct measurements,'' Q. J. R. Meteorol. Soc. **107**, 203–230 (1981).
[26] J. C. Wyngaard, Y. Izumi, and S. A. Collins, Jr., ''Behavior of the refractive-index structure parameter near the ground,'' J. Opt. Soc. Am. **61**, 1646–1650 (1971).
[27] L. R. Tsvang, ''Microstructure of temperature fields in the free atmosphere,'' Radio Sci. **4**, 1175–1177 (1969).
[28] ISO 9613-1, 1993, Acoustics—Attenuation of sound during propagation outdoors—Part 1: Calculation of the attenuation of sound by atmospheric absorption.''
[29] *Encyclopedia of Astron. & Astrophys*, Vol. 3, editor-in-chief, P. Murdin.
[30] D. L. Fried, ''Optical resolution through a randomly inhomogeneous medium for very long and very exposures,'' J. Opt. Soc. Am. **56**, 1372–1379 (1966).
[31] Y. Ito, ''Design of a tri-monostatic Doppler sodar system,'' in *Acoustic Remote Sensing Applications*, edited by S. P. Singal (Narosa, 1997), pp. 85–104.
[32] X. Fan, ''Model HK-11 Doppler sodar—its principle and application,'' in *Proceedings of the Fifth International Symposium on acoustic remote sensing of the atmosphere and oceans*, 1990, pp. 131–134.

On the sound field from a moving source in a viscous medium

Michael J. Buckingham[a)]
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0238

(Received 23 February 2003; accepted for publication 15 September 2003)

Based on a one-dimensional model, a wave-theoretic analysis of the sound field from a moving, harmonic source in a viscous medium is developed. When the source is inbound to the receiver, the field consists of an attenuated propagating wave with Döppler up-shifted frequency. A similar wave is present when the source is outbound from the receiver but with down-shifted frequency. Also present on the outbound run is an evanescent wave that appears at the instant the source passes the receiver. The evanescent wave is very highly attenuated and exists only in the presence of both source motion and dissipation. Expressions are formulated for the frequency and attenuation coefficient of the two propagating waves and the evanescent wave. The attenuation of the propagating waves scales with the square of the frequency, which is a characteristic of viscous dissipation; and the attenuation is strongly asymmetrical, being significantly higher on approach than departure. The asymmetry in the attenuation arises partly from the upward and downward Döppler shifts in the frequency on approach and departure, respectively. In addition, the attenuation is skewed by the factor $(1\pm\beta)^{-1}$, where the lower (upper) sign applies on approach (departure) and β is the Mach number of the source. At a Mach number of 0.85, the ratio of the inbound to outbound attenuation is 2000. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1624068]

PACS numbers: 43.28.Py [GG] Pages: 3112–3118

I. INTRODUCTION

Sound from a moving source is usually discussed on the implicit assumption that the medium supporting the acoustic field is nondispersive and nondissipative.[1–5] In such a situation, the familiar Döppler effect is observed, taking the form of an up-shift or down-shift in received frequency, according to whether the source is inbound to or outbound from the receiver.

The Döppler frequency shift has been exploited recently in a series of underwater acoustics experiments conducted off the coast of La Jolla, southern California[6,7] in which a propeller-driven light aircraft was used as an acoustic source for measuring the speed of sound in a marine sediment. The acoustic coupling across the air–sea interface and the ocean–sediment interface is sufficiently good for the sound of the aircraft's propeller, which takes the form of a series of harmonics spanning the frequency range from about 80 Hz to 1 kHz, to be detectable not only on hydrophones in the water column but also on a receiver buried approximately 75 cm deep in the sediment. The technique that has been developed exploits the difference in Döppler-shifted frequencies on aircraft approach and departure. This Döppler difference frequency is inversely proportional to the speed of sound in the local medium where the receiver is located, that is, the sediment in the case of a buried sensor. It has been found that the difference-frequency technique yields the low-frequency sediment sound speed with reasonable precision.[6,7]

Like the sound speed, the acoustic attenuation in a marine sediment is also of considerable interest, although an entirely satisfactory technique for extracting it from propeller-noise data has not yet been developed. The challenge of determining the attenuation in the sediment using aircraft noise involves several questions concerning the effects of dissipation on the sound field from a moving source. Before dealing with the complexities of the atmosphere-ocean-sediment environment, however, a much simpler problem is addressed in this article, with a view to establishing some of the fundamental properties of waves from a moving sound source in a dissipative medium.

A simple one-dimensional scenario is considered (Fig. 1) in which a moving, nonaccelerating, harmonic acoustic source and a stationary receiver are located in a homogeneous, isotropic viscous fluid. The source is assumed to follow a rectilinear path passing through the receiver. A full wave-theoretic analysis of the received field is developed, which leads to a solution involving three waveforms. On the inbound run, a propagating Döppler up-shifted wave appears at the receiver showing an attenuation that is greater than

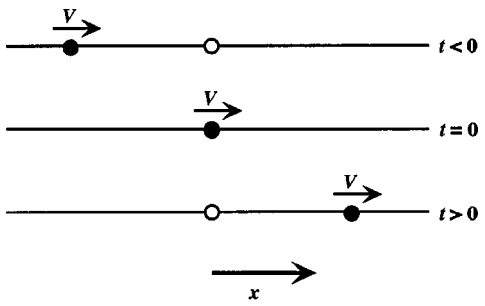


FIG. 1. Geometry of the one-dimensional moving-source (solid circle, speed V) and stationary-receiver (open circle) problem. For negative times the source is inbound to the receiver, they are coincident at time zero, and the source is outbound from the receiver for positive times.

[a)]Also affiliated to the: Institute of Sound and Vibration Research, The University, Southampton SO17 1BJ, England. Electronic mail: mjb@mpl.ucsd.edu

would be observed in the absence of source motion. Outbound, a propagating Döppler down-shifted wave appears, but with attenuation less than that seen when the source is stationary. The attenuation, of course, scales with the viscosity of the medium, and the asymmetry in attenuation, inbound versus outbound, is a direct result of the motion of the source. In addition to the two propagating waves, an evanescent wave also exists, appearing at the instant the source passes the stationary receiver. The evanescent wave is extremely short-lived and appears only when source motion and viscosity are both present. Expressions are given for the wave speeds and attenuations of all three waves.

II. THE WAVE EQUATION

If ϕ is the velocity potential of the acoustic field, Q is the source strength, and x is the direction of the line connecting the source and receiver, then the wave equation describing the one-dimensional field in the viscous medium[8] is

$$\frac{\partial^2\phi}{\partial x^2}-\frac{1}{c_0^2}\frac{\partial^2\phi}{\partial t^2}+\gamma\frac{\partial^3\phi}{\partial x^2\partial t}=-Q\,\delta(x-Vt)f(t), \tag{1}$$

where c_0 is the speed of sound in the medium in the low-frequency limit, $\delta(...)$ is the Dirac delta function, which accounts for the source motion, $f(t)$ is the time series of the source, and $V\geqslant 0$ is the speed of the source. According to Eq. (1), the source, located at $x=Vt$, is coincident with a receiver at $x=0$ when $t=0$. As shown in Fig. 1, the source is inbound to the receiver when $t<0$ and outbound when $t>0$. The third term on the left of Eq. (1) represents the effects of viscous dissipation and γ is the dissipation coefficient, which scales with the dynamic viscosity, μ:

$$\gamma=\frac{4\mu}{3\rho c_0^2}, \tag{2}$$

where ρ is the density of the medium.

To solve for the field, ϕ, two Fourier transforms are applied to Eq. (1), one with respect to time, t, and the second to distance, x. The Fourier transform and its inverse are, respectively,

$$\phi_q=\int_{-\infty}^{\infty}\phi e^{-iqz}\,dz \tag{3a}$$

and

$$\phi=\frac{1}{2\pi}\int_{-\infty}^{\infty}\phi_q e^{iqz}\,dq, \tag{3b}$$

where $i=\sqrt{-1}$ and the transform variable appearing as a subscript denotes the transformed field, which is a convenient notation when several integral transforms are applied in sequence to a partial differential equation, as is the case here.

On Fourier transforming with respect to x, Eq. (1) reduces to

$$-p^2\left(1+\gamma\frac{\partial}{\partial t}\right)\phi_p-\frac{1}{c_0^2}\frac{\partial^2\phi_p}{\partial t^2}=-Qe^{-ipVt}f(t), \tag{4}$$

where the transform variable p is the wave number. A further Fourier transform, with respect to time, t, yields an algebraic expression for the doubly transformed field:

$$\phi_{p\omega}=-\frac{QF(\omega+pV)}{k^2-p^2(1+i\omega\gamma)}, \tag{5}$$

where the transform variable ω is the angular frequency, $F(\omega)$ is the temporal Fourier transform of the source waveform, $f(t)$, and $k=\omega/c_0$ is the acoustic wave number. Note that the effect of source motion is fully accounted for by the shifted frequency in the Fourier transform of the source function appearing in the numerator of Eq. (5).

To perform the inverse Fourier transforms necessary to recover the velocity potential, ϕ, the source function, $f(t)$, must be specified. Assuming an harmonic source of angular frequency Ω, the source function may be expressed as

$$f(t)=f_0\cos(\Omega t+\psi), \tag{6}$$

where f_0 is the amplitude and ψ is the phase of the wave. Accordingly, the Fourier transform of the source function, $F(\omega)$, is

$$F(\omega)=A\,\delta(\omega-\Omega)+A^*\delta(\omega+\Omega), \tag{7a}$$

where the asterisk denotes complex conjugation and complex coefficient is

$$A=\pi f_0 e^{i\psi}. \tag{7b}$$

On substituting Eq. (7a) into Eq. (5), the expression for the doubly transformed field takes the form

$$\phi_{p\omega}=-Q\frac{A\,\delta(\omega-\Omega+pV)+A^*\delta(\omega+\Omega+pV)}{k^2-p^2(1+i\omega\gamma)}. \tag{8}$$

The presence of the delta functions in the numerator of Eq. (8) allows the inverse Fourier transform with respect to angular frequency, ω, to be performed immediately:

$$\begin{aligned}\phi_p&=-\frac{Q}{2\pi}\\&\quad\times\int_{-\infty}^{\infty}\frac{[A\,\delta(\omega-\Omega+pV)+A^*\delta(\omega+\Omega+pV)]}{k^2-p^2(1+i\omega\gamma)}e^{i\omega t}\,d\omega\\&=-\frac{Q}{2\pi}\left\{\frac{Ae^{i(\Omega-pV)t}}{(\Omega-pV)^2/c_0^2-p^2[1+i(\Omega-pV)]}\right.\\&\quad\left.+\frac{A^*e^{-i(\Omega+pV)t}}{(\Omega+pV)^2/c_0^2-p^2[1-i(\Omega+pV)]}\right\}.\end{aligned} \tag{9}$$

Application of the remaining inverse Fourier transform, with respect to wave number p, yields an integral expression for the velocity potential:

$$\phi = -\frac{Q}{2\pi^2}\,\mathrm{Re}\left\{ A e^{i\Omega t} \int_{-\infty}^{\infty} \frac{e^{ip(x-Vt)}}{iV\gamma p^3 - p^2(1 - V^2/c_0^2 + i\Omega\gamma) - 2pV\Omega/c_0^2 + \Omega^2/c_0^2}\, dp \right\}, \tag{10}$$

where Re denotes the real part of the expression in parenthesis.

The denominator of the integrand in Eq. (10) is a cubic in the wave number p and therefore has three roots. Note that if either V or γ goes to zero, representing, respectively, a motionless source or an inviscid medium, the third-order term vanishes and the denominator reduces to a quadratic. The two roots of the quadratic represent simple poles in the complex p-plane. For the case of a moving source in an inviscid medium, one of these poles yields a Döppler-shifted propagating wave with up-shifted frequency when $t<0$ and the other a propagating wave with down-shifted frequency when $t>0$. In the other situation, with a motionless source in a viscous medium, the two poles yield identical propagating waves, one for the case when the source position is positive and the other for when it is negative relative to the receiver.[8] These waves exhibit frequency dispersion and attenuation, the latter scaling as ω^2 at low frequencies ($\omega\gamma \ll 1$) and $\omega^{1/2}$ at high frequencies ($\omega\gamma \gg 1$). Thus, for source motion alone or dissipation alone the solution in Eq. (10) correctly represents familiar, well-understood behavior. When source motion and dissipation are both present, and only then, the coefficient of the cubic term in the denominator of the integrand in Eq. (10) is nonzero, making the situation a little more complicated.

III. THE CUBIC DENOMINATOR

The presence of the cubic in Eq. (10) suggests that three types of wave must be present in the field detected at the receiver. Two of these wave forms will be the familiar Döppler up-shifted (on approach) and down-shifted (on departure) propagating waves, each of which will be attenuated due to the viscosity of the medium. We may anticipate that the attenuation of the up- and down-shifted waves will be asymmetrical, to a degree that is determined by the speed of the source. The third waveform will turn out to be extremely short-lived in most circumstances, appearing at the instant the source passes the receiver and subsequently decaying rapidly in time and space.

To examine the properties of the three waves, it is necessary to find the roots of the cubic in the wave number p. For this purpose, it is convenient to express the denominator in Eq. (10) in the form

$$D = i\beta L p^3 - p^2(1 - \beta^2 + iqL) - 2q\beta p + q^2, \tag{11}$$

where

$$\beta = \frac{V}{c_0} \tag{12}$$

is the Mach number of the source,

$$q = \frac{\Omega}{c_0} \tag{13}$$

is the wave number of the source (with units of inverse length), and

$$L = c_0\gamma \tag{14}$$

is a viscous dissipation length. Notice that for most practical situations the dimensionless product qL is very much less than unity, a fact that is used below in deriving approximate roots for the cubic in Eq. (10).

By writing the cubic in the standard form

$$D = ap^3 + 3bp^2 + 3cp + d = 0, \tag{15}$$

the coefficients may be identified as

$$a = i\beta L, \tag{16a}$$

$$b = -(1 - \beta^2 + iqL)/3, \tag{16b}$$

$$c = -2q\beta/3, \tag{16c}$$

and

$$d = q^2. \tag{16d}$$

The three roots of Eq. (15) can be expressed analytically through a standard procedure,[9] which relies on the substitution $y = (ap + b)$ to eliminate the second-order term from the cubic. With $j = 1$, 2 or 3, the three roots are of the form

$$p_j = \frac{y_j - b}{a}, \tag{17}$$

where

$$y_j = As_j + Bs_j^2, \tag{18}$$

$$A = \left\{ \frac{-G + \sqrt{G^2 + 4H^3}}{2} \right\}^{1/3}, \tag{19}$$

$$B = \left\{ \frac{-G - \sqrt{G^2 + 4H^3}}{2} \right\}^{1/3}, \tag{20}$$

$$H = ac - b^2, \tag{21}$$

and

$$G = a^2 d - 3abc + 2b^3. \tag{22}$$

The s_j in Eq. (18) are the cube roots of unity:

$$s_1 = s_1^2 = 1, \tag{23a}$$

$$s_2 = s_3^2 = e^{2i\pi/3} = \frac{-1 + i\sqrt{3}}{2}, \tag{23b}$$

and

$$s_3 = s_2^2 = e^{4i\pi/3} = \frac{-1 - i\sqrt{3}}{2}. \tag{23c}$$

Although Eqs. (17)–(23) represent the formal solution for the roots of a cubic, the formulation is too cumbersome to provide much insight into the physics of the sound field

from a moving source in a viscous medium. However, a more revealing, approximate form of this solution may be obtained by recognizing that, for subsonic source speeds and realistic levels of viscosity, the magnitude of the coefficient of the third-order term in Eqs. (11) and (15) is small compared with unity: $|a|=\beta L \ll 1$. It is appropriate, therefore, to expand the solutions for the three roots in Eq. (17) to first-order in the small coefficient, a. (*N.B.* This requires that G^2 and H^3 be expanded to third-order in a.) To this level of approximation, the three roots are found to be

$$p_1 = -\frac{3b}{a} + \frac{c}{b} - \frac{a}{9b^3}(bd-3c^2), \tag{24}$$

$$p_2 = -\frac{1}{2b}\left[c - i\sqrt{\frac{4bd-3c^2}{3}}\right] + \frac{a}{18b^3}\left[(bd-3c^2) + \frac{i3\sqrt{3}c(bd-c^2)}{\sqrt{4bd-3c^2}}\right], \tag{25}$$

and

$$p_3 = -\frac{1}{2b}\left[c + i\sqrt{\frac{4bd-3c^2}{3}}\right] + \frac{a}{18b^3}\left[(bd-3c^2) - \frac{i3\sqrt{3}c(bd-c^2)}{\sqrt{4bd-3c^2}}\right]. \tag{26}$$

The latter two expressions, for p_2 and p_3, are Taylor series in a, whereas Eq. (24) for p_1 is a Laurent series, possessing a term in a^{-1}. When the expressions for the coefficients in Eqs. (16) are substituted into Eqs. (24)–(26), a little algebra yields the following:

$$p_1 = \frac{N^2-\beta^2}{i\beta L} + \frac{2q\beta}{N^2-\beta^2} - \frac{i\beta L q^2(N^2+3\beta^2)}{(N^2-\beta^2)^3}, \tag{27}$$

$$p_2 = \frac{q}{N+\beta}\left[1 + \frac{i\beta qL}{2N(N+\beta)^2}\right], \tag{28}$$

and

$$p_3 = -\frac{q}{N-\beta}\left[1 - \frac{i\beta qL}{2N(N-\beta)^2}\right], \tag{29}$$

where

$$N = \sqrt{1+iqL}. \tag{30}$$

Note that the expression for p_1 in Eq. (27) goes to $-i\infty$ when either the viscosity reduces to zero or the source is stationary. In general, p_1 possesses a large imaginary component, indicating that this pole represents an evanescent wave. The remaining two poles, p_2 and p_3, which also possess imaginary components but with relatively small magnitudes, represent attenuated propagating waves with down-shifted ($t>0$) and up-shifted ($t<0$) Döppler frequencies, respectively.

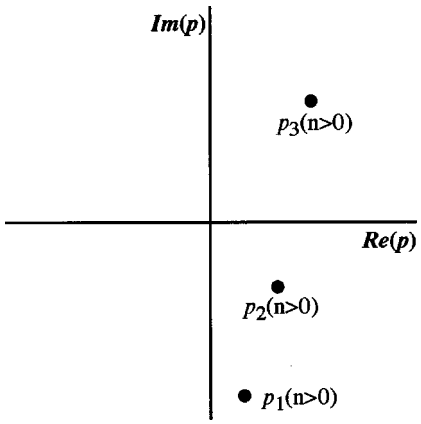


FIG. 2. The complex p-plane showing the positions of the three poles of the integrand in Eq. (10).

IV. THE VELOCITY POTENTIAL

To evaluate the Fourier inversion integral in Eq. (10), it is necessary to identify the positions of the poles in the complex p-plane. From Eqs. (27)–(29), poles p_1 and p_2 lie in the fourth quadrant, while p_3 is in the first quadrant, as shown in Fig. 2. Now the field expression in Eq. (10) can be expressed in the form

$$\phi(t) = \frac{Q}{2\pi^2\beta L} \times \mathrm{Re}\left\{ iAe^{i\Omega t}\int_{-\infty}^{\infty} \frac{e^{-ipVt}}{(p-p_1)(p-p_2)(p-p_3)}\,dp \right\}, \tag{31}$$

where x, representing the receiver position, has been set to zero. For negative times, when the source is inbound to the receiver, a D-contour of infinite radius around the top half-plane, enclosing p_3, yields the solution

$$\phi(t<0) = -\frac{Q}{\pi\beta L}\,\mathrm{Re}\left\{ A\frac{e^{i(\Omega t - p_3 Vt)}}{(p_3-p_1)(p_3-p_2)} \right\}. \tag{32}$$

On the outbound track, when $t>0$, the D-contour around the lower half-plane must be taken, enclosing p_1 and p_2, in which case the expression for the field is

$$\phi(t>0) = \frac{Q}{\pi\beta L}\,\mathrm{Re}\left\{ A\frac{e^{i(\Omega t - p_2 Vt)}}{(p_2-p_1)(p_2-p_3)} + A\frac{e^{i(\Omega t - p_1 Vt)}}{(p_1-p_2)(p_1-p_3)} \right\}. \tag{33}$$

Since $qL \ll 1$ for realistic values of the viscosity, only the leading-order terms need be retained in the factors outside the exponential functions in Eqs. (32) and (33), an approximation which yields

$$\phi(t<0) = -\frac{Q}{2\pi q}\,\mathrm{Re}\{iAe^{i(\Omega t - p_3 Vt)}\} \tag{34}$$

and

$$\phi(t>0) = -\frac{Q}{2\pi q}\mathrm{Re}\{iAe^{i(\Omega t - p_2 V t)}\}$$

$$-\frac{Q\beta L}{\pi(1-\beta^2)^2}\mathrm{Re}\{Ae^{i(\Omega t - p_1 V t)}\}. \quad (35)$$

The asymmetry between the inbound [Eq. (34)] and outbound [Eq. (35)] wave fields is apparent from the presence of the second term on the right of Eq. (35). On approach, the received field consists of a propagating wave with Döppler up-shifted frequency and on departure the corresponding down-shifted propagating wave is present; but at the instant the source passes the receiver an evanescent wave also appears, associated with the pole p_1.

It is worth mentioning that, if the time-derivative of the velocity potential is taken, to yield the pressure field, the amplitude of the propagating wave to leading-order in qL is found to scale as $(1\pm\beta)^{-1}$, where the minus and plus signs apply to the inbound and outbound fields, respectively. Thus, the source motion raises (lowers) the amplitude of the inbound (outbound) pressure field, a result that is familiar from the elementary theory of sound from a moving source in a nondispersive medium.[10]

V. THE PROPAGATING WAVE

In a viscous medium, the inbound and outbound propagating waves from the moving source exhibit Döppler-shifted frequencies and also attenuation. The shifted frequencies are obtained from the real part of the coefficient of t in the arguments of the exponential functions for the propagating waves [Eqs. (34) and (35)]. The attenuation coefficient is the magnitude of the imaginary part of the wave numbers p_2 (outbound) and p_3 (inbound) in Eqs. (28) and (29), respectively.

To first order in qL, the expressions for the poles p_2 and p_3 are

$$p_{2,3} \approx \pm q\left[\frac{1}{(1\pm\beta)} - \frac{iqL}{2(1\pm\beta)^3}\right], \quad (36)$$

where the upper sign yields p_2 (outbound) and the lower sign p_3 (inbound). From the real part of Eq. (36), the received (angular) frequencies on approach (minus sign) and departure (plus sign), to first-order in qL, are easily shown to be

$$\Omega_{\mathrm{in}} \approx \frac{\Omega}{(1-\beta)} \quad (37a)$$

and

$$\Omega_{\mathrm{out}} \approx \frac{\Omega}{(1+\beta)}. \quad (37b)$$

Thus, to this order of approximation, the inbound and outbound frequency shifts due to source motion are the same as would be observed in an inviscid medium.

Turning to the imaginary part of Eq. (36), the attenuation coefficient of the pressure field when the source is inbound to the receiver is

$$\alpha_{\mathrm{in}} \approx \frac{q^2 L}{2(1-\beta)^2}\frac{1}{(1-\beta)}, \quad (38a)$$

and, when the source is outbound,

$$\alpha_{\mathrm{out}} \approx \frac{q^2 L}{2(1+\beta)^2}\frac{1}{(1+\beta)}. \quad (38b)$$

According to these expressions, the source motion affects the amplitudes of the propagating waves arriving at the receiver by introducing an asymmetrical modification to the viscous attenuation. It is evident that, even for modest source speeds, the attenuation may be significantly higher on approach than departure.

Part of the asymmetry may be attributed to the square of the frequency, $[\Omega/(1\pm\beta)]^2$, which appears as a factor in Eqs. (38). Such a square-law frequency scaling is a familiar characteristic of viscous dissipation. From the up- and down-shifts in frequency, the attenuation is enhanced on approach and reduced on departure. The asymmetrical effect of the Döppler-shifted frequency on the inbound and outbound attenuations is reinforced by a further asymmetry, represented by the factors $1/(1\mp\beta)$ appearing in Eqs. (38). [*N.B.* These extra factors derive directly from the third-order term in the cubic in Eq. (11), as may be appreciated by neglecting this term and solving the resulting quadratic for $p_{2,3}$. Identical results to those in Eqs. (38) are obtained, except that the factors $1/(1\mp\beta)$ are absent.] The full asymmetry between the attenuation on approach and departure is evident from the ratio

$$\frac{\alpha_{\mathrm{in}}}{\alpha_{\mathrm{out}}} = \frac{(1+\beta)^3}{(1-\beta)^3} = 2.42, \quad (39)$$

where the numerical value corresponds to a Mach number $\beta=0.147$, representative of a light-aircraft source traveling at $V=50$ m/s in an atmosphere with sound speed $c_0=340$ m/s. Figure 3 shows the expression in Eq. (39) for the ratio of inbound and outbound attenuations plotted as a function of the Mach number, β. Note that at a Mach number of 0.85, representative of the cruise speed of a commercial jet transport, the attenuation coefficient as the aircraft approaches an observer is a factor of 2000 greater than that as it departs. This asymmetry in the level of the perceived sound is, of course, in addition to any that may arise from the aspect dependence of the source itself, which in the case of a jet or indeed a propeller aircraft may be quite pronounced.

The absolute values of the inbound and outbound attenuation coefficients depend on the dynamic viscosity and the density of the medium. For air[11] at 20 °C, $\mu=1.8\times10^{-5}$ kg/m s and $\rho=1.21$ kg/m^3. At a frequency of 1 kHz and with $\beta=0.147$, Eqs. (38) then yield $\alpha_{\mathrm{in}}=1.6\times10^{-5}$ Nepers/m (1.4×10^{-4} dB/m) and $\alpha_{\mathrm{out}}=6.6\times10^{-6}$ Nepers/m (5.8×10^{-5} dB/m). In fact, these values are likely to be underestimates of the inbound and outbound attenuation, since thermal conduction losses give rise to an "effective" viscosity, which is approximately 50% higher than the dynamic viscosity.[12,13] If, in addition, molecular relaxation effects are included in the effective viscosity, its value increases by about one order of magnitude, raising the

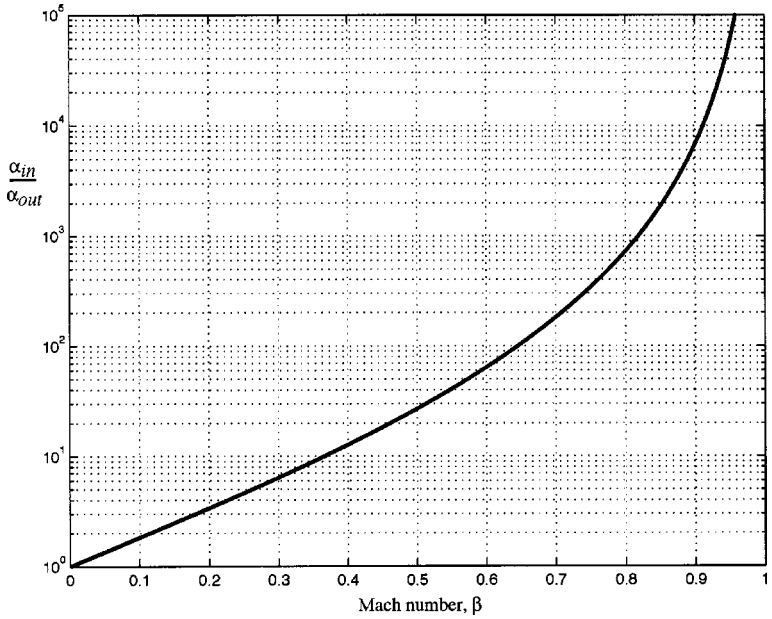


FIG. 3. Ratio of the attenuation of the inbound and outbound propagating waves as a function of Mach number [Eq. (38)].

attenuations at 1 kHz to $\alpha_{\text{in}} \approx 2.4 \times 10^{-4}$ Nepers/m (2.1×10^{-3} dB/m) and $\alpha_{\text{out}} \approx 9.9 \times 10^{-5}$ Nepers/m (8.7×10^{-4} dB/m).

VI. THE EVANESCENT WAVE

The second term on the right of Eq. (35) represents an evanescent wave that appears when the source passes the receiver and which persists very briefly after time $t=0$. Under typical atmospheric conditions, the decay of this wave is extremely rapid, the attenuation coefficient being given by the magnitude of the imaginary part of the wave number p_1. From Eq. (27), the Laurent series for p_1, to first-order in qL, is

$$p_1 \approx -i\frac{(1-\beta^2)}{\beta L} + q\frac{(1+\beta^2)}{\beta(1-\beta^2)} - i\frac{\beta q^2 L(3+\beta^2)}{(1-\beta^2)^3}. \quad (40)$$

For subsonic source speeds ($\beta<1$), only the leading-order term of the attenuation coefficient is significant:

$$\alpha_{\text{evan}} \approx \frac{(1-\beta^2)}{\beta L}, \quad (41)$$

which is independent of the source frequency, Ω.

Unlike the propagating waves, where the attenuation is proportional to the dissipation constant, γ, the attenuation of the evanescent wave, α_{evan} scales as the reciprocal of γ. Since α_{evan} also scales inversely with the Mach number, β, it follows that in the absence of either source motion ($\beta=0$) or dissipation ($L=0$), the attenuation coefficient diverges to infinity and the evanescent wave vanishes.

Even in the presence of both source motion and dissipation, the attenuation of the evanescent wave is extremely high. Taking values of $\beta=0.147$ and $L=5.9\times10^{-8}$ m, as used above to represent a source such as a light aircraft mov-

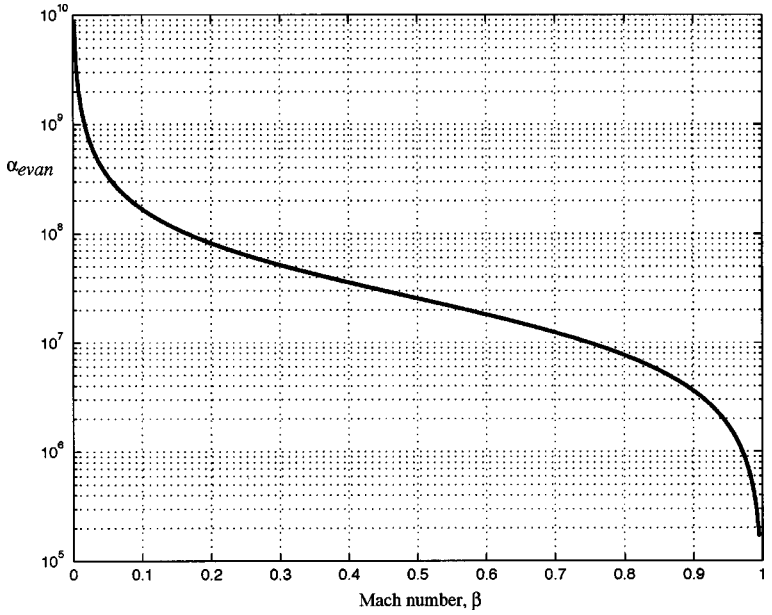


FIG. 4. Attenuation (nepers/m) of the evanescent wave as a function of Mach number [Eq. (41)], with $L=5.9\times10^{-8}$ m, appropriate to air.

ing through the atmosphere, we find that $\alpha_{\text{evan}}=1.1\times10^{8}$ Nepers/m $(9.9\times10^{8}$ dB/m). Obviously, such heavy attenuation means that the evanescent wave would be almost impossible to detect in most practical circumstances. When the Mach number is very close to unity, the numerator of Eq. (41) falls to near zero and the attenuation coefficient is reduced accordingly, suggesting, on the face of it, that this may be a situation where the evanescent could be detected. However, this is unlikely to be the case, as can be seen in Fig. 4 showing α_{evan} as a function of Mach number; even when the Mach number is as high as $\beta=0.99$, the attenuation is still in the region of 3.4×10^{5} Nepers/m $(3\times10^{6}$ dB/m). Such high attenuation is a direct result of the low value of L for a typical atmosphere.

In view of its excessively high attenuation, it is almost meaningless to talk of the frequency of the evanescent wave. Nevertheless, as a mere formality, an expression for the (angular) frequency of the evanescent wave is included here, as deduced from Eqs. (27) and (35):

$$\Omega_{\text{evan}}=2\Omega\,\frac{\beta^{2}}{1-\beta^{2}}. \tag{42}$$

Although this frequency depends on the Mach number of the source, it is independent of the level of viscous dissipation.

VII. CONCLUDING REMARKS

The properties of the sound field from a moving source in a viscous medium have been investigated theoretically in this article. Based on a simple, one-dimensional model, three types of wave from the moving source are shown to exist in the viscous fluid: an attenuated propagating wave with upshifted frequency, which is present as the source approaches the receiver; an attenuated propagating wave with downshifted frequency, which is present as the source recedes from the receiver; and an evanescent wave, which appears at the instant the source passes the receiver and which persists for a very brief time thereafter.

From a theoretical point of view, the evanescent wave is an interesting feature of the received field in that it exists only when source motion and dissipation are both present. When either the Mach number of the source or the viscous dissipation is zero, the attenuation escalates to infinity and the evanescent wave vanishes. Although not zero, dissipation in the atmosphere is sufficiently weak for the evanescent wave to be highly attenuated, so much so that the wave is something of a curiosity and unlikely to be of significance in any practical observation of the sound field from a subsonic source.

A more important effect of the combined presence of source motion and viscous dissipation is the asymmetry introduced between the two propagating waves. Although viscosity has a negligible effect on the Döppler-shifted frequencies, the source motion couples into the attenuation of the two propagating waves in an asymmetrical fashion. Each attenuation coefficient exhibits typical viscous characteristics, scaling as the square of the Döppler-shifted frequency, which is higher on approach than departure. But this is not all. A further asymmetry is introduced into the attenuation coefficients in the form of the factor $(1\mp\beta)^{-1}$, where β is the Mach number of the source and the upper sign applies on approach, the lower on departure. The combined effect of both forms of asymmetry is an attenuation coefficient for the propagating waves that is higher on approach than departure by the factor $[(1+\beta)/(1-\beta)]^{3}$. This factor increases monotonically with β, taking values in the range 2 to 2000 for Mach numbers between 0.1 and 0.85.

ACKNOWLEDGMENTS

This work was supported by Dr. J. Simmen and Dr. E. Livingston, Ocean Acoustics Code, the Office of Naval Research, under Grant No. N00014-93-1-0054.

[1] J. W. Strutt (Lord Rayleigh), *The Theory of Sound*, 2nd ed. (Dover, New York, 1945).
[2] A. B. Wood, *A Textbook of Sound*, 3rd ed. (Bell, London, 1964).
[3] M. J. Lighthill, *Waves in Fluids* (Cambridge U.P., Cambridge, 1978).
[4] A. D. Pierce, *Acoustics: An Introduction to its Physical Principals and Applications* (McGraw-Hill, New York, 1981).
[5] H. Schmidt and W. A. Kuperman, "Spectral and modal representations of the Doppler-shifted field in ocean waveguides," J. Acoust. Soc. Am. **96**, 386–395 (1994).
[6] M. J. Buckingham, E. M. Giddens, J. B. Pompa, F. Simonet, and T. R. Hahn, "Sound from a light aircraft for underwater acoustics experiments?" Acta Acust. united with Acust. **88**, 752–755 (2002).
[7] M. J. Buckingham, E. M. Giddens, F. Simonet, and T. R. Hahn, "Propeller noise from a light aircraft for low-frequency measurements of the speed of sound in a marine sediment," J. Comput. Acoust. **10**, 445–464 (2002).
[8] M. J. Buckingham, "Acoustic pulse propagation in dispersive media," in *New Perspectives on Problems in Classical and Quantum Physics. Part II. Acoustic Propagation and Scattering—Electromagnetic Scattering*, edited by P. P. Delsanto and A. W. Sáenz (Gordon and Breach, Amsterdam, 1998), Vol. 2, pp. 19–34.
[9] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).
[10] D. G. Crighton, A. P. Dowling, J. E. Ffowcs Williams, M. Heckl, and F. G. Leppington, *Modern Methods in Analytical Acoustics: Lecture Notes* (Springer-Verlag, London, 1992).
[11] D. R. Lide (ed.), *CRC Handbook of Chemistry and Physics* (CRC, Boca Raton, 1993).
[12] L. B. Evans, H. E. Bass, and L. C. Sutherland, "Atmospheric absorption of sound: theoretical predictions," J. Acoust. Soc. Am. **51**, 1565–1575 (1971).
[13] H. E. Bass, H.-J. Bauer, and L. B. Evans, "Atmospheric absorption of sound: analytical expressions," J. Acoust. Soc. Am. **52**, 821–825 (1972).

Coupled perturbed modes over a sloping penetrable bottom

C. J. Higham[a] and C. T. Tindle
Department of Physics, University of Auckland, Private Bag 92019, Auckland, New Zealand

(Received 11 July 2003; accepted for publication 22 September 2003)

The development of coupled perturbed mode theory—the combination of conventional coupled modes and perturbation theory for problems of sound speed range dependence—is extended using a transformation of the normal modes to include a range-dependent penetrable bottom. The result is a fast, direct, and accurate range-dependent modal solution of the acoustic field in the water column. The method is applied to upslope propagation in shallow water wedge waveguides. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1625935]

PACS numbers: 43.30.Bp [AT] Pages: 3119–3124

I. INTRODUCTION

The normal mode solution of underwater acoustic problems is exact for range-independent environments. Most real problems, however, involve some range dependence, which requires an extension to the standard theory. One successful approach is Evans's[1] coupled mode theory that breaks range-dependent environments into stepwise range-independent sections. The acoustic field is found by matching the modal solution in each step at the step endpoints, and reducing the step size until the solution converges. Coupled mode theory is considered exact, but requires a repeated solution of the wave equation for the normal modes at each range step. It is computationally intensive and is mainly used to obtain reference solutions to benchmark problems.

The combination of perturbation theory and coupled mode theory eliminates the repeated solution of the wave equation at each range step. This greatly improves the overall efficiency of coupled mode calculations with little loss in accuracy. The authors have previously successfully applied the method to deep water oceanic fronts[2] and shallow water internal solitary waves.[3] The present description expands the method to include a range-dependent penetrable bottom. Where, in order to accommodate the sudden change in density at the sea floor within perturbation theory, a transformation is introduced to remove the gradient discontinuity of the normal mode functions.

The relevant coupled mode and perturbation theory is outlined in Sec. II and tested on two shallow water wedges described in Sec. III. Also described in Sec. III is the treatment (within a coupled perturbed mode scheme) of modal cutoff and truncated mode sets with resonant continuous modes. A discussion and conclusions follow in Sec. IV.

II. PERTURBATION THEORY AND COUPLED MODES

A. Perturbation theory

If a set of normal modes is known for a given environment, then perturbation theory can be used to solve for the set of normal modes of a slightly different environment. For a change in sound speed profile the derivation is straightforward and is outlined in Ref. 2. However, to include a change in density structure such as occurs over a sloping penetrable bottom, modification to the modal equation is required.

1. Modifying the modal equation

In a horizontally stratified environment, where sound speed $c(z)$ and density $\rho(z)$ are assumed to be functions of depth z only, the normal modes $U_n(z)$ are solutions of the depth-separated component of the wave equation:[4]

$$\rho(z)\frac{d}{dz}\left[\frac{1}{\rho(z)}\frac{dU_n(z)}{dz}\right]+[q(z)-k_n^2]U_n(z)=0, \tag{1}$$

where $q(z)=\omega^2/c^2(z)$ is defined for convenience (ω is the angular frequency of the acoustic source), and the normal mode eigenvalues k_n are equivalent to the wave numbers for horizontal propagation. With appropriate boundary conditions the normal modes are orthogonal with weight function $\rho^{-1}(z)$ and satisfy

$$\int_0^\infty \frac{U_n(z)U_m(z)}{\rho(z)}dz=\begin{cases}1, & \text{for } n=m,\\ 0, & \text{for } n\neq m.\end{cases} \tag{2}$$

For a small change in the sound speed profile, $c(z)\rightarrow c(z)+\Delta c(z)$, say, perturbation theory assumes the normal modes and eigenvalues have a similar and associated change, i.e., $U_n(z)\rightarrow U_n(z)+\Delta U_n(z)$ and $k_n\rightarrow k_n+\Delta k_n$. These new values must also satisfy Eq. (1): the modal equation. Substituting these quantities into Eq. (1) and expanding, expressions for $\Delta U_n(z)$ and Δk_n can be found by utilizing the properties of the normal modes. The process should, in principle, be similar for a small change in the density profile: $\rho(z)\rightarrow\rho(z)+\Delta\rho(z)$. In reality this depends on the form of $\rho(z)$.

Most waveguide models take density to be constant within the water column and piecewise constant in the bottom. The simplest description to begin with assumes isodensity in the bottom:

$$\rho(z)=\begin{cases}\rho_w, & \text{for } z<h,\\ \rho_b, & \text{for } z>h,\end{cases} \tag{3}$$

where ρ_w and ρ_b are the densities in the water column and bottom, respectively, $\rho_b>\rho_w$, and h is the seafloor depth. Equation (3) forms the basis of the discussion and construction that follows.

[a]Electronic mail: cjh@phy.auckland.ac.nz

At $z=h$, Eq. (3) has a density jump. This feature, coupled with continuity considerations, leads to a gradient discontinuity of the normal modes at this boundary. The depth of this ''kink'' in the mode shapes varies for range-dependent bathymetry. The resulting change in the mode shapes is not a smooth incremental change, and is not suitable for a perturbation expansion of Eq. (1).

The gradient discontinuity of the mode shapes can be removed by seeking a transformation $z \to \tilde{z}$, defined by

$$\frac{dU_n(z)}{dz} = \alpha(\tilde{z};z)\,\frac{dU_n(\tilde{z};z)}{d\tilde{z}}, \tag{4}$$

where $\alpha(\tilde{z};z)$ is chosen so that the derivative of the transformed mode is continuous. An inspection of Eq. (4) yields the scaling relationship

$$d\tilde{z} = \alpha(\tilde{z};z)\,dz. \tag{5}$$

The transformation may therefore be considered a description that stretches (or compresses) the normal modes $U_n(z)$, according to $\alpha(\tilde{z};z)$, to create the modified set $U_n(\tilde{z};z)$ with no gradient discontinuity at the sea floor.

To determine $\alpha(\tilde{z};z)$, Eqs. (3) and (4) are substituted into the boundary condition which requires continuity of a vertical displacement

$$\frac{1}{\rho(h^-)}\left[\frac{dU_n(z)}{dz}\right]_{z=h^-} = \frac{1}{\rho(h^+)}\left[\frac{dU_n(z)}{dz}\right]_{z=h^+},$$

applied at the water column and bottom boundary $z=h$; i.e.,

$$\frac{\alpha(\tilde{z};h^-)}{\rho_w}\left[\frac{dU_n(\tilde{z};z)}{d\tilde{z}}\right]_{z=h^-} = \frac{\alpha(\tilde{z};h^+)}{\rho_b}\left[\frac{dU_n(\tilde{z};z)}{d\tilde{z}}\right]_{z=h^+}. \tag{6}$$

Since the gradients of the new normal modes $U_n(\tilde{z};z)$ are assumed to be continuous at $z=h$, the square bracketed terms in Eq. (6) must have the same value, and so can be dropped from the expression. This leaves

$$\alpha(\tilde{z};h^-)/\rho_w = \alpha(\tilde{z};h^+)/\rho_b. \tag{7}$$

By taking $\alpha(\tilde{z};h^-)=1$, the mode shapes within the water column remain unchanged, and Eq. (7) then gives

$$\alpha(\tilde{z};z) = \begin{cases} 1, & \text{for } z<h, \\ \rho_b/\rho_w, & \text{for } z>h. \end{cases} \tag{8}$$

The transformation is thus

$$\tilde{z} = \alpha(\tilde{z};z)[z-h]+h, \tag{9}$$

and the $U_n(\tilde{z};z)$ are simply the original modes with the extensions into the bottom stretched by the factor ρ_b/ρ_w.

From Eq. (5) it follows that the modal equation, Eq. (1), now becomes

$$\rho(\tilde{z};z)\,\alpha(\tilde{z};z)\,\frac{d}{d\tilde{z}}\left[\frac{\alpha(\tilde{z};z)}{\rho(\tilde{z};z)}\,\frac{dU_n(\tilde{z};z)}{d\tilde{z}}\right] + [q(\tilde{z};z)-k_n^2]U_n(\tilde{z};z) = 0, \tag{10}$$

where the ratio $\alpha(\tilde{z};z)/\rho(\tilde{z};z)$ can be factored out using Eqs. (3) and (8), which give $\alpha(\tilde{z};z)/\rho(\tilde{z};z) = \rho_w^{-1}$. It is noted that the discontinuity at $z=h$ is removable by defining $\alpha(\tilde{z};h)/\rho(\tilde{z};h) = \rho_w^{-1}$, since

$$\lim_{z\to h}[\alpha(\tilde{z};z)/\rho(\tilde{z};z)] = \rho_w^{-1}.$$

The modified modal equation is hence

$$\alpha^2(\tilde{z};z)\,\frac{d^2U_n(\tilde{z};z)}{d\tilde{z}^2} + [q(\tilde{z};z)-k_n^2]U_n(\tilde{z};z) = 0. \tag{11}$$

Although constructed for a density profile consisting of a single layer bottom, Eq. (11) holds for multilayered bottoms, each layer of a different constant density. The appropriate expression for $\alpha(\tilde{z};z)$ is derived in a similar fashion to that given above.

2. Solving the modified modal equation

To derive the relevant perturbation equations it is convenient to define $X(\tilde{z};z) = [\alpha^2(\tilde{z};z)]^{-1}$ and to rewrite Eq. (11) as

$$\frac{d^2U_n(\tilde{z};z)}{d\tilde{z}^2} + X(\tilde{z};z)[q(\tilde{z};z)-k_n^2]U_n(\tilde{z};z) = 0. \tag{12}$$

With appropriate boundary conditions this modified modal equation is a classical Sturm–Liouville eigenvalue problem with well-known properties. Two of these properties are important in a perturbation expansion of Eq. (12): (i) the modified normal modes are orthogonal with weight function $X(\tilde{z};z)$, i.e.,

$$\int_0^\infty X(\tilde{z};z)\,U_n(\tilde{z};z)\,U_m(\tilde{z};z)\,d\tilde{z} = \begin{cases} 1, & \text{for } n=m, \\ 0, & \text{for } n\neq m; \end{cases} \tag{13}$$

(ii) the modes form a complete set meaning that any arbitrary function, such as $\Delta U_n(\tilde{z};z)$, can be expressed as a sum of the normal modes:

$$\Delta U_n(\tilde{z};z) = \sum_m a_{nm}U_m(\tilde{z};z), \tag{14}$$

where the a_{nm} are expansion coefficients.

Suppose the set of modified normal modes $U_n(\tilde{z};z)$ and eigenvalues k_n are known for some particular local sound speed profile $c(\tilde{z};z)$ and local density profile $\rho(\tilde{z};z)$. If for a given range increment the changes to the sound speed and density profiles are small, then perturbation theory assumes the normal modes change to $U_n(\tilde{z};z)+\Delta U_n(\tilde{z};z)$ and the eigenvalues change to $k_n+\Delta k_n$. The changes in sound speed and density structure can be couched in terms of $q(\tilde{z};z)+\Delta q(\tilde{z};z)$ and $X(\tilde{z};z)+\Delta X(\tilde{z};z)$.

The new values must also satisfy Eq. (12). The new Eq. (12) can be solved using Eqs. (13) and (14) to obtain expressions for the Δk_n and the a_{nm}. The steps involved are outlined in the Appendix. The results are

$$\Delta k_n = \frac{1}{2k_n}\left[\int_0^\infty \Delta q(\tilde{z};z)\,U_n^2(\tilde{z};z)\,d\tilde{z} - \int_0^\infty (q(\tilde{z};z)-k_n^2)\,\Delta X(\tilde{z};z)\,U_n^2(\tilde{z};z)\,d\tilde{z}\right], \tag{15}$$

$$a_{nm}=\frac{1}{k_n^2-k_m^2}\left[\int_0^\infty \Delta q(\tilde{z};z)U_n(\tilde{z};z)U_m(\tilde{z};z)d\tilde{z} - \int_0^\infty (q(\tilde{z};z)-k_n^2)\Delta X(\tilde{z};z) \times U_n(\tilde{z};z)U_m(\tilde{z};z)d\tilde{z}\right], \quad (16)$$

where Eq. (16) applies for $n\neq m$. The a_{nn} are found by satisfying Eq. (13); details are given in the Appendix.

Together, Eqs. (14), (15), and (16) describe corrections to the wave numbers and mode shapes for changes in sound speed *and* density structure. The normal modes and eigenvalues are thus now known for two slightly different local sound speed and density profiles—achieved by numerically solving the modal equation only once.

This process can be repeated for multiple range increments, each with some small change in sound speed and density structure. By reversing the transformation given by Eq. (9) so as $U_n(\tilde{z};z)\rightarrow U_n(z)$, these perturbed normal modes can be readily incorporated into the standard coupled mode formalism.

B. Coupled mode theory

For a harmonic point source (angular frequency ω) and cylindrical geometry, the acoustic pressure $p(r,z)$ of the jth range-independent step can be written[4] as

$$p^{(j)}(r,z)\approx r^{-1/2}\sum_n A_n^{(j)}U_n^{(j)}(z)e^{ik_n^{(j)}(r-r_{j-1})}, \quad (17)$$

where the jth step is defined for range $r\in[r_{j-1},r_j]$. The equation is a one-way coupled mode description of the acoustic field, where the mode coefficients A_n can be expressed recursively, by matching pressure and displacement across the step interface using the ''approximate single-scatter'' approach:[5]

$$A_n^{(j+1)}=\sum_m C_{nm}^{(j)}A_m^{(j)}e^{ik_m^{(j)}(r_j-r_{j-1})}, \quad (18)$$

where

$$C_{nm}^{(j)}=\frac{1}{2}\left[\int_0^\infty \frac{U_n^{(j+1)}(z)U_m^{(j)}(z)}{\rho^{(j+1)}(z)}dz + \frac{k_m^{(j)}}{k_n^{(j+1)}}\int_0^\infty \frac{U_n^{(j+1)}(z)U_m^{(j)}(z)}{\rho^{(j)}(z)}dz\right]. \quad (19)$$

The approximate single-scatter formulation addresses the problem of energy conservation common to one-way descriptions. It has been shown by Porter *et al.*,[5] for negligible backscatter, to agree almost perfectly with the exact solution.

Since it is assumed that the change in environment is small between adjacent range steps, then perturbation theory allows the $U_m^{(j)}(z)$ and the $U_n^{(j+1)}(z)$ of Eq. (19) to be expanded as

$$U_m^{(j)}(z)=U_m^{(j+1)}(z)+\sum_\ell b_{m\ell}^{(j+1)}U_\ell^{(j+1)}(z), \quad (20)$$

$$U_n^{(j+1)}(z)=U_n^{(j)}(z)+\sum_\ell f_{n\ell}^{(j)}U_\ell^{(j)}(z), \quad (21)$$

where the second term in each equation satisfies Eq. (14). The coefficients $f_{n\ell}^{(j)}$ and $b_{m\ell}^{(j+1)}$ are easily determined using matrix manipulations, if the mode shapes of adjacent range steps are known.

The substitution of Eq. (20) into the first integral and Eq. (21) into the second integral of Eq. (19) leaves both integrals independent of adjacent range steps. The analytic evaluation of Eq. (19) is then possible using mode orthogonality as in Eq. (2), and yields a single expression for the mode coefficients,

$$A_n^{(j+1)}=\frac{1}{2}\left[\left(1+\frac{k_n^{(j)}}{k_n^{(j+1)}}\right)A_n^{(j)}e^{ik_n^{(j)}(r_j-r_{j-1})} + \sum_m b_{mn}^{(j+1)}A_m^{(j)}e^{ik_m^{(j)}(r_j-r_{j-1})} + \frac{1}{k_n^{(j+1)}}\sum_m f_{nm}^{(j)}k_m^{(j)}A_m^{(j)}e^{ik_m^{(j)}(r_j-r_{j-1})}\right]. \quad (22)$$

For the initial range step, $j=1$, the mode coefficients are[4]

$$A_n^{(1)}=\sqrt{\frac{2\pi}{k_n}}\frac{e^{i\pi/4}}{\rho(z_s)}U_n(z_s), \quad (23)$$

where z_s is the depth of the source.

III. WEDGE WAVEGUIDES

Shallow water wedge waveguides, because of their range-dependent bathymetry, feature both sound speed and density range dependence; the density structure (as a function of depth) is usually modelled by Eq. (3). A wedge waveguide therefore provides a suitable test for the perturbation approximation of Sec. II.

A. Modal cutoff

An important feature of shallow water acoustics is the possibility of modal cutoff: the transition of a normal mode between the discrete (''trapped'') and continuous mode spectrum. Waveguides support a fixed number of trapped modes for a given seafloor depth and source frequency. For upslope wedge propagation the number of trapped modes decreases as the water column depth decreases; each trapped mode is lost from the discrete mode spectrum at its cutoff depth.

The energy transfer that characterises cutoff is large and significant. The mode shapes change rapidly and do not satisfy the requirements of perturbation theory. This difficulty is overcome by using exact modes to update the coupled mode calculation when a mode is near cutoff, and using perturbed modes elsewhere. As detailed in Ref. 3, the small update interval about cutoff can be determined by comparing the wave numbers k_n to the cutoff wave number $k_c=\omega/c_b$ (c_b is the sound speed in the bottom), as the coupled perturbed mode calculation advances with range.

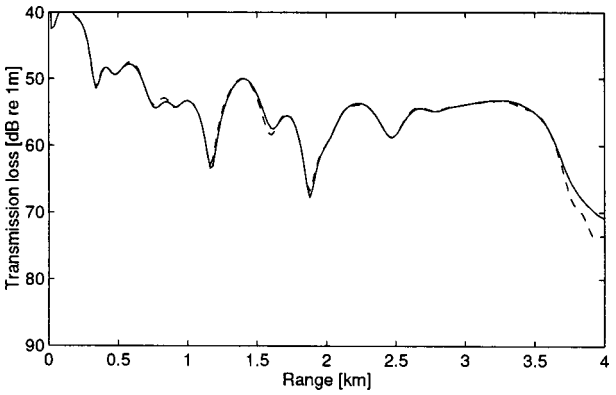


FIG. 1. Transmission loss for benchmark 2: exact (solid line) and approximate (dashed line) one-way solutions. The approximate solution includes exact mode updates about cutoff.

B. Penetrable wedge calculations

The ASA (Acoustical Society of America) penetrable wedge benchmark 2 was posed in 1986, as one of a series of benchmarks for testing the accuracy of different modeling schemes. It consists of a homogeneous water column ($c_w = 1500$ m/s, $\rho_w = 1000$ kg/m^3) of decreasing depth, bounded by a pressure release sea surface and a homogeneous bottom ($c_b = 1700$ m/s, $\rho_b = 1500$ kg/m^3). The water depth decreases linearly from 200 m at the source range (25 Hz source at $r_s = 0$ m, $z_s = 100$ m) to zero depth at $r = 4$ km.

An accurate modal solution to benchmark 2 was presented by Jensen and Ferla,[6] who created a proper Sturm–Liouville problem by truncating the bottom with a pressure release false floor. (The floor depth was set so as to adequately sample the continuous mode spectrum.) The solution output was given as a plot of transmission loss, as a function of range, for a fixed depth receiver.

Shown in Fig. 1 is the benchmark 2 transmission loss for a 30 m receiver. The solid line is an approximate single-scatter one-way solution that uses exact normal modes and wave numbers. This reference solution agrees within a linewidth with a full two-way coupled mode solution, kindly provided by Finn Jensen. A total of 50 modes were used to determine the reference solution, obtained with a false floor depth of 2000 m. The modes were numerically calculated at each of 200 stair steps used to approximate the sloping bottom. Following Jensen and Ferla,[6] restricted source beamwidth solutions were required to determine the farfield so as to avoid false floor reflections; coupling was allowed between all 50 source modes.

At the source range, three normal modes are trapped within the water column. As the solution advances in range, each of the three trapped modes is lost from the discrete mode spectrum. The cutoff ranges are approximately 800 m for mode 3, 2100 m for mode 2, and 3350 m for mode 1. As mentioned above, for small intervals about these ranges it is necessary to use exact modes and wave numbers in the coupled perturbed mode approximation, as perturbation theory is unable to handle the sudden changes in mode shape that occurs about cutoff. The one-way perturbation approximation to benchmark 2 (dashed line, Fig. 1) includes 14 field updates using exact modes and wave numbers about the three cutoff ranges. This is only 7% of the total 200 range steps used to represent the bathymetry.

Like the reference solution, the approximate solution uses 50 modes to determine the field. However, a total of 57 modes are retained in the perturbation calculations, Eqs. (15) and (16). This ensures that the continuum is well sampled and the a_{nm} are sufficiently accurate for high mode numbers. The parameter $[k_n^2 - k_m^2]^{-1}$ of Eq. (16) shows that coupling is strongest among modes close in mode number. The truncated set of 50 modes is effectively ''padded'' with the addition of modes 51–57 to satisfy coupling among the high-order modes (modes 40–50).

The overall accuracy of the perturbation approximation in Fig. 1 is very good. The approximate solution overlays the curve of the exact solution until the receiver enters the bottom. The apparent solution divergence toward the wedge apex is a function of the receiver depth, the shoaling bottom, and the modified mode shapes $U_n(\tilde{z};z)$.

In the bottom the modes $U_n(\tilde{z};z)$ are essentially stretched versions of the normal modes of the original problem. Stretching the mode shapes, through the transformation described by Eqs. (8) and (9), also displaces the false floor boundary condition downward. The displacement of the false floor is a linear function of the distance between the seafloor depth and the false floor depth of the original problem: the greater the separation the greater the displacement. For upslope propagation the depth of the wedge decreases and the separation between the seafloor and false floor increases. The displaced lower boundary condition, therefore, shifts progressively farther downward with each advancing range step.

The perturbation scheme in Sec. II makes no provision for the sloping nature of the displaced lower boundary. Instead it implicitly assumes that the mode shapes expand adiabatically in the bottom between range steps. This is a crude approximation of the field in the bottom and leads to the solution divergence seen in Fig. 1 for ranges beyond 3.6 km. However, as the receiver enters the bottom at a range of 3.4 km, this inaccuracy in the bottom is of little importance—most solutions are only of interest within the water column. The coupled perturbed mode description accurately describes the field within the water column.

The low-frequency 25 Hz source used in benchmark 2 is not typical of the sources used in experimental research; sources used are typically of the order of a few hundred hertz. The ASA, at the time of establishing the benchmark problems, chose a 25 Hz source for the wedge benchmarks, as high-frequency problems were computationally prohibitive for the combination of numerical schemes and the then available computer processing power. Below, a more realistic source frequency of 250 Hz is considered to test the perturbation approximation. The new wedge environment is identical to benchmark 2, except for the change in source frequency.

The transmission loss plot (receiver depth 30 m) for this new problem is illustrated in Fig. 2. Again, the perturbation approximation (dashed line) is compared with an exact one-way solution (solid line) to verify its accuracy. To avoid false floor reflections contaminating the result, the bottom attenu-

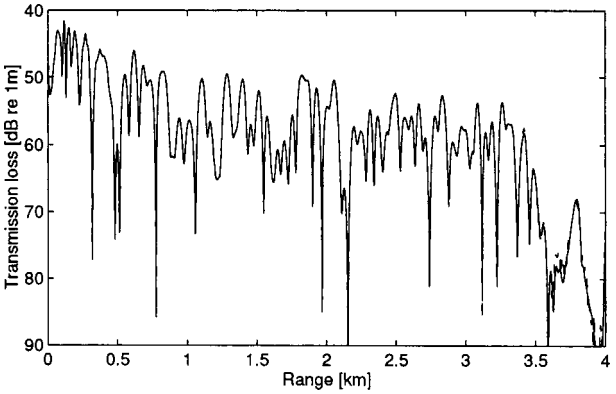


FIG. 2. Transmission loss for 250 Hz wedge: exact (solid line) and approximate (dashed line) one-way solutions. The approximate solution includes exact mode updates about cutoff and for high-order modes at resonance.

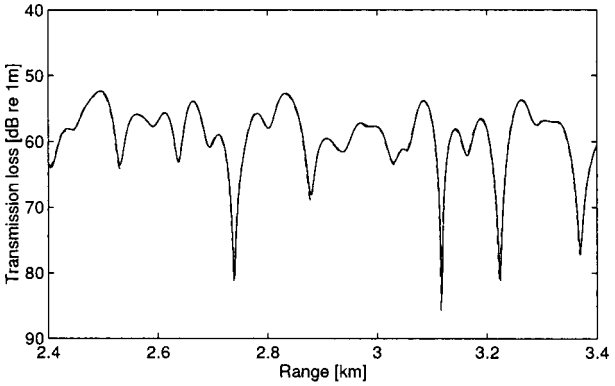


FIG. 3. Expanded view (2.4–3.4 km) of the transmission loss curves in Fig. 2: exact (solid line) and approximate (dashed line) one-way solutions.

ation of 0.5 dB/λ is included using an approximate perturbational treatment.[4]

Both the solutions of Fig. 2 include 100 modes in the mode sum, Eq. (17), with the false floor set at a depth of 400 m. As with benchmark 2, to improve the accuracy of the high-order modes in the perturbation approximation, additional modes (modes 101–117) are retained in the perturbation calculations. These additional modes are increasingly poorly approximated with each range step increment, as there is no mode coupling with the neglected part of the continuum. In general this does not matter, as the additional modes have little acoustical energy within the water column. However, truncation of the continuous modes can lead to a loss of accuracy when one of the continuous modes is near resonance.

A continuous mode is oscillatory everywhere, both within the water column and the bottom. It is said to be resonant when the gradient of the mode shape at the seafloor depth h is zero, and the mode shape has a maximum amplitude in the water column relative to the amplitude in the bottom.[7] The water-borne amplitude of a resonant mode can therefore be larger than most of its neighbors, yielding a corresponding larger contribution to the mode sum.

In a wedge, resonant modes are a function of the changing water depth with range. If a continuous mode is about to become resonant, then sizable transfers of energy occur between itself and nearby modes. This is reflected by the changing amplitude within the water column, of the resonant mode and its nearby neighbors. This change is handled properly by Eq. (16), provided the mode sum is not truncated and the resonant mode has nearby modes into which it can couple. However, for a truncated mode set it is necessary to update the perturbation solution with exact modes and wave numbers when a high-order mode is near resonance.

For the coupled perturbed mode solution of Fig. 2, the gradients of modes 101–117 at $z=h$ are monitored (as the solution advances in range) to check for mode resonances. If a mode from this set is at resonance, the field is updated with exact mode values; the perturbation calculations then continue from the next range step. With these extra mode updates included, the perturbation solution gives a very good approximation to the exact solution.

The 25 Hz problem did not require mode updates for this truncated continuum artifact.

Both curves in Fig. 2 use 2000 stair steps to approximate the sloping bottom. It was found that 2000 steps and 100 modes were needed for the exact one-way solution to converge. The solution is accurate for ranges greater in value than 200 m, the water depth at the source. The perturbation approximation includes 206 field updates with exact mode functions for high-order mode resonances and mode cutoffs (there are 31 trapped modes at the source range). This is roughly 10% of the total number of range steps. Little is lost in the efficiency of the coupled perturbed mode scheme with the inclusion of mode updates for resonant modes.

As well as being efficient, the perturbation solution of Fig. 2 is also accurate. The only discrepancy with the exact solution is for ranges where the receiver is in the bottom ($r \geqslant 3.4$ km), where the perturbation approximation is invalid. An expanded view of the last 1000 m of range before the receiver hits the bottom is shown in Fig. 3. The dashed line of the coupled perturbed mode result is in excellent agreement with the reference solution (solid line).

IV. CONCLUSIONS

A novel method has been described for the inclusion of a sloping penetrable bottom in a coupled perturbed mode formulation. The method essentially stretches the mode shapes in the bottom, such that the mode shape gradient discontinuity at the seafloor disappears. This then yields a modified modal equation that is straightforward to solve with perturbation theory. The new theory is applicable to field calculations within the water column.

The coupled perturbed mode scheme removes the need to solve the wave equation at each range step by expanding the normal modes and eigenvalues of each advancing range step in terms of the mode functions of the preceding step. This greatly improves the efficiency of the modal solution, whilst still providing good solution accuracy.

For shallow water range-dependent waveguides, modal cutoff is important. The large transfer of energy that occurs at cutoff is too great for perturbation theory and requires field updates. Updates are also necessary to accurately describe the mode coupling of high-order resonant modes. These oc-

casional updates, using exact normal modes and eigenvalues, improve the accuracy of the solution with little effect on the solution efficiency.

The method has been applied to the ASA benchmark 2 problem at 25 and 250 Hz. The approximate solutions to both test cases agree well with the reference solutions. The results show coupled perturbed modes provides an accurate means of determining the acoustic field in shallow water wedge waveguides, where density, as well as sound speed, is range-dependent.

ACKNOWLEDGMENT

The authors thank Finn Jensen for kindly supplying the two-way COUPLE data for benchmark 2.

APPENDIX: PERTURBATION THEORY FOR MODIFIED NORMAL MODES

To apply perturbation theory to Eq. (12), it is assumed that $q(\tilde{z};z)$ and $X(\tilde{z};z)$ change by $\Delta q(\tilde{z};z)$ and $\Delta X(\tilde{z};z)$, respectively, and that this leads to corresponding changes $\Delta U_n(\tilde{z};z)$ and Δk_n in $U_n(\tilde{z};z)$ and k_n. The modified modal equation becomes

$$(U_n+\Delta U_n)''+(X+\Delta X)[q+\Delta q-(k_n+\Delta k_n)^2] \times(U_n+\Delta U_n)=0, \tag{A1}$$

where the double prime represents the second derivative with respect to $\tilde{z}$. Subtracting Eq. (A1) from Eq. (12), first-order perturbation theory yields

$$(\Delta U_n)''+X[q-k_n^2]\Delta U_n+\Delta X[q-k_n^2]U_n +X[\Delta q-2k_n\Delta k_n]U_n=0. \tag{A2}$$

Since the $U_n(\tilde{z};z)$ form a complete set, the $\Delta U_n(\tilde{z};z)$ can be expanded in terms of them as in Eq. (14). Substituting Eq. (14) into Eq. (A2), the first two terms of the new expression may be simplified to $X(\tilde{z};z)\Sigma_m a_{nm}[k_m^2-k_n^2]U_m(\tilde{z};z)$ using Eq. (12), and Eq. (A2) then becomes

$$X\sum_m a_{nm}[k_m^2-k_n^2]U_m+\Delta X[q-k_n^2]U_n +X[\Delta q-2k_n\Delta k_n]U_n=0. \tag{A3}$$

Finally, by applying in turn the operators $\int_0^\infty(\cdot)U_n(\tilde{z};z)d\tilde{z}$ and $\int_0^\infty(\cdot)U_m(\tilde{z};z)d\tilde{z}$ to Eq. (A3), mode orthogonality reveals

$$\Delta k_n=\frac{1}{2k_n}\left[\int_0^\infty \Delta q(\tilde{z};z)U_n^2(\tilde{z};z)d\tilde{z} -\int_0^\infty (q(\tilde{z};z)-k_n^2)\Delta X(\tilde{z};z)U_n^2(\tilde{z};z)d\tilde{z}\right], \tag{A4}$$

$$a_{nm}=\frac{1}{k_n^2-k_m^2}\left[\int_0^\infty \Delta q(\tilde{z};z)U_n(\tilde{z};z)U_m(\tilde{z};z)d\tilde{z} -\int_0^\infty (q(\tilde{z};z)-k_n^2)\Delta X(\tilde{z};z) \times U_n(\tilde{z};z)U_m(\tilde{z};z)d\tilde{z}\right] \quad (n\neq m). \tag{A5}$$

The a_{nn} are found by requiring the perturbed modes to satisfy the normalization integral of Eq. (13):

$$\int_0^\infty (X+\Delta X)(U_n+\Delta U_n)^2\,d\tilde{z}=1. \tag{A6}$$

Subtracting Eq. (13) from an expanded Eq. (A6), and retaining the second-order perturbation terms, leaves

$$\int_0^\infty [2XU_n\,\Delta U_n+X(\Delta U_n)^2+\Delta X\,U_n^2 +2\,\Delta X\,U_n\,\Delta U_n]d\tilde{z}=0. \tag{A7}$$

Substituting Eq. (14) into Eq. (A7), Eq. (13) then yields

$$a_{nn}=-\frac{1}{2}\left[\sum_{m\neq n} a_{nm}^2+\int_0^\infty \Delta X(\tilde{z};z)U_n^2(\tilde{z};z)d\tilde{z} +2\sum_{m\neq n} a_{nm}\int_0^\infty \Delta X(\tilde{z};z)U_n(\tilde{z};z)U_m(\tilde{z};z)d\tilde{z}\right] \times\left[1+\int_0^\infty \Delta X(\tilde{z};z)U_n^2(\tilde{z};z)d\tilde{z}\right]^{-1}. \tag{A8}$$

[1] R. B. Evans, "A coupled mode solution for acoustic propagation in a waveguide with stepwise depth variations of a penetrable bottom," J. Acoust. Soc. Am. **74**, 188–195 (1983).

[2] C. T. Tindle, L. M. O'Driscoll, and C. J. Higham, "Coupled mode perturbation theory of range dependence," J. Acoust. Soc. Am. **108**, 76–83 (2000).

[3] C. J. Higham and C. T. Tindle, "Coupled perturbed modes and internal solitary waves," J. Acoust. Soc. Am. **113**, 2515–2522 (2003).

[4] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics*, AIP Series in Modern Acoustics and Signal Processing (American Institute of Physics, Woodbury, 1994).

[5] M. B. Porter, F. B. Jensen, and C. M. Ferla, "The problem of energy conservation in one-way models," J. Acoust. Soc. Am. **89**, 1058–1067 (1991).

[6] F. B. Jensen and C. M. Ferla, "Numerical solutions of range-dependent benchmark problems in ocean acoustics," J. Acoust. Soc. Am. **87**, 1499–1510 (1990).

[7] C. T. Tindle, A. P. Stamp, and K. M. Guthrie, "Virtual modes and the surface boundary condition in underwater acoustics," J. Sound Vib. **49**, 231–240 (1976).

Effect of ocean currents on the performance of a time-reversing array in shallow water

Karim G. Sabra and David R. Dowling[a)]
Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109-2133

(Received 13 February 2003; accepted for publication 22 September 2003)

Active acoustic time reversal may be accomplished by recording sounds with an array of transducers—a time-reversing array (TRA) or time-reversal mirror (TRM)—and then replaying the recorded and time-reversed sounds from the same array to produce back-propagating waves that converge at the location(s) of the remote sound source(s). Future active sonar and underwater communication systems suitable for use in unknown shallow ocean waters may be developed from the automatic spatial and temporal focusing properties of TRAs. However, ocean currents affect time reversal because they alter acoustic reciprocity in the environment. This paper presents a theoretical and computational investigation into how ocean currents influence TRA retrofocusing in shallow ocean environments for various array orientations. The case of TRA retrofocusing in a three-dimensional range-independent sound channel with a steady horizontal ocean current is covered here, based on a normal-mode propagation model valid for low Mach number currents. The main finding is that in the presence of ocean currents (typically <1 m/s), a TRA performs well (the associated retrofocus amplitude decay is less than 1 dB) except that a retrofocus shift (up to a few wavelengths at 500 Hz at a range of 2.5 km) may occur due to the differing interaction between the ocean current profile and each acoustic normal mode. In addition, TRA performance is predicted to depend on the array orientation relative to the ocean current direction, especially for horizontal arrays. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1625929]

PACS numbers: 43.30.Bp, 43.30.Es, 43.30.Vh [WLS] Pages: 3125–3135

I. INTRODUCTION

Active acoustic time reversal is a technique for focusing sounds recorded in complex unknown environments back to their remote point(s) of origin. It is based on the invariance of the lossless wave equation to reversing the sign of the time variable. Active acoustic time reversal involves recording sounds with an array of transmitting and receiving transducers, reversing the direction of time of the recordings, and then transmitting each time reversed recording from the transducer that received it. The sound waves produced by the array's time reversed transmission commonly propagate backwards through the acoustic environment and converge at the location(s) where they originated to produce a time-reversed, but otherwise undistorted, replica of the original sound(s). The transducer array that accomplishes this task—a time-reversing array (TRA) or time-reversal mirror (TRM)—may be of nearly any size or shape and can operate in any frequency range. TRAs perform well when absorption losses are small, the acoustic environment is reciprocal, signal-to-noise ratios are high, and the array possesses sufficient aperture. Remote sound focusing via time reversal, sometimes referred to as retrofocusing, is of interest in many applications of underwater acoustics because time reversal is often effective when the sound source location(s), transducer locations, and environmental characteristics are all unknown and other sound focusing technologies cannot be applied.

Most prior analysis and measurements of TRA performance in underwater environments have involved stationary vertical linear arrays in environments lacking medium motion. Recent studies of the time-reversal process include its experimental implementation in the ocean,[1] quantification of the spatial and temporal properties of the retrofocused field,[2,3] quantification of its sensitivity to ocean dynamics,[4] and ambient noise.[5,6] This paper extends this body of work by addressing the effects of steady depth-dependent ocean currents, and array orientation, on TRA performance in shallow water. Shallow ocean currents are ubiquitous,[7] and they can affect TRA performance by altering acoustic reciprocity. Prior research suggests that monitoring TRA performance changes may provide information about current or flow-field speed and direction. For example, uniform flow in a free space environment causes a cross-range TRA retrofocus shift.[8] TRAs have also been used to measure vortex flow[9,10] or temperature and velocity profiles[11] in ultrasonics experiments. Furthermore, current-induced acoustic reciprocity loss has been shown to influence matched-field processing (MFP) in shallow water[12] so it is likely to influence the time-reversal process in a similar manner given the common technical heritage of MFP and acoustic time-reversal.[1,13]

Practical implementation of the TRA technology for commercial or military sonar systems may require use of more nearly horizontal arrays that are bottom mounted or towed. Thus, the combined effects of ocean currents and orientation on the performance of the time-reversal process are important for applications yet are largely unquantified. Array orientation is known to play a role in stationary TRA-retrofocusing performance in shallow water, and performance findings for stationary horizontal TRAs[14] parallel those in MFP studies.[15,16] This paper extends these

a)Electronic mail: drd@engin.umich.edu

stationary-environment array-performance studies to include the effects of ocean currents.

The complicated interaction of sound with a moving multipath environment has been thoroughly investigated. Early ray acoustics models were limited to specific cases,[17,18] but general formulations were subsequently developed using the parabolic approximation[19–21] or normal modes.[22–24] Recently, more refined theoretical formulations have been developed notably for matched-field current tomography in shallow water environments.[25–29] These previous studies highlight the influence of ocean currents on acoustic propagation and the difficulty of quantitatively predicting this influence for multipath environments.

In this paper, the effect of ocean currents on the TRA retrofocus field is illustrated by using a first-order-in-Mach-number normal-mode formulation in a three-dimensional (3D) but range-independent sound channel having a steady depth-dependent horizontal current.[26,29] This perturbation approach was adopted because the Mach number of ocean currents is typically less than 10^{-3}. The main findings for typical ocean current speeds (<1 m/s) are (i) TRAs perform well except for a retrofocus shift that may occur because the ocean current interacts differently with each acoustic normal-mode, and (ii) TRA performance depends on source-array geometry relative to the current direction.

The remainder of this paper is divided into three sections. In the next section, closed-form analytical results for the retrofocus field created by a TRA in a waveguide with an ocean current are developed. Here, the differing performance of vertical and horizontal TRAs is considered and compared. In the third section, numerical simulations of the effect of ocean currents on the time-reversal process are presented for various source-array ranges and array configurations. A summary of the findings and conclusions drawn from this study are presented in the final section.

II. THEORETICAL ANALYSIS OF TRA PERFORMANCE IN A MOVING MEDIUM

A. Normal-mode formulation of the pressure field in a moving medium

The formal development starts with the modal representation of the acoustic pressure field from a harmonic point source in a moving medium[26,29] using the Cartesian coordinate system, (x,r,z)—cross range, range, and depth, respectively—defined in Fig. 1. The range axis originates at the source location $(x_s=0,r_s=0,z_s)$, lies on the ocean surface, and points toward the center of the TRA when viewed from above. The cross-range axis also lies on the ocean surface but is orthogonal to the range axis. The depth axis starts from the surface and points downward. The M TRA transducer locations are (x_j,r_j,z_j). Based on this Cartesian system, a polar coordinate system for source–receiver pairs, which yields more compact analytical expressions, can be defined as $(R_{s,j},\theta_{s,j})$, where the subscripts s, j refer to the source–receiver pair, and $\theta_{s,j}$ is measured from the cross-range axis,

$$R_{s,j}=\sqrt{r_j^2+x_j^2}, \tag{1}$$

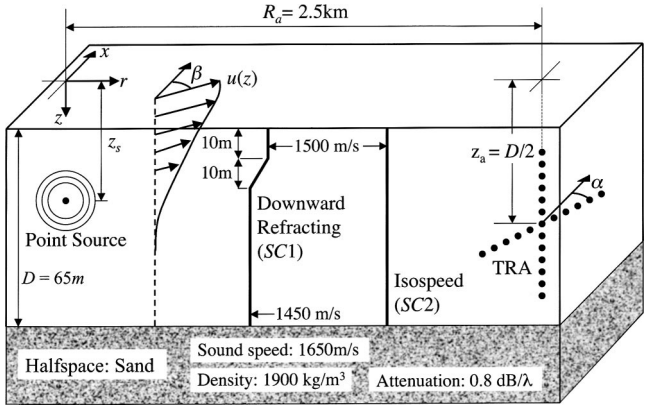


FIG. 1. Computational range-independent shallow water sound channel with geometric and environmental parameters.

$$\theta_{s,j}=\tan^{-1}(r_j/x_j).$$

Here, a three-dimensional (3D) range-independent layered environment is considered along with a steady horizontal ocean current having Cartesian components $u_x(z)=u(z)\cos(\beta)$ and $u_r(z)=u(z)\sin(\beta)$ where $u(z)$ is the current speed profile, and β is the angle between the current direction and the cross-range axis (see Fig. 1). For a given source–receiver pair (s,j), the ocean current component parallel to the source–receiver direction, $u_\parallel=u(z)\cos(\theta_{s,j}-\beta)$, and the ocean current component perpendicular to this direction, $u_\perp=u(z)\sin(\theta_{s,j}-\beta)$, have different effects on the amplitude and phase of the field generated by a harmonic point source. There is a first order effect of $u_\perp$ on propagating mode amplitudes, but it is of order 10^{-2} dB and can be neglected because it does not accumulate with increasing range.[26] The effect of $u_\perp$ on acoustic mode phases is of order $k_0R_{s,j}M_\perp^2$,[26] where k_0 is a representative wave number and $M_\perp$ is a ratio of representative values of $u_\perp$ and sound speed c_0: $M_\perp=u_\perp/c_0$. For most underwater acoustic situations, $k_0R_{s,j}M_\perp^2$ is much less than unity ($u_\perp\approx 1$ m/s, $c_0\approx 1500$ m/s, and $f=500$ Hz yield $k_0R_{s,j}M_\perp^2\leqslant 0.1$ for $R_{s,j}\leqslant 100$ km). Even so, these second-order $u_\perp$ phase contributions do not depend on the propagation direction so they cancel out during the back and forth propagation steps of time reversal. Thus, the influence of $u_\perp$ for any source–receiver pair is neglected in the following development.

For a range $R_{s,j}$ large compared to the ocean depth, $u_\parallel$ primarily influences the propagating-mode phases through changes in the modal wave number $k_n(\theta_{s,j})$ that depend on the propagation direction relative to the current flow direction. When $M_\parallel$, the ratio of a representative value of $u_\parallel$ and c_0, is small compared to unity, and small sound speed fluctuations along the propagation path are included, perturbation theory[26] produces

$$k_n^{\pm}(\theta_{s,j})=k_n^0\mp\langle k_\parallel\rangle_n+\langle\Delta k\rangle_n, \tag{2}$$

where

$$\langle\Delta k\rangle_n=\int_0^\infty dz\,\omega^2\,\frac{c(z)-c_0(z)}{\rho k_n^0}\,\frac{}{c_0(z)^3}\,\Psi_n^0(z)^2, \tag{3}$$

$$\langle k_\parallel\rangle_n=\langle k_u\rangle_n\cos(\theta_{s,j}-\beta), \tag{4}$$

$$\langle k_u\rangle_n = \int_0^\infty \frac{dz}{\rho\omega} u(z)\left[\left(\frac{d\Psi_n^0(z)}{dz}\right)^2 + k_n^2 \Psi_n^0(z)^2\right], \tag{5}$$

$n=1,2,3,...$ is the mode number index, ω is the acoustic frequency (rad/s), $c(z)$ is the sound speed profile at the time of interest, and k_n^0 and $\Psi_n^0(z)$ are the mode propagation constant and mode shape in the long-time-average environment having sound speed profile $c_0(z)$ and density profile $\rho(z)$.[36] Here, the $\Psi_n^0(z)$ are normalized so that

$$\int_0^\infty \frac{dz}{\rho}(\Psi_n^0(z))^2 = 1. \tag{6}$$

The positive and negative superscripts of the mode propagation constant $k_n^\pm$ in Eq. (2) correspond to normal modes propagating toward positive and negative r. The wave number corrections, $\langle k_\parallel\rangle_n$ and $\langle \Delta k\rangle_n$ correspond to the depth-averaged parallel-flow contribution and sound speed fluctuation effects on the nth mode, respectively.

With the above results, a first-order normal-mode formulation for the acoustic pressure field generated by a harmonic source and recorded by the jth transducer of the array is given by

$$G_{s,j}^\pm(\omega) = \sum_n \frac{e^{+i\pi/4}}{|8\pi k_n^0 R_{s,j}|^{1/2}} \Psi_n^0(z)\Psi_n^0(z_s) \times \exp(ik_n^\pm(\theta_{s,j})R_{s,j}) \tag{7}$$

when $\omega R_{s,j}/c_0$ is much greater than unity. In Eq. (7), $k_n^\pm$ depends on the mode number and frequency. Thus, each modal phase will have a different sensitivity to a given current profile $u(z)$. Among the terms in Eq. (2) for $k_n^\pm$, only $\langle k_\parallel\rangle_n$ depends on the direction of propagation. This is the nonreciprocal propagation factor added by the presence of the ocean current.

B. Retrofocused-field created by a TRA in a moving medium

In this section, Eq. (7) is used to determine the retrofocus field created by a TRA in 3D when a current is present. Here, the range axis r is specifically oriented from the source location $(r_s=0, x_s=0, z_s)$ to the geometric center of the TRA $(R_a, x_a=0, z_a)$ (see Fig. 1). The range-directed component of the ocean current is assumed to be positive, that is $u_r(z) \geqslant 0$ and $0 \leqslant \beta \leqslant \pi$.

In the situation of interest here, the standard harmonic monopole TRA formulation[1,13] for the retrofocus field P_f at frequency ω is

$$P_f(r_f, x_f z_f; \omega) = \sum_{j=1}^M [G_{s,j}^+(\omega)]^* \cdot G_{j,f}^-(\omega), \tag{8}$$

where the array has M transducers, * denotes a complex-conjugate, $G_{s,j}^+$ is the Green's function for the forward propagation (occurring downstream) from the source to the jth TRA element, and $G_{j,f}^-$ is the Green's function for the backward propagation (occurring upstream) from the jth TRA element to (r_f, x_f, z_f), a field point lying near the source location. Substituting Eq. (7) into Eq. (8), the expression for the retrofocus field becomes

$$P_f(r_f, x_f, z_f; \omega) = \sum_{j=1}^M \sum_{n=1}^N \sum_{l=1}^N \frac{\Psi_l^0(z_f)\Psi_n^0(z_s)}{8\pi\rho(z_s)} \frac{\Psi_l^0(z_j)\Psi_n^0(z_j)}{\rho(z_j)} \times \frac{\exp(-ik_n^+(\theta_{s,j})\sqrt{r_j^2+x_j^2} + ik_l^-(\theta_{s,j})\sqrt{(r_j-r_f)^2+(x_j-x_f)^2})}{\sqrt{k_n^0\sqrt{r_j^2+x_j^2}\cdot k_l^0\sqrt{(r_j-r_f)^2+(x_j-x_f)^2}}}. \tag{9}$$

The phase arising from the current accumulates during both the forward and backward propagation[9,10] while the phase arising from sound speed fluctuations is essentially cancelled[9,10,13] when the field point is near enough to the source location. Thus, for the remainder of this paper, only current effects will be considered.

Based on simplification of Eq. (9), the accumulated phase of the retrofocused field, in the vicinity of the source location, is on the order of $\Gamma = 2M_\parallel R_a\omega/c_0$, where R_a is the distance from the source to the center of the TRA. For $f=500$ Hz, $u=0.15$ m/s ($M_\parallel = 10^{-4}$) and $c_0=1500$ m/s, then $\Gamma > 0.1$ rad for $R_a > 240$ m. Thus, TRAs may be sensitive to ocean currents in many situations of practical interest.

Further simplification of Eq. (9) can be achieved once a source-TRA geometry is specified. Vertical and horizontal linear TRAs are considered in the next two sections.

C. Vertical TRA

This section will present formal results for the effects of ocean currents on the performance of a vertical linear TRA. The phenomena highlighted here are also active in the horizontal linear TRA case.

The point of departure for the vertical array analysis is Eq. (9) with the jth array element located at $(x_j=0, r_j=R_a, z_j)$. The mode orthonormality property for a vertical water-column-spanning array with a sufficient number of transducers[1,14] having a uniform vertical spacing Δ_z is

$$\sum_{j=1}^M \frac{\Psi_n(z_j)\Psi_l(z_j)}{\rho(z_j)} \simeq \frac{1}{\Delta_z}\int_0^H \frac{\Psi_n(z)\Psi_l(z)}{\rho(z)}dz = \frac{\delta_{ln}}{\Delta_z}. \tag{10}$$

Using this property and neglecting the small contribution of the current effect in the nonreciprocal modal phase within the retrofocus zone itself and the small amplitude adjustment in the cylindrical square-root-of-range spreading factor, and setting $\theta_{s,j}=\pi/2$ for each source–receiver pair, Eq. (9) can be simplified for a vertical linear TRA:

$$P_f(r_f,x_f=0,z_f;\omega)=\sum_{n=1}^{N}\frac{\Psi_n(z_f)\Psi_n(z_s)}{8\pi\rho(z_s)\Delta_z}\frac{\exp(i(2R_a\sin(\beta)\langle k_u\rangle_n-k_n^0(\omega)r_f))}{k_n^0(\omega)R_a}. \tag{11}$$

This expression shows that the location of maximum retrofocus acoustic intensity, will depend on $u_r(z)=u(z)\sin(\beta)$ through $\langle k_u\rangle_n$ and on which modes are excited by the source. For example, if the nth and lth modes are predominantly excited by the source, then the maximum-intensity location, $r_{f,\max}$, will be shifted from the source location ($r_s=0$), by an amount

$$r_{f,\max}=\frac{2R_a\sin(\beta)(\langle k_u\rangle_l-\langle k_u\rangle_n)}{k_l^0-k_n^0}. \tag{12}$$

Thus $r_{f,\max}$ might be positive or negative depending on modal characteristics and modal interactions with the current profile $u(z)$. For instance, if $k_l^0>k_n^0$ and $u(z)>0$ across the water depth, the sign of $r_{f,\max}$ can be negative if $\langle k_u\rangle_n \geqslant\langle k_u\rangle_l$. This situation may occur when the location of the maxima of the current profile through the depth coincides with more antinodes of mode n than mode l. Therefore, in some environments for some source depths, the retrofocus location may not be pushed downstream by the current profile but may be shifted upstream instead. This horizontal shift could be compensated using techniques based on the waveguide invariant,[31] if its value is known.

And, as final note, Eq. (11) also shows that P_f is independent of the cross-range current component. Such independence is expected at this level of approximation since all the array's transducers lie on the same azimuth from the source. Such independence is generally not expected for a horizontal array.

D. Horizontal TRA

In this section, the TRA is horizontal, linear, and located at depth z_a. The source-to-center-of-array range is R_a and the spacing Δ_H between the M array elements is constant for an overall array length of $L=(M-1)\Delta_H$. The element spacing, array length, and array depth were chosen based on suggestions in prior MFP[15,16] and TRA studies.[14] The TRA's orientation angle, measured from the x axis, is α with $\alpha=0$ ($\pi/2$) placing the array broadside (endfire) to the source (see Fig. 1). Thus, the coordinates of the jth TRA element are

$$\left(x_j=\left(j-\frac{M+1}{2}\right)\Delta_H\cos(\alpha),\right.$$

$$\left.r_j=R_a+\left(j-\frac{M+1}{2}\right)\Delta_H\sin(\alpha),z_a\right). \tag{13}$$

Unfortunately, no simplification of Eq. (9) based on mode orthogonality is possible for a horizontal array having elements at constant depth. However, horizontal arrays are commonly treated by separating the double modal sum in Eq. (9) into diagonal ($l=n$) and off-diagonal ($l\neq n$) terms[15] because the off-diagonal terms will tend to cancel with each other while the diagonal terms will tend to reinforce each other. In the case of TRA retrofocusing in a moving medium, the additional phase terms due to the ocean current have no features that tend to change the mutual cancellation of the off-diagonal terms. Thus, the influence of the ocean current will be evaluated only for the diagonal terms of Eq. (9). Using Eq. (13), and approximating the amplitude decay associated with the square-root cylindrical range spreading by $1/(k_nr_j)$ for small values of r_f and x_f in Eq. (9), the following expression for the diagonal terms of the retrofocused field can be obtained:

$$[P_f(r_f,x_f,z_f,\omega)]_{l=n}=\sum_{n=1}^{N}\left[\frac{\Psi_n(z_f)\Psi_n(z_s)}{8\pi\rho(z_s)}\cdot\frac{\exp(i(2R_a\sin(\beta)\langle k_u\rangle_n-k_n^0(\omega)r_f))}{k_n^0(\omega)R_a}\right]\frac{|\Psi_n(z_a)|^2}{\rho(z_a)}\cdot F_n(\alpha,\beta), \tag{14}$$

where

$$F_n(\alpha,\beta)=\sum_{j=1}^{M}\frac{\exp\left(i\left(2\langle k_u\rangle_n\cos(\alpha-\beta)-\dfrac{k_n^0(\omega)x_f\cos(\alpha)}{R_a+\Delta_H\left(j-\dfrac{M+1}{2}\right)\sin(\alpha)}\right)\Delta_H\left(j-\dfrac{M+1}{2}\right)\right)}{1+\Delta_H\left(j-\dfrac{M+1}{2}\right)\Big/R_a}. \tag{15}$$

Although more difficult to interpret, this result shares some common features with the vertical array result. In Eq. (14), the terms inside the square brackets in the modal sum are similar to those in Eq. (11). Therefore if the retrofocus exists, it may be shifted in range when the ocean current component in the range direction, $u(z)\sin(\beta)$, is nonzero, just like the vertical TRA case. However, each term in the modal sum is multiplied by two additional factors (1) an array-depth dependent modal

amplitude factor $|\Psi_n(z_a)|^2/\rho(z_a)$, and (2) a range and orientation dependent modal phase and amplitude correction function $F_n(\alpha,\beta)$. The first factor creates an additional weighting by the vertical structure of the propagating modes, independently of ocean currents, and is responsible for persistent retrofocus side lobes in depth.[14] The second factor is more critical for the present discussion of TRA performance because the phase in Eq. (15) explicitly depends on the cross range location, x_f, the current magnitude $u(z)$ through $\langle k_u\rangle_n$, the current direction, and the TRA's orientation. This term produces cross-range retrofocus shifts.

When the source-array range is much larger than the spatial extent of the array [i.e., $\Delta_H(j-(M+1)/2)/R_a \ll 1$], the previous expression of $F_n(\alpha)$ can be simplified, yielding

$$F_n(\alpha,\beta) \simeq \frac{\sin(M\Delta_H(\langle k_u\rangle_n \cos(\alpha-\beta) - k_n^0(\omega)x_f\cos(\alpha)/(2R_\alpha)))}{\sin(\Delta_H(\langle k_u\rangle_n \cos(\alpha-\beta) - k_n^0(\omega)x_f\cos(\alpha)/(2R_a)))}. \tag{16}$$

This simplified form of F_n depends on R_a, α, $u(z)$, β, M, Δ_H, the mode shapes, and their horizontal wave numbers. Thus, effective prediction of horizontal TRA performance in the presence of nontrivial currents requires both environmental and geometric information.

Here, the current can produce both range and cross-range retrofocus shifts ($r_{f,\max}$ and $x_{f,\max}$, respectively) from the initial source location. As for the vertical array, $r_{f,\max}$ depends *only* on $u_r(z)=u(z)\sin(\beta)$ and is set through the interference of the modal phases. The cross-range shift, $x_{f,\max}$, is set by $F_n(\alpha,\beta)$ and depends on the TRA's orientation angle α, and the relative orientation of the TRA with respect to the ocean current direction, $\alpha-\beta$. Equations (14) and (16) show that only TRAs having some horizontal aperture may be sensitive to $u_x(z)=u(z)\cos(\beta)$.

The simplified expression for $F_n(\alpha,\beta)$, Eq. (16), suggests two particular TRA orientations might yield interesting results. First, when the TRA is endfire to the source ($\alpha=\pi/2$), any non-zero u_r will systematically degrade TRA performance because it will induce an uncompensated systematic phase shift at the retrofocus that is different for each element of the array. Thus, the TRA retrofocus field amplitude will drop by an amount set by the product of $\langle k_u\rangle_n$ and the array parameters M and Δ_H. This result is similar to the case of a stationary source and a towed endfire array.[30] Second, if the array and current are perpendicular ($\alpha=\beta\pm\pi/2$), then the cross-range retrofocus shift is zero.

III. NUMERICAL SIMULATIONS

This section presents numerical simulations of shallow ocean TRA retrofocusing with nontrivial ocean currents based on Eq. (9). The simulation results cover a vertical and two horizontal array configurations in the generic range-independent shallow water sound channel depicted in Fig. 1. The primary simulation goal was to evaluate TRA performance as a function of the ocean current magnitude and direction for a few specific cases.

A. Computational approach

Two range-independent shallow water sound channels are considered here. The first one (SC1) has a downward refracting sound speed profile. The second one (SC2) has constant sound speed; its results are included where appropriate for comparison with SC1 results. The depth, $D=65$ m, and bottom properties are identical for SC1 and SC2, and the Cartesian coordinates defined in the preceding section are also used here.

The normal mode program KRAKEN[32,33] was used to compute the mode shape functions, $\Psi_n(z)$, and horizontal wave numbers, k_n^0, at the frequencies of interest. At 500 Hz, SC1 and SC2 support 21 and 18 propagating modes, respectively. The first 12 modes of both sound channels are illustrated in Fig. 2. Although there are mode-shape differences in every case, the most important modal difference is the trapping of the first nine modes of SC1 below the thermocline compared to the modes of SC2. Thus, the higher order modes in SC1 are more likely to be affected by a surface current than the low order modes, while for SC2 nearly all the modes may be affected by a surface current.

In addition to the two sound channels, generic flow profiles for tidal and wind-driven currents were considered. A parabolic profile[34] was used for the tidal current with a 1/7 power law fitted near $z=D$ to enforce a no-slip boundary condition at the ocean bottom:[35]

$$u(z) = U_s\left(1-0.89\left(\frac{1}{2}\left(\frac{D-z}{D}\right)^2 - \frac{D-z}{D} + 1/2\right)\right) \times \left(\frac{D-z}{D}\right)^{1/7}, \tag{17}$$

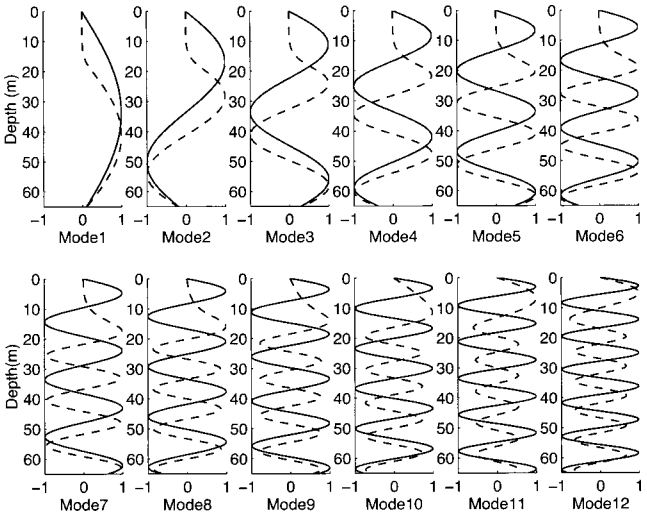


FIG. 2. Shapes of the first 12 propagating modes at 500 Hz for the downward refracting sound channel, SC1 (dashed lines), and the isospeed sound channel, SC2 (plain lines).

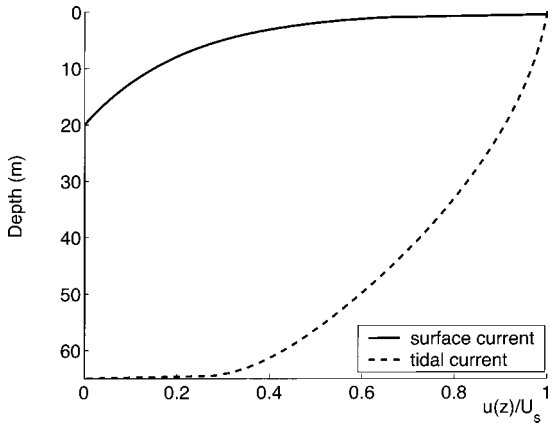


FIG. 3. Flow profiles for tidal and wind-driven (surface) ocean currents.

where U_s is the current speed at the ocean surface. This current profile is shown as the dashed curve in Fig. 3.

Wind-driven currents primarily affect near-surface water. In analogy with flat-plate turbulent boundary layer flow, the surface current was assumed to decrease logarithmically with depth below the wind-driven layer:[7]

$$u(z)=U_s, \quad z<z_0,$$
$$u(z)=U_s\cdot\left(1-\frac{\log(z/z_0)}{\log(z/z_c)}\right), \quad z_0\leqslant z\leqslant z_c, \tag{18}$$
$$u(z)=0, \quad z>z_c$$

where again U_s is the current speed at the ocean surface, and z_0 is a roughness length scale depending on the surface waveforms and the wind shear stress. Here, an empirical relationship[7] was used: $z_0=\eta U_s^2$ for $0\leqslant U_s\leqslant 2$ m/s, with $\eta=0.19\ \mathrm{s^2/m}$. In Eq. (18) z_c, the depth at which the wind-driven flow falls to zero, was set to $z_c=20$ m, to match the lower end of the thermocline in SC1. This current profile is shown as the solid curve in Fig. 3.

The next two sections present vertical and horizontal TRA simulations results for SC1 and SC2 for both current profiles.

B. Vertical TRA

The results of Sec. II C suggest that a retrofocus range shift, $r_{f,\max}$, may occur when a vertical TRA is operated in an ocean sound channel having a nontrivial current, and that retrofocus shift should increase with the range directed current speed, $U_s\sin(\beta)$. The differences in the SC1 and SC2 mode shapes highlighted above suggest that such TRA retrofocus shifting in the two channels may differ depending on the current type and the source depth. Therefore, a total of eight subcases (two sound channels, two current types, and shallow or deep source depths) are potentially interesting. Thus, to simplify the presentation of these results, only a single vertical TRA centered in the water column having 21 elements with uniform 3 m spacing is considered. In addition, the source-array range is fixed at $R_a=2.5$ km. For conciseness, results for shallow, $z_s=6$ m, and deep, $z_s=54$ m, sources are presented simultaneously because there is little or no side lobe overlap under these conditions. Results for simple 500 Hz tones and 500-Hz-center-frequency pulses (Gaussian windowed sine waves, 99% signal energy bandwidth 258 Hz, 3 dB down signal energy bandwidth 139 Hz) are included in the line plots described below. For the harmonic (pulse) source signal the retrofocus location is the field point where the TRA produced field achieves its peak intensity (energy).

To properly launch the comparisons, Fig. 4 shows the retrofocused field amplitudes at 500 Hz (in dB relative to the louder deep retrofocus peak) in depth vs range coordinates in the vicinity of the shallow and deep sources when $U_s=0$ for SC1 [Fig. 4(a)], and SC2 [Fig. 4(b)]. The array is off to the right in both cases. These results represent best-case TRA performance since acoustic reciprocity holds when the current is absent. The difference in shallow and deep retrofocus amplitude results from the greater importance of attenuation on the higher order propagating modes for TRA retrofocusing on the shallow source.

Simulation results with nonzero current are provided in Fig. 5 for SC1 and in Fig. 6 for SC2. Parts (a) and (c) of these figures plot the retrofocus range shift $r_{f,\max}$ vs $U_s\sin(\beta)$ for the tidal and surface currents, respectively, with the curves marked by black dots corresponding to the signal pulses. Parts (b) and (d) of these figures show the retrofocus field amplitude at 500 Hz (in dB) when $U_s\sin(\beta)=2$ m/s for the tidal and surface currents, respectively. The dB references for these figures is the same as for Fig. 4 but each panel has its own grayscale settings to provide sufficient contrast.

A comparison of the various subcases, both with and without currents, shows that currents may cause a slight TRA

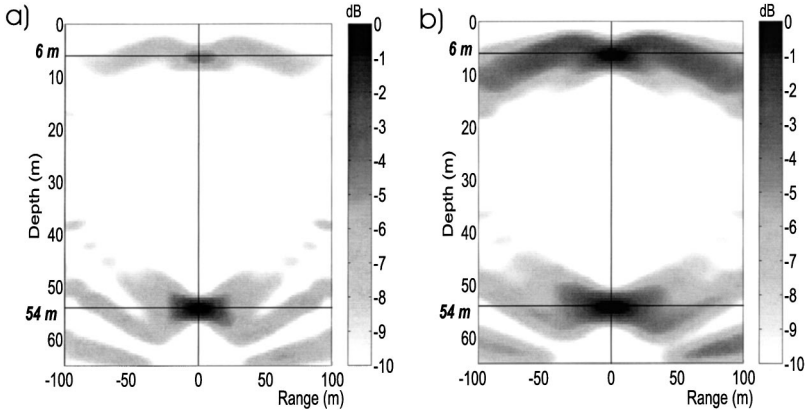


FIG. 4. Vertical TRA retrofocused energy (in dB) in the vicinity of shallow ($z_s=6$ m) and deep ($z_s=54$ m) sources for SC1 (a) and SC2 (b) without a current. Retrofocused energy values in each sound channel are normalized by the deep sources energy peak. The array is located to the right of each frame at $R_a=2.5$ km.

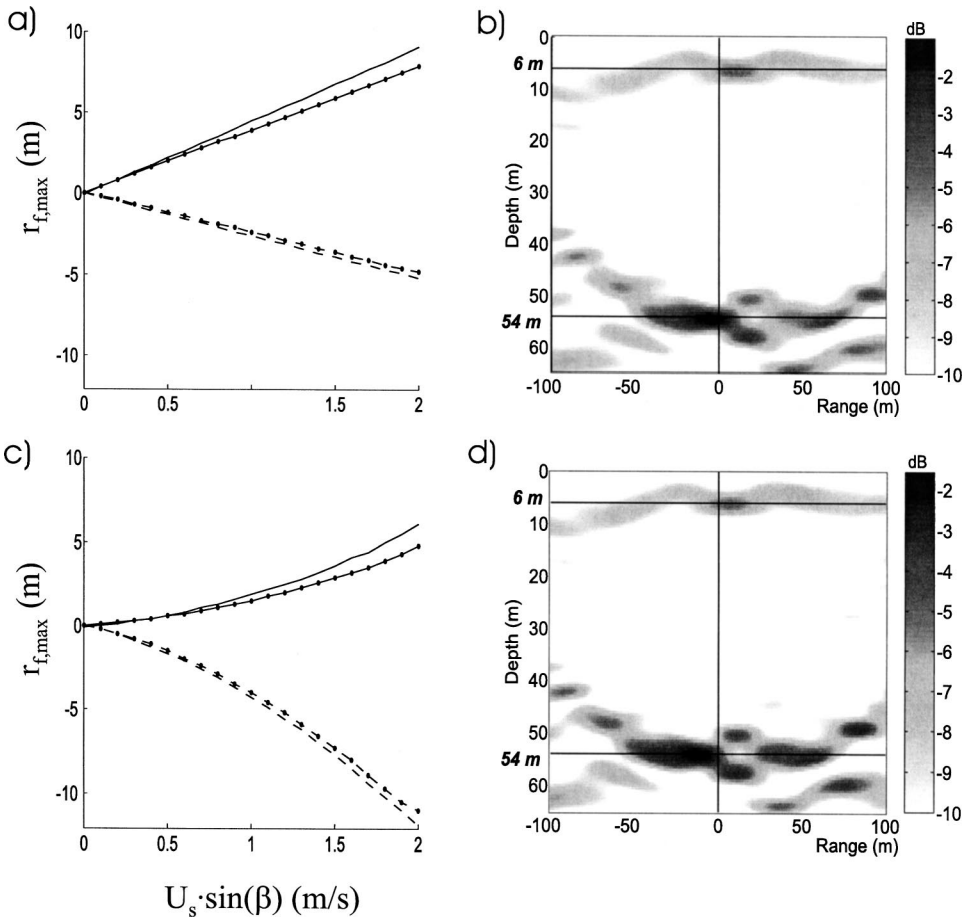


FIG. 5. Vertical TRA retrofocused energy (in dB) in the vicinity of shallow ($z_s = 6$ m) and deep ($z_s = 54$ m) sources in the downward refracting sound channel, SC1, for both current types. Range direction retrofocus shifts, $r_{f,\max}$, are plotted in (a) for the tidal current and in (c) for the surface current vs $U_s \sin(\beta)$. In (a) and (c), the solid lines are for shallow sources, the dashed lines are for deep sources, the lines without dots are for a 500 Hz harmonic signal, and the lines with black dots are for a 500 Hz center-frequency broadband pulse. Sample retrofocus fields showing results for both deep and shallow sources at $U_s \sin(\beta) = 2$ m/s are provided in (b) for the tidal current and in (d) for the surface current.

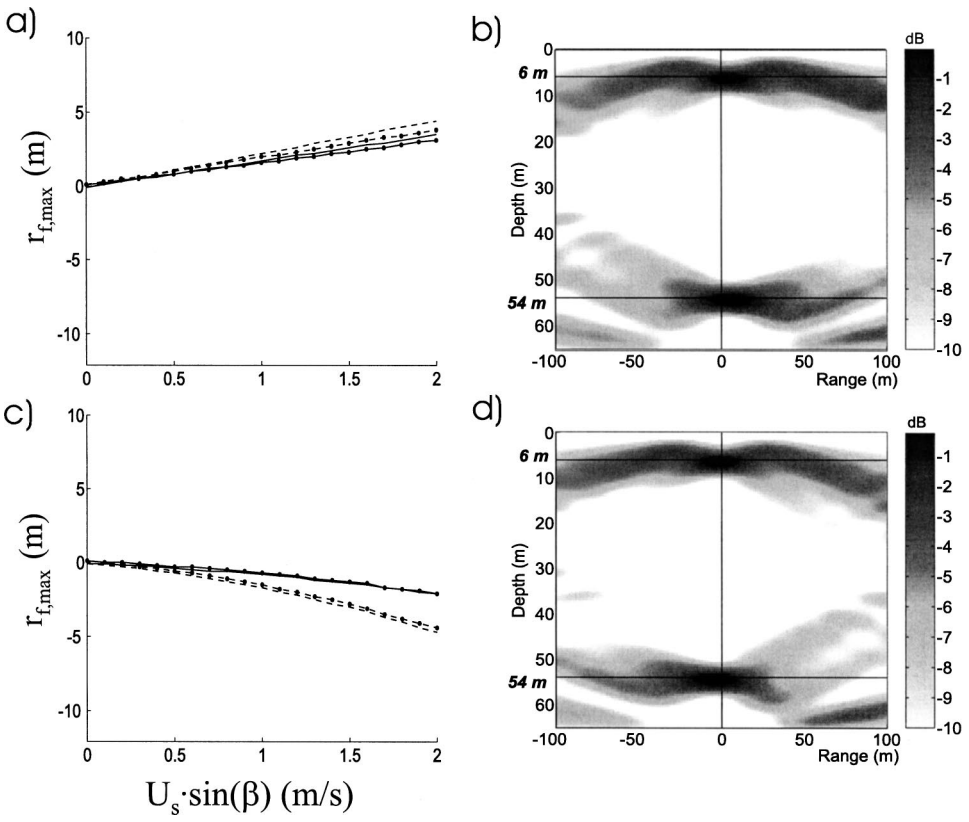


FIG. 6. Same as Fig. 6 except these results are for the isospeed sound channel, SC2.

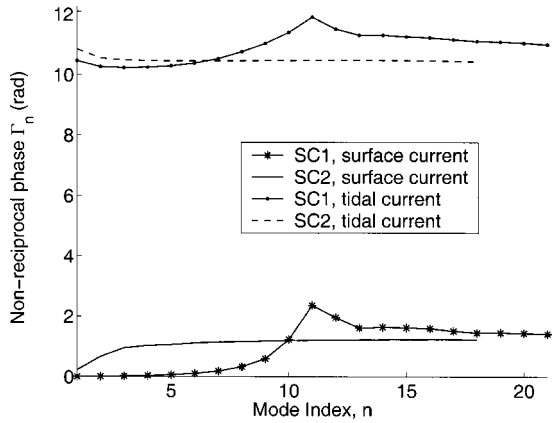


FIG. 7. Values of $\Gamma_n = 2R_a\langle k_u\rangle_n$, the nonreciprocal modal phase for both types of ocean currents [see Eq. (2)] for the propagating modes of SC1 (downward refracting) and SC2 (isospeed) when $R_a = 2.5$ km.

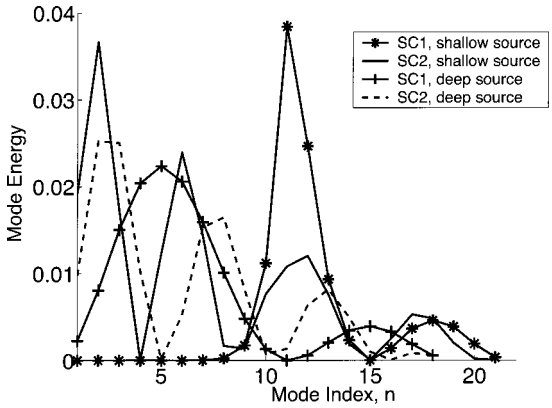


FIG. 8. Relative energy of the normal modes, $\Psi_n^0(z_s)^2 \exp(2\zeta_n R_a)$, excited by the shallow and deep sources in the sound channel SC1 and SC2 after the back and forth propagation when $R_a = 2.5$ km. The modal energies shown here depend on the chosen source level (arbitrary), the source depths (6 m or 54 m), and mode characteristics.

retrofocus amplitude decay, and an upstream or downstream retrofocus range shift. The amplitude decay, which is accompanied by broadening of the focal spot, arises because the mode-specific current-induced phase decreases the intended constructive interference of the dominant modes forming the TRAs retrofocus field. These figures also show that current-induced retrofocus shifts are weakly dependent on signal bandwidth in this frequency range (approximately 400 to 600 Hz), so only harmonic signal results at 500 Hz are presented in the rest of this paper.

The retrofocus range shifts, and their surprising propensity to be of either sign, are caused by the dichotomy between low and high order modes, and variations in the nonreciprocal current-induced modal phases. For each of the four sound-channel and current combinations, Fig. 7 displays the values of the nonreciprocal modal phase $\Gamma_n = 2R_a\langle k_u\rangle_n$, at 500 Hz and $R_a = 2.5$ km. The tidal current extends deeper into the sound channel and thus generates the larger values of Γ_n. A WKB expression[36] for the mode shape $\Psi_n^0(z)$ and Eq. (5) can be used to find

$$\Gamma_n \approx 2R_a \sin(\beta)\frac{\omega}{c_{\text{av}}^2}\int_0^\infty \frac{dz}{\rho(z)}u(z)\Psi_n^0(z)^2, \tag{19}$$

where c_{av} is a depth-averaged value of the speed of sound. This approximation highlights the dependency of Γ_n on the integral of the product $u(z)\Psi_n^0(z)^2$, and provides a means for explaining the numerical results. Good TRA retrofocusing occurs when all the source-excited modes are returned to the original source location in phase. Figure 7 shows that the variations in Γ_n are monotonic for SC2 but not for SC1. This occurs because the thermocline in SC1 induces a dichotomy between the high and low order modes. Thus, shallow and deep sources that excite the modes differently can be expected to lead to different retrofocus shifts as well.

Figure 8 shows the relative energy in the modes, $\Psi_n^0(z_s)^2 \exp(2\zeta_n R_a)$, excited by the shallow and deep sources in SC1 or SC2 after the back and forth propagation. Here, ζ_n is the modal attenuation caused by bottom absorption. Based on Figs. 7 and 8, the variations of Γ_n across a subset of modes determine the retrofocus location. More precisely, if two modes dominate and Γ_n decreases (increases) for increasing mode index n, then the retrofocus range shift, $r_{f,\max}$ given by Eq. (12), is positive (negative) since k_n^0 decreases with increasing n. In SC2, both sources excite low and high order modes (see Fig. 8) and Γ_n is a monotonic function of n, slightly decreasing for the tidal current and slightly increasing for the surface (see Fig. 7). Therefore the retrofocus is advected downstream for the tidal flow [Figs. 6(a) and 6(b)] and upstream for the surface flow [Figs. 6(c) and 6(d)].

For the downward refracting sound channel (SC1), a dichotomy between the high and low order modes induces different retrofocus behavior for shallow and deep sources. The high order modes ($n \geqslant 11$) penetrate above the thermocline while the low order modes ($n \leqslant 10$) do not (see Fig. 2). Thus, the low order modes are not strongly excited by the shallow source (see Fig. 8), but the first near-surface antinode of the lower order modes approaches the surface with increasing n. Therefore $\langle k_u\rangle_n$ and Γ_n are generally increasing functions of the mode index n up through $n \approx 10$ or 11 (see Fig. 7) because both current types have maximum flow speed at the ocean surface. Conversely, the high-order modes in SC1 are more strongly excited by the shallow source (see Fig. 8) and the main antinode of these modes diminishes in importance (see Fig. 2) with increasing mode index n. Consequently, $\langle k_u\rangle_n$ and Γ_n are decreasing functions of the mode index n for $n \geqslant 11$. Hence, the retrofocus is advected downstream for the shallow source, and upstream for the deep source, for both the tidal [Figs. 5(a) and 5(b)] and surface currents [Figs. 5(c) and 5(d)].

C. Horizontal TRA

In this section, to facilitate comparisons, the TRA again operates at 500 Hz, has $M = 21$ elements, and is linear with its center lying 2.5 km from the source. However, it is deployed horizontally at constant depth, $z_a = D/2 = 32.5$ m, and, to prevent side lobe issues unrelated to currents from obscuring the intended results, the deep and shallow point sources of the prior subsection are replaced by a single source at $z_s = 27$ m. Numerical simulations were performed for two TRA's orientation angles: $\alpha = \pi/2$ (an endfire array), and $\alpha = \pi/3$ (an array between broadside and endfire). In

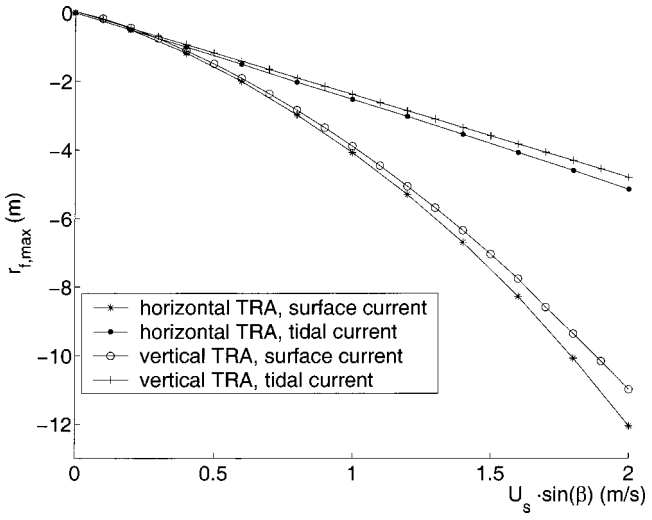


FIG. 9. Retrofocus range shift, $r_{f,\max}$ vs $U_s \sin(\beta)$ in the downward refracting sound channel SC1 for tidal and surface currents when $z_s=27$ m and $R_a=2.5$ km. The horizontal TRA is linear, at depth $z_a=D/2=32.5$ m and oriented at $\alpha=\pi/3$. Vertical TRA results in same environment are shown for comparison.

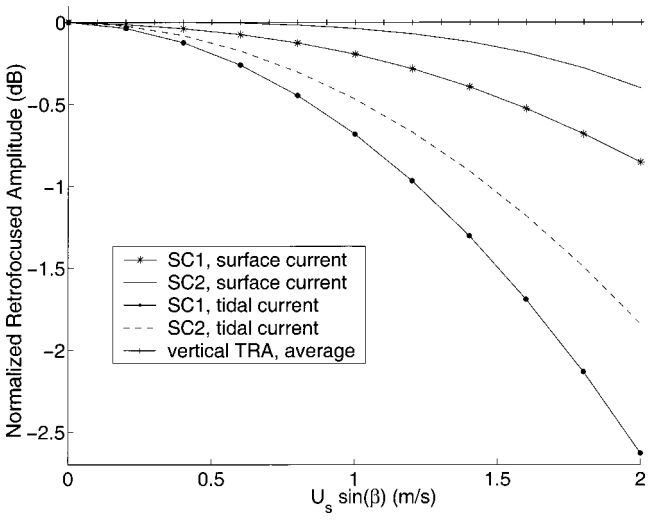


FIG. 11. Decay of the retrofocus amplitude, P_f, at the shifted retrofocus location $(r_{f,\max}, x_{f,\max}, z_s)$ versus $U_s \sin(\beta)$ when $z_s=27$ m and $R_a=2.5$ km. The horizontal TRA is linear, at depth $z_a=D/2=32.5$ m, and oriented endfire to the source. Average vertical array results for the four sound-channel/current-profile possibilities are shown for comparison.

both cases, the array length L is 500 m and the element spacing is $\Delta_H=25$ m. These choices are consistent with prior studies of horizontal arrays.[14–16] As in the preceding section, TRA performance is evaluated for both surface and tidal ocean currents in both sound channels, thus yielding eight environmental combinations of sound channel type, ocean current profile, and TRA orientation angle.

Figures 9 and 10 illustrate the influence of the ocean current on the effective retrofocus location $(r_{f,\max}, x_{f,\max}, z_s)$. The first of these displays $r_{f,\max}$ versus the range-direction current speed, $U_s \sin(\beta)$, for the sound channel SC1. The cross range current component, $U_s \cos(\beta)$ has little influence on $r_{f,\max}$. Only results for $\alpha=\pi/3$ and SC1 are shown here because the results for an endfire TRA configuration were almost identical to those shown. In addition, for a given current-profile/sound-channel pair, the behavior of the retrofocus range shift is similar to that of the vertical TRA with the deep source; the retrofocus occurs upstream since the source depth $z_s=27$ m lies below the thermocline. Thus only SC1 results are shown on Fig. 9. Differences in $r_{f,\max}$ for vertical and horizontal arrays are mainly due to the extra weighting factor, $|\Psi_n^0(z_a)|^2/\rho(z_a)$, appearing in the horizontal TRA formulation.

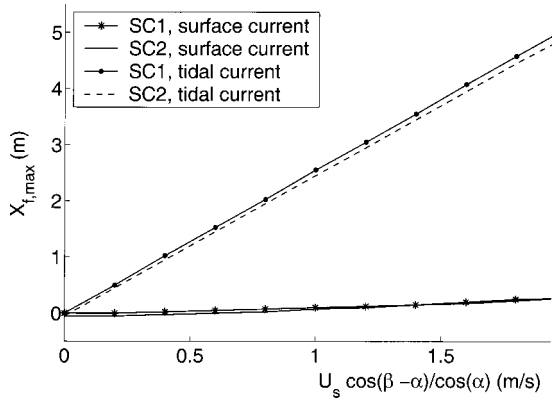


FIG. 10. Retrofocus cross-range shift, $x_{f,\max}$ vs a normalized current magnitude $U_s \cos(\beta-\alpha)/\cos(\alpha)$ when $z_s=27$ m and $R_a=2.5$ km. The TRA is linear, horizontal, at depth $z_a=D/2=32.5$ m and oriented at $\alpha=\pi/3$. The tidal current produces larger cross-range shifts.

Figure 10 shows $x_{f,\max}$ versus a normalized current magnitude $U_s \cos(\beta-\alpha)/\cos(\alpha)$, suggested by inspection of Eq. (16). Again, only results for $\alpha=\pi/3$ are displayed because $x_{f,\max}=0$ for an endire source-array orientation. These simulations show that $x_{f,\max}$ is nearly proportional to $U_s \cos(\beta-\alpha)/\cos(\alpha)$ and occurs in the downstream direction as expected from Eq. (16) and prior work.[8] In addition, the tidal current produces larger $x_{f,\max}$ values than the surface current because its greater depth extent ensures that $\langle k_u \rangle_n$ for any mode is larger than for the surface current.

Figure 11 displays the normalized retrofocus amplitude P_f, at $(r_{f,\max}, x_{f,\max}, z_s)$ versus $U_s \sin(\beta)$ for both sound channels, both types of currents, and for endfire and vertical TRAs at $R_a=2.5$ km. In each case, the TRA retrofocus amplitude is normalized by its maximum value in the absence of a current (i.e., $U_s=0$). In addition, the $\alpha=\pi/3$ TRA simulations (not shown) predict less amplitude decay for each case than the endfire results shown, but not as little as the vertical array. Several important observations can be made from this figure. The decay of P_f remains small (less than one decibel) even when the ocean current speed reaches 1 m/s. For a given sound channel, the decay of the retrocus amplitude for the endfire configuration is more pronounced for a tidal current than for a surface current. And, the vertical array has the best performance (cross line).

IV. SUMMARY AND CONCLUSIONS

Practical underwater applications of the acoustic time reversal are likely to face the influence of ocean currents which alter acoustic reciprocity in the environment. Fortunately, ocean current Mach numbers are small, ($\approx 10^{-3}$), so the perturbation approach described herein is likely to be

adequate for predicting TRA performance in the ocean. In an ocean waveguide, the differing interaction between each propagating acoustic normal mode and the type of ocean current may create a retrofocus shift, which depends on the source-array geometry, sound channel characteristics, and ocean current properties.

The research reported here leads to three main conclusions. First, for all TRA orientations, the retrofocus shift in the source-array range direction is proportional to the range-directed component of the current, and occurs upstream or downstream from the initial source location depending on the detailed influence of the ocean current on each acoustic mode. Low or high order modes may have different sensitivities to surface or tidal currents so deep or shallow sources may produce TRA retrofoci shifted in opposite directions in the same sound channel with the same current. Second, only horizontal TRAs away from the endfire orientation appear to be sensitive to cross-range ocean currents. When a cross range retrofocus shift does occur, it depends on the horizontal TRA's orientation angle with respect to the source-to-center-of-array direction and on its orientation with respect to the ocean current direction. However any cross range focus shift always occurs in the downstream direction, and no cross-range retrofocus shift should be observed when the TRA is oriented perpendicular to the ocean current. This property might be used to determine or monitor the primary direction of an unknown ocean current. Third the decay of the retrofocused field induced by modal phase shifts accumulated along the range direction in the presence of oceanic currents, is mild. However, this flow-induced decay is more critical for horizontal TRAs orientation close to the endfire configuration.

At first thought, TRA retrofocus shifts might be used to estimate ocean current properties. However, it appears that the utilization of TRAs as a tool for estimating ocean currents is contingent on the knowledge of the normal mode shapes supported by the ocean sound channel. In a multipath environment, an ocean current affects each mode differently. Therefore, the resulting retrofocus field depends on numerous ocean-current/normal-mode interactions. The differing nonreciprocal phases accumulated in the range direction during the back and forth acoustic propagation between the source and the TRA contain information about the ocean-current properties. Classical ocean tomography techniques[37] are not successful in inverting for current properties when acoustic rays cannot be identified separately. One may think of using a vertical TRA and extracting the normal modes at the array location once the recording is done,[38,39] and then backpropagating the different modes separately toward the source in order to identify nonreciprocal phases for each mode. This would allow the separation of the various contributions of the ocean current before proceeding to inversion for ocean-current properties (such as magnitude or depth dependency). On the other hand, existing techniques, such as matched-field tomography,[25,40,41] have been proven to efficiently address the inversion for an ocean current once the multipath environment is known. Such efforts require large calculations using different types of objective functions and extensive parametric searches over the ocean-current profile characteristics (magnitude, direction, depth-dependency, etc.). Nonreciprocal modal phase measurements issued from preliminary time-reversal experiments might be used to restrict the initial search space for matched-field tomography.

ACKNOWLEDGMENTS

The research was sponsored by the Ocean Acoustics Program of the Office of Naval Research under Grant No. N00014-96-1-0040. Portions of this work were presented at the 145th ASA meeting, Cancun, Mexico, December 2002.

[1] W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and D. R. Jackson, ''Phase-conjugation in the ocean: experimental demonstration of an acoustic time reversal mirror,'' J. Acoust. Soc. Am. **103**, 25–40 (1998).
[2] P. Roux and M. Fink, ''Time reversal in a waveguide: Study of the temporal and spatial focusing,'' J. Acoust. Soc. Am. **107**, 2418–2429 (2000).
[3] S. Kim, G. F. Edelmann, W. A. Kuperman, W. S. Hodgkiss, and H. C. Song, ''Spatial resolution of time-reversal arrays in shallow water,'' J. Acoust. Soc. Am. **110**, 820–829 (2001).
[4] M. R. Dungan and D. R. Dowling, ''Computed narrowband time-reversing array retrofocusing in a dynamic shallow ocean,'' J. Acoust. Soc. Am. **107**, 3101–3112 (2000).
[5] S. R. Khosla and D. R. Dowling, ''Time-reversing array retrofocusing in noisy environments,'' J. Acoust. Soc. Am. **109**, 538–546 (2001).
[6] K. G. Sabra, S. R. Khosla, and D. R. Dowling, ''Broadband time-reversing array retrofocusing in noisy environments,'' J. Acoust. Soc. Am. **111**, 823–830 (2002).
[7] K. F. Bowden, *Physical Oceanography of Coastal Waters* (Wiley, New York, 1983), pp. 16–69, 124–149.
[8] D. R. Dowling, ''Phase-conjugate array focusing in a moving medium,'' J. Acoust. Soc. Am. **94**, 1716–1718 (1993).
[9] P. Roux and M. Fink, ''Experimental evidence in acoustics of the violation of time-reversal invariance induced by vorticity,'' Europhys. Lett. **32**, 25–29 (1995).
[10] S. Manneville, J. H. Robres, A. Maurel, P. Petitjeans, and M. Fink, ''Vortex dynamics investigation using an acoustic technique,'' Phys. Fluids **11**, 3380–3389 (1999).
[11] A. Maurel and G. Bouchet, ''Simultaneous characterization of temperature and velocity profiles using time reversal: Numerical study of free convection,'' Europhys. Lett. **53**, 584–590 (2001).
[12] M. H. Orr and P. C. Mignerey, ''Matched-field processing gain degradation caused by tidal flow over continental shelf bathymetry,'' J. Acoust. Soc. Am. **111**, 2615–2620 (2002).
[13] D. R. Jackson and D. R. Dowling, ''Phase conjugation in underwater acoustics,'' J. Acoust. Soc. Am. **89**, 171–181 (1991).
[14] M. R. Dungan and D. R. Dowling, ''Orientation effects on linear time-reversing array retrofocusing in shallow water,'' J. Acoust. Soc. Am. **112**, 1842–1852 (2002).
[15] C. W. Bogart and T. C. Yang, ''Source localization with horizontal arrays in shallow water: Spatial sampling and effective aperture,'' J. Acoust. Soc. Am. **96**, 1677–1686 (1994).
[16] S. L. Tantum and L. W. Nolte, ''On array design for matched-field processing,'' J. Acoust. Soc. Am. **107**, 2101–2111 (2000).
[17] L. A. Stallworth and M. J. Jacobson, ''Acoustic propagation in an isospeed channel with uniform tidal current and depth change,'' J. Acoust. Soc. Am. **48**, 382–391 (1970).
[18] L. A. Stallworth and M. J. Jacobson, ''Sound transmission in an isospeed ocean channel with depth-dependent current,'' J. Acoust. Soc. Am. **51**, 1738–1750 (1972).
[19] L. N-Phu and F. Tappert, ''Parabolic equation modeling of the effects of ocean currents on sound transmission and reciprocity in the time domain,'' J. Acoust. Soc. Am. **78**, 642–648 (1985).
[20] J. S. Robertson, W. L. Siegmann, and M. J. Jacobson, ''Current and current shear effects in the parabolic approximation for underwater sound channels,'' J. Acoust. Soc. Am. **77**, 1768–1780 (1985).
[21] J. S. Robertson, W. L. Siegmann, and M. J. Jacobson, ''Acoustical effects of ocean current shear structures in the parabolic approximation,'' J. Acoust. Soc. Am. **82**, 559–573 (1987).
[22] V. E. Ostashev, ''Discrete spectrum of the sound field of a point source in a stratified moving medium,'' Sov. Phys. Acoust. **32**, 303–306 (1986).

[23] O. A. Godin, ''Discrete spectrum of a sound field in a moving medium,'' Sov. Phys. Acoust. **36**, 355–358 (1990).
[24] M. B. Porter and E. L. Reiss, ''Numerical method for acoustical normal modes for shear flows,'' J. Sound Vib. **100**, 91–105 (1985).
[25] P. Ellisseeff and H. Schmidt, ''Acoustic propagation through a low Mach number, stratified flow,'' J. Acoust. Soc. Am. **101**, 1936–1944 (1997).
[26] L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Medium. 2: Point Sources and Bounded Beams*, 2nd ed. (Springer, Berlin, Heidelberg, 1999), pp. 121–192.
[27] O. A. Godin, ''A wide-angle, energy-conserving parabolic equation for sound in moving medium,'' in *Theoretical and Computational Acoustics'97*, edited by Y.-C. Teng, E.-C. Shang, Y.-H. Pao, M. H. Schultz, and A. D. Pierce (World Scientific, Singapore, 1999), pp. 329–340.
[28] D. Y. Mikhin, ''Energy-conserving and reciprocal solutions for high-order parabolic equations,'' J. Acoust. Soc. Am. **105**, 1362–1363 (1999).
[29] O. A. Godin, ''Coupled-mode sound propagation in a range-dependent, moving fluid,'' J. Acoust. Soc. Am. **111**, 1984–1995 (2002).
[30] K. G. Sabra and D. R. Dowling, ''Broadband performance of a moving time-reversing array,'' J. Acoust. Soc. Am. **114**, 1395–1405.
[31] H. C. Song, W. A. Kuperman, and W. S. Hodgkiss, ''A time-reversal mirror with variable range focusing,'' J. Acoust. Soc. Am. **103**, 3234–3240 (1998).
[32] M. Porter and E. L. Reiss, ''A numerical method for ocean-acoustic normal modes,'' J. Acoust. Soc. Am. **76**, 244–252 (1984).
[33] M. Porter, ''The KRAKEN normal mode program,'' Report No. SM-245, SACLANT Undersea Research Centre, La Spezia, Italy, 1991.
[34] D. Prandle, ''The vertical structure of tidal currents and other oscillatory flows,'' Cont. Shelf Res. **1**, 191–207 (1982).
[35] P. K. Kundu, *Fluid Mechanics* (Academic, San Diego, 1990), pp. 441–461.
[36] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP Press, Woodbury, NY, 1994), pp. 41, 272–275, 314–315, 320–323.
[37] W. Munk, P. Worcester, and C. Wunsch, *Ocean Acoustic Tomography* (Cambridge University Press, Cambridge, 1995), Chap. 3.
[38] P. Hursky, W. S. Hodgkiss, and W. A. Kuperman, ''Matched-field source localization with data-derived modes,'' J. Acoust. Soc. Am. **107**, 1355–1366 (2001).
[39] T. B. Neilsen and E. K. Westwood, ''Extraction of acoustic normal mode depth functions using vertical line array data,'' J. Acoust. Soc. Am. **111**, 748–756 (2002).
[40] O. A. Godin, D. Yu. Mikhin, and D. R. Palmer, ''Monitoring ocean currents in the coastal zone,'' Atmos. Oceanic Phys. **36**, 131–142 (2000).
[41] P. Ellisseeff, H. Schmidt, M. Johnson, D. Herold, N. R. Chapman, and M. McDonald, ''Acoustic tomography of a coastal front in Haro Strait, British Columbia,'' J. Acoust. Soc. Am. **106**, 169–184 (1999).

Depth-dependent target strengths of gadoids by the boundary-element method

David T. I. Francis
Department of Electronic, Electrical and Computer Engineering, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom

Kenneth G. Foote[a)]
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 5 October 2002; revised 24 July 2003; accepted 25 August 2003)

The depth dependence of fish target strength has mostly eluded experimental investigation because of the need to distinguish it from depth-dependent behavioral effects, which may change the orientation distribution. The boundary-element method (BEM) offers an avenue of approach. Based on detailed morphometric data on 15 gadoid swimbladders, the BEM has been exercised to determine how the orientation dependence of target strength changes with pressure under the assumption that the fish swimbladder remains constant in shape and volume. The backscattering cross section has been computed at a nominal frequency of 38 kHz as a function of orientation for each of three pressures: 1, 11, and 51 atm. Increased variability in target strength and more abundant and stronger resonances are both observed with increasing depth. The respective backscattering cross sections have been averaged with respect to each of four normal distributions of tilt angle, and the corresponding target strengths have been regressed on the logarithm of fish length. The tilt-angle-averaged backscattering cross sections at the highest pressure have also been averaged with respect to frequency over a 2-kHz band for representative conditions of insonification. For all averaging methods, the mean target strength changes only slightly with depth. [DOI: 10.1121/1.1619982]

PACS numbers: 43.30.Gv, 43.30.Sf [RAS] Pages: 3136–3146

I. INTRODUCTION

Numerous studies witness to the importance of fish target strength.[1–5] This is, for example, a key quantity in the echo integration method of determining the numerical density of fish.[6] It is also important in the echo counting method of density determination, as it appears in the sampling-volume term.[7]

In general, the target strength depends on fish size, orientation relative to the direction of incidence, acoustic frequency, biological state, depth, and depth history. A large number of studies have attempted to elucidate some of these dependences. References 4 and 8–12 are illustrative.

Determination of the depth dependence has been problematical for several reasons. *In-situ* observation based on free-swimming, unconfined fish is difficult for want of control over the fish, especially that of behavior through the orientation distribution, which may change with depth. When confined, the influence of captivity as a stressor must be suspected. The degree to which the fish controls the surface tension on the swimbladder wall may be significant, thus potentially affecting the target strength.[13,14]

Notwithstanding these comments, the depth dependence of target strength from swimbladdered fish lacking *rete mirabile* seems clear: Boyle's law, or the inverse relationship of ambient pressure and volume, is operative. The mass of gas in the swimbladder is constant, and the swimbladder volume diminishes with depth, affecting the target strength.[11] For swimbladdered fish possessing *rete mirabile*, the swimbladder volume can be regulated, presumably to maintain a state of fixed buoyancy independent of depth. Given the similar mass densities of fish flesh and sea water and similar compressibilities, such fish may be expected to maintain their swimbladders in a state of constant volume. If the swimbladder shape also remains constant, the target strength may be expected to remain more or less constant, with any difference in target strength being due to the increased mass density of gas with depth.

It is the present aim to investigate the effect of increased mass density of swimbladder gas with depth on the target strength of swimbladdered fish that maintain a constant gas volume. The particular fish are members of the gadoid family, specifically pollack (*Pollachius pollachius*) and saithe (*Pollachius virens*) for which the orientation dependence of target strength has already been measured[15] and swimbladder morphometries, performed by Ona's method,[16] are available. The approach is by theoretical modeling with the boundary-element method, applied in a preceding study to the same specimens, but under assumption there of an empty swimbladder volume corresponding to an ideal pressure-release boundary.[17]

II. BOUNDARY-ELEMENT METHOD (BEM)

In the acoustic boundary-element method, the Helmholtz equation $(\nabla^2 + k^2)p = 0$, where k is the wavenumber, is recast as an integral equation in which the acoustic pressure p

a)Author to whom correspondence should be addressed. Electronic mail: kfoote@whoi.edu

at any point is expressed in terms of the pressure and normal displacement (or velocity) fields on the surface S of the scatterer.[18–20] To solve for the unknown fields, this integral equation is evaluated at each node associated with the set of discrete elements that collectively span S, thus producing a system of simultaneous equations for the pressures and displacements at those nodes. The solution to this system may then be used to determine the pressure at any other point using a numerical form of the original integral equation.

The exterior form of the standard integral equation is known to suffer from singularities at certain critical frequencies.[21] Solutions to the integral equation may be nonunique. At critical frequencies, the aberrant solutions are physically inadmissible,[22] which raises the problem of distinguishing these in numerical solutions.

Various methods exist to overcome the problem of critical frequencies, for example, those described in Refs. 23–28. The method available to the authors is based on Burton and Miller's approach,[24] in the particular form developed in Ref. 29 and applied in Ref. 17. This incorporates a second integral equation, the normal derivative form, which is obtained by differentiating the standard form with respect to the normal direction at the surface. The two integral equations are combined by adding a multiple α of the normal derivative form to the standard form. The value of this coupling parameter was not specified by Burton and Miller; later authors have made various suggestions, e.g., Refs. 30–32, the consensus being that α should have a non-zero imaginary part and vary inversely as the wavenumber. Terai's recommendation[31] that $\alpha=-\mathrm{i}/k$ has been adopted in the present work. The combined integral equation approach, in the particular formulation used here, is referred to as the partial Helmholtz gradient formulation (pHGF), while that of the standard integral equation is referred to simply as the standard formulation (SF).

Interestingly, inclusion of a fluid region interior to the scattering surface S, as in the current work, may be sufficient in itself to render the standard integral equation reliable at all frequencies, although this remains unproven. Because of the higher degrees of singularity of integrands in the pHGF, which are integrable, there is a nominal penalty to be paid: lesser precision relative to that achieved with the SF at noncritical frequencies.

To achieve the highest precision and accuracy, therefore, the pHGF is used to identify possible critical frequencies. In their apparent absence, the SF is used to describe the numerical results. The general pHGF is now introduced; the SF is derived from this by equating α to zero.

A. Exterior equations

The system of equations, in matrix form, is given by

$$\mathbf{Ap}=\mathbf{Bu}-\mathbf{p}_{\mathrm{inc}}-\alpha\frac{\partial\mathbf{p}_{\mathrm{inc}}}{\partial n} \tag{1}$$

where $\mathbf{p}$ is the pressure field and $\mathbf{u}$ is the normal component of the displacement field due to the incident pressure field $\mathbf{p}_{\mathrm{inc}}$, with time variation $\exp(\mathrm{i}\omega t)$ understood, where ω is the angular frequency. The coefficients of the matrices $\mathbf{A}$ and $\mathbf{B}$ are assembled from local matrices pertaining to each element of the mesh. In performing the assembly, it is necessary to distinguish between the *global* label of a node, which is the label in the complete, global mesh, and the *local* node, which is the label of the node relative to an element on which it lies. With the calculation point of the Helmholtz integral taken at global node i, with position $\mathbf{r}_i$, integration over element m provides the following coefficients in the standard formulation:

$$a_{mn}^{(1)}(\mathbf{r}_i)=\int_{S_m}N_n(\mathbf{q})\frac{\partial G(\mathbf{r}_i,\mathbf{q})}{\partial n_q}\,\mathrm{d}S_q, \tag{2a}$$

$$b_{mn}^{(1)}(\mathbf{r}_i)=\rho\omega^2\int_{S_m}N_n(\mathbf{q})\cos\theta_{mn}G(\mathbf{r}_i,\mathbf{q})\,\mathrm{d}S_q, \tag{2b}$$

where ρ is the fluid density, $\omega=ck$, c is the speed of sound, $\mathbf{q}$ is the position vector of the integration point on the element surface S_m, G is the Green's function, given by $G(\mathbf{r}_i,\mathbf{q})=e^{-\mathrm{i}k|\mathbf{r}_i-\mathbf{q}|}/4\pi|\mathbf{r}_i-\mathbf{q}|$, n is the local nodal label, and $N_n(\mathbf{q})$ ($n=1,2,...,6$ for triangular elements, $n=1,2,...,8$ for quadrilateral elements) are the shape functions, which are of the standard second-order, or quadratic, form.[33] The factor $\cos\theta_{mn}$ is included to allow for the deviation θ_{mn} of the normal to the element m at local node n from the mean normal at that node. This deviation is inherent in the quadratic formulation. The mean is weighted according to the differential area of each element at the node where they meet.

The normal derivative form of the Helmholtz integral equation, calculated at the centroids $\overline{\mathbf{r}}_l$ of the elements, similarly provides coefficients as follows:

$$a_{mn}^{(2)}(\overline{\mathbf{r}}_l)=\int_{S_m}N_n(\mathbf{q})\frac{\partial^2 G(\overline{\mathbf{r}}_l,\mathbf{q})}{\partial n_r\partial n_q}\,\mathrm{d}S_q, \tag{3a}$$

$$b_{mn}^{(2)}(\overline{\mathbf{r}}_l)=\rho\omega^2\int_{S_m}N_n(\mathbf{q})\cos\theta_{mn}\frac{\partial G(\overline{\mathbf{r}}_l,\mathbf{q})}{\partial n_r}\,\mathrm{d}S_q, \tag{3b}$$

where the normal derivative is evaluated at the centroid. These are combined with the previous coefficients by adding a multiple of $a_{mn}^{(2)}(\overline{\mathbf{r}}_l)$ or $b_{mn}^{(2)}(\overline{\mathbf{r}}_l)$, respectively, for all elements l on which global node i lies, i.e.,

$$a_{mn}(\mathbf{r}_i)=a_{mn}^{(1)}(\mathbf{r}_i)+\alpha_i\sum_{l:i\varepsilon S_l}a_{mn}^{(2)}(\overline{\mathbf{r}}_l), \tag{4a}$$

$$b_{mn}(\mathbf{r}_i)=b_{mn}^{(1)}(\mathbf{r}_i)+\alpha_i\sum_{l:i\varepsilon S_l}b_{mn}^{(2)}(\overline{\mathbf{r}}_l), \tag{4b}$$

where the combination factor α_i is taken to be $-\mathrm{i}/kM_i$, following Terai's recommendation,[31] but allowing for the number M_i of elements meeting at node i.[29–31]

The coefficients $a_{mn}(\mathbf{r}_i)$ and $b_{mn}(\mathbf{r}_i)$ are assembled into the global matrices $\mathbf{A}$ and $\mathbf{B}$ by summing the coefficients that correspond to the same global node, thus

$$A_{ij}=\sum_{m,n:C(m,n)=j}a_{mn}(\mathbf{r}_i)-\beta(\mathbf{r}_i)\delta_{ij}, \tag{5a}$$

$$B_{ij}=\sum_{m,n:C(m,n)=j}b_{mn}(\mathbf{r}_i), \tag{5b}$$

where $C(m,n)$ is the global node label of local node n on element m. The quantity $4\pi\beta(\mathbf{r})$ is the solid angle occupied

by the fluid region surrounding the point **r**.[31] For **r** entirely within the fluid region, $\beta=1$; for **r** on the surface S separating the fluid region from its neighbor, $\beta=0.5$ provided that S is smooth at **r**. For non-smooth surfaces such as those defined by the boundary-element meshes, methods of determining β are described in Ref. 29.

The source terms in Eq. (1) are evaluated thus:

$$\mathbf{p}_{\mathrm{inc}}(\mathbf{r}_i)+\alpha_i\sum\frac{\partial p_{\mathrm{inc}}(\overline{\mathbf{r}}_l)}{\partial n_r},$$

where the summation is performed for all elements l on which the global node i lies.

B. Interior equations

The fundamental equation in the interior of the swimbladder resembles that of Eq. (1) but without the exciting pressure field $\mathbf{p}_{\mathrm{inc}}$. Thus,

$$\mathbf{A}_1\mathbf{p}-\mathbf{B}_1\mathbf{u}=0. \tag{6}$$

The matrices $\mathbf{A}_1$ and $\mathbf{B}_1$ resemble the respective matrices **A** and **B**, but use the properties of the internal gas rather than those of the external fluid. In addition, the normal direction is oriented into the gas, hence it is reversed with respect to the normal direction in the exterior. The solid angle is similarly referred to the interior, hence

$$\beta_1=1-\beta, \tag{7}$$

at any point on S. With these conditions, a set of equations similar to those of Eqs. (2)–(5) can be developed.

C. Simultaneous solution

Pressure and normal displacement are continuous across the water–gas interface S, i.e., the nodal pressures and normal displacements in Eq. (1) are identical to those in Eq. (6). There are therefore as many unknown values of pressure and displacement as there are independent equations, namely $2N$, where N is the number of nodes on S. The simultaneous solution of Eqs. (1) and (6) is derived directly:

$$\mathbf{u}=(\mathbf{B}-\mathbf{A}\mathbf{A}_1^{-1}\mathbf{B}_1)^{-1}\left(\mathbf{p}_{\mathrm{inc}}+\alpha\frac{\partial\mathbf{p}_{\mathrm{inc}}}{\partial n}\right), \tag{8}$$

and

$$\mathbf{p}=\mathbf{A}_1^{-1}\mathbf{B}_1\mathbf{u}. \tag{9}$$

D. Scattered field

The scattered pressure at an exterior point **r** is obtained from the standard integral equation by calculating coefficients similar to $a_{mn}^{(1)}(\mathbf{r}_i)$ and $b_{mn}^{(1)}(\mathbf{r}_i)$, but with $\mathbf{r}_i$ replaced by the position vector **r**:

$$a_j^{(3)}(\mathbf{r})=\sum_{m,n:C(m,n)=j}\int_{S_m}N_n(\mathbf{q})\frac{\partial G(\mathbf{r},\mathbf{q})}{\partial n_q}\,\mathrm{d}S_q, \tag{10a}$$

$$b_j^{(3)}(\mathbf{r})=\rho\omega^2\sum_{m,n:C(m,n)=j}\int_{S_m}N_n(\mathbf{q})\cos\theta_{mn}G(\mathbf{r},\mathbf{q})\,\mathrm{d}S_q, \tag{10b}$$

and then

$$p(\mathbf{r})=\mathbf{a}^{(3)}(\mathbf{r})\cdot\mathbf{p}-\mathbf{b}^{(3)}(\mathbf{r})\cdot\mathbf{u}, \tag{11}$$

where the dot-product operations are indicated. The backscattering amplitude at finite range r is

$$f(r)=\frac{r|p(\mathbf{r})|}{|p_{\mathrm{inc}}|}. \tag{12}$$

The farfield backscattering amplitude f is the limit of $f(r)$ as r approaches infinity. Expressions for the backscattering cross section σ and target strength TS are derived by substituting $f(r)$, or f, in the following equations:

$$\sigma=4\pi|f|^2 \tag{13}$$

and

$$\mathrm{TS}=10\log\left[\frac{\sigma}{4\pi r_0^2}\right], \tag{14}$$

where r_0 is a reference distance, assumed here to be 1 m.

E. Numerical evaluation techniques

The elements used here are quadrilaterals and triangles of the quadratic isoparametric type, in which both the geometric and acoustic quantities are interpolated from the nodal values using quadratic shape functions, the nodes being situated at the vertices and mid-sides.[33] Experience suggests that good representation of the acoustic variables is obtained if the lengths of the sides of the elements are less than one-third of a wavelength.[17] The accuracy of geometrical representation depends on the degree of undulation of the surface, but it should be noted that the quadratic interpolation allows the sides and faces of the elements to be curved. Further details of the formulation and equations can be found in Ref. 29.

In evaluating the coefficients in Eqs. (1) and (6), Gauss quadrature is used.

III. INDEPENDENT VALIDATION OF BEM

The BEM has been validated previously for application to surface-adapted gadoids, but assuming that the swimbladder acts as a void.[17] The BEM was also tested against the example of scattering by a spherical void for which a series solution is available.

For application to problems of scattering by a swimbladdered fish at depth, the method is tested against the analytical (series) solutions[34,35] for two test cases. The object in both cases is a constant-volume gas-filled sphere of diameter 50 mm. In the first case, bistatic scattering for an incident wave of frequency 50 kHz is described over the angular range 180° as measured from the forward direction at each of two pressures, 1 and 51 atm, corresponding to the nominal depths of 0 and 500 m. The values of density ρ and sound speed c were taken to be 1025 kg/m^3 and 1470 m/s, respectively, in water; 1.247 kg/m^3 and 337.4 m/s, respectively, in air at 1 atm; and 63.597 kg/m^3 and 337.4 m/s, respectively, in air at 51 atm. The mass density of swimbladder gas is assumed to change in proportion to pressure to maintain a constant volume. For the BEM, the mesh representing the sphere was formed from elements delineated by lines of latitude and

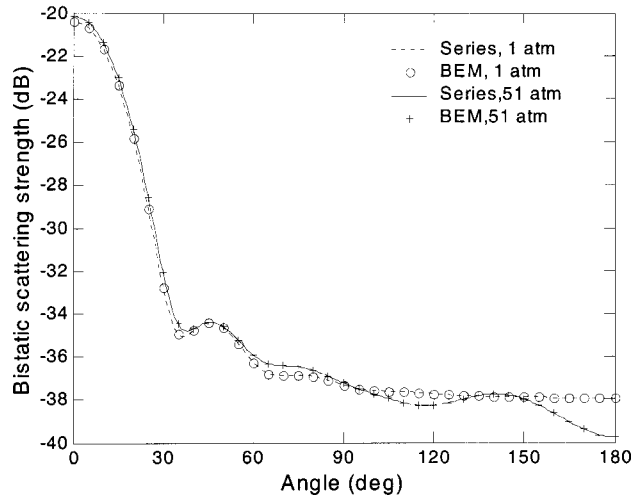


FIG. 1. Bistatic scattering strength in decibels as a function of scattering angle for a plane wave of frequency 50 kHz incident on a spherical air bubble of diameter 50 mm, at pressures of 1 and 51 atm. The analytical (series) solution is shown by the continuous and broken lines, while the BEM predictions are shown as discrete points. The scattering angle is relative to the direction of incidence, so that 0° represents the forward direction.

longitude at 2.25° intervals. By invoking rotational symmetry, the problem was reduced to one of solving for the surface pressures and displacements on only one segment, consisting of 80 elements between adjacent meridians. The results shown in Fig. 1 indicate an excellent agreement between the BEM and the analytical solution. Results with the pHGF and SF are indistinguishable.

In the second test case, the target strength of a similar gas-filled sphere is computed over the frequency range 1 Hz to 50 kHz, again at pressures of 1 and 51 atm, with densities and sound speeds as given above. The target strength was initially calculated using the analytical solution at 1-Hz intervals. In the neighborhood of each of the numerous peaks and troughs identified from these initial results, further calculations were then performed at increasing resolution, down to 0.001 Hz as required, in order to identify the peak or trough TS values to within 0.1 dB. For the BEM, the longer computing times made it impractical to compute solutions to the same resolution of 1 Hz over the whole frequency range. Instead, the target strength was calculated initially at 100-Hz intervals, and then at finer resolution around the peaks and troughs as identified in the analytical solution, again with the aim of determining the predicted peak or trough TS value to within 0.1 dB.

The same mesh as described in the first test case was used in the frequency range 25–50 kHz. For lower frequencies, where the condition that the nodal separation should be less than $\frac{1}{6}$ of the wavelength allows for larger elements, similar meshes were used but at intervals in latitude and longitude of 4.5° for the range 12.5–25 kHz and 9° for the range 0–12.5 kHz.

The results are plotted in Fig. 2(a) for a pressure of 1

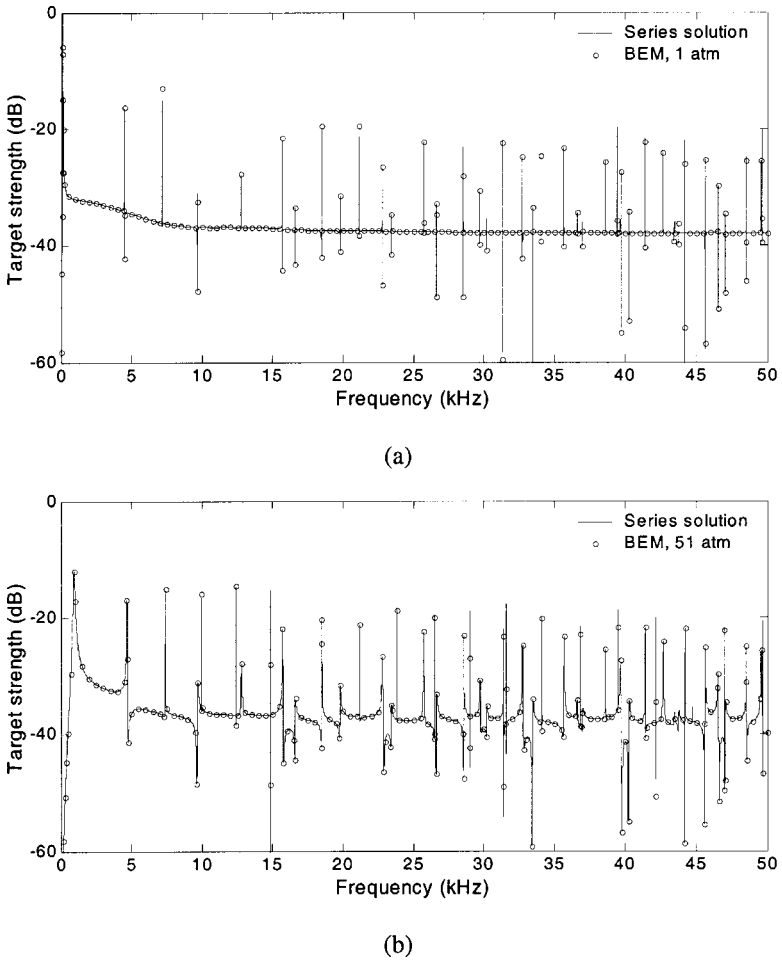


FIG. 2. Target strength of a spherical air bubble of diameter 50 mm in water at (a) 1 atm and (b) 51 atm. The analytical solution is shown by a continuous line, the BEM predictions as discrete points. In (a), the first peak, at 0.13 kHz with a TS of 5.1 dB, is truncated.

TABLE I. Properties of the 15 swimbladders and the meshes used to represent them.

Fish no.	Species	Fish length (cm)	Fish mass (g)	Meshes		Swimbladder		Nodal separation		
				n_{elem}	n_{nodes}	Surface area (cm^2)	Volume (cm^3)	95% limit (cm)	99% limit (cm)	Max (cm)
201	Pollack	31.5	195	1168	3364	33.01	6.91	0.111	0.120	0.221
202	Pollack	44.0	533	1389	4041	58.83	16.33	0.126	0.137	0.177
204	Pollack	35.5	321	1078	3116	42.39	10.03	0.131	0.141	0.172
205	Pollack	39.0	380	1107	3181	45.75	11.34	0.132	0.143	0.193
206	Pollack	35.0	287	1159	3347	31.37	7.75	0.104	0.117	0.146
207	Pollack	44.5	635	1487	4363	65.24	19.15	0.124	0.134	0.161
209	Saithe	38.5	385	1501	4387	43.29	10.08	0.100	0.106	0.139
213	Pollack	34.5	259	1039	2935	34.11	7.83	0.123	0.133	0.161
214	Pollack	39.0	406	1164	3362	44.14	10.15	0.125	0.134	0.153
215	Pollack	37.0	332	1076	3092	38.89	8.75	0.124	0.134	0.174
216	Pollack	36.5	343	1062	3060	43.33	10.85	0.131	0.140	0.164
217	Pollack	34.5	253	1662	4840	32.15	6.57	0.081	0.088	0.134
218	Pollack	32.5	257	1327	3879	29.75	6.27	0.092	0.100	0.139
219	Pollack	35.5	292	1039	3005	35.74	8.15	0.120	0.127	0.153
220	Saithe	38.0	406	1321	3857	44.32	10.46	0.106	0.113	0.132

atm, and Fig. 2(b) for a pressure of 51 atm. For clarity, the BEM predictions, shown as discrete points, are plotted at intervals of 500 Hz, with the addition of the peak and trough values. The SF results are shown for giving slightly better agreement with the exact series solution. This agreement is excellent away from the peaks and in the values of the frequencies at which the peaks and troughs occur; the main discrepancies occur in the predictions of the peak and trough values at high-Q resonances. At these frequencies the system of equations solved in Eq. (8) may be less well conditioned, and numerical errors in the coefficients and in the solution of the system of equations become more significant.

While the original purpose of this test case was to investigate the performance of the BEM, the results in Fig. 2 provide insight into the effect of pressure on the target strength response when the interior fluid is included in the analysis. No resonances appear when the bubble is treated as a void, which is illustrated in Ref. 17, Fig. 2. At 1 atm, resonances appear, but with very narrow bandwidths, typically less than 1 Hz for the case considered. These bandwidths increase with increasing pressure. At 51 atm, there is a notable effect, with deviations from the smooth line of the solution for a void in the frequency bands between neighboring resonances.

IV. SWIMBLADDER MORPHOMETRY

The swimbladder morphometric data were derived from a study performed in 1980 on surface-adapted gadoids[15] by Ona's method of cryomicrotoming.[16] The data were reduced to a set of curvilinear quadrilaterals and triangles spanning each swimbladder surface, as described in Ref. 17. The statistics of the nodes at which the fundamental equations (8) and (9) were solved are described in Table I, which also summarizes information on the specimens themselves: species, length, mass, and so forth. An example of one of the meshes, that for specimen 217,[36] is shown in Fig. 3.

As earlier noted,[17] the BEM is considered valid for nodal spacings less than one-sixth of the acoustic wavelength. It is important to note that, in the case of gas-filled swimbladders, the relevant wavelength is that of the acoustic field inside the swimbladder. Relative to the wavelength λ in the exterior immersion medium with sound speed c, the interior wavelength λ_1 is

$$\lambda_1 = \frac{\lambda c_1}{c}, \tag{15}$$

where c_1 is the speed of sound in the gas. Thus, whereas the nodal spacings were adequate for computations at 120 kHz in the preceding work on scattering by voids, the nodal spacings are only adequate here for computations up to about 40 kHz.

V. COMPUTATIONAL PARAMETERS

The mass density of the gas in the swimbladder at atmospheric pressure is assumed to be that of air at standard condition, namely 1.247 kg/m^3. At ambient pressure P atm, the mass density is assumed to be $1.247\,P$ kg/m^3. The speed of sound in the gas is assumed to be 337.4 m/s, independent of depth.

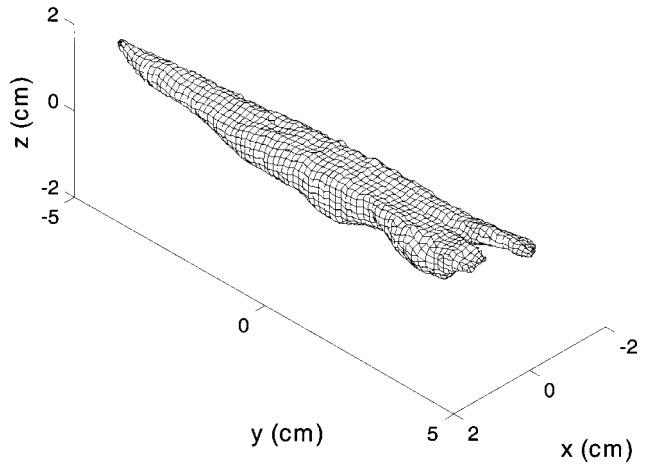


FIG. 3. Boundary-element mesh of the swimbladder to specimen 217 of Ref. 36. The meshed swimbladder length is 108 mm and the mesh has 1662 elements and 4840 nodes.

The mass density of the surrounding fish flesh is assumed to be identical to that of the seawater immersion medium, namely 1025 kg/m^3. For the assumed temperature of 5 °C and salinity of 35 ppt, the nominal sound speed in the immersion medium is 1470 m/s.

The numbers of elements and nodes for each swimbladder are given in Table I. Statistical measures of the distance between nearest-neighbor nodes are given in the same table. As mentioned, this limits the upper frequency of computation to about 40 kHz, on the basis that the length of element sides, nominally double the nodal separation, should be less than one-third of a wavelength in air.

The backscattering cross section given by Eq. (13) is averaged with respect to the orientation distribution, then expressed in the logarithmic domain according to Eq. (14). Each of four normal distributions of tilt angle are considered: N(0,0), N(0,5), N(0,10), and N(−4.4,16)°. The last distribution is derived from *in-situ* photographic observations on cod (*Gadus morhua*).[37]

Because of the effect of perspective,[38] the effective standard deviations for the four cases are 2.5, 5.5, 10.2, and 16°, respectively.[36] In terms of the normal distribution g of tilt angle θ, with mean $\bar{\theta}$ and effective standard deviation s_θ, the average backscattering cross section at frequency ν is

$$\bar{\sigma}(\nu) = \int g(\theta)\sigma(\nu,\theta)\, d\theta \Big/ \int g(\theta)\, d\theta, \tag{16}$$

where the integration is performed over the range $[\bar{\theta} - 3s_\theta, \bar{\theta} + 3s_\theta]$.

Because of the presence of gas in the swimbladder cavity, resonances occur, even at high frequencies beyond the low-frequency breathing-mode resonance.[39] These could be troublesome, particularly at greater depths since the greater bandwidths of the resonances, observed in the results for the spherical bubble, make it more likely that a particular frequency will fall within such a resonance. However, there are no infinite-duration, single-frequency signals in practice; sonar measurements are performed with finite-duration signals, hence with bandwidth. Realizable receivers also have an associated bandwidth of processing. These effects have been dealt with in Ref. 40 through the following operational average of the backscattering cross section as a function of tilt angle θ for a downward-pointing transducer, here expressed in terms of the tilt-angle-averaged cross section:

$$\bar{\bar{\sigma}} = \int |SH|^2 \bar{\sigma}(\nu)\, d\nu \Big/ \int |SH|^2\, d\nu, \tag{17}$$

where S is the transmit signal spectrum and H is the receiver frequency response function, all functions of frequency ν. In the computations reported here, S is the Fourier spectrum corresponding to the signal $s(t) = \cos(2\pi\nu_0 t)\mathrm{rect}(t/\tau)$, where ν_0 is the center frequency of the transmit signal, τ is the signal duration, assumed to be 0.64 ms, and rect(x) is 0 for $|x| > 0.5$ and 1 for $|x| \leqslant 0.5$. Thus, $S(\nu) = (\sin y)/y$, where $y = (\nu - \nu_0)\tau/2$. The function H is described in Fig. 4 for $\nu_0 = 38$ kHz.

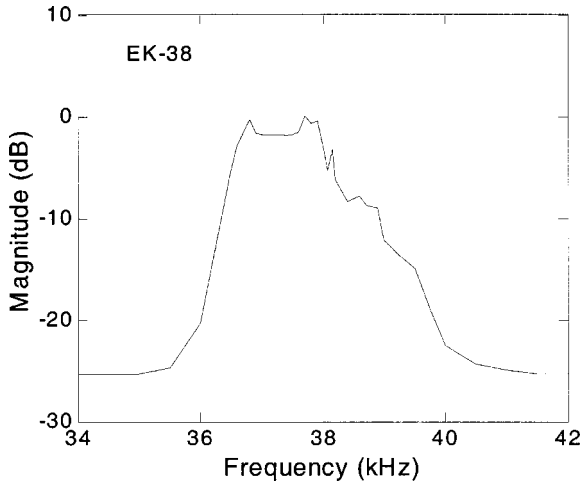


FIG. 4. Magnitude of the frequency response function of the receiver of the EK-38 echo sounder, from Fig. 4 of Ref. 40.

VI. RESULTS AND DISCUSSION

A. Comparison of computations with standard and partial Helmholtz gradient formulations

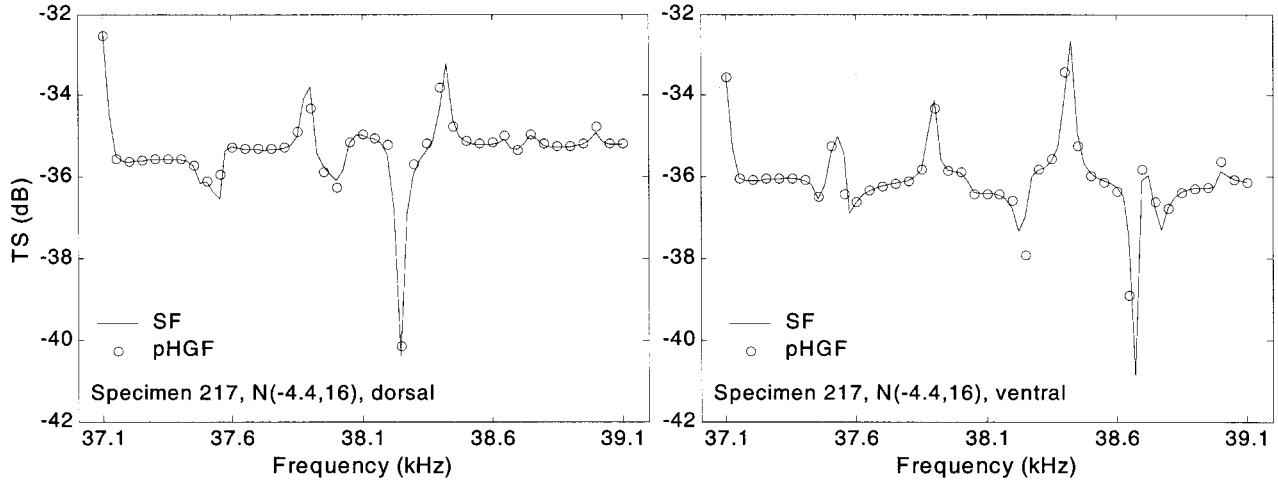
Results with the two approaches are illustrated in Fig. 5 for specimens 217 and 219. The resolution of the pHGF results is 50 Hz; that of the SF results is 25 Hz. Evidently, the agreement is very strong, but with a few differences where the pHGF results appear to be slightly more dispersed, suggesting a somewhat greater variability. As mentioned in Sec. II, this is believed to be due to the higher degrees of singularity of integrands in the pHGF compared with those in the SF, witness Eq. (3a) compared with Eq. (2b), with corresponding loss of numerical precision. By reason of assumed greater precision and accuracy, the SF is used to derive the computational results presented below.

Agreement of the corresponding results for the two approaches is poorer for some other specimens, especially when their nodal spacing approaches the λ/6-limit of applicability, where λ is the wavelength of sound in the gaseous interior of the swimbladder. At 38.1 kHz, with speed of sound in the gas of 337.4 m/s, λ/6=0.144 cm. Referring to Table I, 95% of the nodal spacings for specimen no. 219 are less than 0.120 cm. This is within 10% of the corresponding value for nodal spacing with the coarsest mesh, which applies to specimen 205.

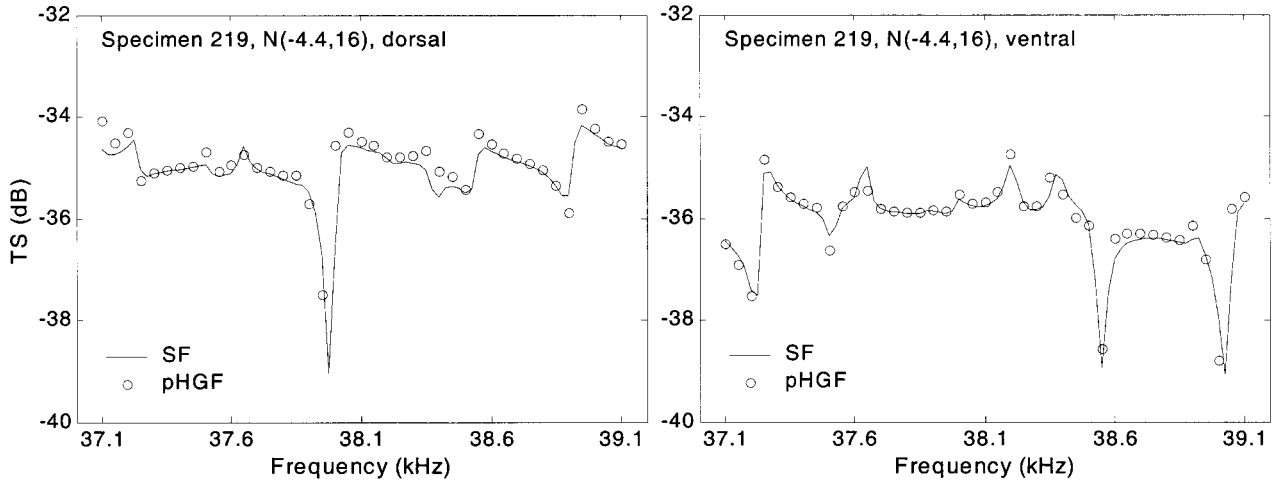
B. Depth dependence of target strength for a single specimen

The tilt-angle dependence of target strength has been computed for each specimen at each of three pressures: 1, 11, and 51 atm. The dependences for the swimbladder modeled as a void[17] and at pressures of 1 and 11 atm are in very close agreement.

The computed tilt-angle dependence of target strength is shown for a single specimen, a 34.5-cm-long pollack, in Fig. 6. Also shown are the dependences for the case of a void of the same shape[17] and for actual measurements of the whole fish with intact swimbladder.[36]



(a)



(b)

FIG. 5. Predicted target strength versus frequency for (a) specimen 217 and (b) specimen 219 in the range 37.1–39.1 kHz, at a pressure of 51 atm. Results obtained using the standard formulation (SF) of the BEM, at intervals of 25 Hz, are shown by the continuous lines; results from the partial Helmholtz gradient formulation (pHGF) are shown as discrete points, at intervals of 50 Hz.

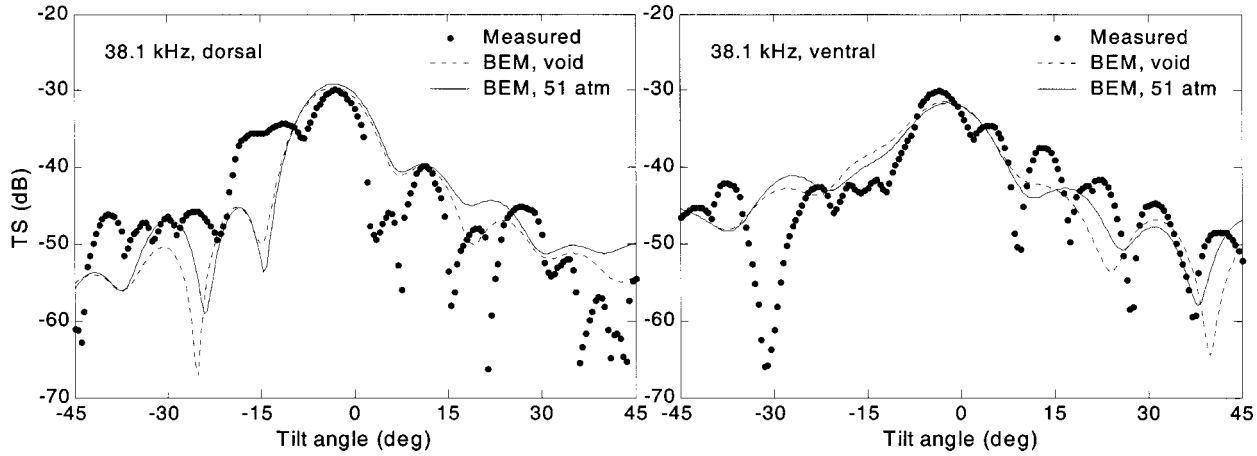


FIG. 6. BEM computations of target strength as a function of tilt angle for swimbladder specimen 217, treated as a void and as gas-filled at a depth of 500 m, compared against direct measurements at 2.5-m depth. The functions are shown for both dorsal and ventral aspects at a frequency of 38.1 kHz.

TABLE II. Regression analyses of target strength on fish length, based on BEM computations of TS for 15 specimens in dorsal aspect. Averaging is performed with respect to each of four normal distributions of tilt angle at each of four pressures: 0, 1, 11, and 51 atm. The tilt-angle-averaged results at 51 atm are additionally averaged with respect to frequency, using Eq. (17), and presented in the final row for each tilt-angle distribution. Results are expressed through the regression coefficient b in Eq. (18) and standard error SE of the regression. The correlation coefficient ρ of underlying target strength values in each set is also shown.

Frequency	Tilt angle (degrees)		Pressure	BEM computed TS functions		
kHz	Mean	s.d.	(atm)	ρ	b	SE
38.1	0.0	0.0	0	0.933	−61.36	1.00
38.1	0.0	0.0	1	0.933	−61.36	1.00
38.1	0.0	0.0	11	0.930	−61.38	1.00
38.1	0.0	0.0	51	0.747	−61.30	1.37
37.1–39.1	0.0	0.0	51	0.922	−61.46	1.00
38.1	0.0	5.0	0	0.945	−62.44	0.66
38.1	0.0	5.0	1	0.945	−62.44	0.65
38.1	0.0	5.0	11	0.938	−62.45	0.66
38.1	0.0	5.0	51	0.709	−62.40	1.09
37.1–39.1	0.0	5.0	51	0.930	−62.55	0.67
38.1	0.0	10.0	0	0.946	−64.18	0.47
38.1	0.0	10.0	1	0.947	−64.18	0.46
38.1	0.0	10.0	11	0.938	−64.19	0.47
38.1	0.0	10.0	51	0.664	−64.14	0.98
37.1–39.1	0.0	10.0	51	0.930	−64.28	0.49
38.1	−4.4	16.0	0	0.941	−65.76	0.40
38.1	−4.4	16.0	1	0.942	−65.76	0.39
38.1	−4.4	16.0	11	0.931	−65.76	0.41
38.1	−4.4	16.0	51	0.610	−65.68	0.96
37.1–39.1	−4.4	16.0	51	0.924	−65.81	0.43

The effect of pressure on the orientation dependence of target strength of the other 14 specimens is similar, if differing in the precise details of deviations from the respective void case and measurements.

C. Target strength—length regressions

The large number of computations that have been performed for the 15 swimbladders have been combined by averaging. In particular, the backscattering cross section computed with the BEM, and illustrated in Fig. 6 for a single specimen, have been averaged with respect to the tilt-angle distributions according to Eq. (16). The target strength corresponding to $\overline{\sigma}$ has then been computed according to Eq. (14), and the regression equation

$$\mathrm{TS} = 20 \log l + b \tag{18}$$

computed by the method of least squares, where l is the total fish length in centimeters. Results are shown in Table II for dorsal aspect and in Table III for ventral aspect. Scatter diagrams of target strength and fish length have been prepared for the tilt-angle distribution N(−4.4,16)° for a void and at a pressure of 51 atm. They are presented in Fig. 7.

In general, the modeled target strengths display increasing variability with increasing depth. The physical explanation for this is presently unclear, but based on the second validation example, in Fig. 2, with evidence of resonances at frequencies above the ordinary very-low-frequency breathing-mode resonance, it is speculated that stronger resonances are excited more easily at greater pressures. Thus, TS values are likely to be elevated or depressed depending on the proximity of a resonance or anti-resonance, respectively, for each specimen. Physically, the potential of the swimbladder under pressure to store acoustic energy and act as a resonator, or absorber, of incident acoustic energy seems to increase with depth. The condition that the swimbladder maintain a constant volume ensures that the mass of enclosed gas increases with ambient pressure, hence depth.

The results so far described have been for calculations of TS at single frequencies. However, as mentioned in Sec. V, sonar measurements use signals of finite duration and therefore possess bandwidth. In order to reproduce this characteristic more faithfully in the BEM predictions, further computations were undertaken in a frequency band around 38.1 kHz at the highest pressure of 51 atm, where the effect of resonance is likely to be most significant. Equation (17) provides a basis for then determining a frequency-averaged TS. A spectral band of ±1 kHz about the center frequency was found to be sufficient to account for 95% of the scattered energy. Within this band, BEM predictions were made at intervals of 50 Hz. The integral in the numerator of Eq. (17) was evaluated numerically, first as a Riemann summation, and second by fitting a cubic spline to the TS response and integrating using Simpson's rule with an interval of 10 Hz. The results of the latter method were virtually identical to those of the former, the difference in the eventual value of b in Eq. (18), for example, being generally less than 0.02 dB.

The frequency-averaged results are included in Tables II and III, as the final line in each set of results for a given tilt distribution. Scatter diagrams and regression lines are plotted

TABLE III. As Table II, but in ventral aspect.

Frequency kHz	Tilt angle (degrees) Mean	Tilt angle (degrees) s.d.	Pressure (atm)	BEM computed TS functions ρ	BEM computed TS functions b	BEM computed TS functions SE
38.1	0.0	0.0	0	0.427	−64.84	1.85
38.1	0.0	0.0	1	0.425	−64.84	1.84
38.1	0.0	0.0	11	0.451	−64.91	1.87
38.1	0.0	0.0	51	0.494	−64.88	1.89
37.1–39.1	0.0	0.0	51	0.453	−64.79	1.75
38.1	0.0	5.0	0	0.684	−65.03	1.15
38.1	0.0	5.0	1	0.682	−65.03	1.14
38.1	0.0	5.0	11	0.703	−65.09	1.17
38.1	0.0	5.0	51	0.653	−65.12	1.43
37.1–39.1	0.0	5.0	51	0.736	−64.95	1.02
38.1	0.0	10.0	0	0.875	−65.79	0.66
38.1	0.0	10.0	1	0.874	−65.79	0.66
38.1	0.0	10.0	11	0.878	−65.84	0.70
38.1	0.0	10.0	51	0.823	−65.83	0.93
37.1–39.1	0.0	10.0	51	0.907	−65.69	0.57
38.1	−4.4	16.0	0	0.950	−66.74	0.41
38.1	−4.4	16.0	1	0.948	−66.74	0.41
38.1	−4.4	16.0	11	0.941	−66.79	0.47
38.1	−4.4	16.0	51	0.875	−66.74	0.74
37.1–39.1	−4.4	16.0	51	0.952	−66.62	0.39

in Fig. 8 for the tilt distribution N(−4.4,16)°, allowing a direct comparison with Fig. 7.

In the dorsal aspect, the effects of averaging over the bandwidth of the signal are to slightly decrease the TS level as given by the value of b in Eq. (18), increase the correlation, and decrease the variability when compared with the corresponding single-frequency results. In the ventral aspect, the TS level is slightly increased. The correlation is increased

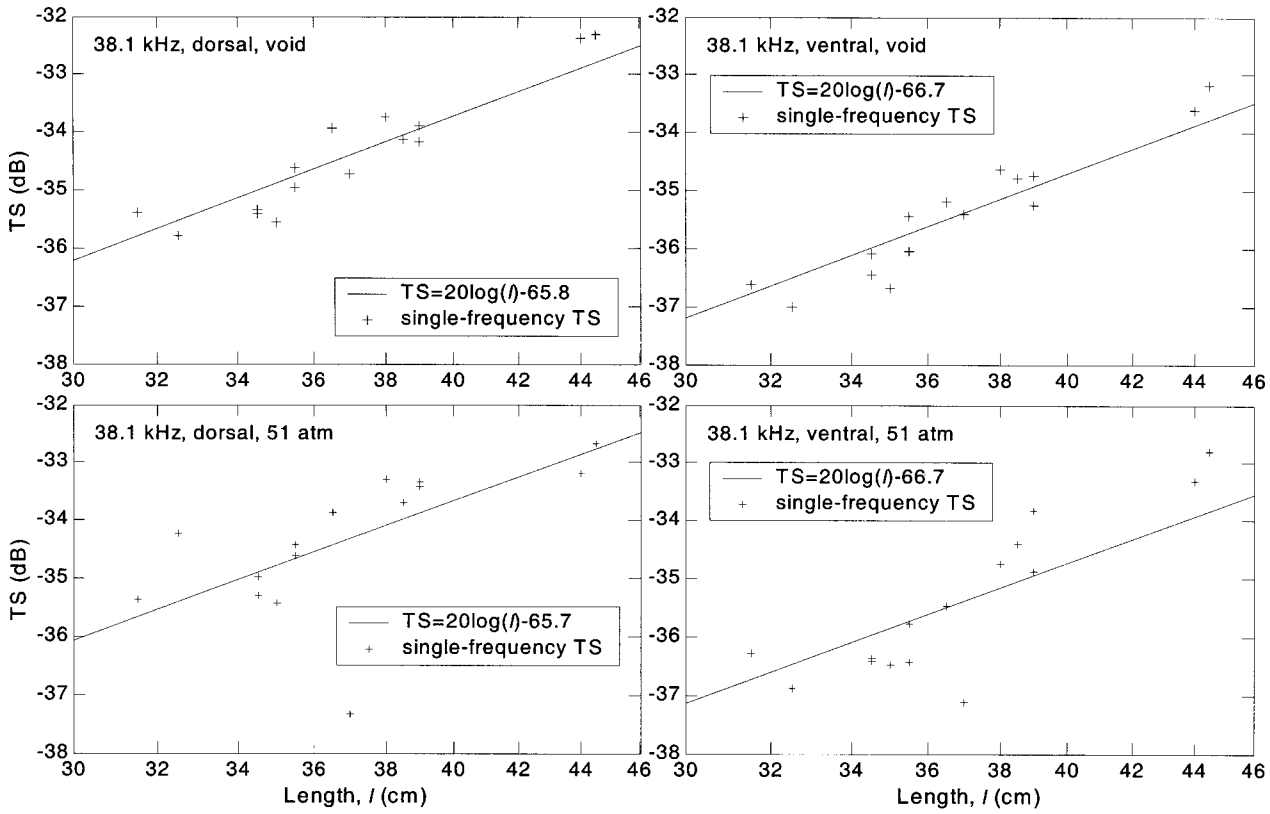


FIG. 7. Scatter diagram of target strength versus length l, expressed on a logarithmic scale, at the single frequency of 38.1 kHz for the tilt distribution (−4.4,16), for a void, and at a pressure of 51 atm, for dorsal and ventral aspects. The regression equation $TS = 20 \log l + b$ is shown by the continuous line in each case.

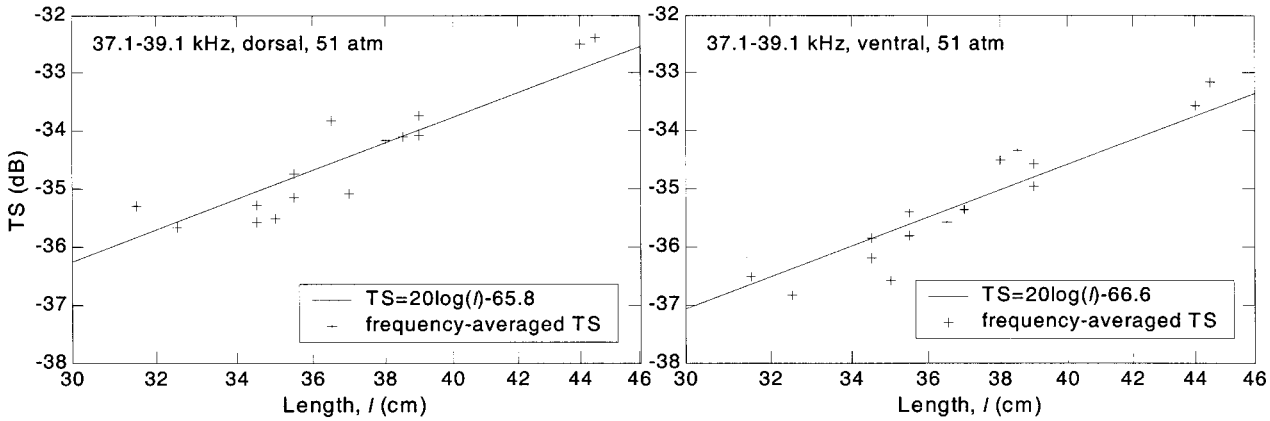


FIG. 8. As Fig. 7 at 51 atm, but for target strength averaged over the frequency band 37.1–39.1 kHz, using Eq. (17).

in all cases but one, and the variability is decreased. These weak trends probably reflect the particular locations and magnitudes of resonance frequencies in the 2-kHz band of averaging. The sample size of 15 specimens is simply too small, and the four orientation distributions are too dependent, to permit drawing stronger conclusions.

Since the depth dependence of target strength is weak, both the echo integration and echo counting methods can be applied as they are at present with at most minor corrections. More importantly, possible observed differences in *in-situ* target strength of depth-adapted gadoids may be attributed to behavior, as manifested through the orientation distribution, rather than to the simple effect of depth acting on the swimbladder volume.

D. Future work

Exercise of the BEM for the test case of a spherical gas-filled bubble revealed low- and high-frequency resonances, if without inclusion of internal damping. Only the lowest-frequency resonance seems to have been studied, suggesting future areas of experimental and theoretical investigation.

The size and shape of each swimbladder have been assumed to be constant, independent of depth. Insofar as the computations are meant to explore the depth dependence of target strength, these would describe situations in which the fish are fully depth-adapted, with inflated swimbladders. Under some circumstances of vertical migration, the rate of change in depth is so rapid that compensation by the *rete mirabile* lags behind, leading to situations of negative or positive buoyancy, as the fish is migrating respectively downwards or upwards. Examination of the change in target strength under uncompensated or partially compensated depth changes would be valuable, but requires a separate investigation. A prerequisite is knowledge of the manner of swimbladder form change with pressure change, which presently is mostly speculative.

Other swimbladder shapes and types should be amenable to computation with the BEM. A study of the Atlantic herring swimbladder, lacking *rete mirabile*, would be of particular value given that the fish is acoustically surveyed for purposes of estimating stock abundance.

Another problem waiting to be addressed is that of the induced surface velocity field on the swimbladder wall and its possible relationship to auditory function. The BEM has already been used in a preliminary study to generate the surface-velocity field,[41] and should shed light on the relationship of external acoustic stimuli and their transmission to the presumed organ of hearing.

VII. CONCLUSIONS

The effect of depth on the target strength of depth-adapted gadoids has been modeled by the boundary-element method. The mean target strength, based on the averaged backscattering cross section, has been found to change insignificantly with depth. There is, however, increased variability in the orientation dependence of target strength. In addition, it is evidently easier to excite high-frequency resonances with increasing depth.

Thus, for applications of the echo counting and integration methods[6] in acoustic surveys of gadoid abundance, there need be no change either in execution of the surveys or in interpretation of forthcoming echo data. This assumes that the gadoid swimbladder remains fully inflated at all depths of measurement. In the event that depth excursions occur or are undertaken without concurrent compensation to maintain a constant state of inflation, the target strength may be expected to change, very possibly to a significant degree depending on the extent and rapidity of the depth excursion.

ACKNOWLEDGMENTS

This work began with sponsorship by the European Commission through its RTD-program, Contract No. MAS3-CT95-0031 (BASS), and was completed with support by the Office of Naval Research, Contract No. N000140310368. Dr. P. J. Harris, University of Brighton, is thanked for discussions on the Burton–Miller formulation. Professor A. D. Pierce is thanked for discussions on the question of uniqueness. Dr. V. R. Starczak is thanked for assistance in executing preliminary BEM computations. A. Norton is thanked for much valued assistance in the technical editing. This is Woods Hole Oceanographic Institution Contribution No. 10839.

[1] R. H. Love, ''Measurements of fish target strength: a review,'' Fish. Bull. U.S. **69**, 703–715 (1971).
[2] L. Midttun, ''Fish and other organisms as acoustic targets,'' Rapp. P.-V. Reun.-Cons. Int. Explor. Mer **184**, 25–33 (1984).
[3] K. G. Foote, ''Summary of methods for determining fish target strength at ultrasonic frequencies,'' ICES J. Mar. Sci. **48**, 211–217 (1991).
[4] S. McClatchie, G. Macaulay, R. F. Coombs, P. Grimes, and A. Hart, ''Target strength of an oily deep-water fish, orange roughy (*Hoplostethus atlanticus*) I. Experiments,'' J. Acoust. Soc. Am. **106**, 131–142 (1999).
[5] K. G. Foote, ''Target strength of fish,'' in *Encyclopedia of Acoustics*, edited by M. J. Crocker (Wiley, New York, 1997), pp. 493–500.
[6] D. N. MacLennan, ''Acoustical measurement of fish abundance,'' J. Acoust. Soc. Am. **87**, 1–15 (1990).
[7] K. G. Foote, ''Acoustic sampling volume,'' J. Acoust. Soc. Am. **90**, 959–964 (1991).
[8] O. Nakken and K. Olsen, ''Target strength measurements of fish,'' Rapp. P.-V. Reun.-Cons. Int. Explor. Mer **170**, 52–69 (1977).
[9] Y. Miyanohana, K. Ishii, and M. Furusawa, ''Measurements and analyses of dorsal-aspect target strength of six species of fish at four frequencies,'' Rapp. P.-V. Reun.-Cons. Int. Explor. Mer **189**, 317–324 (1990).
[10] S. McClatchie, J. Alsop, and R. F. Coombs, ''A re-evaluation of relationships between fish size, acoustic frequency, and target strength,'' ICES J. Mar. Sci. **53**, 780–791 (1996).
[11] T. Mukai and K. Iida, ''Depth dependences of target strength of live kokanee salmon in accordance with Boyle's Law,'' ICES J. Mar. Sci. **53**, 245–248 (1996).
[12] K. Sawada, Z. Ye, R. Kieser, G. A. McFarlane, Y. Miyanohana, and M. Furusawa, ''Target strength measurements and modeling of walleye pollock and Pacific hake,'' Fish. Sci. **65**, 193–205 (1999).
[13] R. H. Love, ''Resonant acoustic scattering by swimbladder-bearing fish,'' J. Acoust. Soc. Am. **64**, 571–580 (1978).
[14] C. Feuillade and R. W. Nero, ''A viscous-elastic swimbladder model for describing enhanced-frequency resonance scattering from fish,'' J. Acoust. Soc. Am. **103**, 3245–3255 (1998).
[15] K. G. Foote, ''Linearity of fisheries acoustics, with addition theorems,'' J. Acoust. Soc. Am. **73**, 1932–1940 (1983).
[16] E. Ona, ''Physiological factors causing natural variations in acoustic target strength of fish,'' J. Mar. Biol. Assoc. U.K. **70**, 107–127 (1990).
[17] K. G. Foote and D. T. I. Francis, ''Comparing Kirchhoff-approximation and boundary-element models for computing gadoid target strengths,'' J. Acoust. Soc. Am. **111**, 1644–1654 (2002).
[18] L. H. Chen and D. G. Scheikert, ''Sound radiation from an arbitrary body,'' J. Acoust. Soc. Am. **35**, 1626–1632 (1963).
[19] G. Chertock, ''Sound radiation from vibrating surfaces,'' J. Acoust. Soc. Am. **36**, 1305–1313 (1964).
[20] L. G. Copley, ''Integral equation method for radiation from vibrating bodies,'' J. Acoust. Soc. Am. **41**, 807–816 (1967).
[21] L. G. Copley, ''Fundamental results concerning integral representations in acoustic radiation,'' J. Acoust. Soc. Am. **44**, 28–32 (1968).
[22] A. D. Pierce, ''Variational formulations in acoustic radiation and scattering,'' in *Underwater Scattering and Radiation*, edited by A. D. Pierce and R. N. Thurston (Academic, San Diego, 1992), pp. 195–371.
[23] H. A. Schenck, ''Improved integral formulation for acoustic radiation problems,'' J. Acoust. Soc. Am. **44**, 41–58 (1968).
[24] A. J. Burton and G. F. Miller, ''The application of integral equation methods to the numerical solutions of some exterior boundary problems,'' Proc. R. Soc. London, Ser. A **323**, 201–210 (1971).
[25] A. F. Seybert, B. Soenarko, F. J. Rizzo, and D. J. Shippy, ''An advanced computational method for radiation and scattering of acoustic waves in three dimensions,'' J. Acoust. Soc. Am. **77**, 362–368 (1985).
[26] T. W. Wu, A. F. Seybert, and G. C. Wan, ''On the numerical implementation of a Cauchy principal value integral to insure a unique solution for acoustic radiation and scattering,'' J. Acoust. Soc. Am. **90**, 554–560 (1991).
[27] T. W. Wu and A. F. Seybert, ''A weighted residual formulation for the CHIEF method in acoustics,'' J. Acoust. Soc. Am. **90**, 1608–1614 (1991).
[28] D. J. Segalman and D. W. Lobitz, ''A method to overcome computational difficulties in the exterior acoustics problem,'' J. Acoust. Soc. Am. **91**, 1855–1861 (1992).
[29] D. T. I. Francis, ''A gradient formulation of the Helmholtz integral equation for acoustic radiation and scattering,'' J. Acoust. Soc. Am. **93**, 1700–1709 (1993).
[30] W. L. Meyer, W. A. Bell, M. P. Stallybrass, and B. T. Zinn, ''Prediction of the sound field radiated from axisymmetric surfaces,'' J. Acoust. Soc. Am. **65**, 631–638 (1979).
[31] T. Terai, ''On calculation of sound fields around three dimensional objects by integral equation methods,'' J. Sound Vib. **69**, 71–100 (1980).
[32] K. A. Cunefare and G. Koopman, ''A boundary element method for acoustic radiation valid for all wavenumbers,'' J. Acoust. Soc. Am. **85**, 39–48 (1989).
[33] O. C. Zienkiewicz and R. L. Taylor, *The Finite Element Method*, 4th ed. (McGraw–Hill, London, 1989), Vol. 1.
[34] V. C. Anderson, ''Sound scattering from a fluid sphere,'' J. Acoust. Soc. Am. **22**, 426–431 (1950).
[35] R. R. Goodman and R. Stern, ''Reflection and transmission of sound by elastic spherical shells,'' J. Acoust. Soc. Am. **34**, 338–344 (1962).
[36] K. G. Foote, ''Rather-high-frequency sound scattering by swimbladdered fish,'' J. Acoust. Soc. Am. **78**, 688–700 (1985).
[37] K. Olsen, ''Orientation measurements of cod in Lofoten obtained from underwater photography, and their relation to target strength,'' Counc. Meet. Int. Counc. Explor. Sea 1971/B:17.
[38] K. G. Foote, ''Averaging of fish target strength functions,'' J. Acoust. Soc. Am. **67**, 504–515 (1980).
[39] National Defense Research Committee, ''Physics of sound in the sea'' (Department of the Navy, Washington, D.C., 1969), Chap. 28.
[40] K. G. Foote, ''Optimizing copper spheres for precision calibration of hydroacoustic equipment,'' J. Acoust. Soc. Am. **71**, 742–747 (1982).
[41] D. T. I. Francis and K. G. Foote, ''Acoustic excitation of the fish swimbladder,'' Bioacoustics **12**, 265–267 (2002).

Blind prediction of broadband coherence time at basin scales

John L. Spiesberger
Department of Earth and Environmental Science, 240 South 33rd Street, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6316

Frederick Tappert[a)]
Applied Marine Physics, Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida 33149

Andrew R. Jacobson
Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, New Jersey 08544-0710

(Received 28 March 2003; revised 19 September 2003; accepted 3 October 2003)

A blind comparison with data is made with a model for the coherence time of broadband sound (133 Hz, 17-Hz bandwidth) at 3709 km. Coherence time is limited by changes in the ocean because the acoustic instruments are fixed to the Earth on the bottom of the sea with time bases maintained by atomic clocks. Although the modeled coherence time depends a bit on the difficult problem of correctly modeling relative signal-to-noise ratios, normalized correlation coefficients of the broadband signals for the data (model) are 0.90 (0.83), 0.72 (0.59), and 0.51 (0.36) at lags of 2, 4.1, and 6.2 min, respectively. In all these cases, observed coherence times are a bit longer than modeled. The temporal evolution of the model is based on the linear dispersion relation for internal waves. Acoustic propagation is modeled with the parabolic approximation and the sound-speed insensitive operator. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1629305]

PACS numbers: 43.30.Re, 43.30.Zk, 43.30.Qd [RAS] Pages: 3147–3154

I. INTRODUCTION

It would be valuable to have oceanographic and acoustic models that blindly predict the broadband coherence time of sound over basin scales in the ocean. ''Blind'' means that the models are run with parameters that are taken from their expected values in the literature without any tuning to fit measurements. If such models could be constructed, they could be used to design systems, optimally process data for signals from different hypothetical locations, and serve as guides for developing theories. More specifically, reliable models for coherence time would be used by the surveillance community in the military. They would be used by ocean tomographers interested in designing sources and emitted waveforms whose receptions could gain the required signal-to-noise ratios through coherent integration. One would be able determine prior to an experiment the regimes where frequencies were sufficiently low so that propagation would be coherent for hours at a time if required. They would be used by those interested in designing and using wireless acoustic modems where one would be able to know if one could enjoy a coherent rather than incoherent communication system. As will be seen below, theories that have been applied to predict fluctuations of broadband sound at basin scales have serious discrepancies with data by one or more orders of magnitude. The existence of a reliable computational model (rather than a theory) will help guide the development of better theories. Problems with blind comparisons are useful for discovering deficiencies in models and their inputs, if any. So, for all these reasons, a blind prediction for the broadband coherence time of sound is made for a 3709-km section in the eastern North Pacific (Fig. 1) using a Monte Carlo approach. The situation is complicated by the interaction of sound with the bottom near the source and receiver. Despite these complications, it is found that the predictions for coherence time are similar to the measurements.

Blind experiments are important in most fields of science. For example, blind predictions for contemporary climate do not work[1–3] but rather need to be forced to fit current conditions. Blind clinical trials sometimes show the efficacy of a medical treatment, and sometimes do not.[4] This study may be the first blind comparison for coherence time of broadband sound over basin scales, which makes the resemblance with measurements more surprising considering the complicated nature of the transmission.

As will be seen, blind predictions for coherence time appear to be technically inconsistent with measurements at the 95%-confidence limit. Such agreement would perhaps be too much to hope for. But, the predictions are close enough to the measurements so that they appear to have utility.

One factor favoring the value of the present comparison is that the source and receiver are on the bottom. So, all the temporal changes of the signal are due to temporal changes within the ocean. Perhaps the most complicated aspect of the comparison is that the coherence time of the model depends on an accurate representation of the relative signal-to-noise ratios of the paths for each transmission. This sensitivity is small, but predictions still depend on this factor. It is possible to use the data to obtain accurate estimates of the signal-to-noise ratios, but the relative amplitudes of the model and data naturally differ on a sample-by-sample basis. Amplitudes do not obey a principle analogous to Fermat's in which travel time is insensitive to first-order changes in path. Thus,

[a)]Deceased.

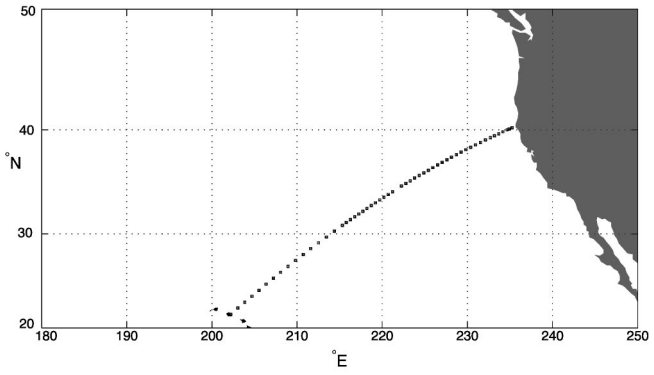


FIG. 1. Plan view of the Kaneohe source experiment. The source is located about 5 miles north of Kaneohe Bay, Oahu. The receiver near the coast of Northern California is one of many U.S. Navy SOSUS stations used to receive these transmissions. Circles indicate the positions of 56 stations where conductivity, temperature, and depth were measured with a CTD in July 1988 by the Naval Oceanographic Office.

the problem of getting the relative amplitudes right is generic to all practitioners, and is exacerbated by the fact that sound interacts with the sub-bottom where the geoacoustic properties are poorly known. Such cases deserve attention as there are so many examples where sources and receivers sit on the bottom, and predicting coherence time is valuable.

Time variability in the models is derived from the linear dispersion relation for internal waves obeying a Garrett–Munk spectrum.[5] The acoustic model is based on the parabolic approximation with the sound-speed insensitive operator.[6] It is remarkable because it appears to offer a nearly perfect solution for the travel times of broadband signals over basin scales over all launch angles without any tunable parameters.

Two studies[7,8] investigated something related to but different than coherence time of broadband sound at basin scales. In Ref. 7, the autocorrelation function is computed for travel time. The autocorrelation of a wavefront's travel time may not be sufficient to compute an integration time if phase changes of the wavefront cannot be deduced from arrival times of peaks. In other words, Ref. 7 reports an upper limit of 2 h for the *incoherent* quantity of travel time based on the amplitude (not phase) of a wavefront. Another study reports a coherent integration time of 764 s for this 3250-km transmission in the North Pacific for a source at 75 Hz and 37.5-Hz bandwidth.[9] No theory or model is used to compare with this observation.

The existing theory used in Refs. 7 and 8 to predict acoustic fluctuations may not be suitable to predict the coherence time of broadband sound. In these references, modeled energy of internal waves is adjusted so that theoretical and measured variances of travel time are the same. These reports say that the theory predicts fluctuations to be in the fully saturated regime, while instead most of the observations are reported in the unsaturated to partly saturated regimes. The reports go on to say that observations of pulse spread are two orders of magnitude different than predicted from theory. The cause for the mismatch is attributed to the fact that the theory is designed for single-frequency emissions, which the authors say is inadequate to predict broadband propagation.

The present investigation differs from those studies because acoustic data are not used here to tune a theory to fit the data, and, in fact, no theory is used. Instead, a Monte Carlo approach is used to blindly predict acoustic fluctuations with the parabolic approximation. This approximation includes the physics of finite-wavelength propagation for broadband signals. Because the prediction fits the data rather well, it appears that this and possibly other blind predictions would prove useful for improving and testing theories for the coherence time of broadband sound at long distances.

II. DATA AND MODELS

A. Data

The Kaneohe source is mounted on the bottom at 183-m depth on the north coast of Oahu at 21.512 35° N, 202.228 49° E. Continuous transmissions occur for 5 days in late 1983 to Sound Surveillance Stations (SOSUS) stations, one of which is at a distance of 3709.21 km and a depth of 1433 m near northern California at 40.078 56° N, 234.887 97° E (Fig. 1). The 183-dB *re*: 1 μPa @1 m phase-modulated signal has a center frequency of 133 Hz and a bandwidth of 17 Hz. Phase is modulated with a 511-digit maximal shift register sequence every 8 carrier cycles. The period of the signal is $511*8/133 \cong 30.7$ s. The received signal is beamformed, complex demodulated, low-pass filtered to suppress the double-frequency component, correlated with a replica to implement a matched filter, and coherently averaged over four sequence periods to boost the signal-to-noise ratio. Replica correlation compresses 30.7 s of energy along each acoustic path into a single pulse of 1/17 Hz = 0.06-s resolution without sidelobes. The output consists of records at $4 \times 30.7 \cong 123$-s intervals. Each record contains $511 \times 4 = 2044$ complex demodulates with 0.015-s separations. Atomic clocks maintain time stability at the source and receiver with a fractional frequency error of about 10^{-11}. The bathymetry in the source region is measured with an extensive SEABEAM survey.[10] Further details are discussed elsewhere.[11]

B. Model

Except for a few differences discussed below, the modeling of internal waves has been described before in detail.[12] Reference 12 includes successful comparisons of the modeling with the power spectral density of vertical displacement of water as a function of horizontal wave number, standard deviation of vertical displacement of water as a function of depth, and horizontal correlation length scale as a function of depth. A brief summary is provided here.

The sound-speed field is taken from an eddy-resolving conductivity and temperature versus depth (CTD) survey in July 1988 (Fig. 1). Most of the CTD stations extend to 2-km depth, and some to 4500 m. For comparison, the ray construction for the propagation shows lower turning depths at 2.5 km and above.[13]

The data we have do not resolve internal waves. Sound speeds below the depth of the CTD stations are taken from Levitus' climatology for spring.[14] A range-dependent three-dimensional field of internal waves is synthesized along the geodesic. Sound-speed perturbations associated with internal waves are generated by assuming that vertical displacements of water lead to adiabatic changes in the speed of sound. The acoustic propagation models described below utilize the vertical slice along the geodesic through the three-dimensional field. The internal waves have the full spectrum given by Garrett and Munk.[5] The longest horizontal wavelength represented is 80 km. The boundary conditions are zero for the vertical modes at the top and bottom; thus, the WKB approximation is not used. About 50 vertical modes are numerically computed in each 80- by 80-km horizontal region. The geodesic runs through the centers of about 50 such regions constructed by dividing the 3709-km length by the 80-km size of each region. The attendant perturbations in sound speed are smoothed at region boundaries to avoid discontinuities. Frequencies and modes of internal waves are precomputed and the linear dispersion relation is used to synthesize the field at the geophysical times desired.

Sound interacts with the bottom near the source and receiver, and this is modeled with a geoacoustic bottom.[12] Geoacoustic parameters near the receiver are modeled differently for steep and flat arriving energy, so each encounters a different reflection coefficient due to a different effective density in the sediments. Modeling the bottom in these two ways provides a way to account for the observed relative levels (Ref. 12). The geoacoustic parameters near the receiver are probably not known well enough to modify these parameters for any justifiable reason.

The requirement of this paper is to choose inputs for models that are not tuned to observations. So, instead of trying to match the relative amplitudes in the data with two different geoacoustic sets of parameters,[12] only one set of geoacoustic parameters is chosen near the receiver. The chosen geoacoustic parameters will greatly underestimate the amplitudes at the end of the reception. The same parameters near the receiver are chosen as before (Table CII, Ref. 12) except the ratio of speeds at the top of the sediment to the bottom of the water column is 1.02, and the sediment density is 1.7 kg/m^3. These parameters are chosen without regard to their effects on the model.

The sound-speed insensitive parabolic approximation[6] is used to model the propagation at each of many frequencies. An inverse Fourier transform is used to synthesize the broadband impulse response. That response is complex demodulated to produce samples at 0.0152-s intervals. This is similar to the data interval. A running average of 4 complex demodulates is then used to mimic the matched filter for the data. The computational grid[12] is small enough to yield convergence within a few decibels at the receiver. The parabolic approximation includes acoustic absorption that depends on frequency in the standard way.

Models are synthesized from realizations of the internal wave field at 123-s intervals over a geophysical time of 3.4 and 2 h, respectively, for normal and half-normal energy of the Garrett–Munk spectrum.[5] Thus, there are 100 and 59 realizations, respectively, for the normal and half-normal runs. Each realization requires 5 h on an AMD Athlon 1800+ processor, so all model realizations require 795 h. Fifty-nine realizations at the half-normal level appear to be sufficient to examine the sensitivity of the model predictions.

III. RESULTS

A. Incoherent averages

The long-term stability or instability of features in the data can often be investigated by averaging the receptions during long periods of time. If the data were phase coherent for many hours at a time, one would naturally coherently average the complex records together, and look for stable features between separate averaging periods. Since the data in this experiment are not phase coherent over periods like hours, the stability of features can be investigated using an incoherent average, i.e., an average that discards acoustic phase. The incoherent average is one way to form this average. Averages are made over intensity, and then a square root can be taken to yield an amplitude scale.

The incoherent average for the mth demodulate, $a[m]$, is formed from N_{rec} records as

$$a[m]=\left[\frac{1}{N_{\mathrm{rec}}}\sum_{r=1}^{N_{\mathrm{rec}}}\frac{\|d[m,r]\|^2}{\sigma[r]^2}\right]^{1/2},\quad m=1,2,\ldots,M, \tag{1}$$

where the mth complex demodulate of the rth record is $d[m,r]$. Note that $a[m]$ is a measure of the expected value of a dimensionless amplitude of $d[m,r]$ because $d[m,r]$ is divided by its standard deviation. The variance of the noise for record r is $\sigma[r]^2$, and is included to give proper weight to records based on their signal-to-noise ratios. When forming incoherent averages from data, $\sigma[r]^2$ is estimated from each data record where signal is not present. This is easy to do because each record consists of about 30.7 s of complex demodulates, only 4 s of which cover the time that the energy is present in significant quantity.[11] When forming incoherent averages from model realizations, $\sigma[r]^2$ is set to unity.

Noise is added to each model record in the following way prior to forming the incoherent average. The average signal-to-noise ratio in the data is not stationary from one record to the next. For data record i, the average signal-to-noise is estimated between the travel times of 2504.2 to 2505.2 s. Then, 5 s of noise from the data are taken from record i, and added to the ith model realization with the same signal-to-noise ratio as the modeled travel times from 2504.2 and 2505.2 s.

An addition of 0.367 s to modeled travel times aligns them with the data on 29 Nov. 1983 (Fig. 2). Rossby waves are likely responsible[15] for some of the 0.367 s. In order to achieve a match between the duration of the energy between model and data, a previous investigation finds it necessary to add internal wave and mesoscale components to the climatological averages of sound speed.[12] It appears that the model duration in Fig. 2 is similar to the data, but the model amplitudes are too low at the end near 2507.5 s, just like that found before for similar values of geoacoustic parameters near the receiver.[12] If two models are used with different sets

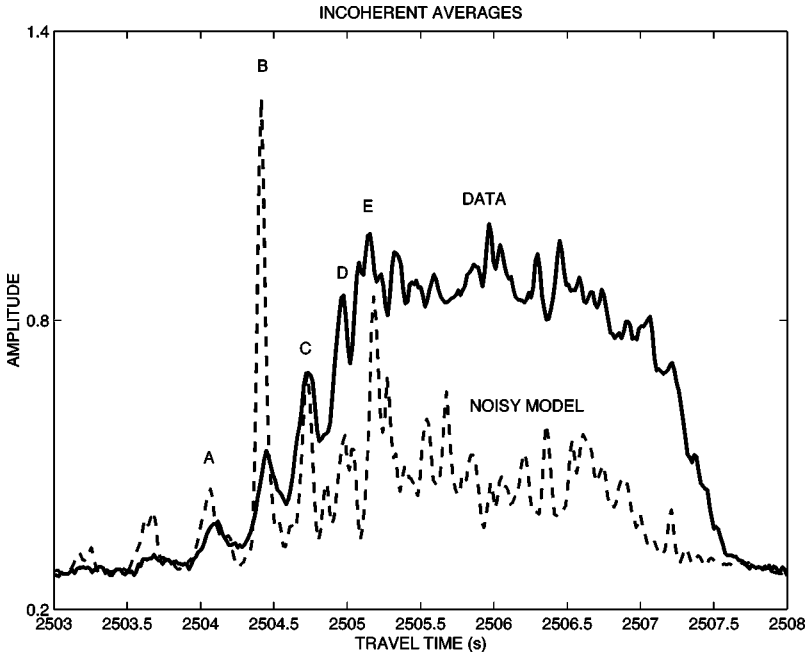


FIG. 2. Incoherent averages for the data at Northern California on 29 Nov. 1983 (Fig. 1) compared with an incoherent average of model realizations. Paths A–E can be tracked throughout the intermittent transmissions covering 6 years, 1983–1989. The time evolution of the modeled sound-speed field is determined by the linear dispersion relation of internal waves used by Garrett and Munk (Ref. 5). Modeled internal waves are superimposed on the mesoscale resolving CTD section (Fig. 1). The incoherent average of the data consists of 703 records at 123-s intervals, which covers a day. The incoherent average from the model consists 100 runs separated by 123 s each, which is a duration of 3.4 h, 0.367 s is added to modeled travel times to align with the data. Rossby waves are probably responsible for some of the 0.367 s (Ref. 15).

of geoacoustic parameters for flat and steeply arriving energy, amplitudes better resemble the observations.[12]

Even if two sets of geoacoustic parameters are used to better predict amplitudes, there are still differences between model and data. Four possibilities for this difference are listed. First, the models only cover 3.4 h of geophysical time, and the data cover a day. Second, the daily incoherent averages of the data are not stationary (Fig. 3 in Ref. 13). Third, the geoacoustic parameters for the bottom are imperfect. Fourth, the sound-speed field for the model is imperfect. A further discussion of amplitude differences is beyond the scope of this paper.

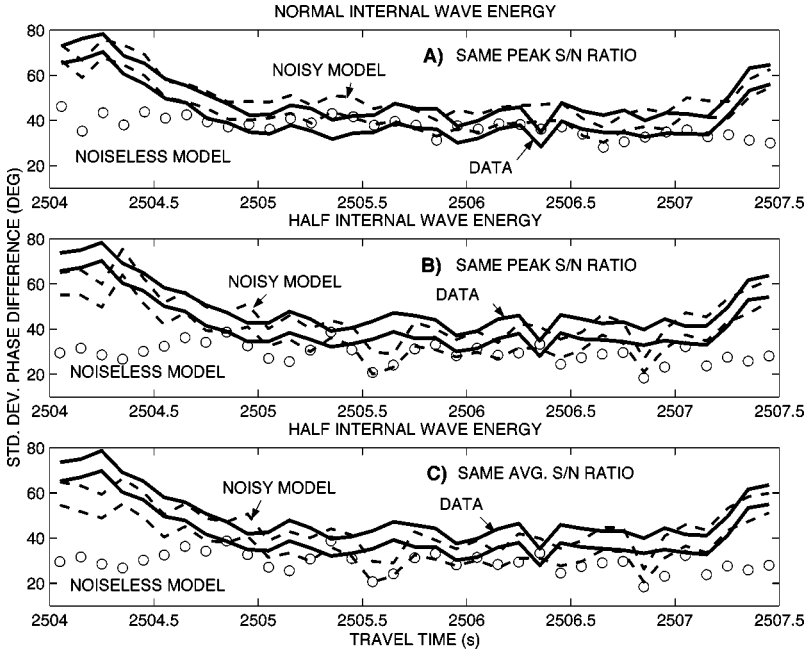


FIG. 3. The 95%-confidence limits, via the bootstrap (Ref. 16), for the standard deviation of change in acoustic phase at 123-s intervals across a 3709-km section (Fig. 1) for the data, noiseless, and noisy models. The standard deviation is estimated in 35 windows of 0.1-s duration each between 2504 and 2507.5 s of travel time. The comparison is made using data records starting at 1 Dec. 1983 18:13:12 (GMT). The model is the sound-speed insensitive parabolic approximation (Ref. 6). The sound-speed field has a mesoscale (Fig. 1) and a time-evolving internal wave field based on the linear dispersion relation for internal waves. (A) The model uses the normal Garrett–Munk (Ref. 5) energy spectrum where the peak signal-to-noise ratio in each data window sets the peak signal-to-noise ratio in each model window. Models are generated from 100 realizations of internal waves at 123-s intervals. Models are compared to 100 consecutive data records. (B). Same as (A) except the model has one-half the Garrett–Munk energy level and uses 59 realizations of the internal wave field at 123-s intervals. The models are compared to 59 consecutive data records. (C) Same as (B) except the signal-to-noise ratio of the average of the five largest intensities in each 0.1-s data window sets the signal-to-noise ratio of the five largest intensities in each model window. Model results are given with and without acoustic noise. The 95% limits are not given for the noiseless model.

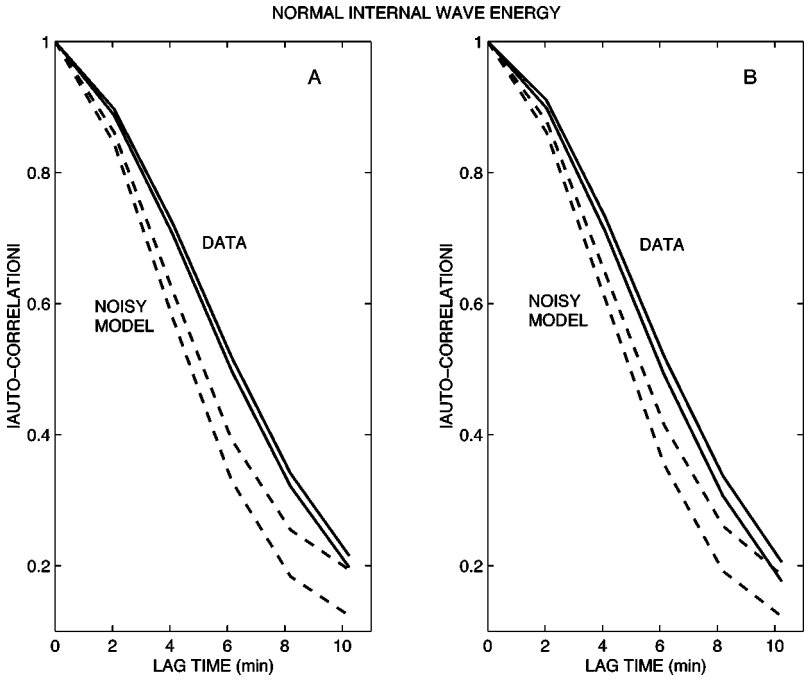


FIG. 4. Coherence time via Eq. (2) of the data and models with 95% confidence limits computed via the bootstrap (Ref. 16). The 95%-confidence limits are indicated by two curves for the data and two curves for the model. An average signal-to-noise ratio in each data record is used to set the same in each model record to produce the ''noisy model'' results. The comparisons are made with normal energy in the Garrett–Munk spectrum of internal waves using the linear dispersion relation of these waves to evolve the fields at 123-s intervals for 100 model realizations. The model realizations are compared with 100 consecutive data records. (A) Coherence time for energy between 2504 and 2507.5 s (Fig. 2). (B) Same but for energy between 2505.5 and 2506.6 s.

B. Phase differences

Differences in acoustic phase between adjacent records at the same travel time are estimated by first unwrapping phases by choosing the angle closest to the prior one, and then taking the difference in angle between the unwrapped phases. The largest value for the standard deviation of these differences is given by the standard deviation of a uniformly distributed random variable on the interval $[0,2\pi]$ which is $\sqrt{(2\pi)^2/12}$ radians or 104 deg. The standard deviation of the phase differences is plotted (Fig. 3) for travel times between 2504 and 2507.5 s because this is when the signal arrives (Fig. 2).

Depending on the precise time window used for computation of the Fourier transform of the time series, the scintillation index has values between 0.8 to 1.4 at 133 Hz. Phase differences for noisy models are about 20 to 40 deg larger than noiseless models at travel times near 2504 and 2507.5 s because the signal-to-noise ratio is small (Fig. 3). As the signal-to-noise ratio rises, the differences between noiseless and noisy models decreases to 5 to 10 deg. The standard deviation of phase difference is not a strong function of travel time in the noiseless model case or in the data, except for the data when the signal-to-noise ratio is lower near 2504 and 2507.5 s. The standard deviations of phase differences are significantly less than 104 deg, which would be that due to independently distributed uniform random noise. We conclude that a blind comparison between model and data yields statistically compatible results for most of the record.

Turning to a nonblind prediction, it is found that halving the energy in the Garrett–Munk spectrum of internal waves tends to decrease the standard deviations by about 10 deg in the noiseless models (panels A and B, Fig. 3). The standard deviations of the noisy models decrease in places by perhaps 10 deg, but the change is smaller because noise is not allowing the standard deviations to decrease as much as that from the effects of reducing the energy in the internal waves.

The standard deviation of the noisy model decreases by about 5 deg or less when the average rather than peak signal-to-noise ratio is imposed on the model (panels B and C, Fig. 3) at the one-half energy level for internal waves. Given the sensitivities of the comparison to the manner in which noise is added to the model, it does not appear to be productive to use further analysis to determine which energy level gives the best fit with the data.

C. Coherence time

Coherence time (Fig. 4) is estimated using an autocorrelation function

$$A(p)=\left|\frac{\Sigma_{r=1}^{R}\Sigma_{m=1}^{M}d[m,r]d^{*}[m,r+p]}{\Sigma_{r=1}^{R}\Sigma_{m=1}^{M}d[m,r]d^{*}[m,r]}\right|, \tag{2}$$

where the superscript * denotes complex conjugate, and the vertical bars denote the modulus of the complex number. 95% confidence limits for $A(p)$ are estimated using the bootstrap.[16] The 95%-confidence limits are indicated in Fig. (4) by two curves for the data and two curves for the model. R pairs of records are used for the estimate and M complex demodulates are used for two windows of arrival time. The first is from 2504 to 2507.5 s, which encompasses most of the energy (Fig. 2). The second is from 2505.5 to 2506.5 s, a region where the signal-to-noise ratios are high and the relative amplitudes in the data and model are relatively flat. In both cases, an average of the largest N intensities in the data in each record is used to estimate an average intensity signal-to-noise ratio to set the same in the model. N is 200 and 50 for the wider and narrower windows, respectively. A smaller value of 50 is used for the latter because that window contains only 66 samples. Correlation values in Fig. 4 drop by amounts of between 0 and 0.3 for the noisy model if the peak signal-to-noise ratio in each data window is used to set the peak signal-to-noise ratio in each model window on a record-

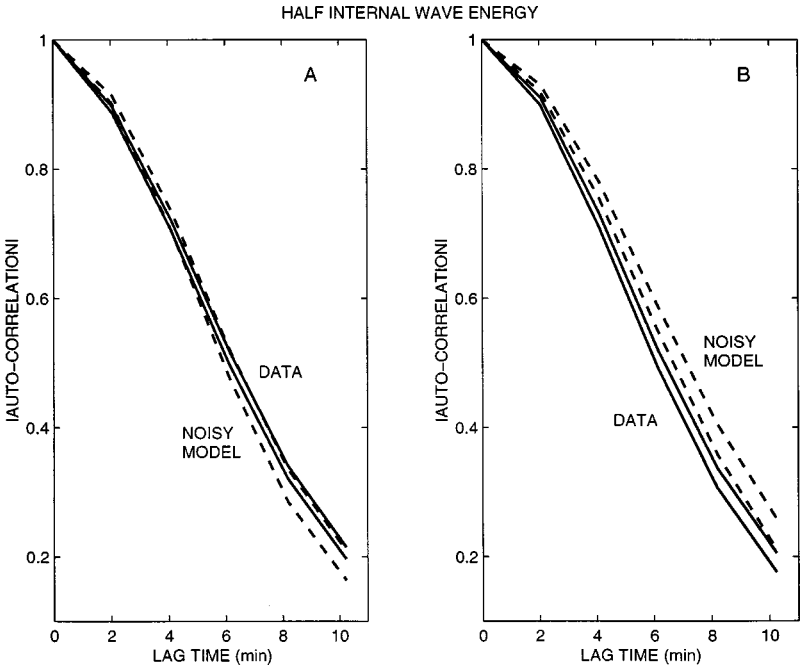


FIG. 5. Same as Fig. 4 except for half-normal energy in the Garrett–Munk field of internal waves (Ref. 5). In this case, 59 model realizations are compared with 59 data records.

by-record basis. Evidently, specifying the model's signal-to-noise ratio from the highest peak in the data results in an overall smaller signal-to-noise ratio in the model than setting the model's signal-to-noise ratio according to the highest N intensities in the data.

Figure 4 represents a blind comparison of model with data. Because some of the literature[7,8] discusses the possibility that a half-normal energy of internal waves may better fit observations, a prediction based on this lower energy is made here (Fig. 5). This is not a blind comparison because the expected energy level for internal waves is not used. The noisy model looks like observed coherence times for the longer data window, but not for the shorter one.

IV. BATHYMETRY AND GEOACOUSTIC BOTTOM: EFFECTS ON COHERENCE TIME

One might wonder whether the rather close agreement between modeled and observed coherence time is fortuitous. Lines of reasoning based on the modeling and data support the conclusion that the agreement is not fortuitous. In time, however, the strongest test for the validity of a new scientific finding or approach involves many scientists who apply the technique to many different experiments and obtain similar results.

The fact that the blind modeling yields an answer close to observed coherence times seems to be a reason to believe that the agreement is not fortuitous. There are many ways the modeling could have gone astray. Acoustic models that do not couple modes or that do not yield accurate travel times for a wide range of acoustic launch angles would seem to yield highly inaccurate impulse responses. Indeed, the sound-speed insensitive parabolic approximation,[6] the bathymetry, and geoacoustic parameters yield a good match to the impulse response.[12,13] The bathymetry was carefully measured by SEABEAM near the source and by the Navy near the receiver. The set of geoacoustic parameters used here is selected from the best available unclassified values in the literature.[12] There are undoubtedly bathymetric profiles and geoacoustic parameters that would not yield a good match with the data. It is often difficult to get a model to agree with observations without tuning or fitting.[1–3] The fact that the modeled impulse response looks like the data indicates that the models have the acoustic paths about right, and lends evidence to support that the bathymetry and geoacoustic parameters are reasonable.

Going further, let us continue to hypothesize that uncertainties in bathymetry or in the geoacoustic parameters lead to significant changes in predicted coherence time. In other words, one is invoking a time-independent process to explain changes in a time-dependent phenomenon. The only thing in the modeling that has time dependence is the evolving internal wave field. If the internal waves are frozen, then one obtains an infinite coherence time at the receiver. Let us see where the turning on of ocean fluctuations leads us, taking into account the observations. Much of the discussion that follows has appeared before.[10]

The only way this time-independent phenomenon can affect this time-dependent phenomenon is if temporal fluctuations in the water column lead to significant changes in the acoustic travel times by changing the *paths* by which sound travels. The evidence against significant changes in path is significant for the following reasons.

First, changes in travel time due to changes in path are guaranteed to be of second-order importance because of Fermat's principle. The zero-order change in travel time is an integral of the fluctuations of sound speed due to internal waves along a frozen ray path. The first-order change is due to changes in path due to those fluctuations, which is zero because of Fermat's principle. The second-order change is due to changes in path. Fermat's principle thus supports the notion that modeled coherence time is insensitive to plausible changes in the geoacoustic parameters with respect to the set of values used in this paper.

Second, sound interacts with the bottom and sub-bottom with different angles of incidence in this experiment at both the source and receiver. This means that the sonic paths themselves sense different effective bathymetric and geoacoustic values, and all the sonic energy interacts with the bottom in this experiment.[12] The evidence in this paper (Figs. 3, 4) demonstrates that the coherence time of sound is not sensitive to which acoustic energy is being analyzed. This means that actual different bathymetric and geoacoustic values that are important for the different paths are not important in changing their coherence times. Thus, the data themselves provide a sensitivity analysis that points to a lack of sensitivity in the calculations for coherence time.

Third, experimental evidence supports the hypothesis that any changes in path geometries lead to small changes in travel time. The travel times of the peaks of the five stable arrivals all change by the same amount at the same time by up to 1/2 second, within measurement error.[10] If these paths change significantly, it is hard to see why their travel time changes would be the same since the points at which they reflect from the bottom are all different from one another, and it would be expected that their travel times would thus change in a discordant manner.

Fourth, evidence based on detecting small tidal signals suggests that any path changes lead to very small changes in travel time. The barotropic and internal tides generated by flat-topped seamounts several thousands of kilometers from the source can be accurately estimated at five SOSUS stations, despite the fact that the tidal signals amount to only about 10 ms of travel time. The time series of these 10-ms tidal oscillations are very clean, showing little evidence of noise. The ability to detect the small signals is due to the use of the phase and amplitude of the acoustic signals. Indeed, the measurements are made with an accuracy of 135 μs at 2-min intervals over several months.[10]

These lines of reasoning support a conclusion that the agreement between modeled and observed coherence time is not sensitive to plausible changes in the bathymetry and the geoacoustic values in this experiment.

V. CONCLUSIONS

Blind predictions for the coherence times of broadband signals at basin scales are similar to but a little less than observed (Fig. 4). The blind prediction utilizes the the Garrett–Munk spectrum of internal waves[5] and a sound-speed insensitive parabolic approximation.[6] A blind prediction for the standard deviations of broadband phase differences at 2-min intervals is statistically consistent with observations (Fig. 3). The results above depend somewhat, but not much, on the method used to assign a signal-to-noise ratio to each complex sample from the model.

Despite the fact that the overall relative amplitudes in the model and data differ, especially near the coda, the predictions for coherence time do not change much (e.g., panels A and B in Fig. 4) when different windows in arrival time are used for the blind comparison. This is fortunate because it is usually difficult for practitioners to model amplitudes for sources and receivers on the bottom. The difficulty in getting the relative amplitudes right probably comes from the lack of accurate geophysical parameters for the sub-bottom, limited ways that the acoustic model can incorporate those parameters, a lack of a stability principle for amplitudes, and the fact that the ocean's change from day-to-day leads to changes in the relative amplitudes of the paths (Fig. 3, Ref. 13).

Departing from a blind prediction, a nonblind prediction is made for internal waves at one-half normal energy. The comparison with the standard deviation of phase at 2-min intervals agrees with the data at both half and normal energy (Fig. 3). The comparison with coherence time agrees with the data for one data window and disagrees with another data window (Fig. 5). We believe that the analysis in this paper is insufficient to decide if the half-normal energy better fits the data than the case with normal energy. In order to decide this issue, we would feel more comfortable if there were more degrees of freedom from which to make comparisons. For example, it would be desirable to have a month of continuous results from models and data.

Considering the discrepancies between blind predictions and observations in some fields such as the study of the Earth's climate,[1–3] it is remarkable that blind predictions for sound comes so close to reality. It is worthwhile to try blind predictions for different experiments.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research Contract No. N00014-00-C-0317. John Spiesberger thanks Bruce Einfalt for programming expertise and Michael Wolfson for previous collaborative research during which some of the computer software used here was developed. We thank the reviewers for catalyzing the perspective of how the results in this paper fit into the field of sound transmission in the ocean and for questions which led us to improve its quality, including the addition of Sec. IV.

[1] S. Manabe, R. J. Stouffer, M. J. Spelman, and K. Bryan, "Transient responses of a coupled ocean–atmosphere model to gradual changes of atmospheric CO_2. I. Annual mean response," J. Clim. **4**, No. 8, 785–818 (1991).

[2] S. Manabe and R. J. Stouffer, "Century-scale effects of increased atmospheric CO_2 on the ocean–atmosphere system," Nature (London) **364**, 215–218 (1993).

[3] J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson, "Climate Change 2001: The Scientific Basis," Contribution of Working Group I to the *Third Assessment Report of the Intergovernmental Panel on Climate Change* (IPCC) (Cambridge University Press, Cambridge, 2001).

[4] *Fundamentals of Clinical Trials*, 3rd ed., edited by L. M. Friedman, C. D. Furberg, and D. L. DeMets (Springer, New York, 1998).

[5] C. Garrett and W. Munk, "Space-time scales of internal waves," Geophys. Fluid Dyn. **2**, 225–264 (1972).

[6] F. Tappert, J. L. Spiesberger, and L. Boden, "New full-wave approximation for ocean acoustic travel time predictions," J. Acoust. Soc. Am. **97**, 2771–2782 (1995).

[7] J. A. Colosi, E. K. Scheer, S. M. Flatte, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, P. F. Worcester, B. M. Howe, J. A. Mercer, R. C. Spindel, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, "Comparisons of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern North Pacific Ocean," J. Acoust. Soc. Am. **105**, 3202–3218 (1999).

[8] S. M. Flatte, J. A. Colosi, M. A. Dzieciuch, and P. F. Worcester, "Acoustic observations of internal-wave strength in the Mid-Pacific in 1989 and 1996," J. Acoust. Soc. Am. **100**, 2582 (1996).

[9] P. F. Worcester, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, B. M. Howe, J. A. Mercer, R. C. Spindel, J. A. Colosi, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, ''A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3185–3201 (1999).
[10] J. L. Spiesberger, ''An updated perspective on basin-scale tomography,'' J. Acoust. Soc. Am. **109**, 1740–1742 (2001).
[11] J. L. Spiesberger, P. B. Bushong, K. Metzger, and T. G. Birdsall, ''Ocean acoustic tomography: Estimating the acoustic travel time with phase,'' IEEE J. Ocean. Eng. **14**, 108–119 (1989).
[12] M. A. Wolfson and J. L. Spiesberger, ''Full wave simulation of the forward scattering of sound in a structured ocean: A comparison with observations,'' J. Acoust. Soc. Am. **106**, 1293–1306 (1999).
[13] J. L. Spiesberger and F. D. Tappert, ''Kaneohe acoustic thermometer further validated with rays over 3700 km and the demise of the idea of axially trapped energy,'' J. Acoust. Soc. Am. **99**, 173–184 (1996).
[14] S. Levitus, ''Climatological atlas of the world ocean,'' in *NOAA Professional Paper 13* (U.S. Government Printing Office, Washington, DC, 1982).
[15] J. L. Spiesberger, H. E. Hurlburt, M. Johnson, M. Keller, S. Meyers, and J. J. O'Brien, ''Acoustic thermometry data compared with two ocean models: The importance of Rossby waves and ENSO in modifying the ocean interior,'' Dyn. Atmos. Oceans **26**, 209–240 (1998).
[16] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman and Hall, New York, 1993).

Equivalence of Gaussian and piston ultrasonic transducer voltages

Oleg I. Lobkis[a)]
Department of Theoretical and Applied Mechanics, University of Illinois, Urbana, Illinois 61801

D. E. Chimenti
Center For Nondestructive Evaluation and Aerospace Engineering and Engineering Mechanics Department, Iowa State University, Ames, Iowa 50011-3020

(Received 18 October 2002; revised 11 July 2003; accepted 11 August 2003)

A plane-wave decomposition of collimated beams and electromechanical reciprocity relations are used to demonstrate fundamental differences and unusual similarities about transducer fields and transducer voltages at ultrasonic frequencies under various conditions. It is shown that the voltage induced by a transmitting acoustic piston transducer (constant particle velocity over the transducer surface) radiating into an ideal fluid medium on a second identical piston transducer, operating as a receiver, is nearly identical to the voltage observed when the two transducers have instead a Gaussian radial surface velocity distribution. The strong similarity in induced voltage begins when the two devices are separated by only several acoustic wavelengths, still well within the nearfield of both transducers, and the similarity increases with separation. Cases of transducer velocity distributions different from piston, such as Lorentzian or triangular, yield results nearly identical to the piston case. Transducers of differing size are also treated. It is further shown that an ''equivalent transducer'' can be derived for any combination of radiators, whose field is identical to the voltage measured using the pair. Moreover, with this concept the calculation of voltage in two-probe experiments is as simple as well-known approximations to the Rayleigh integral for a single transducer. These results have substantial consequences for calculations, either analytical or numerical, that predict the voltage measured in two-transducer experiments. [DOI: 10.1121/1.1616927]

PACS numbers: 43.35.Bf [YHB] Pages: 3155–3166

I. INTRODUCTION

Plane wave decomposition of the field of a finite acoustic radiator is a well-established method used routinely to describe mathematically the spatial and temporal variations of transducer beams, often in the calculation of acoustic fields of ultrasonic piston radiators. The idea is to represent the collimated beam as a spatial superposition of plane waves, each having a different spatial frequency and amplitude. The infinite coherent sum of these plane wave components is then equivalent to the original beam, as is well known. In most real experimental situations, however, the field of an acoustic transducer (or the field after interaction with an immersed structure) is seldom directly measured; more typically, the field is sampled by sensing the resulting acoustic pressure with a second, often identical, reciprocal transducer operating as a receiver.

We calculate here the voltage induced on a second transducer at ultrasonic frequencies resulting from the radiation of a transmitting transducer, and we demonstrate several novel features of the transducer voltage at the receiver as a function of the relative transducer positions, their sizes, and the distribution of particle velocities on their surfaces. Surprisingly, the voltage induced by a transmitting piston transducer on a second identical piston receiving transducer is nearly identical to the voltage observed when the particle velocities of two transducers have a radial Gaussian dependence. The strong similarity in the induced voltage is already apparent when the two ultrasonic transducers are separated by only a dozen or so acoustic wavelengths from each other, that is, while they are still well within each other's nearfield. As the separation increases, the two lateral voltage dependences become more and more similar, until the combined directivity functions of the piston fields can be justifiably replaced by Gaussian directivity functions. The origin of this phenomenon is in the averaging effect of the second piston radiator, when it acts as a coherent detector of the acoustic field.

Among many authors, Brekhovskikh[1,2] and Godin[3,4] have undertaken analyses of bounded acoustic beams using a spectrum of plane waves. In addition to these calculations, many applications involving collimated acoustic beams incident on fluid–solid interfaces have relied on similar methods.[5–13] Following the pioneering investigations of Schoch,[5] the initial rigorous treatment of interacting acoustic beams was performed for fluid–solid halfspaces by Bertoni and Tamir,[7] exploiting a similar theoretical calculation for electromagnetic waves. Their method was extended to plates by Ra *et al.*[8] Others[9] have used the Bertoni formulation as the basis of further investigations, and Schmidt and Jensen[10] demonstrated a full-field solution for Rayleigh waves, also based on spectral integrals over plane waves. Hybrid ray-wave schemes have been proposed by Ruan and Felsen.[11] In Ref. 12, the method was extended to curved layered plates, and Lobkis *et al.*[13] exploited reciprocity to develop an accu-

a)Electronic mail: lobkis@uiuc.edu

rate model of the voltage detected on a finite piston receiving transducer.

Moreover, plane-wave decomposition is also widely used in optics and microwave engineering[14,15] to represent aperture-limited beams of coherent electromagnetic radiation. In fact, expansions in plane waves are applicable to beams of any scalar wave phenomenon. In the case of acoustics at ultrasonic frequencies, however, beam effects take on special importance because the devices used to generate *and* detect acoustic waves are phase-sensitive across their aperture; that is, they are coherent transmitters and detectors. This fact is especially significant on detection at finite receivers, where amplitude summing occurs over the face of the receiver, and the resulting detected voltage is influenced by a combination of the transmitter and receiver directivity functions, as we have shown earlier.[13]

The entire literature on the topic of transducer radiation is quite extensive, however, and we cannot possibly review it fully here.[16–21] Most of the emphasis of this prior work is on the pressure fields of transducers or diffraction corrections to be used when estimating intrinsic attenuation from measurements. The transient behavior of acoustic transducers has also been actively investigated by many authors. Our concern in this article is quite different in that we seek the voltage induced in monochromatic cw excitation on a second transducer acting as a receiver and, in particular, the lateral dependence of that voltage, at any range, as the central axes of the transducers are displaced from each other. We extend the earlier preliminary results to the case of transducers of arbitrary dimension and position and analyze the problem of transducer voltage for a variety of particle velocity distributions on the transducer face. We find that in general only the rectangular velocity distribution (appropriate for most piezoelectric devices) yields the well-known highly variable acoustic field, whereas its voltage behavior is nearly triangular in the extreme nearfield, smoothing to nearly Gaussian after a range equivalent to about only one transducer diameter. Moreover, there is a receiver diameter (different from the transmitter) for which the voltage at the transducer center behaves like the field of the smaller device. This is an unusual and, as far as we can tell, original result of our calculation. We demonstrate this, and related phenomena, numerically in the last section of the article.

II. THEORETICAL DEVELOPMENT

Let two planar transducers (transmitter and receiver) be located facing each other in an acoustic coupling fluid, such as water. The range distance between them is z, and their acoustical axes are displaced by a distance r, as shown in Fig. 1(a). A closely related and essentially equivalent geometry is that of two transducers deployed on one side (or opposite sides in transmission) of a uniform plate, as seen in Fig. 1(b). The transducers have radii of a and b, respectively. Because we are dealing here with ultrasonic frequencies, the spatial dimensions of the transducers are much greater that the acoustic wavelength λ.

Although the two geometries illustrated in Fig. 1 are equivalent and lead to physically equivalent results, the first is easier to analyze because the problem can be conveniently

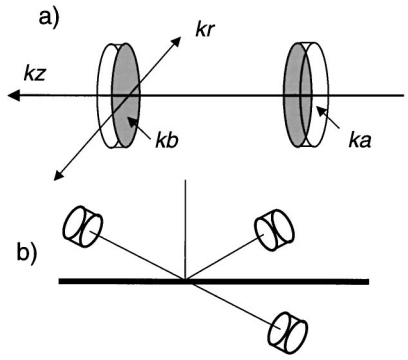


FIG. 1. Schematic diagrams of two-transducer experiments; (a) transducers facing; (b) reflection from, or transmission through, a uniform plate.

solved in cylindrical coordinates. The addition of the plate to the problem will introduce a reflection (or transmission) coefficient. Completely identical results in the two geometries [Figs. 1(a) and (b)] are achieved when the reflection coefficient is -1 or the transmission coefficient is 1 for all incident angles. For simplicity in illustrating our results we analyze the case of Fig. 1(a), remembering that all results derived in this case are fully applicable to the geometry of Fig. 1(b).

A. Incident field Ψ_i

According to the Rayleigh integral formula for the field of a planar transducer, the particle velocity potential of the incident field can be expressed as

$$\Psi_i(\rho,z)=\frac{1}{2\pi}\int\int_{S_1}\mathcal{V}_1(\rho_1)\frac{\exp(i\kappa R)}{R}dS_1, \tag{1}$$

where $\mathcal{V}_1(\rho_1)$ is the axisymmetric particle velocity distribution on the transmitter surface S_1. Here, ρ_1 is the radial polar coordinate measured from the transmitter center, κ is the wavenumber in the fluid, and R is the distance between an arbitrary observation point (ρ,z) in cylindrical coordinates and some point on the transmitter surface S_1. The amplitude of $\mathcal{V}_1(\rho_1)$ can depend on the local radial coordinate ρ_1 in one of several different ways. The subscript 1 refers always to the transmitter; later, subscript 2 will refer to the receiver. The geometry of the calculation is shown in Fig. 2. The distances R and x are given by

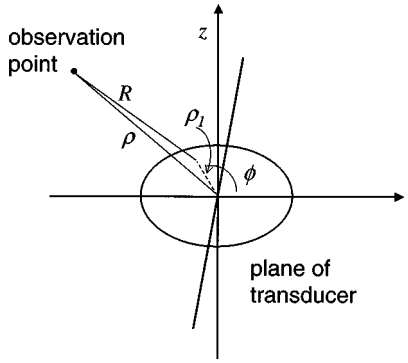


FIG. 2. Geometry of the calculation leading to Eq. (7). R the is distance between an arbitrary observation point and some point on the transducer, ρ and z are cylindrical coordinates of the observation point, and ρ_1 and ϕ_1 are the local polar coordinates on the transmitter surface.

$$R=\sqrt{x^2+z^2}, \quad x^2=\rho^2+\rho_1^2-2\rho\rho_1\cos\phi_1. \tag{2}$$

The spherical wave $\exp(i\kappa R)/R$ can be expanded in terms of cylindrical functions,

$$\frac{\exp(i\kappa R)}{R}=\frac{\exp(i\kappa\sqrt{x^2+z^2})}{\sqrt{x^2+z^2}} =\int_0^\infty \frac{\exp(iz\sqrt{\kappa^2-\tau^2})}{i\sqrt{\kappa^2-\tau^2}}J_0(x\tau)\tau\,d\tau, \tag{3}$$

where $J_0(x\tau)$ is the zeroth-order Bessel function, and τ is a wavenumber integration variable.

The 2D integral expression in Eq. (1) can be simplified for an axially symmetric transducer velocity distribution by introducing the relation in Eq. (3) for the spherical wave $\exp(i\kappa R)/R$. Then, we invert the order of integration and integrate over the transmitter surface, where $dS_1=\rho_1\,d\rho_1\,d\phi_1$, as illustrated in Fig. 2. The result is

$$\Psi_i(\rho,z)=\frac{1}{2\pi}\int_0^\infty \frac{\exp(iz\sqrt{\kappa^2-\tau^2})}{i\sqrt{\kappa^2-\tau^2}}\tau\,d\tau \times\int_0^\infty \mathcal{V}_1(\rho_1)_1\rho_1\,d\rho_1\int_0^{2\pi}J_0(\tau x)d\phi_1, \tag{4}$$

where the coordinate x is given as before. The integral over the azimuthal transmitter angle ϕ_1 is given by $2\pi J_0(\rho\tau)J_0(\rho_1\tau)$.[22] The resulting expression for the field is

$$\Psi_i(\rho,z)=\int_0^\infty \frac{\exp(iz\sqrt{\kappa^2-\tau^2})}{i\sqrt{\kappa^2-\tau^2}}J_0(\rho\tau)\tau\,d\tau \times\int_0^\infty \mathcal{V}_1(\rho_1)J_0(\tau\rho_1)\rho_1\,d\rho_1, \tag{5}$$

where the integral over the local radial variable ρ_1 may be recognized as the Hankel transform of the particle velocity profile function $\mathcal{V}_1(\rho_1)$.[23] This realization permits us to write

$$\Psi_i(\rho,z)=\int_0^\infty \frac{\exp(iz\sqrt{\kappa^2-\tau^2})}{i\sqrt{\kappa^2-\tau^2}}H_\tau[\mathcal{V}_1(\rho_1)]J_0(\rho\tau)\tau\,d\tau, \tag{6}$$

where H_τ denotes the Hankel transform in the wavenumber variable τ. The results in Eqs. (5) and (6) are equivalent to a generalized version of King's solution[24] of the Rayleigh integral for planar transducers. If we now particularize the velocity distribution to one that is rectangular [i.e., $\mathcal{V}_1=\text{const}$ if $\rho_1\leqslant a$, and $\mathcal{V}_1(\rho_1)=0$ for $\rho_1>a$], the Hankel transform becomes $\mathcal{V}_1 aJ_1(a\tau)/\tau$, and Eq. (6) reduces to

$$\Psi_i(\rho,z)=\mathcal{V}_1 a\int_0^\infty \frac{\exp(iz\sqrt{\kappa^2-\tau^2})}{i\sqrt{\kappa^2-\tau^2}}J_1(a\tau)J_0(\rho\tau)d\tau, \tag{7}$$

which is exactly King's solution. In the following section we will employ the general form of the field solutions Eqs. (5) and (6) to calculate properties of field and voltage for arbitrary velocity distributions. Later, we will apply these results to several special cases.

B. Receiver voltage and reciprocity

According to the Auld–Kino[25,26] electromechanical reciprocity relation, we can write the voltage $V(z,r)$ induced on a receiving transducer (device 2) that samples the acoustic field produced by the transmitter (device 1) in the following manner (see the Appendix),

$$V(r,z)=\int\int_{S_2}\mathcal{V}_2(\rho_2)\Psi_i(z,\rho)dS_2, \tag{8}$$

where $\mathcal{V}_2(\rho_2)$ is the surface velocity distribution on the receiving transducer (when it operates as a transmitter), Ψ_i is the incident field produced by device 1, and dS_2 is the differential surface area on the receiver, given by $dS_2=\rho_2\,d\rho_2\,d\phi_2$.

The receiver is a large planar surface, and it will indeed perturb the incident field. For large values of ka, however, such as those in a typical ultrasonics experiment, the incident and total fields are simply proportional to each other, except for a narrow annulus on the transducer edge, which is about $1/\sqrt{ka}$ of the radius of the transducer. Because of the rapid variation of the phase and amplitude of the field near the transducer edge, an integration over transducer surface will suppress the edge effect as $1/(ka)$. Typical values of ka for piston transducers in routine ultrasonics measurements are in the hundreds. We reasonably assume that we can neglect this effect. For off-axis positioning of the transducers the edge effect can be stronger because of the higher incident field near transducer edge, but the large values of ka allow the effect to be neglected in this case also.

The geometry for the receiver is entirely equivalent to that for the transmitter, as was shown in Fig. 2. The radial distance from the transmitter axis to a point on the receiver is given analogously by

$$\rho=\sqrt{r^2+\rho_2^2-2r\rho_2\cos\phi_2}. \tag{9}$$

Now, replace the incident field Ψ_i in Eq. (8) by its representation from Eq. (6) to find

$$V(r,z)=\int_0^\infty \frac{\exp(iz\sqrt{\kappa^2-\tau^2})}{i\sqrt{\kappa^2-\tau^2}}H_\tau[\mathcal{V}_1(\rho_1)]J_0(r\tau)\tau\,d\tau \times\int_0^\infty \mathcal{V}_2(\rho_2)\rho_2\,d\rho_2\int_0^{2\pi}J_0(\tau\rho(\phi_2))d\phi_2. \tag{10}$$

Following the procedure outlined earlier for integration over the transmitter, the integral over the local receiver coordinate ϕ_2 is $2\pi J_0(r\tau)J_0(\rho_2\tau)$, and the subsequent integral over the remaining local receiver coordinate ρ_2 is the Hankel transform of the receiver velocity distribution $\mathcal{V}_2(\rho_2)$. With these simplifications we finally arrive at the result

$$V(r,z)=2\pi\int_0^\infty \frac{\exp(iz\sqrt{\kappa^2-\tau^2})}{i\sqrt{\kappa^2-\tau^2}} \times H_\tau[\mathcal{V}_1(\rho_1)]H_\tau[\mathcal{V}_2(\rho_2)]J_0(r\tau)\tau\,d\tau, \tag{11}$$

where (z,r) are the cylindrical coordinates of the receiver center with respect to the transmitter center. Similar expressions have been derived by other workers as well.

Of course, the receiver voltage is the result of a coherent sum over its face of the incident field Ψ_i and is therefore expected to be quite different from the acoustic field detected at the center of the receiving transducer, as can be verified by a comparison of Eqs. (6) and (11). These two quantities will approach each other only as the radius of the receiver shrinks to zero, i.e., as the receiver approximates a point detector. In that case, the velocity distribution approaches a delta function $\mathcal{V}_2(\rho_2) \rightarrow \delta(\rho_2)$, and $H_\tau[\mathcal{V}_2(\rho_2)]$ is constant.

C. The "equivalent transducer"

We have seen that the field and voltage of finite radiators are, in general, quite different. There does exist, however, a subtle relationship between the voltage on the receiver and the field of the transmitter. To examine this behavior in more detail, we introduce the concept of an ''equivalent transducer.'' What is meant here is the equivalence between the actual situation—finite transmitter and finite receiver of some dimensions and velocity distribution—and a conceptually equivalent transmitter, of some radius and having some velocity distribution, detected by a point receiver. Does such an equivalent transducer (ET) exist, and, if so, how can its properties be calculated? To find the ET, we imagine that the *voltage V* of Eq. (11) could be the *field* Ψ of an equivalent transducer located at the transmitter and measured (by a point detector) at the center of the actual receiver.

We can find the particle velocity distribution on the surface of an equivalent transmitter by differentiating the voltage with respect to the coordinate z and evaluating the resulting expression on the surface of the transmitter,

$$\mathcal{V}_{\mathrm{ET}}(r) = \frac{d}{dz} V(r,z)\big|_{z=0} = 2\pi \int_0^\infty H_\tau[\mathcal{V}_1(\rho_1)] H_\tau[\mathcal{V}_2(\rho_2)] J_0(r\tau)\, \tau\, d\tau. \tag{12}$$

The integral on the right in the expression above can be recognized as the inverse Hankel transform of the product of the Hankel transforms of the actual transmitter and receiver velocity distributions $\mathcal{V}_1(\rho_1)$ and $\mathcal{V}_2(\rho_2)$. [The kernels of the Hankel transform and its inverse are both $J_0(\tau r)$, where the integration is carried either over $r\,dr$ or $\tau\,d\tau$.] This fact about Hankel transforms permits us to rewrite the integral of Eq. (12) in formal notation as

$$\mathcal{V}_{\mathrm{ET}}(r) = 2\pi H_r^{-1}\{H_\tau[\mathcal{V}_1(\rho_1)] H_\tau[\mathcal{V}_2(\rho_2)]\}(r). \tag{13}$$

The expression above represents precisely the general relationship between $\mathcal{V}_1(\rho_1)$, $\mathcal{V}_2(\rho_2)$ and $\mathcal{V}_{\mathrm{ET}}$ that we sought. With it, it is possible in principle, assuming the Hankel transforms exist, to compute an equivalent velocity distribution for any two actual distributions. Even in the case when no analytical form of the Hankel transform can be found, the expression on the right in Eq. (13) can always be computed numerically.

For any particular velocity distributions $\mathcal{V}_1(\rho_1)$ and $\mathcal{V}_2(\rho_2)$ it can be convenient to rewrite Eq. (12) in a different form. We first calculate the τ-variable integral over the product of Bessel functions $J_0(\rho_1\tau)J_0(\rho_2\tau)J_0(r\tau)$. As a result of this manipulation, the equivalent transducer velocity distribution can be expressed as a 2D integral

$$\mathcal{V}_{\mathrm{ET}}(r) = 4\int_0^\infty \int \frac{\mathcal{V}_1(\rho_1)\mathcal{V}_2(\rho_2)\rho_1\rho_2\, d\rho_1\, d\rho_2}{\sqrt{(r^2-(\rho_1-\rho_2)^2)((\rho_1+\rho_2)^2-r^2)}}, \tag{14}$$

where the additional condition $|\rho_1-\rho_2|<r<\rho_1+\rho_2$ restricts the region of integration to the plane (ρ_1,ρ_2).

With this result, the understanding of the differences between the voltage and field can be approached far more easily. For each pair of transducers having some specific velocity distributions $\mathcal{V}_1$ and $\mathcal{V}_2$, we can find an equivalent distribution $\mathcal{V}_{\mathrm{ET}}$ such that the *field* of the ET is exactly the same as the *voltage* of the actual transducer pair. For example, we can use the Rayleigh formula to calculate the voltage,

$$V(r,z) = \frac{1}{2\pi}\int_{S_{\mathrm{ET}}}\int \mathcal{V}_{\mathrm{ET}}(S_{\mathrm{ET}})\frac{\exp(i\kappa R)}{R}\, dS_{\mathrm{ET}}, \tag{15}$$

where we integrate only over the surface of the ET and obtain the voltage in a single step. Recall that in Eq. (1) we find only the field with this one integration. In that prior formulation, additional work was needed to find the voltage. With the ET, however, jumping to the voltage can be done in a single integration step.

As an example, consider that the farfield voltage [or field of an ET for $\kappa(a+b)^2<z$] can be easily calculated from Eq. (15). Expanding the phase function $\exp(i\kappa R)$ in the farfield, and calculating the resulting integral over the angular variable ϕ_1, the voltage can be represented as a one-dimensional Hankel transform (see, for example, Ref. 27, pp. 225–226)

$$V(r,z) = 2\pi\frac{\exp(i\kappa R_c)}{R_c}\int_0^\infty \mathcal{V}_{\mathrm{ET}}(\rho_1) J_0\left(\frac{\kappa r \rho_1}{R_c}\right)\rho_1\, d\rho_1 = 2\pi\frac{\exp(i\kappa R_c)}{R_c} H_\tau[\mathcal{V}_{\mathrm{ET}}(\rho_1)]\big|_{\tau=\kappa r/R_c}, \tag{16}$$

where $R_c=\sqrt{z^2+r^2}$ is the distance between transducer centers. After substitution of the expression in Eq. (13) for the ET velocity distribution and taking into account the Hankel transform properties, the farfield voltage becomes

$$V(r,z) = 2\pi\frac{\exp(i\kappa R_c)}{R_c} \times H_\tau[\mathcal{V}_1(\rho_1)] H_\tau[\mathcal{V}_2(\rho_2)]\big|_{\tau=\kappa r/R_c}. \tag{17}$$

This expression is valid for arbitrary velocity distributions $\mathcal{V}_1$ and $\mathcal{V}_2$. Therefore, the farfield voltage for a pair of the transducers is simply the product of each individual transducer field distribution.

A careful examination of Eqs. (10) and (11) reveals that this phenomenon of the ET is actually only a direct consequence of reciprocity. The roles of the two transducers in Eq. (11) are completely equivalent and interchangeable. Therefore, any two transducers whose velocity distributions yield a Hankel-transform product equivalent to the product in Eq.

(11) for the physical devices actually employed in the measurement will give the same voltage. In particular, when one of the pair of virtual transducers is a point receiver, the other device will be a single, finite ''equivalent transducer'' whose velocity distribution, calculated in Eq. (13), yields a voltage at the receiver that is identical to the voltage of the physical pair. But because the virtual receiver is a point detector, the voltage $V_{\mathrm{ET}}(r,z)$ of the ET is also its field $\Psi_{\mathrm{ET}}(r,z)$. In principle, there is any number of transducer pairs whose Hankel transform products yield identical contributions to the integral in Eq. (11) and whose velocity profiles, therefore, could be considered equivalent to the ones actually used in a given measurement.

We are now prepared to study examples of transducers having particular axially symmetric velocity distributions of interest.

III. APPLICATION TO TRANSDUCER VELOCITY PROFILES

In this section we apply the results of the preceding theoretical development to several special cases of transducer velocity profiles (all having axial symmetry) for which analytical solutions are possible.

A. Gaussian beams

This case is perhaps the most important of all from the standpoint of its impact on theoretical modeling because of the widespread use of Gaussian beams in the modeling of finite-aperture acoustic phenomena over the past 30 years. Beginning with the work of Bertoni and Tamir[7] and extending to many other calculations,[9–11] including recent ones,[28–33] Gaussian beams have been favored in acoustics because of their analytical convenience and ease of numerical computation, but not necessarily because they were considered a good approximation for piston fields. In this section we will see that when the voltage function, instead of the field, is considered, highly apodized Gaussian beams are entirely acceptable replacements for piston beams.

We represent the velocity distribution of a Gaussian-beam transducer using the conventional formulation

$$\mathcal{V}_1(\rho_1)=\mathcal{V}_1\exp\left[-\rho_1^2/a^2\right] \quad \text{and} \quad \mathcal{V}_2(\rho_2)=\mathcal{V}_2\exp\left[-\rho_2^2/b^2\right] \tag{18}$$

for the transmitter and receiver, respectively. To be a good approximation to a Gaussian beam, the physical transducer diameter must be large enough that the velocity amplitude at the edge is small compared with its value at the center, or the transducer must be otherwise apodized.[34] Then, the integration limit representing the Hankel transform can be extended to infinity, yielding

$$H_\tau[\mathcal{V}_1(\rho_1)]=\mathcal{V}_1\int_0^\infty \exp\left[-\rho_1^2/a^2\right]J_0(\tau\rho_1)\rho_1\,d\rho_1$$

$$=\frac{\mathcal{V}_1 a^2}{2}\exp\left[-\tau^2 a^2/4\right]. \tag{19}$$

Inserting this intermediate result into Eq. (13), we obtain

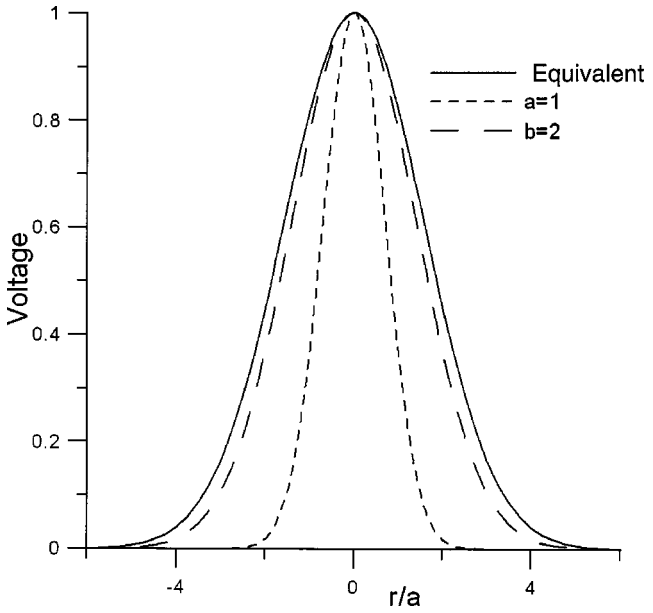


FIG. 3. Velocity profiles for two Gaussian-beam transducers and their ''equivalent transducer.''

$$\mathcal{V}_{\mathrm{ET}}=\frac{\pi}{2}a^2b^2\mathcal{V}_1\mathcal{V}_2\int_0^\infty \exp\left[-(a^2+b^2)\tau^2/4\right]J_0(r\tau)\tau\,d\tau$$

$$=\frac{\pi a^2b^2}{a^2+b^2}\mathcal{V}_1\mathcal{V}_2\exp\left[-r^2/(a^2+b^2)\right]. \tag{20}$$

We have demonstrated here that the acoustic system consisting of a Gaussian-beam transmitter of beam waist a and a receiver of beam waist b behaves the same (has the same lateral and range dependence) as the combination of an equivalent Gaussian-beam transducer of waist $\sqrt{a^2+b^2}$ and a point receiver, as illustrated in Fig. 3. Here, the profile of each Gaussian-beam transducer is shown along with the profile of the equivalent transducer, also a Gaussian in this case. Alternatively, one may say that the *field* of a single Gaussian-beam transducer of waist $\sqrt{a^2+b^2}$ is equivalent to the *voltage* of a pair of Gaussian-beam transducers of beam waists a and b, respectively. So, in this particular case, both the original two radiators and the ET have Gaussian-beam profiles, as we would expect.

B. Rectangular velocity distributions

From a practical viewpoint the rectangular velocity distribution is by far the one most often met with in experiments. Its behavior, therefore, is particularly valuable to examine in detail. Such a comparison between the field values of a piston radiator (rectangular velocity profile) and the voltage, as measured by a second identical probe, is shown in Fig. 4 for $ka=30$. In frame (a) is shown the behavior quite close to the face of the transmitter, at a range of $kz=6$, or alternatively when $z/\mathcal{R}=0.04$, where $\mathcal{R}$ is the Rayleigh distance a^2/λ. Here, the field is characterized by rapid oscillations owing to successive interference of edge and face waves, whereas the voltage function at the same range is quite smooth. Because the range is so close to zero, the voltage is a monotonically decreasing function. In frame (b) at

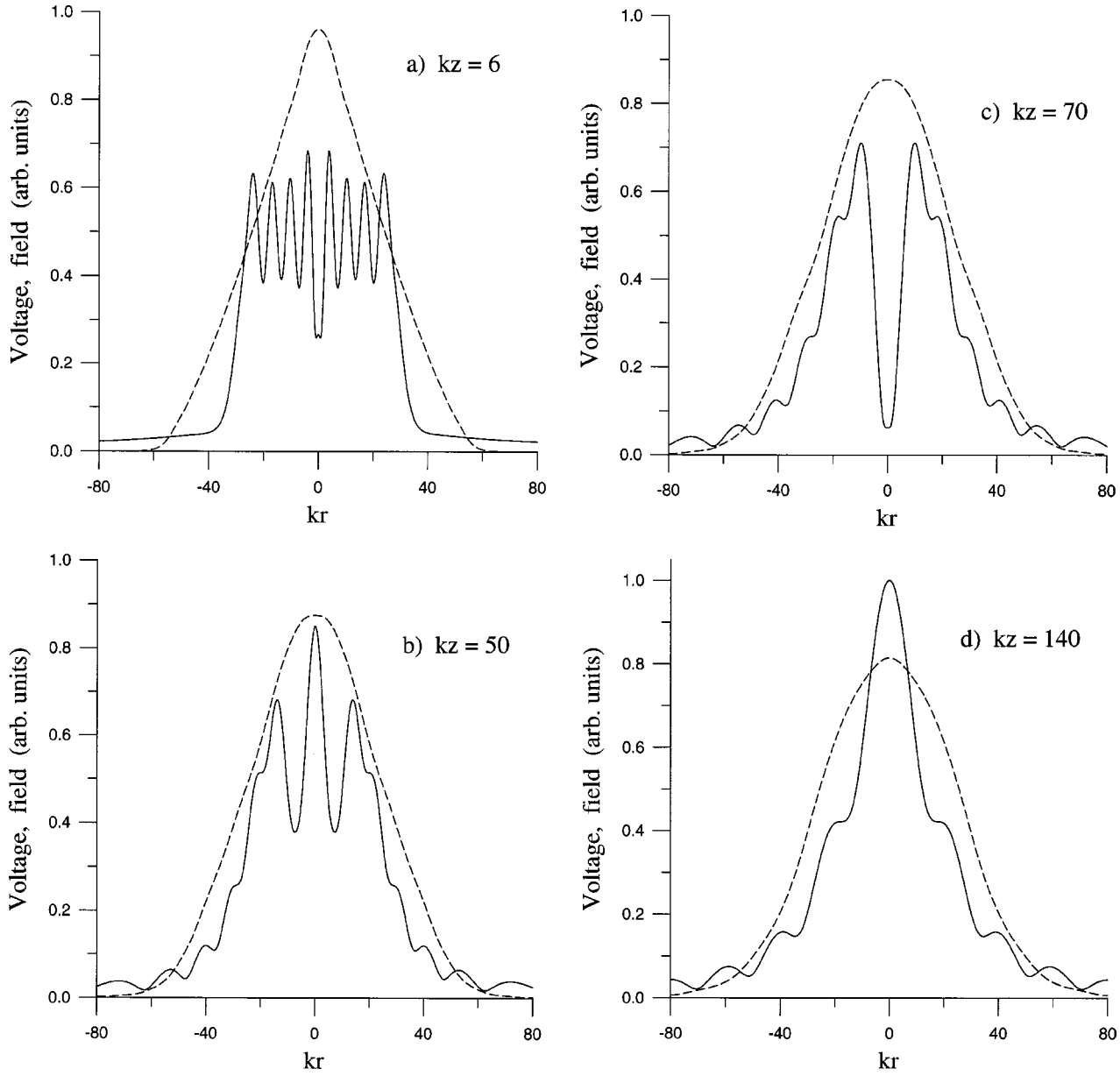


FIG. 4. Comparison of field (solid line) and voltage (dashed line) functions for piston radiator transducers with $ka=30$ at various ranges. (a) $z/\mathcal{R}=0.04$; (b) $z/\mathcal{R}=0.35$; (c) $z/\mathcal{R}=0.52$; (d) $z/\mathcal{R}=1.03$.

$kz=50$ or $z/\mathcal{R}=0.35$, the field is still rapidly varying and the voltage already resembles a Gaussian function. In frame (c), at nearly one-half the Rayleigh distance ($kz=70$ or $z/\mathcal{R}=0.52$), there is the last on-axis pressure null before the farfield regime. The voltage at this point is nearly Gaussian in its functional form. Finally, in frame (d) at one Rayleigh distance ($kz=140$ or $z/\mathcal{R}=1.03$), the field and voltage are still quite different. Note that all points calculated in Fig. 4 are in the nearfield, with the exception of frame (d), which lies on the near- and farfield boundary.

Now we examine this case analytically. The velocity profile is given by a step function,

$$\mathcal{V}_1(\rho_1)=\begin{cases}\mathcal{V}_1, & \rho_1<a,\\ 0, & \rho_1>a,\end{cases}\qquad \mathcal{V}_2(\rho_2)=\begin{cases}\mathcal{V}_2, & \rho_2<b,\\ 0, & \rho_2>b.\end{cases} \tag{21}$$

The Hankel transforms are straightforward because the profile is a constant from $r=0$ to radius a or b,

$$H_r[\mathcal{V}_1(\rho_1)]=\mathcal{V}_1\int_0^a J_0(r\tau)\tau\,d\tau=\mathcal{V}_1 aJ_1(a\tau)/\tau, \tag{22}$$

and a similar expression for the transducer of radius b.

Then, for the equivalent velocity profile of the two rectangular transducers we have

$$\mathcal{V}_{\mathrm{ET}}=2\pi ab\mathcal{V}_1\mathcal{V}_2\int_0^\infty J_1(a\tau)J_1(b\tau)J_0(r\tau)\frac{d\tau}{\tau}. \tag{23}$$

This expression can be evaluated analytically (apart from constant multipliers) with the following result,

$$\mathcal{V}_{\mathrm{ET}}\propto\begin{cases}\pi b^2, & r<|a-b|,\\ a^2\arccos p+b^2\arccos q-\Omega, & |a-b|<r<a+b,\\ 0, & r>a+b,\end{cases} \tag{24}$$

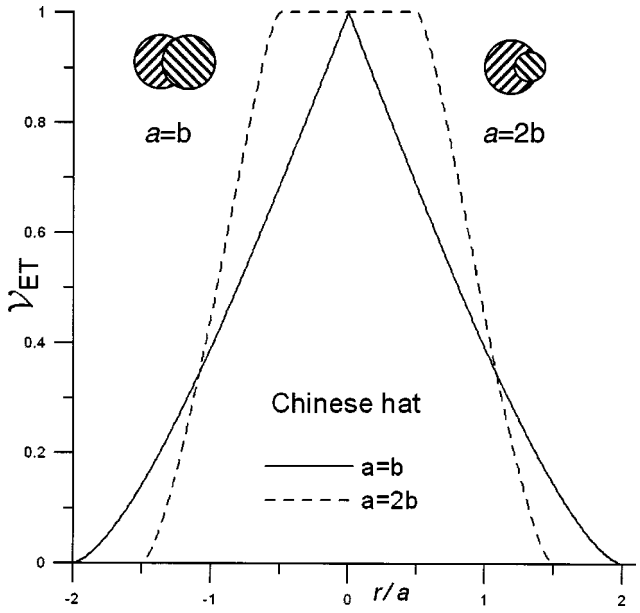


FIG. 5. Effective velocity profiles for an equivalent transducer to a pair of rectangular-beam transducers. (a) For identical transducer radii $a=b$, the equivalent profile is nearly the so-called ''Chinese hat,'' or triangular distribution; (b) for $a>b$ the profile is first rectangular, then gradually decreasing. The insets show schematically the geometric overlap functions for the cases when $a=b$ and when $b<a$.

where $p=(r^2+a^2-b^2)/(2ar)$ and $q=(r^2-a^2+b^2)/(2br)$. The term Ω is given by

$$\Omega=\sqrt{(r^2-(a-b)^2)((a+b)^2-r^2)}/2. \tag{25}$$

Although the expression in Eq. (24) appears complicated, its plotted value is simple and easily interpreted. The result is shown in Fig. 5 for the two possible cases. Frame (a) shows the velocity profile for an ET when both piston radiators have the same radius $b=a$. In frame (b), for $b<a$, the dependence changes only in that the velocity is constant to a radius equal to the absolute difference between the piston radii $|a-b|$. Physically, this profile corresponds to the simple overlap function of two disks, illustrated schematically in the insets for $a=b$ and $a\neq b$. This problem is also found in optics.[35]

The results developed leading to Eq. (24) are formally identical to the calculation given by Daly and Rao for the transient behavior of piston radiators in the *time domain*.[36] In their case, the purpose is to find a spatially averaged velocity-potential impulse response, but the manipulations they show lead to formally identical final results. In our case, the radially decreasing velocity profile has the shape of a nearly linear function. The unusual and unexpected result here is that the equivalent source we calculate to yield the voltage described above should be a planar transducer of size $a+b$ with a spatial distribution of its particle velocity according to Eq. (24) and as shown in Fig. 5. The mathematical form of this radial velocity function is nearly triangular, and the value drops from a maximum at the transducer center to zero at the rim $(r=a+b)$. This is the so-called ''Chinese hat'' function, a term coined by Bracewell[37] in conjunction with a related problem in optics, and cited by Daly and Rao.[38]

C. Arbitrary velocity distribution

We now consider an even more drastic variation in the transducer velocity profile, namely the one of an arbitrary velocity distribution. Let one transducer have a rectangular distribution and the other be arbitrary. Which is transmitter and which is receiver is irrelevant owing to reciprocity. To be definitive, let us assume that for the rectangular probe, $\mathcal{V}_2(\rho_2<a)=1$ and $\mathcal{V}_2(\rho_2>a)=0$. The second transducer is assumed to have an arbitrary velocity distribution $\mathcal{V}_1(\rho_1)$. The integral over variable ρ_2 in Eq. (14) can be calculated, and the resulting velocity distribution for the ET is

$$\mathcal{V}_{ET}(r)=2\pi\int_0^a \mathcal{V}_1(\rho_1)\rho_1\,d\rho_1-2\int_{a-r}^{a}\mathcal{V}_1(\rho_1)\times\arccos\frac{a^2-r^2-\rho_1^2}{2r\rho_1}\rho_1\,d\rho_1,\quad r<a,$$

$$=2\int_{r-a}^{a}\mathcal{V}_1(\rho_1)\arccos\frac{\rho_1^2+r^2-a^2}{2r\rho_1}\rho_1\,d\rho_1,\quad r>a. \tag{26}$$

To simplify this expression, we introduce the average velocity $S(\rho)$ on the disk of size ρ, which is determined as

$$S(\rho)=\frac{2}{\rho^2}\int_0^{\rho}\mathcal{V}_1(\rho_1)\rho_1\,d\rho_1. \tag{27}$$

The average velocity S can be simply calculated for any velocity distribution $\mathcal{V}_1$. For example, for the normalized rectangular distribution of size a, the average velocity within the transducer radius is just $S(\rho<a)=1$, and outside this radius $S(\rho>a)=a^2/\rho^2$.

Equation (26) can be integrated by parts, isolating terms identical to those in Eq. (27). Then, using Eq. (27) in Eq. (26) and making a variable substitution of $\rho_1^2=a^2+r^2-2ar\cos\phi$ leads to an expression for the equivalent transducer voltage in terms of average velocity S as

$$\mathcal{V}_{ET}(r)=a\int_0^{\pi}S(\sqrt{a^2+r^2-2ar\cos\phi})(a-r\cos\phi)d\phi. \tag{28}$$

Equation (28) shows that the velocity distribution on the ET is determined mostly by the average velocity $S(\rho)$ on the surface of the real transducer.

An even more interesting case exists when the velocity amplitude of the second transducer increases (instead of decreases) from the center of the transducer to the edge. For example, we might postulate a transducer of the same radius a, but with a parabolic velocity profile $\mathcal{V}_1(\rho_1<a)=(\rho_1/a)^2$ and $\mathcal{V}_1(\rho_1>a)=0$. The velocity of this hypothetical transducer is zero at the center and unity at the circumference, where the velocity then drops instantaneously to zero. The average velocity S on its surface is $S(\rho<a)=(\rho/a)^2/2$ and $S(\rho>a)=(a/\rho)^2/2$. The integral of Eq. (28) can be evaluated analytically, and the resulting ET velocity is

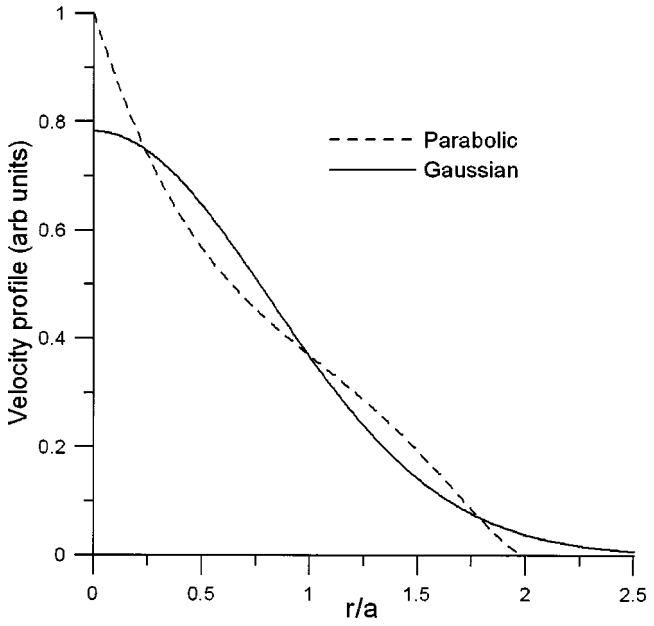


FIG. 6. Velocity profile for an equivalent transducer to a set of rectangular and parabolic transducers.

$$\mathcal{V}_{\mathrm{ET}}(r)=(a^2+r^2)\arccos\frac{r}{2a}-\frac{r(6a^2+r^2)}{8a^2}\sqrt{4a^2-r^2},$$

$$r<2a,$$

$$=0, \quad r>2a. \tag{29}$$

The result is shown in Fig. 6. One can see that the ET in this case is a probe of radius $2a$ having a well-apodized velocity distribution.

IV. SIMILARITY OF GAUSSIAN AND PISTON VOLTAGES

A. Multiple point receivers

In the previous sections we have shown that the received voltage for a transmitter–receiver transducer pair with arbitrary velocity distributions is mathematically equivalent to the acoustic field of some ''equivalent transducer'' whose velocity distribution is determined by Eq. (13).

Before we compare calculated fields and voltages from expressions derived above, let us examine the averaging effect at the receiver of several discrete point receivers deployed along a line coincident with the diameter of a piston receiver. Either three or five point receivers are arrayed from the center of the local receiver coordinate system to the edge

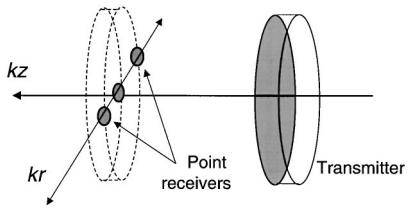


FIG. 7. Geometry of a set of point receivers to illustrate averaging effect of finite receiver. Virtual piston receiver is shown in dashed lines for comparison.

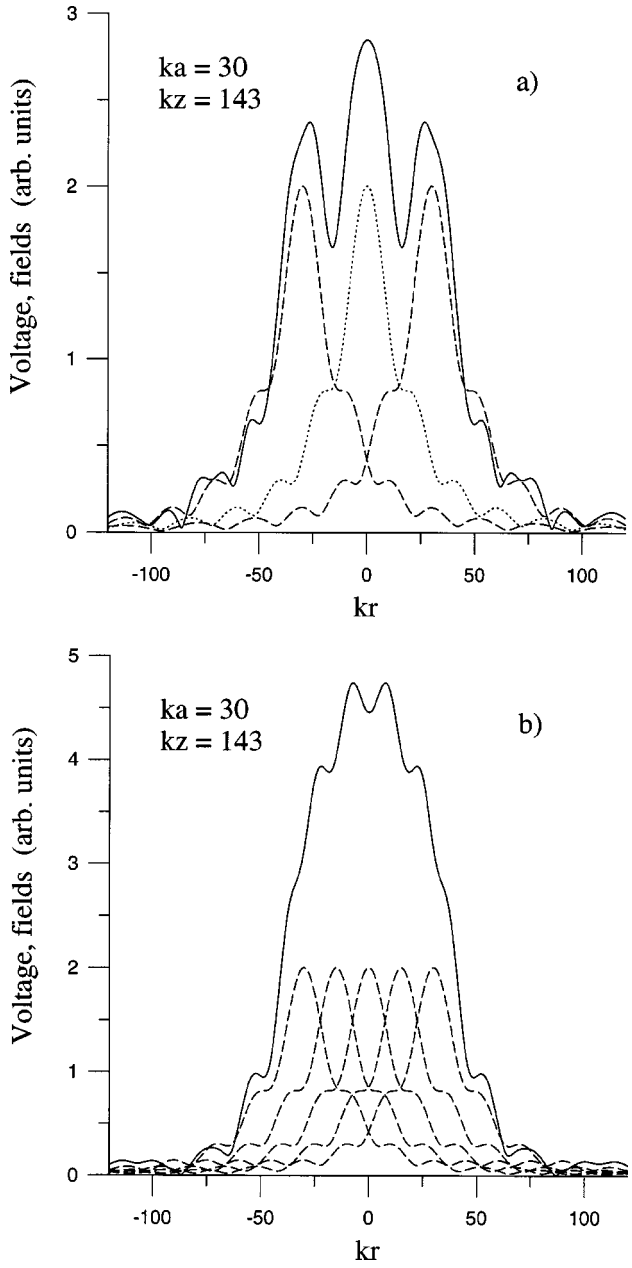


FIG. 8. (a) Fields of three point receivers measuring piston radiator (dashed and dotted lines) and their coherent sum (solid line). (b) Same function for five point receivers at a range of one Rayleigh distance.

of where an identical piston receiver would be located. The geometry is illustrated schematically in Fig. 7, where a virtual piston receiver is shown in dashed lines for comparison. Three receivers are illustrated, but we will also consider the result with five point receivers similarly deployed. The analytical framework for this simple demonstration has been laid in calculations performed in Sec. II. Figure 8(a) shows the voltage profile when a piston radiator at a range of one Rayleigh distance ($kz=ka^2/\lambda$) is detected by three coherent point receivers. The field of the central point receiver is plotted as a dotted curve, and the signals of the left and right point receivers are plotted as dashed curves. The coherent sum is the solid curve. Clearly, the sum of the three receivers is smoother than the any of the identical field curves. A further demonstration is seen in Fig. 8(b), where a total of five

symmetrically placed point receivers monitors the field of a piston radiator. Signals from the point receivers are recorded as dashed curves, each one displaced by its respective distance from the central axis. The coherent sum is shown in the solid curve and is convincingly smoother than the field function. Extending this summation analytically to the entire face of a finite receiver yields the results demonstrated in the following section.

B. Piston transducer pairs

To demonstrate the close similarity of Gaussian and piston transducer voltages, it will be sufficient to show that the acoustic fields of the transducers with Gaussian velocity distributions and those with the ''Chinese hat'' velocity distributions, of Eqs. (20) and (24), are similar. Rigorously speaking, these two voltage functions are not entirely identical for any values r and z because the Gaussian and ''Chinese hat'' functions are simply different. But we want to determine the intervals of r and z where these two acoustic fields are very close to each other. The important theoretical consequence of this similarity is that we can substitute the voltage of a pair of rectangular transducers by Gaussian ones in this range. Such a substitution vastly simplifies all voltage calculations with almost no loss of accuracy.

We compare here the acoustic fields of two equivalent transducers, one of them with the ''Chinese hat'' velocity distribution $\mathcal{V}_{\mathrm{ch}}$ of Eq. (20), and the other with a Gaussian velocity distribution $\mathcal{V}_{\mathrm{G}}(\rho_1)=D\exp(-\rho_1^2/c^2)$. Showing that these two functions are closely similar is equivalent to showing that the behavior of a pair of pistons is similar to that of a pair of Gaussian transducers. The parametric Gaussian beam constants $D=D(a,b)$ and $c=c(a,b)$ can be found by minimizing the spectral integrals on the transducers' surfaces at the plane $z=0$. This operation is equivalent to minimization of the function

$$\chi^2(D,c)=\int_0^\infty (\mathcal{V}_{\mathrm{ch}}(a,b,\rho_1)-\mathcal{V}_{\mathrm{G}}(D,c,\rho_1))^2\rho_1\,d\rho_1. \tag{30}$$

We consider first the axial fields of these two transducers. For $r=0$ the integrand of Eq. (15) does not depend on the angular variable and the ϕ-integral is simply equal to 2π. Integrating by parts, we can present the axial field of the ET as

$$V(0,z)=\mathcal{V}_{\mathrm{ET}}(\rho_1)\exp\,(i\kappa\sqrt{z^2+\rho_1^2})\big|_0^\infty - \int_0^\infty \frac{d\mathcal{V}_{\mathrm{ET}}(\rho_1)}{d\rho_1}\exp\,(i\kappa\sqrt{z^2+\rho_1^2})\,d\rho_1. \tag{31}$$

For the ''Chinese hat'' ET (or for the pair of two rectangular transducers) this equation transforms to

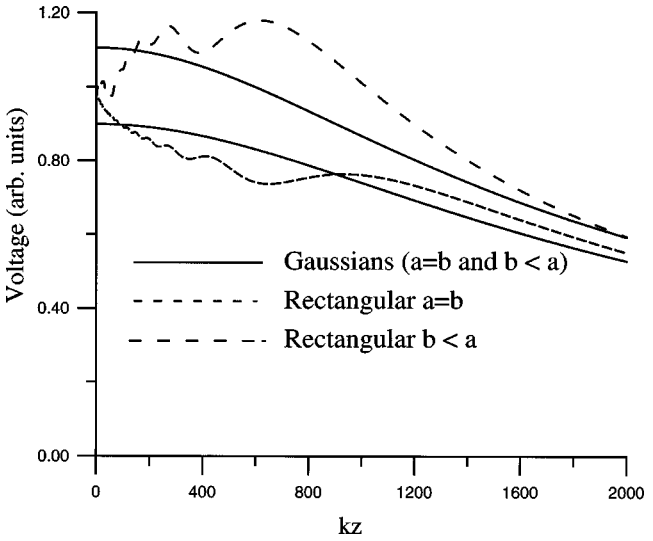


FIG. 9. Voltage comparison for equivalent ''Chinese hat'' function and Gaussian voltage function versus axial transducer position for $\kappa a=\kappa b=50$ and $\kappa a=50>\kappa b=30$.

$$V_r(0,z)=C\left\{\exp(i\kappa z)-\frac{1}{\pi b^2}\times\int_{a-b}^{a+b}\sqrt{(\rho_1^2-(a-b)^2)((a+b)^2-\rho_1^2)}\times\exp(i\kappa\sqrt{z^2+\rho_1^2})\,\frac{d\rho_1}{\rho_1}\right\}, \tag{32}$$

where C is an arbitrary constant, and the subscript V_r denotes the voltage of a pair of rectangular transducers.

The above equation is the so-called ''diffraction-correction integral,''[16] which is widely used to obtain accurate attenuation estimates in experiments with two rectangular transducers. For the Gaussian ET, Eq. (31) can be written as

$$V_{\mathrm{G}}(0,z)=\mathrm{const}\,D\left\{\exp(i\kappa z)-\frac{2}{c^2}\int_0^\infty \exp(-\rho_1^2/c^2+i\kappa\sqrt{z^2+\rho_1^2})\rho_1\,d\rho_1\right\}, \tag{33}$$

where the subscript V_{G} denotes the voltage of a Gaussian transducer. Plots of $V_r(0,z)$ and $V_{\mathrm{G}}(0,z)$ are presented in Fig. 9 for $\kappa a=\kappa b=50$ and for differing values of $\kappa a=50$ and $\kappa b=30$. Clearly, the axial fields of these two transducers are different, especially for the case $a\neq b$. Only the decrease in farfield amplitude of the pair of two identical transducers $(a=b)$ resembles the one for the Gaussian transducers. The significance here is that the Gaussian beam as a transducer model can not be used for accurate estimation of diffraction corrections for piston beams.

But for many purposes the absolute value of the amplitude on the axis of the transducer is irrelevant. For example, in the investigation of the bounded beams interacting with planar objects (see Fig. 1), only the lateral distribution of the incident beam is important. The structure of the reflected

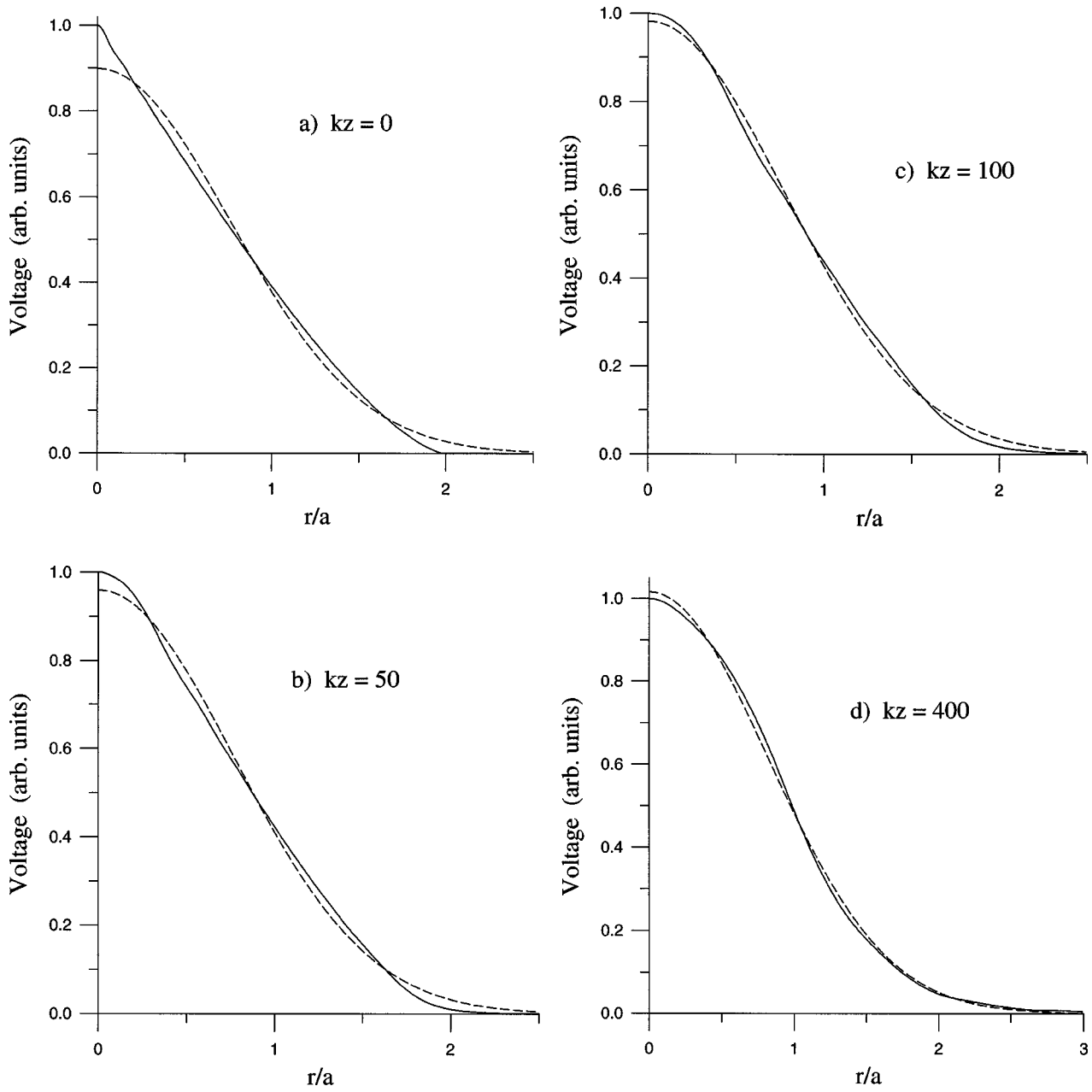


FIG. 10. Voltage comparison for equivalent ''Chinese hat'' function (solid line) and Gaussian voltage function (dashed line) versus lateral transducer position for $\kappa a=\kappa b=50$. (a) $kz=0$; (b) $kz=50$; (c) $kz=100$; (d) $kz=400$ ($\kappa a=50$ and $\kappa b=30$ Gaussian voltage versus).

field (the relation between specular and nonspecular components) is determined by the lateral distribution of the beam near the plate surface.

We compare now the lateral field distribution $V_r(r)$ and $V_G(r)$ for ''Chinese hat'' and Gaussian transducers for different values of z. The explicit forms of both lateral fields are known only for $z=0$ and in the far field. For $z=0$ (the lateral distributions are simply ''Chinese hat'' and Gaussian) the ET fields are not identical (see Fig. 5). The major similarity of these two functions is that they decrease monotonically from the center of the transducer to the edge, where the lateral distribution in the farfield can be calculated from Eq. (17). Using the explicit form of the Hankel transform for the rectangular and for the Gaussian transducers, as in Eqs. (21) and (18), we can present the farfield behavior as

$$V_r(r)\propto\left(\frac{J_1(\kappa ar/R_c)}{\kappa ar/R_c}\right)\left(\frac{J_1(\kappa br/R_c)}{\kappa br/R_c}\right),$$
$$V_G(\rho)\propto\exp\left(-\frac{\kappa^2c^2r^2}{4R_c^2}\right). \tag{34}$$

For the case where $a=b$, the almost complete equivalence of these two distributions was considered earlier by the current authors.[13] Here, we allow a and b to differ, and the larger the ratio of b to a, the more these two farfield lateral functions diverge.

We find that the lateral voltage distribution of two rectangular transducers transforms from a ''Chinese hat'' form to almost Gaussian as z grows from 0 to its farfield value. We can follow this transformation quantitatively for any value of

z in the near- and farfield using Eq. (11). The resulting acoustic fields of the equivalent transducers, however, are so closely approximate, especially beyond several wavelengths, that piston transducer pairs can be substituted by Gaussian transducer pairs for most practical cases with almost no loss of accuracy.

An example of this substitution and the near equivalence of the two voltage profile functions is shown in Fig. 10. The dashed curve is the voltage of two Gaussian transducers, and the solid curve is the voltage profile of two pistons, both pairs at various ranges. In Fig. 10(a) we have $kz=0$, but at zero range the two curves are clearly not identical. In frame (b) where $kz=50$ (or about $\frac{1}{3}$ Rayleigh distance), the two curves are much closer in functional form. Frame (c) at $kz=100$ shows the near equivalence of these two voltage functions, although the range is still only $\frac{2}{3}$ of the Rayleigh distance. In frame (d) at $kz=400$, well into the farfield, the similarity is striking, demonstrating why substitution of a Gaussian voltage for a rectangular one in analytical models is well justified. In all cases of Fig. 10 the Gaussian transducer fields have been chosen with beam waist values that minimize the deviation between the two functions for $z=0$. A robust scheme to accomplish this parametrization for arbitrary z has been presented by several authors.[13,28]

V. CONCLUSIONS

We have shown that plane-wave decomposition of collimated beams and electromechanical reciprocity relations are sufficient to reveal unusual similarities about transducer fields and transducer voltages under various conditions. We have shown that the voltage induced by a transmitting piston transducer radiating into an ideal fluid medium on a second identical piston transducer operating as a receiver is nearly identical to the voltage observed when the two transducers have instead a Gaussian radial surface velocity distribution. The strong similarity in induced voltage is evident at separation distances of only a dozen or so acoustic wavelengths. This similarity increases with separation, becoming nearly identical beyond about one Rayleigh distance. Other transducer velocity distributions yield nearly identical results. The voltages of transducers with arbitrary velocity profiles have also been calculated analytically. When the diameters of the transducers differ, the similarity in the voltage functions between pistons and Gaussians is not as dramatic.

We have also shown that an ''equivalent transducer'' can be derived for any combination of radiators, whose field is identical to the voltage measured using the pair. Using this concept, the calculation of farfield voltage in two-probe experiments or on-axis pressure variations at any range is as simple as the application of well-known approximations to the Rayleigh integral. Taken together, these results have substantial consequences for calculations, both analytical and numerical, that predict the voltage measured in two-transducer experiments. In particular, the substitution of Gaussian beams for piston fields greatly simplifies the calculation of the voltage and is a highly accurate approximation at any practical range.

APPENDIX: VOLTAGE ON RECEIVER

We show in this Appendix an equivalence between Eq. (8) and the Auld–Kino formula for a complicated transmitter-plate-receiver geometry, as shown in Fig. 1(b). This result is also valid for the geometry of Fig. 1(a) because of the symmetry of these two cases. According to the Auld–Kino formula the normalized signal on the receiver (transducer 2) is equal to

$$V=\frac{1}{4P}\int_{S_p}(\mathbf{u}_2^i\psi_1-\mathbf{u}_1\psi_2^i)\mathbf{n}\,dS_p, \tag{A1}$$

where $\mathbf{u}_1$ and ψ_1 are the total (incident plus reflected or transmitted) particle velocity and its potential and on the plate surface S_p, which is created by the transmitter (transducer 1), and $\mathbf{u}_2^i$ and ψ_2^i are the incident (in the absence of the plate) particle velocity and its potential from transducer 2, when it operates as transmitter. Here, P is the incident power and $\mathbf{n}$ is the inward-directed normal to the plate surface. The integral in Eq. (A1) is calculated over the entire plate surface (top and bottom). It is straightforward to show that the integral is equal to zero over bottom of the plate, because the transmitted wave from transducer 1 and the incident wave from transducer 2 propagate in the same z direction. The remaining integral over the top surface of the plate S_t can be rewritten as

$$V=\frac{1}{4P}\int_{S_t}\left(\frac{\partial\psi_2^i}{\partial z}\psi_1-\frac{\partial\psi_1}{\partial z}\psi_2^i\right)dS_t. \tag{A2}$$

According to the Rayleigh integral formula for the field of a planar transducer, the particle velocity potential for the incident field of transducer 2 on the plate surface can be expressed as

$$\psi_2^i(S_t)=\frac{1}{2\pi}\int_{S_2}\mathcal{V}_2(S_2)\frac{\exp(i\kappa r)}{r}dS_2, \tag{A3}$$

where $\mathcal{V}_2(S_2)$ is the particle velocity on the receiver surface, when it operates as transmitter, and r is the distance between the transmitter and the plate. After substitution of Eq. (A3) into (A2), and changing the order of integration, we get

$$V=\frac{1}{4P}\int_{S_2}\mathcal{V}_2(S_2)dS_2\frac{1}{2\pi}\int_{S_t}\left(\psi_1\frac{\partial}{\partial z}\frac{\exp(i\kappa r)}{r}-\frac{\partial\psi_1}{\partial z}\frac{\exp(i\kappa r)}{r}\right)dS_t. \tag{A4}$$

According to Kirchhoff–Helmholtz's integral theorem (Ref. 27, pp. 181–182) the integral over the plate surface is equal to the incident field on the receiver surface $\psi_1^i(S_2)$, multiplied by 2. So, finally, we obtain

$$V=\frac{1}{2P}\int_{S_2}\mathcal{V}_2(S_2)\psi_1^i(S_2)dS_2, \tag{A5}$$

which is exactly our Eq. (8), where the factor $1/2P$ is omitted as irrelevant for our purposes.

[1] L. M. Brekovskikh, *Waves in Layered Media* (Academic, New York, 1975).

[2] L. M. Brekovskikh, ''Reflection of wave beams and pulses,'' Usp. Fiz. Nauk **50**, 539–579 (1953).

[3] O. A. Godin, ''Cause of the discrepancies between the results of different

theories of the Goos-Hanchen effect,'' Sov. Phys. Acoust. **31**, 18–21 (1985).
[4] L. M. Brekovskikh and O. A. Godin, *Acoustics of Layered Media II* (Springer-Verlag, Berlin, 1992), Chap. 2.
[5] A. Schoch, ''Schall Durchgang durch Platten (Sound transmission through plates),'' Acostica **2**, 1–17 (1952) (in German).
[6] B. R. Horowitz and T. Tamir, ''Lateral displacement of a light beam at a dielectric interface,'' J. Opt. Soc. Am. **61**, 586–594 (1971).
[7] H. L. Bertoni and T. Tamir, ''Unified theory of Rayleigh-angle phenomena for acoustic beams at liquid-solid interfaces,'' Appl. Phys. **2**, 157–172 (1973).
[8] J. W. Ra, H. L. Bertoni, and L. B. Felsen, ''Reflection and transmission of beams at a dielectric interface,'' SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **24**, 396–413 (1973).
[9] J. Pott and J. G. Harris, ''Scattering of an acoustic Gaussian beam from a fluid-solid interface,'' J. Acoust. Soc. Am. **76**, 1829–1838 (1984).
[10] H. Schmidt and F. B. Jensen, ''A full-wave solution for propagation in multilayered viscoelastic media with application to Gaussian beam reflection,'' J. Acoust. Soc. Am. **77**, 813–825 (1985).
[11] Y. Z. Ruan and L. B. Felsen, ''Reflection and transmission of beams at a curved interface,'' J. Opt. Soc. Am. A **3**, 566–579 (1986).
[12] D. E. Chimenti, J.-G. Zhang, S. Zeroug, and L. B. Felsen, ''Interaction of acoustic beams with fluid-loaded elastic structures,'' J. Acoust. Soc. Am. **95**, 45–59 (1994).
[13] O. I. Lobkis, A. Safaeinili, and D. E. Chimenti, ''Precision ultrasonic reflection studies in fluid-coupled plates,'' J. Acoust. Soc. Am. **99**, 2727–2736 (1996).
[14] R. K. Luneburg, *Mathematical Theory of Optics* (California U.P., Berkeley, 1964).
[15] P. R. Karmel, *Introduction to Electromagnetic and Microwave Engineering* (Wiley, New York, 1998).
[16] A. Williams, ''The piston source at high frequencies,'' J. Acoust. Soc. Am. **23**, 1–6 (1951).
[17] H. Seki, A. Granato, and R. Truell, ''Diffraction effects in the ultrasonic field of a piston source and their importance in the accurate measurement of attenuation,'' J. Acoust. Soc. Am. **28**, 230–238 (1956).
[18] P. A. Laura, ''Directional characteristics of vibrating circular plates and membranes,'' J. Acoust. Soc. Am. **40**, 1031–1033 (1966).
[19] J. Zemanek, ''Beam behavior within the nearfield of a vibrating piston,'' J. Acoust. Soc. Am. **49**, 181–191 (1971).
[20] A. Williams, ''Integrated signal on a circular piston receiver centered in a piston beam,'' J. Acoust. Soc. Am. **48**, 285–289 (1970).
[21] T. Rhyne, ''Radiation coupling of a disk to a plane and back or a disk to a disk,'' J. Acoust. Soc. Am. **61**, 318–324 (1977).
[22] C. J. Tranter, *Bessel Functions with Some Physical Applications* (Hart, New York, 1969).
[23] P. M. Morse and H. Feschbach, *Methods of Theoretical Physics* (McGraw–Hill, New York, 1953), Chap. 8.
[24] L. V. King, ''On the acoustic radiation field of the piezo-electric oscillator and the effect of viscosity on transmission,'' Can. J. Res. **N11**, 125–155 (1934).
[25] B. A. Auld, ''General electromechanical reciprocity relations applied the calculation of elastic wave scattering coefficients,'' Wave Motion **1**, 3–10 (1979).
[26] G. S. Kino, ''The application of reciprocity theory to scattering of acoustic waves by flaws,'' J. Appl. Phys. **49**, 3190–3199 (1978).
[27] A. D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications* (McGraw–Hill, New York, 1981).
[28] R. B. Thompson and E. F. Lopes, ''The effects of focusing and refraction on Gaussian ultrasonic beams,'' J. Nondestruct. Eval. **4**, 107–123 (1984).
[29] T. D. K. Ngoc and W. G. Mayer, ''Influence of plate mode structure and Gaussian beam profile characteristics on ultrasonic reflection and transmission,'' IEEE Trans. Sonics Ultrason. **SU-29**, 112–114 (1982).
[30] I. T. Lu, L. B. Felsen, and J. M. Klosner, ''Observables due to beam-to-mode conversion of a high-frequency Gaussian P-wave input in an aluminum plate in vacuum,'' J. Acoust. Soc. Am. **87**, 42–53 (1990).
[31] S. Zeroug and L. B. Felsen, ''Nonspecular reflection of two- and three-dimensional acoustic beams from fluid-immersed plane-layered elastic structures,'' J. Acoust. Soc. Am. **95**, 3075–3098 (1994).
[32] S. Zeroug, F. E. Stanke, and R. Burridge, ''A complex-transducer-point model for emitting and receiving ultrasonic transducers,'' Wave Motion **24**, 21–40 (1996).
[33] H. Zhang, D. E. Chimenti, and S. Zeroug, ''Transducer misalignment effects in beam reflection from elastic structures,'' J. Acoust. Soc. Am. **104**, 1982–1991 (1998).
[34] M. A. Breazeale, L. Adler, and G. W. Scott, ''Interaction of ultrasonic waves incident at the Rayleigh angle onto a liquid-solid interface,'' J. Appl. Phys. **48**, 530–537 (1977).
[35] J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics* (Wiley, New York, 1978).
[36] C. J. Daly and N. A. H. K. Rao, ''Time- and frequency-domain descriptions of spatially averaged one-way diffraction for an unfocused piston transducer,'' Ultrasonics **37**, 209–221 (1999).
[37] R. N. Bracewell, *The Fourier Transform and its Applications*, 3rd ed. (McGraw–Hill, Boston, 2000).
[38] C. J. Daly and N. A. H. K. Rao, ''A spatially averaged impulse response for an unfocused piston transducer,'' J. Acoust. Soc. Am. **105**, 1563–1566 (1999).

Gas-phase generation of photoacoustic sound in an open environment

Serdar H. Yönak[a] and David R. Dowling[b]
Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109-2133

(Received 27 January 2003; revised 12 September 2003; accepted 30 September 2003)

The photoacoustic effect is commonly exploited for molecular spectroscopy, nondestructive evaluation, and trace gas detection. Photoacoustic sound is produced when a photoactive material absorbs electromagnetic radiation and converts it to acoustic waves. This article focuses on the generation of photoacoustic sound from thermal expansion of photoactive gases due to unsteady heating from a laser light source, and extends the work of prior studies on photoacoustic sound generation in an open environment. Starting with the forced free-space wave equation, a simple model is constructed for photoacoustic sounds produced by both acoustically distributed and compact gas clouds. The model accounts for laser absorption through the Lambert–Beer law and includes the effects of photoactive gas cloud characteristics (shape, size, and concentration distribution), but does not include molecular diffusion, thermal conduction, convection, or the effects of acoustic propagation through sound-absorbing inhomogeneous media. This model is compared to experimentally measured photoacoustic sounds generated by scanning a 10.6-μ carbon dioxide (CO_2) laser beam through small clouds of a photoactive gas, sulfur hexafluoride (SF_6). For the current investigation, the photoactive gas clouds are formed either by low flow-rate calibrated leak sources or by a laminar jet emerging from a 1.6-mm-diam tube. Model-measurement comparisons are presented over a 3- to 160-kHz bandwidth. Signal pulse shapes from simple gas cloud geometries are found to match calculated results when unmeasured gas cloud characteristics within the model are adjusted. [DOI: 10.1121/1.1628250]

PACS numbers: 43.35.Ud, 43.38.Zp [RR] Pages: 3167–3178

I. INTRODUCTION

Photoacoustics is the generation of sound due to unsteady heating of a photoactive material by a source of light or invisible electromagnetic radiation. It was discovered more than a century ago (Bell, 1880, 1881). Since then, photoacoustics has been used for spectroscopic studies and nondestructive evaluation of solids, liquids, and gases (see Rosencwaig, 1980). Photoacoustic signal characteristics in solids and thin films are discussed in Bresse and Hutchins (1989) and Rogers *et al.* (2000). Liquid-phase photoacoustics and potential applications to medical diagnosis are presented in Hoelen and de Mul (1999).

Although the photoacoustic effect can occur without thermal expansion of the photoactive material (Diebold *et al.*, 2002), gas-phase photoacoustic sound is primarily the result of radiation-induced thermal expansion of a photoactive gas. In this study, photoacoustic sound was generated by the illumination of a photoactive gas cloud with radiation (typically infrared) that is absorbed by the gas molecules via quantum mechanical transitions to elevated vibration or rotation states. At room temperature and pressure, the excited photoactive gas molecules quickly lose their excess energy and thermally equilibrate with the surrounding gas through the action of molecular collisions. However, this energy transfer from the radiation source to the constituents of the photoactive gas cloud increases the temperature of the photoactive gas cloud. In unconfined environments, this radiation-induced gas-phase temperature rise causes the gas cloud to expand. When the illumination is unsteady (i.e., the radiation source has nonzero first temporal derivative), the photoactive gas-cloud expansion will be unsteady and may launch acoustic pressure waves. Illumination variations from lower (higher) to higher (lower) power levels produce compression (expansion) waves. The expansion waves occur when the radiation level decreases because this decreases the rate of gas cloud growth, and this growth must fall to zero when the radiation source is turned off. In addition, thermal conduction of the radiation-induced heat to a solid (or liquid surface) cools the gas cloud and will cause it to contract. Thus, a listener (or microphone) at some distance from the region of unsteady radiant heating might detect a rise and fall in acoustic pressure due to the growth and contraction of the photoactive gas cloud.

In gas-phase spectroscopic investigations employing photoacoustics, a cylindrical chamber is typically filled with a photoactive gas, or mixture of gases, and the light source is modulated to excite an acoustic resonance of the chamber in order to increase signal levels and sensitivity (Kreuzer and Patel, 1971; Kreuzer, 1977). This approach is successful enough to detect trace gas contaminants in the atmosphere (Dewey *et al.*, 1973; Dewey, 1974; Goldan and Goto, 1974; Claspy, 1977). Here, the acoustic waves are confined and the test gases are uniform or uniformly mixed. The main topic of this paper is the generation of unconfined photoacoustic sound from spatially inhomogeneous distributions of photoactive gas. The goal is to formulate the process with suffi-

[a]Currently employed by Jaguar Cars Ltd.
[b]Author to whom correspondence should be addressed. Electronic mail: drd@engin.umich.edu

cient accuracy so that photoacoustic pulse shapes from isolated photoactive gas clouds can be predicted in open environments comprised mostly of air.

Prior work on gas-phase photoacoustics in open environments has involved remote detection and ranging of gas clouds (Brassington, 1982). This approach was specialized to the case of small gas clouds formed near leaks sites on commercial components carrying a pressurized photoactive tracer gas to produce a noncontacting leak testing technique (McRae and Dewey, 1992; McRae, 1994). The potential broad application of such gas-phase photoacoustic leak test techniques spurred further developments using microphone arrays and matched-field signal processing to acoustically determine leak locations (Yönak and Dowling, 1999, 2001, 2002).

In this paper, measurements made as part of these recent leak-test studies are compared to an open-environment model of gas-phase photoacoustic signal generation. The radiation source is a carbon dioxide (CO_2) laser tuned to 10.6 μ. This wavelength corresponds to the P(20) rotational line for CO_2 which is one of the strongest transitions for a CO_2 laser (Hecht, 1992). The photoactive gas is sulfur hexafluoride, SF_6, which has a strong absorption line coincident with the laser output (Braker and Mossman, 1980) that corresponds to a vibrational transition (Hanna, 1981; McQuarrie, 1983). Illumination unsteadiness is introduced by rapidly sweeping (or scanning) the laser beam through a stationary cloud of the photoactive gas via a rotating prismatic mirror. This leads to a time-periodic illumination pattern whose characteristics are determined by the size and shape of the laser beam cross section, the power in the laser beam, the number of mirror facets, and the rotational speed at which they are spun. Thus, in the experiments reported here, photoacoustic sound is generated at integer multiples of the laser beam scan rate.

Basic open-environment photoacoustic signal characteristics are summarized in Yönak and Dowling (1999) and Huang *et al.* (2003). This paper extends their work by examining the additional details of photoacoustic sound generation and investigating the factors that influence photoacoustic amplitudes and temporal pulse shapes. Such information may lead to new applications for photoacoustic technology as well as guide research and improvements in established areas.

Here gas-phase photoacoustic sound generation is modeled using a forced wave equation, the Lambert–Beer radiation-absorption law, and simple photoactive gas clouds whose geometry could be surmised, at least approximately, without detailed gas-phase advection-diffusion simulations or sophisticated auxiliary (i.e., nonacoustic) experimental measurements. First, this model is developed and evaluated for an acoustically distributed gas cloud that could be visualized using the optical shadow photography method. In this context, an acoustically distributed (or noncompact) gas cloud has a laser absorption path length or characteristic size that is comparable to or larger than the acoustic wavelengths of interest (i.e., kL near unity or greater, where k is the acoustic wave number and L is the absorption path length or characteristic dimension of the photoacoustic source). Second, the model is simplified for acoustically compact photoactive gas clouds ($kL \ll 1$) where shadow visualization was not possible but a nominally hemispherical distribution could be assumed. And third, the model results are compared with some success to measured photoacoustic signal pulses for both cloud geometries. The experimental apparatus used for this investigation was the same as that used by Yönak and Dowling (1999, 2002), except that a single microphone was used to record the photoacoustic signals and the backing plate or cylinder was absent from the noncompact cloud studies.

TABLE I. Gas properties at room temperature and pressure.

Properties	Symbol	Units	Air	SF_6
Absorption constant at 10.6 μm, 1 atm.	β	m^{-1}	⋯	2.78×10^{4}
Density	ρ	kg/m^3	1.20	6.16
Diffusivity into air	D	m^2/s	1.86×10^{-5}	9.2×10^{-6}
Gas constant	R	$m^2/s^2\,K$	287	56.9
Heat capacity at constant pressure	c_p	$m^2/s^2\,K$	1,004	665
Molecular weight	M_w	kg/kmol	29.0	146
Ratio of specific heats	γ	⋯	1.4	1.12
Speed of sound	c	m/s	345	137
Thermal conductivity	κ	W/m·K	0.026	0.014
Thermal diffusivity	α	m^2/s	2.2×10^{-5}	3.4×10^{-6}
Viscosity	μ	kg/m·s	1.82×10^{-5}	1.56×10^{-5}

The remainder of this paper is divided into four sections. Section II describes a simplified acoustic theory for the generation of photoacoustic sound due to the spectroscopic absorption of laser light by an inhomogeneous distribution of photoactive gas having the same acoustic properties as the photo-inactive gas of the surroundings. Sections III and IV present the application of the theory to noncompact and compact gas clouds, respectively, with comparison to experimental results. The final section provides a summary of this paper and the conclusions drawn from this research.

II. THEORY OF PHOTOACOUSTIC SOUND GENERATION BY GAS CLOUDS

This analysis of gas-phase photoacoustic signal generation begins with the lossless forced wave equation (Morse and Ingard, 1968),

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}-\nabla^2 p=\left(\frac{\gamma-1}{c^2}\right)\frac{\partial}{\partial t}H(\vec{x},t), \tag{1}$$

where c is the speed of sound (m/s), p is the acoustic pressure (Pa), γ is the ratio of specific heats, and $H(\vec{x},t)$ is the local rate of heat addition per unit volume (J/m^3). In order to use Eq. (1), the properties (c and ρ=gas density in kg/m^3) of the gas in which the sound waves originate and propagate must be constant, and acoustic absorption losses must be negligible.

For the present study involving nonuniform mixtures of air and SF_6, this equation represents an approximation because the physical properties of these gases at room temperature and pressure are very different. Table I compares the

physical properties of these gases based on prior tabulations (Braker and Mossman, 1980; Incropera and DeWitt, 1996; Van Wylen *et al.*, 1994) and empirical formulas (Reid *et al.*, 1977). Although Eq. (1) does not allow for spatially varying c and ρ, SF_6 absorbs CO_2 laser radiation at 10.6 μ very efficiently. Thus, for the simple SF_6-cloud-laser-microphone geometries considered here, photoacoustic sound waves were assumed to be generated either in regions where the mole fraction of SF_6 was low and the gas properties were close to those of air, or at the outer surface of regions where the mole fraction of SF_6 was high and nearly all of the subsequent acoustic propagation took place in air. Thus, Eq. (1) is the basis for the model development and the physical properties of air are used wherever appropriate. Although this modeling approach limits the accuracy of final results, it does allow a relatively simple analytic formulation.

A second approximation inherent in Eq. (1) is the neglect of acoustic absorption losses. Such losses are important when high-frequency sound waves propagate over long distances. In this investigation the longest path length is 20 cm and the highest measured frequencies lie between 100 and 200 kHz. For low humidity room temperature air, the acoustic absorption losses estimated from a standard attenuation plot (see Kinsler *et al.*, 2000) and a 20-cm path length at 100 and 200 kHz are 0.4 dB ($\sim$5% amplitude loss) and 1.4 dB ($\sim$17% amplitude loss), respectively. Thus, amplitude discrepancies between the results calculated from Eq. (1) and the experimental measurements at frequencies above 100 kHz are expected to be 1 dB or more.

A third approximation made in this modeling effort is the neglect of the residual heating of the gas. For a given laser intensity $I\,(\mathrm{W/m^2})$, exposure time Δt, and radiation absorption depth l_D, conservation of energy (ignoring conduction and convection, and the energy carried by in the photoacoustic wave) for a radiation-absorbing gas volume of unit cross-sectional area suggests that the local gas temperature will increase an amount $\Delta T = I\Delta t/\rho c_p l_D$, where $c_p\,(\mathrm{J/kg\cdot K})$ is the gas specific heat at constant pressure. For a moving collimated laser beam, I is approximately equal to P/d_b^2, where $P\,(\mathrm{W})$ is the laser power and d_b is the laser beam diameter, and $\Delta t \sim d_b'/a$ where $a\,(\mathrm{m/s})$ is the sweep or scan speed of the laser beam. With these replacements, the temperature increment becomes $\Delta T = P/a d_b \rho c_p l_D$, and is $\sim$0.5 °C when evaluated using the noncompact photoacoustic-source experimental parameters for one laser pulse and air properties. If conduction and convection heat losses—or SF_6 gas properties—are included, this temperature rise estimate would be smaller. Thus, neglect of thermal changes in the irradiated gas should be an acceptable approximation, even if the same volume of gas absorbs a few tens of laser pulses.

Although a more exact formulation of the photoacoustic sound generation process based on the field equations for an inhomogeneous sound-absorbing medium with conductive and convective heat transfer and molecular transport is possible, such an effort was not pursued because results from Eq. (1) achieve acceptable agreement with experimental measurements of photoacoustic sound when unmeasured nonacoustic properties of the photoactive gas cloud are adjusted within physically reasonable bounds.

In Eq. (1), The time derivative of H explicitly displays the requirement for the heating rate to be unsteady. The following analysis differs from prior formulations of laser-induced thermal waves and thermal gratings in gases (Paul *et al.*, 1995; Cummings *et al.*, 1995; Schlamp *et al.*, 1999) in its specialization for predicting the photoacoustic pressure waves. In addition, Brassington (1982) and Zharov and Symanovskii (1990) have both investigated photoacoustic signals from large gas clouds, and arrived at similar results. However, Brassington (1982) neglects the time dependence of the photoacoustic sound generation and Zharov and Symanovskii (1990) do not include realistic laser absorption through the gas cloud as described by the Lambert–Beer law. Furthermore, previous research efforts also do not address the acoustic implications of noncompact-gas-cloud photoacoustic sound generation.

The formal solution of Eq. (1) can be cast in terms of a spatial-temporal integration (see Pierce, 1989):

$$p(\vec{x},t) = \left(\frac{\gamma-1}{c^2}\right)\int_{\text{all }\vec{x}_s}\int_{t_s} G(\vec{x},t|\vec{x}_s,t_s) \times\left[\frac{\partial}{\partial t_s}H(\vec{x}_s,t_s)\right]dt_s\,d^3\vec{x}_s, \tag{2}$$

where $G(\vec{x},t|\vec{x}_s,t_s)$ is the time-domain Green's function for the acoustic environment. Here, to be consistent with the above approximations, γ and c are set to air values, and the simple free space Green's function is used,

$$G(\vec{x},t|\vec{x}_s,t_s) = \frac{\delta(t-t_s-|\vec{x}-\vec{x}_s|/c)}{4\pi|\vec{x}-\vec{x}_s|}. \tag{3}$$

The spatial and temporal extent of the heating term will be determined by the photoactive gas cloud characteristics, and the laser beam intensity and size. In general, when a laser beam propagates through a radiation-absorbing gas, in the absence of reradiation, a portion of the laser radiation heats the gas while the remaining portion is transmitted. Under these circumstances, conservation of laser energy requires

$$H(\vec{x},t) = -\vec{\nabla}\cdot\vec{I}(\vec{x},t), \tag{4}$$

where $\vec{I}(\vec{x},t)$ is the laser beam intensity. To evaluate Eqs. (2)–(4), consider the coordinate system shown in Fig. 1 where the laser beam travels parallel to and in the opposite direction of the z axis. This possibly awkward choice of coordinates facilitates later acoustic calculations by allowing the measurement point for photoacoustic sound to lie in the $z>0$ half-space. The laser intensity vector for this situation has only a z component and can be written as

$$\vec{I}(\vec{x},t) = \begin{bmatrix} 0 \\ 0 \\ -I_z(\vec{x},t) \end{bmatrix}. \tag{5}$$

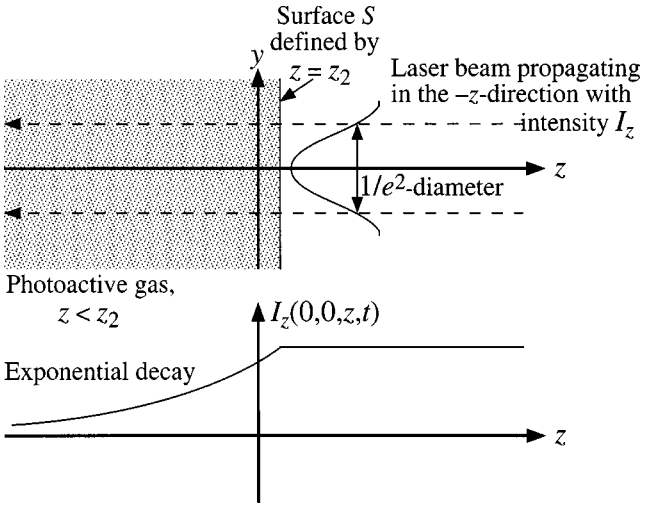


FIG. 1. An illustration of the decrease of laser intensity as it passes through a photoactive gas distribution. The gas present for $z>z_2$ does not absorb the laser radiation. For $z<z_2$ a mixture of an absorbing gas A and the nonabsorbing gas is present. $X_A(x,y,z,t)$ is the mole fraction distribution for gas A within the non absorbing gas. The laser beam profile is axisymmetric about the z axis.

When the photoactive gas of mole fraction X is confined to lie inside a closed surface S (see Figs. 1 and 2), there is no absorption exterior to S so the laser intensity is a function of time and the x and y coordinates alone [i.e., $I_z(x,y,z,t) = I_z(x,y,0,t)$ outside of S]. Inside of S, the photoactive gas absorbs laser radiation. Thus, the laser intensity decreases according to an attenuation integral form of the Lambert–Beer law [adapted here from Lück and Müller (1977)],

$$I_z(x,y,z,t) = I_o(x,y,t)\exp\left\{-\int_z^{z_2}\beta X(x,y,z',t)dz'\right\} \quad \text{for } z_1<z<z_2, \tag{6}$$

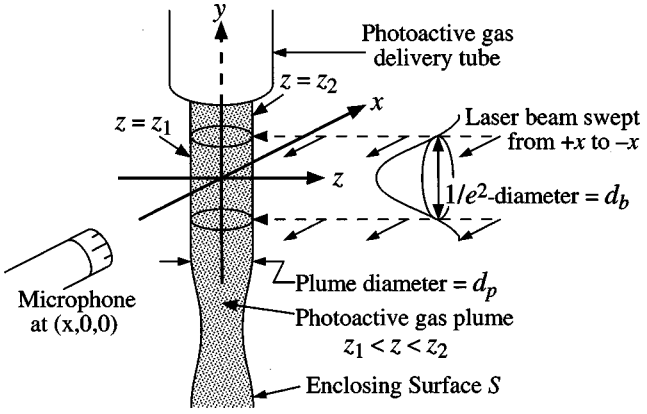


FIG. 2. Photoacoustic sound from a noncompact source is generated by scanning the CO_2-laser beam through the stable cylindrical portion of a vertical laminar plume of SF_6. The direction of the flow is in the negative y direction. The cylindrical portion of the plume has an effective diameter of d_p. The CO_2-laser beam having a Gaussian profile and the $1/e^2$ diameter, d_b, is defined with the dashed lines. The laser beam propagates in the negative z direction and is swept rapidly and repeatedly through the gas cloud from $+x$ to $-x$ with scan speed a.

where I_o is the laser beam's cross-sectional intensity profile for $z>z_2$, β is the spectroscopic gas absorption coefficient, and $z_2(x,y,t)$ and $z_1(x,y,t)$ define the front and back surfaces of S. For $z<z_2$, Eq. (6) implies that the laser beam intensity decreases as it propagates through the photoactive gas and that this decrease will depend exponentially on β. In addition, any bulk motion or diffusion of the absorbing gas described by the spatial distribution $X(x,y,z,t)$ enters in an integral placed in an exponent, thus Eq. (6) is clearly sensitive to gas cloud characteristics. And, finally, if the mole fraction of the absorbing gas is unity, the argument of the exponential function in Eq. (6) reduces to the absorption coefficient times the absorption path length, z_2-z_1.

The radiative heat addition per unit volume of gas can be found from Eqs. (4)–(6):

$$H(x,y,z,t) = \frac{\partial I_z}{\partial z} = \beta X(x,y,z,t) I_o(x,y,t) \times \exp\left\{-\int_z^{z_2}\beta X(x,y,z',t)dz'\right\}. \tag{7}$$

For the remainder of this paper, Eqs. (2), (3), and (7) will be used to predict the photoacoustic sound from acoustically noncompact and compact gas clouds in uniform open environments.

III. AN ACOUSTICALLY NONCOMPACT GAS CLOUD

An acoustic source is not compact when its characteristic size is comparable to or larger than an acoustic wavelength. For these studies, a controlled noncompact photoacoustic sound source was generated by scanning a CO_2 laser beam through a laminar plume of SF_6 emerging from a small tube. The plume-laser scan geometry is shown in Fig. 2. The plume flows downward along the y axis, and the scanning laser beam is swept in the $x-z$ plane perpendicular to the y axis. Although SF_6 is a clear gas, the main flow of such a plume can be visualized using shadow photography (see Liepmann and Roshko, 1957) because SF_6 has a different index of refraction than air (see Fig. 3). Unfortunately, shadow photography is not sensitive enough to visualize the tenuous edges of the SF_6 plume, so low concentrations of SF_6 are likely to be present beyond the ~2-mm plume diameter that is visualized in Fig. 3. However, even if the effective plume diameter is two or three times larger than this visual estimate, the largest length scale, L, defining the size of the acoustic source in the current experiments is the diameter of the CO_2 laser beam ($1/e^2$ diameter of 8.8 mm). For the current experiments, the highest signal frequency is of order 100 kHz. In room temperature air, these parameters yield a $kL=2\pi fL/c\sim 16$, indicating an acoustically noncompact source capable of directional sound radiation.

The intent of this section is to present a model of photoacoustic sound generation for the laminar SF_6 gas plume shown in Fig. 3 when it is scanned by a collimated axisymmetric CO_2 laser beam having power P that is swept rapidly at scan speed a through the nearly cylindrical portion of the plume. The laser beam motion is restricted to constant hori-

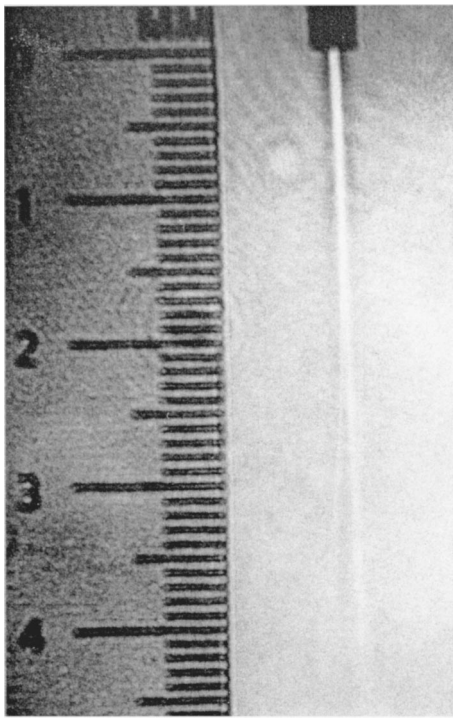


FIG. 3. Shadow photograph image of a laminar plume of SF_6 at Re=380 based on initial plume diameter. The inner diameter of the tube at the top of the picture is 1.6 mm. The light source was a 10 mW red HeNe laser with a 5-mm-diam spherical ball lens mounted directly on the laser. The distance from the laser to the plume was approximately 61 cm and the distance from the laser to the screen on which the image was projected was approximately 5.3 m. The exposure time was 1 s for this picture. The plume was typically stable and axisymmetric for at least 13 initial plume diameters downstream before it became unstable. The center of the scanned CO_2-laser beam cut the plume 1 cm below the tube exit.

zontal travel perpendicular to the beam axis, i.e., the laser beam axis remains parallel to the z axis and moves through the $x-z$ plane from positive to negative x at linear speed a. Thus, the illumination source for the photoactive gas plume can be modeled as a traveling Gaussian intensity profile. The air surrounding the plume is nonabsorbing at the laser wavelength of 10.6 μ. The laser-beam intensity profile is assumed to be Gaussian with a $1/e^2$ diameter$=d_b$.

In these experiments, the laser-plume interaction time is short compared to any unsteady fluid-mechanical or diffusion time scales, so the plume can be treated as a steady spatial mole fraction distribution of SF_6, $X_{SF_6}(x,y,z)$. Unfortunately, an independent means for measuring $X_{SF_6}(x,y,z)$ was not available and detailed calculations for $X_{SF_6}(x,y,z)$ based on an appropriate advection diffusion equation were not attempted. Instead, an approximate plume geometry was used to evaluate the photoacoustic sound generation model. The plume was treated as a perfect cylinder aligned with the y axis having an effective diameter d_p and a uniform interior value of X_{SF_6}. An effective diameter was assigned to the plume because SF_6 diffuses beyond the plume's visually indicated size into the surrounding air at sufficient concentration for significant laser energy absorption outside the visualized plume diameter. Although this simplification of the actual plume geometry limits the extent to which the experiments can be used to validate the photoacoustic sound generation model developed above, it has the advantage of allowing a fully analytic specification of the heating term, H. Thus, the attenuation integral in Eq. (6) can be evaluated and closed-form expressions for I_z can be written for the region exterior to the plume and its shadow ($x^2+z^2>d_p^2/4$ for $z>0$, and $|x|>d_p/2$ for $z\leqslant 0$), the interior of the plume ($x^2+z^2\leqslant d_p^2/4$), and the shadow of the plume ($z\leqslant -\sqrt{d_p^2/4-x^2}$ with $|x|\leqslant d_p/2$) using $z_2=+\sqrt{d_p^2/4-x^2}$ and $z_1=-\sqrt{d_p^2/4-x^2}$:

$$I_z=\begin{cases}\dfrac{8P}{\pi d_b^2}\exp\left(\dfrac{-8y^2}{d_b^2}\right)\exp\left(\dfrac{-8(x+at)^2}{d_b^2}\right)=I_o(x,y,t) & \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{exterior,} & \\ I_o(x,y,t)\exp\{-\beta X_{SF_6}(\sqrt{d_p^2/4-x^2}-z)\} \quad \text{interior,} & \\ I_o(x,y,t)\exp\{-\beta X_{SF_6}(2\sqrt{d_p^2/4-x^2})\} \quad \text{shadow.}\end{cases} \tag{8}$$

Here the constant mole-fraction of SF_6 within the plume is denoted X_{SF_6}. In Eq. (8) the laser intensity does not depend on z outside of the absorbing gas plume so Eqs. (4) or (7) implies

$$H(\vec{x},t)=\begin{cases}0 \quad \text{for } x^2+z^2>d_p^2/4 & \\ \beta X_{SF_6} I_o(x,y,t)\exp\{-\beta X_{SF_6}(\sqrt{d_p^2/4-x^2}-z)\}. & \\ \qquad \text{for } x^2+z^2\leqslant d_p^2/4\end{cases} \tag{9}$$

The time derivative term in Eq. (2) can now be calculated from Eq. (9).

$$\frac{\partial}{\partial t}H(\vec{x},t)=\frac{-16a(x+at)}{d_b^2}H(\vec{x},t). \tag{10}$$

Using Eqs. (10) and (3) together and performing the time integration yields

$$p(\vec{x},t)=\left(\frac{\gamma-1}{4\pi c^2}\right)\frac{\beta X_{SF_6}}{d_b^2}\int_{-d_p/2}^{+d_p/2}dx_s\int_{-\infty}^{+\infty}dy_s\int_{z_1}^{z_2}dz_s \times\frac{-16a(x_s+at')I_o(x_s,y_s,t')\exp\{-\beta X_{SF_6}(\sqrt{d_p^2/4-x_s^2}-z_s)\}}{\sqrt{(x-x_s)^2+(y-y_s)^2+(z-z_s)^2}}, \tag{11}$$

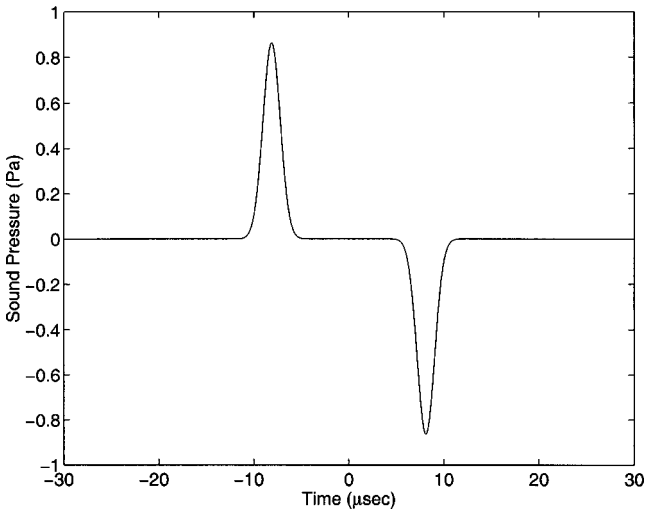


FIG. 4. Typical model result for one sweep of the CO_2-laser beam through the SF_6 gas plume. The results here are for a field point located at $x=-0.2$ m for a laser power of 8.8 W and a laser beam $1/e^2$ diameter of 8.4 mm. The effective diameter of the plume is 6.5 mm and the absorption coefficient is $2.78\times10^4\ \mathrm{m}^{-1}$. The mole fraction of SF_6 within the plume is set to unity. The scan speed of the beam through the gas cloud is 2368 m/s. It should be noted that unlike the experiments no bandwidth limits were placed on this model result.

where $t'=t-|\vec{x}-\vec{x}_s|/c$. Here, x_s, y_s, and z_s represent the spatial coordinates where the heating rate is nonzero. Two of these three spatial integrations are possible when the measurement location $\vec{x}$ resides on the x axis, $\vec{x}=(x,0,0)$, in the Fraunhofer far field of the source, $d_b/|\vec{x}|\ll1$ and $fL^2/4c|\vec{x}|\ll1$ (see Kinsler *et al.*, 2000) where f is the highest sound frequency of interest. These simplifications allow the substitutions $1/|\vec{x}-\vec{x}_s|\approx1/|x|$ and $t'\approx t-(|x|/c)(1-x_s/x)$, so (11) can be reduced to

$$p(x,0,0,t)=\frac{-8\sqrt{2}P(\gamma-1)a}{|x|\pi^{3/2}c^2d_b^3}\int_{-d_p/2}^{d_p/2}[x_s(1+ax/(c|x|))+a(t-|x|/c)]\times\exp\left(\frac{-8[x_s(1+ax/(c|x|))+a(t-|x|/c)]^2}{d_b^2}\right)\times[1-\exp\{-2\beta X_{SF_6}\sqrt{d_p^2/4-x_s^2}\}]dx_s. \quad (12)$$

The remaining x_s integration must be evaluated numerically to find $p(x,0,0,t)$; however, some of the parametric dependencies are apparent in the prefactor to the integral. Increasing the laser power, P, increases the sound amplitude. The prefactor also shows photoacoustic pressure amplitude increasing linearly with the scan speed a and decreasing inversely proportional to the distance to the field point and to the laser beam diameter cubed, but these factors also appear inside the integrand.

Example results from the model are shown in Fig. 4 for one sweep of the laser beam through the gas plume. The parameters used to generate the results in Fig. 4 are given in the figure caption and in Table I. The value for the absorption coefficient, $\beta=2.78\times10^4\ \mathrm{m}^{-1}$, was estimated from a measured low-pressure SF_6 infrared spectrum (Braker and Mossman, 1980) and the Lambert–Beer law for linear absorption from Okabe (1978). This absorption value is very high and suggests that for a laser beam incident on a plume of pure SF_6 at ambient pressure, 99% of the laser radiation would be absorbed in less than 0.17 mm. Interestingly, the photoacoustic sound model, Eq. (12), is not sensitive to β in this parameter range—across approximately a factor of 10—because β is so high that only a thin skin of photoactive gas at the surface of the modeled SF_6 plume undergoes laser heating and produces photoacoustic sound. In the experiments, the zone of laser heating is likely to be much thicker because diffusion of SF_6 into the surrounding air will lead to a smoothly decreasing $X_{SF_6}(x,y,z)$ as $\sqrt{x^2+z^2}$ increases, instead of the modeled-plume's sharp transition from a constant X_{SF_6} to zero once $\sqrt{x^2+z^2}$ exceeds $d_p/2$.

The characteristic diffusion thickness of the actual radiation absorbing layer, l_D, can be estimated from the diffusivity of SF_6 into air, D, and the time, Δt_D, that it takes SF_6 to move from the tube exit to the center of the illumination region: $l_D\sim2\sqrt{D\Delta t_D}$. The factor of 2 is included because air diffuses toward the center of the plume (the y axis) as SF_6 diffuses away from it. Here we can estimate Δt_D from the plume exit velocity, ~0.6 m/s, and the 1-cm distance from the tube outlet to the plane of the laser scan: $\Delta t_D\approx(0.01\,\mathrm{m})/(0.6\,\mathrm{m/s})=0.017\,\mathrm{s}$, so $l_D\sim2\sqrt{(9.2\times10^{-6}\,\mathrm{m^2/s})(0.017\,\mathrm{s})}=0.8\,\mathrm{mm}$, where the value for D has been taken from Table I. This is the value of l_D used to estimate the gas temperature rise ΔT, discussed partway through Sec. II. On this same issue, observations of the shadow image of the SF_6 plume with and without laser irradiation did not reveal perceptible visual changes. Thus, the laser-SF_6 interaction did not change the primary plume flow in a detectable manner.

The general shape of the photoacoustic pulse in Fig. 4 can be deduced from the evolution of the laser-photoactive gas interaction. As the laser beam passes through the SF_6 gas plume, laser radiation is absorbed and the local pressure and temperature within the plume increase. The laser-heated gas expands and launches a compression acoustic wave. Hence, at the field point, the predicted acoustic pressure rises first. A drop in acoustic pressure follows this compression because the photoactive gas expansion slows and then ceases as the moving laser beam passes beyond the SF_6 plume and the laser-induced heating subsides.

The duration of the photoacoustic pulse shown in Fig. 4 can be explained as well. In this case, the laser scan speed (2368 m/s) is far above the ambient speed of sound (345 m/s), so the gas plume is effectively flash illuminated by the laser. Thus, the duration of the photoacoustic pulse is set primarily by the difference in acoustic propagation times from the nearest and furthest parts of the SF_6 plume to the microphone. This propagation time difference is approximately $d_p/c=0.0065\,\mathrm{m}/(345\,\mathrm{m/s})\approx19\,\mu\mathrm{s}$, a value that matches the overall pulse duration shown in Fig. 4.

The experiments used for the model comparisons were performed using a nominal 12 W CO_2 laser tuned to 10.6 μ whose beam was scanned across the SF_6 plume depicted in Fig. 3 with a rotating prismatic mirror. The setup was similar to that described by Yönak and Dowling (1999) with the

aluminum test plate absent and one of the $\frac{1}{4}$-in. microphones moved closer to the SF_6 plume. The geometry for the experiments was identical to that shown in Fig. 3. The laminar SF_6 gas plume discharged downward from a 1.6-mm inner diameter plastic tube with a tube-exit Reynolds number of ~380 (Fig. 2). The tube exit was centered on the y axis 1 cm above the plane of the laser scan (the x-z plane) so that the plume was scanned where it was nearly cylindrical in shape and temporally stable in the laboratory environment. The ambient air temperature used to determine the speed of sound was measured to be 26.5 °C using a bare-wire T-type thermocouple. The scan speed of the laser beam through the gas plume was 2368 m/s using a 12-sided polygonal rotating mirror placed 0.95 m away from the plume and scanning though a 60° angle at a rate of 2380 Hz. At the plume location, the laser power was $P=9.6$ to 9.7 W and the $1/e^2$ beam diameter was $d_b=8.8$ mm.

Photoacoustic sound measurements were made with a single $\frac{1}{4}$-in. microphone placed at $x=-20$ cm. The microphone and plume were positioned to maximize the distance to the nearest flat reflecting surface in the laboratory, in this case the floor, approximately 1 m away. Thus, at the microphone, the reflected wave amplitude should be only about 10% of the directly incident wave. The microphone protection grid was removed in order to take full advantage of the microphone's frequency response out to 100 kHz and beyond. The microphone signal was band pass filtered by the microphone power supply in a 20-Hz to 100-kHz bandwidth using a two-pole Butterworth filter. The signal was subsequently amplified and band-pass filtered again in a 3- to 110-kHz bandwidth using a four-pole Butterworth filter. The signal was then acquired at 1.25 MHz and converted to pressure using the microphone manufacturer's calibration data. In order to provide a fair comparison between the experiment and the model, the model results were digitally filtered in the exactly same manner as the experimental measurements. It is important to note, however, that neither the frequency response of the microphone nor acoustic absorption were taken into account in the model calculations. This will be apparent in the comparisons made between the model and the experiment at frequencies from 100 to 160 kHz which lie beyond the nominal cutoff of the microphone without its protection grid and where acoustic absorption by the air becomes appreciable.

This measurement geometry was chosen to facilitate comparisons with evaluations of Eq. (12). For $|\vec{x}|=20$ cm, a laser beam diameter of $d_b=L=8.8$ mm, and a frequency $f=100$ kHz, the parameters that govern the extent of the photoacoustic source's near field are computed as follows: $L/x=0.044$ and $fL^2/4cx=0.028$. These values suggest that the acoustic approximations leading to Eq. (12) are acceptable at the measurement point.

A comparison between the measurements and the model for a single sweep of the laser beam through the gas plume is shown in Fig. 5(a). The minor fluctuations in the experimental time trace preceding and following the signal pulse in Fig. 5(a) are due to coherent noise and photoacoustic sound reflected from the laboratory walls, floor, and ceiling, none of which were acoustically treated. The experimental observations made here are consistent with those made by Brassington (1982) who used identical model microphones from the same manufacturer. However, Brassington's measurements appear to have been made with the microphone protection grid in place, and—based on the manufacturer's calibration chart—the protection grid reduces the microphone bandwidth by a factor of 2 or so.

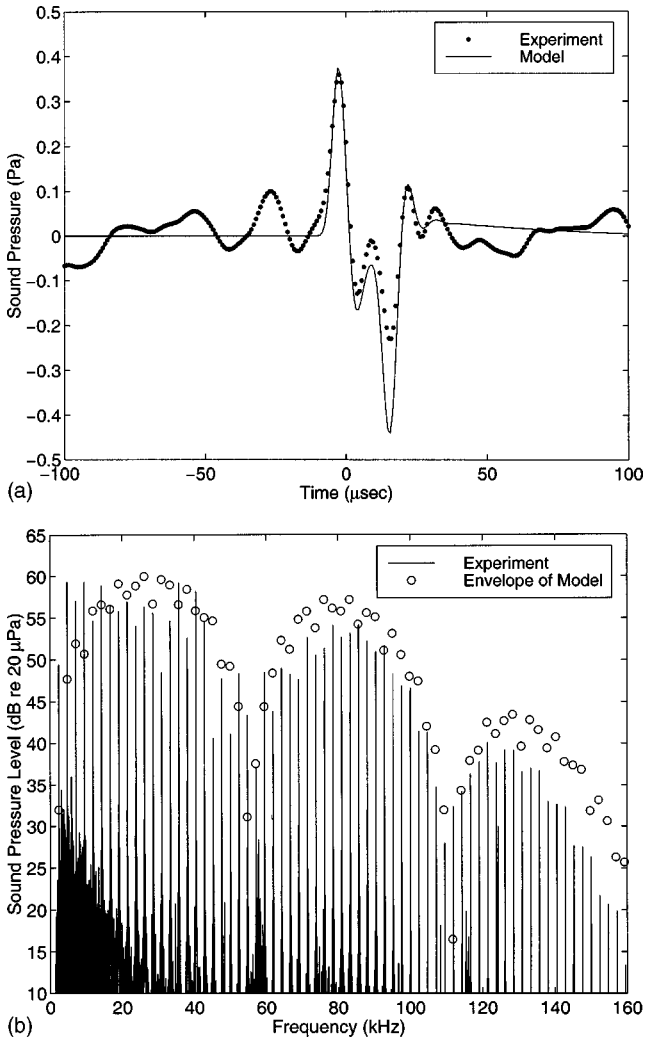


FIG. 5. A comparison of experimental measurements of photoacoustic sound pressure from a large gas cloud to model results in (a) the time domain and (b) the frequency domain. In the model, the effective plume diameter was set to 7.2 mm and the mole fraction of the gas within the plume was set to 0.194.

To evaluate the model for Fig. 5(a), the effective plume diameter and SF_6 mole fraction within the gas plume were adjusted to 7.2 mm and 0.194, respectively, to yield the best agreement between Eq. (12) and the experiment. These parameter adjustments were intended to account for unmodeled diffusive molecular transport. With the fitted plume parameters, there is good agreement between Eq. (12) and measurements for the initial compression, and the overall pulse shape and width. However, the model overpredicts the amplitude of the negative-pressure portion of the photoacoustic pulse. This mismatch is likely due to the model's neglect of the effects of nonuniformities, diffusion, and heat conduction in the region of laser heating, and differences between the actual and assumed SF_6 distributions.

Perhaps the greatest discrepancy between the experimental and theoretical results is the difference between the visually determined plume diameter of 2 mm, or at most 3 mm, and the model-measurement-matching plume diameter of 7.2 mm. When a plume diameter of 3 mm and a mole fraction of unity are used in Eq. (12), there is excellent model-measurement agreement for the initial compression of the photoacoustic pulse, but the pulse duration is underpredicted by more than a factor of 2 and the shape of the second half of the measured pulse is not matched. The plume diameter discrepancy is related to the limitations of shadow imaging and the unknown extent of the radial diffusion of the photoactive gas. Extracting quantitative information from shadow images like Fig. 3 is subject to inherent limitations. The bright and shadow zones in an image are formed where changes in the refractive-index gradients transverse to the direction of light travel are strong. Thus, the shadows shown in Fig. 3 only indicate the width of the core of the SF_6 plume and not its tenuous diffusion-driven outer edge, which will have a larger diameter. Thus, for the present experiments where SF_6 absorbs CO_2 laser radiation at 10.6 μm very strongly, it is likely that the laser-absorbing part of the SF_6 plume is considerably larger than the visually determined 2 to 3 mm. Here it must be noted that even use of an adjusted plume diameter with a sharp ambient air-SF_6 plume boundary is not entirely satisfactory because the actual distance from the region of radiant heating to the plume's axis, and the thickness of this heated region both depend on unmodeled (and unmeasured) features of the spatial distribution of SF_6.

It might be possible to improve upon Eq. (12) using the exact velocity field solution for an axisymmetric laminar jet [see Squire (1951) or Landau and Lifshitz (1959)]. However, matching this point-source solution to the experimental configuration while accounting for gas-mixture-dependent viscosity, diffusivity, and density, as well as the difference between fluid velocity (a vector) and photoactive-gas mole fraction (a scalar), were beyond the scope of this investigation.

It is also interesting to compare the model to the experiments in the frequency domain. The sound pressure level spectrum (in dB re 20 μPa) of the measurements and the model for repeated scans of the laser beam through the SF_6 plume at a scan rate of 2380 Hz is shown in Fig. 5(b). Again the effective plume diameter and mole fraction within the plume are set to 7.2 mm and 0.194, respectively. The open circles correspond to the spectral amplitudes of the model at 2380 Hz and its harmonics. The spiky curve in Fig. 5(b) is the sound pressure level spectrum for the measurements from an experimental time series lasting 0.21 s that encompasses ~500 photoacoustic pulses. Laboratory noise in the measurements, which is not present in the model calculations, is evident at frequencies from 3 to 20 kHz.

As expected from the prior time-domain comparisons, there are differences in the spectra as well. The model underpredicts the sound pressure level at the first four harmonics but this problem is likely the result of coherent noise from the rotating mirror and the unintended effects of reflections of the photoacoustic sound from the walls, floor, and ceiling of the laboratory. From 20 kHz up to 120 kHz or so, the agreement between the model and measurements is generally within 3 dB or so with the model consistently being higher. The additional negative acoustic pressure seen in the time domain results of the model is manifested in the frequency domain as an overprediction of the sound pressure levels. The model's tendency for overprediction might be expected because the air value (not the SF_6 value) was used in the $\gamma-1$ prefactor in Eq. (12). Adjusting this factor for the actual (but unfortunately unknown) gas composition where the photoacoustic sound was generated would likely improve the model-experiment match.

The model more seriously overpredicts the sound pressure level at frequencies above 120 kHz, but this is the region where unmodeled microphone sensitivity loss and acoustic absorption by the air are more prevalent. Between 100 and 200 kHz the microphone response falls off at approximately 20 dB per octave and acoustic absorption in air increases from 0.4 to 1.4 dB. At 150 kHz the microphone response is ~5 dB down from its passband level. When these mitigating factors are considered, the model experiment match is better than that shown in Fig. 5(b) at the highest frequencies measured.

Overall, there is excellent agreement in the general shape of the three-lobe spectral structure seen in the envelope for both the model and experiments. However, the model underpredicts the values in the troughs between the lobes. This is due to the SF_6 gas plume in the model having a sharp interface separating it from the surrounding air where as the SF_6 gas plume in the experiments diffuses into the surrounding air to produce a smooth $X_{SF_6}(x,y,z)$ spatial distribution.

Overall, the photoacoustic sound source model embodied in Eq. (12) may be able to quantitatively predict photoacoustic sound from noncompact photoactive gas clouds when the gas cloud size, the mole fraction distribution within the gas cloud, and any bulk motion or diffusion of the photoactive gas are known. In the study presented here, gas cloud parameters drawn from shadowgraph imagery of a simple cylindrical plume model were insufficient for this task. However, when gas cloud parameters within the photoacoustic model are fitted to the measurements, the model results match the initial compression, pulse duration, and general pulse shape measured in the experiments. Incorporation of more accurate photoactive gas cloud characteristics would presumably further reduce the remaining model-measurement differences.

IV. ACOUSTICALLY COMPACT GAS CLOUDS

An acoustically compact gas cloud is a cloud for which either the radiation absorption path length of the laser beam or the characteristic length L that describes the cloud/laser interaction region is much smaller than the acoustic wavelengths of interest (i.e., $kL \ll 1$). Yönak and Dowling (1999) showed how a half-concentration radius could be found for a point source of photoactive gas in air by solving the radially symmetric species conservation equations in a semi-infinite domain along with the radially symmetric Fick's law of dif-

fusion with constant binary mass-diffusivity. For a point source of SF_6 with a volume flux of $1.19\times10^{-5}\,\mathrm{cm}^3/\mathrm{s}$, this half-concentration radius is approximately $3\times10^{-7}\,\mathrm{m}$. Using a frequency of 100 kHz and the speed of sound in air at room temperature, this yields a kL value of $2\pi(10^5\,\mathrm{Hz})(3\times10^{-7}\,\mathrm{m})/(345\,\mathrm{m/s})=5.5\times10^{-4}$, clearly indicating a compact photoacoustic source in this case. Spatially confined photoacoustic sound sources of this type are important in leak testing.

The experimental geometry used to analyze the photoacoustic sound generated by an acoustically compact gas cloud is essentially the same as that shown in Fig. 2 except the small photoactive gas cloud with effective radius r_l is formed at the origin of coordinates on a flat surface lying in the $x-y$ plane. The ambient temperature used to calculate the speed of sound was measured to be 23.1 °C using a T-type thermocouple. The scan speed of the laser beam through the gas cloud was $a=1267\,\mathrm{m/s}$ using a 12-sided scanner placed 0.508 m away from the plate and scanning though a 60° angle at a rate of 2380 Hz. The laser power measured at the gas cloud location was 11 W. The laser beam diameter was determined from burn patterns at the gas cloud location to be approximately 8 mm. As before, in the modeling, the air is nonabsorbing and the specification of an effective gas cloud radius is intended to partially correct for the lack of a sophisticated diffusion model for the photoactive gas cloud. Unfortunately, the point-source gas-cloud model of Yönak and Dowling (1999) cannot be used in (6) because it predicts that the mole fraction of photoactive gas will decrease inversely with distance from the point source at long distances from the point source. This behavior leads to a fictitious logarithmic singularity in the evaluation of Eq. (2). In reality, even the weakest air currents destroy the inverse distance relationship and prevent the singularity from occurring.

The compact-cloud model development is similar to that in the prior section except that here it is assumed that the tiny gas cloud sees a spatially uniform laser intensity that is only modulated in time. For an effective photoactive gas cloud radius of 0.2 mm and a laser scan speed of 1300 m/s the effects of different portions of the gas cloud being illuminated at different times should only be felt at frequencies approaching $(1300\ \mathrm{m/s})/(0.0002\mathrm{m})\approx6.5$ MHz, well above the roll-off frequency for the microphones used in this study. Hence, the following form for the laser beam intensity I_z inside of the gas cloud (i.e., $x^2+y^2+z^2\leqslant r_l^2$) can be used:

$$I_z=\frac{8P}{\pi d_b^2}\exp\left(\frac{-8a^2t^2}{d_b^2}\right)\exp\{-\beta X_{\mathrm{SF}_6}(\sqrt{r_l^2-x^2-y^2}-z)\}. \tag{13}$$

As seen before in Eq. (8), the laser beam is unattenuated outside the gas cloud, so the photoactive gas heating term becomes

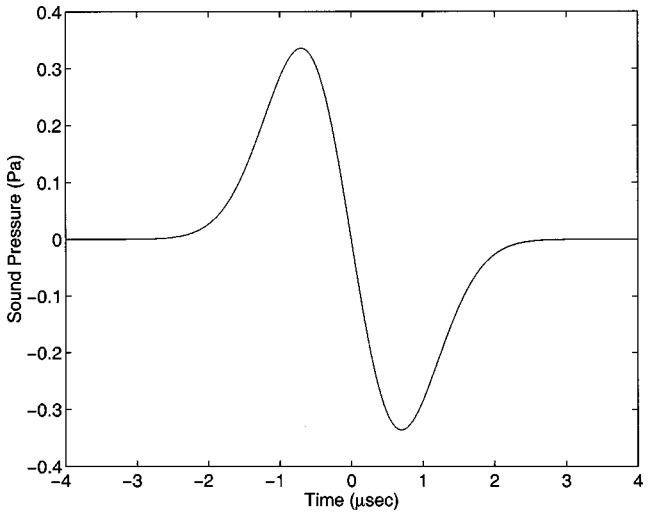


FIG. 6. Typical model results for one sweep of the CO_2-laser beam through a tiny SF_6 gas cloud. The results here are for a field point located at $r=5.1$ cm for a laser power of 11 W and a laser beam $1/e^2$ diameter of 5.5 mm. In the model, the effective radius of the cloud was set to 0.12 mm and the mole fraction of SF_6 within the cloud was set to unity. The scan speed of the beam through the gas cloud is 1963 m/s. It should be noted that unlike the experiments no bandwidth limits were placed on this model result.

$$H(\vec{x},t)=\begin{cases}0 & \text{for } x^2+y^2+z^2>r_l^2,\\ \beta X_{\mathrm{SF}_6}\dfrac{8P}{\pi d_b^2}\exp\left(\dfrac{-8a^2t^2}{d_b^2}\right) & \\ \quad\times\exp\{-\beta X_{\mathrm{SF}_6}(\sqrt{r_l^2-x^2-y^2}-z)\} & \\ \quad\text{for } x^2+y^2+z^2\leqslant r_l^2 & \end{cases}, \tag{14}$$

and its time derivative is simply

$$\frac{\partial}{\partial t}H(\vec{x},t)=\frac{-16a^2t}{d_b^2}H(\vec{x},t). \tag{15}$$

When Eqs. (14) and (15) are combined with Eqs. (2) and (3) and the far-field approximations, the integrations in Eq. (2) can be performed analytically assuming the tiny gas cloud resides on a large flat surface (i.e., it is an acoustically baffled monopole source):

$$p(r,t)=\frac{-128\ Pa^2(\gamma-1)\tau}{d_b^4\pi c^2 r}\exp\left(\frac{-8a^2\tau^2}{d_b^2}\right)\times\left\{\frac{r_l^2}{2}+\left(\frac{r_l}{\beta X_{\mathrm{SF}_6}}+\frac{1}{\beta^2X_{\mathrm{SF}_6}^2}\right)e^{-\beta X_{\mathrm{SF}_6}r_l}+\frac{1}{\beta^2X_{\mathrm{SF}_6}^2}\right\}. \tag{16}$$

Here $\tau=t-r/c$ is the retarded time, which includes the delay between when the acoustic wave is generated and when it reaches the observer at the field point located a distance r from the origin. In addition, Eq. (16) does not include a laser power adjustment for laser beam reflections from the test plate surface because most of the laser energy is predicted to be absorbed by the cloud before reaching the plate when $X_{\mathrm{SF}_6}=1$. Example results from Eq. (16) are shown in Fig. 6 for one sweep of the laser beam through the gas cloud. The

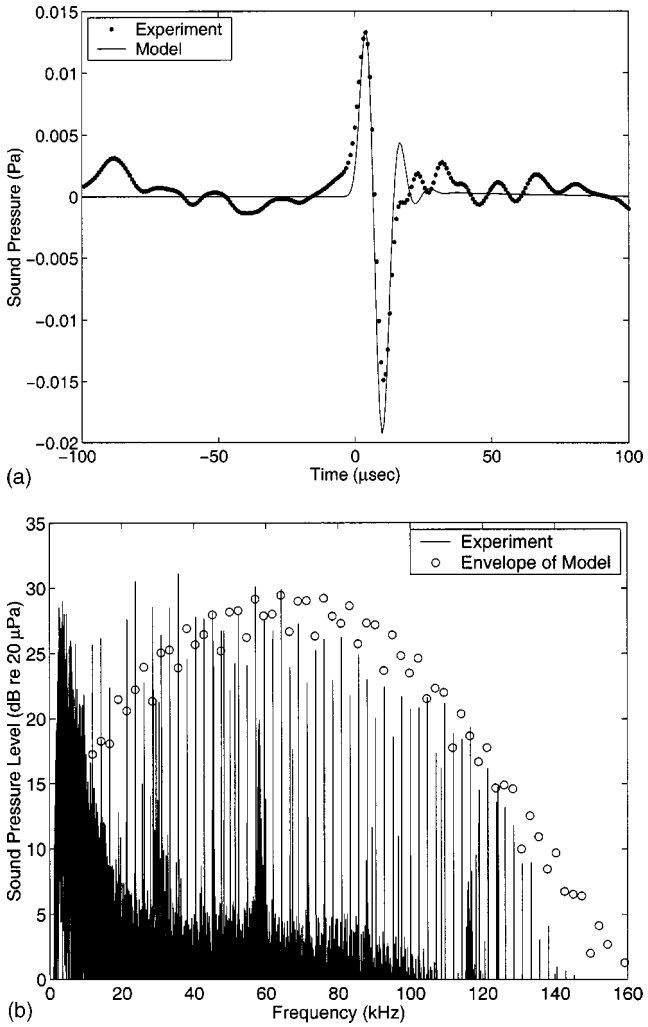


FIG. 7. A comparison of experimental measurements of photoacoustic sound pressure from a small gas cloud to model results in (a) the time domain and (b) the frequency domain. In the model, the effective cloud radius was set to 0.10 mm and the mole fraction of the gas within the cloud was set to unity. A coherent average of ~500 pulses was used to generate the experimental result.

waveform is merely the time derivative of the temporal Gaussian pulse. The bandwidth limits seen in the experiments were not placed on this particular result.

The compact gas cloud measurements were made with a 1.19×10^{-5} cm^3/s calibrated SF_6 source and a single microphone with the protection grid removed placed at $r=5.1$ cm ($x=0$ cm, $y=-1.3$ cm, and $z=4.9$ cm) in such a manner to keep the microphone out of the laser beam path. The nearest reflecting surface other than the baffle on which the SF_6 source was mounted was the untreated surface of an optical table located at $y=-10.2$ cm parallel to the x-z plane. The reflected wave amplitude from the optical tabletop in this case is about 26% and the reflected pulse was calculated to arrive at the microphone at the same time as the direct-path photoacoustic sound generated from the next laser sweep. Thus, measured pressures have been divided by 1.26 to correct for the tabletop reflection.

A comparison between experimental measurements and model results for a single sweep of the laser beam through the gas cloud is shown in Fig. 7(a). Here, the experimental signal-to-noise ratio was enhanced by coherently averaging ~500 photoacoustic pulses to construct the experimental pressure time trace shown in Fig. 7(a). The fluctuations seen ahead of and after the pulse are coherent echoes and noise. For the model calculations, the effective gas cloud radius was adjusted to 0.10 mm and the mole fraction of SF_6 within the gas cloud was assumed to be unity. All other parameters were identical to those listed above. When comparing Figs. 6 and 7(a), the bandpass filtering effects of the measurement system are apparent when looking at the predicted signal pulse, which has a duration about five times greater than its unfiltered counterpart in Fig. 6. The measured signal amplitude from the tiny gas cloud is considerably smaller than that which was measured for the SF_6 plume described in the prior section.

In Fig. 7(a), the model with fitted parameters matches the overall pulse shape and amplitude of the measured signal pulse but two differences are apparent. First, the model slightly underpredicts the measured pulse width. The pulse width and frequency content of the photoacoustic signal in the case of tiny gas clouds is strongly influenced by the laser beam intensity profile. If the laser beam had a broader profile instead of Gaussian, this would tend to favor lower frequencies producing a wider signal pulse than would be expected. Laser output that includes transverse modes other than TEM_{00} could lead to the laser beam having such a non-Gaussian tophat intensity profile. The laser manufacturer states that the beam has an approximately Gaussian intensity profile after traveling 0.6 m or more. The travel distance of the laser beam to the leak source was 0.79 m (limited by the width of the optical table). Hence, some minor deviation of the laser beam intensity profile from Gaussian is possible. Equipment to perform detailed laser-beam profiling of the invisible CO_2-laser beam was not available for this study. And second, the model also underpredicts the magnitude of the negative portion of the signal pulse by approximately 2 dB. However, this error corresponds to approximately a half-dozen-count difference on the data acquisition A/D board and lies well within the combined accuracy limits of the experiment and the model.

A frequency domain comparison of model and experiment is shown in Fig. 7(b) which provides the sound pressure level spectrum of the model result for repeated scans of the laser beam through the gas cloud at a scan rate of 2380 Hz. Ambient noise in the measurements, which is not present in the model calculations, is clearly evident at the frequencies from 3 to 30 kHz. In fact, noise dominates the measured spectrum between 3 and 10 kHz, making comparisons between the model and experiment impossible at the first three signal harmonics. The model significantly underpredicts the sound pressure level at harmonics 4, 5, 6, 8, 9, 11, and 14 and overpredicts at harmonics 7 and 10. The mismatch seen at these signal frequencies is not surprising since the lower frequencies are influenced by coherent reflections from other parts of the experiment, and by correlated noise from the laser beam interacting with the aluminum test plate. Yönak and Dowling (1999, 2002) showed that the correlated background noise level from the laser interacting with the aluminum baffle was in some cases 5 to 10 dB above the ambient

noise level, a phenomenon that may account for nearly all the model-measurement mismatch seen at the lower signal frequencies. As for the noncompact source, the model overpredicts the sound pressure level at frequencies above 130 kHz because it does not include the frequency response of the microphone or acoustic absorption by the air. Only the frequency responses of the known analog filtering devices were taken into account. Still, the agreement between the model and experiments is reasonably good to about 130 kHz and any quantitative accounting for the microphone response and sound absorption at frequencies higher than 130 kHz would further increase the model and experiment agreement.

Overall, within the bandwidth of about 30–130 kHz the model and experiment agree to within approximately 3 to 5 dB, with the model consistently—once again—being higher. The slight underprediction of the measured pulse-width by the model seen in the time domain results favors the higher frequencies and is seen in the overprediction of the sound pressure levels in the frequency domain above approximately 60 kHz. Again, improved laser beam profile knowledge, the use of a realistic gas cloud model, and the possible inclusion of species diffusion and heat conduction would improve the agreement between the model and the experiment.

As embodied in Eq. (16), the photoacoustic source model developed here can potentially predict photoacoustic sound from acoustically compact gas clouds as long as there is *a priori* knowledge of $X_{SF_6}(x,y,z)$. For the present effort, an effective cloud radius of 0.10 mm was chosen to match the amplitude of the experiment. This approximately corresponds to the 0.1% concentration radius for a steady-state mole fraction distribution emanating from a $1.19\times10^{-5}\,\mathrm{cm}^3/\mathrm{s}\,SF_6$ point source (see Yönak and Dowling, 1999). Again the strong absorption of CO_2-laser radiation by the low mole fraction portions of the SF_6 gas cloud deduced while modeling the larger SF_6 gas clouds is found here as well. Unfortunately, the $1.19\times10^{-5}\,\mathrm{cm}^3/\mathrm{s}$ (approximately 1 cm^3 per day) SF_6 source did not release enough SF_6 gas to yield any shadow images.

V. SUMMARY AND CONCLUSIONS

This paper has presented a simplified photoacoustic sound generation model that strives for quantitative accuracy, and two comparisons between this model and photoacoustic experiments conducted in open environments. The model is based on first principles and extends prior work to include Lambert–Beer law laser-radiation absorption by a photoactive gas, and a realistic laser-beam intensity distribution. Although the model does not include transport of heat or molecular species, or wave propagation in sound-absorbing nonuniform media, it does capture noncompact gas cloud photoacoustic features which are absent or unexplained in previous studies. This model was applied to acoustically noncompact and compact gas clouds having presumed simple geometries, and comparisons were made with experimental results in both the time and frequency domains.

Three conclusions can be drawn from this research effort. First of all, the theoretical model presented above is likely to be able to predict photoacoustic sound pulses in open environments when the irradiated photoactive gas cloud characteristics are known. The reasonable model-measurement matches shown in Figs. 5 and 7 are the primary support for this conclusion. Unfortunately, the requisite fitting of the photoactive gas cloud parameters to achieve these matches prevents a stronger statement about the model's accuracy and predictive capabilities.

Second, the fundamental limit to the accuracy of this photoacoustic sound generation model appears to be primarily set by how well the characteristics of the irradiated photoactive gas cloud are known. In this study, adjustment of the idealized gas-cloud characteristics used in the model did improve the match between photoacoustic theory and measurements, and these adjustments are aligned with the anticipated effects of photoactive gas diffusion, a phenomenon not included in the current photoacoustic signal generation model. Here, it was hypothesized that only the very outer fringe of the laminar SF_6 plume contributes to the generation of the photoacoustic sound because SF_6 absorbs CO_2-laser radiation so strongly. Thus, geometrically simple gas cloud models should be larger and have a lower mole fraction than optical measurements or steady-state diffusive models would suggest. However, further improvements in the present photoacoustic model would seem to require more sophisticated ways of handling photoactive gas diffusion. Simple optical measurements (see Fig. 2) or steady-state diffusion models (see Yönak and Dowling, 1999) have proven inadequate. Interestingly, the influence of the various gas-cloud characteristics is not entirely intertwined within the present photoacoustic model. For example, the model-measurement pulse-duration match seen on Fig. 5 was achieved by adjusting only the diameter of the idealized plume. Variations in the idealized plume mole fraction or the absorption coefficient for the photoactive gas had comparatively little effect on the computed results. However, this separation of parametric effects could be anticipated based on the spatial-temporal super-position of unsteady-heat-driven point sources implied by Eq. (2).

And, finally, the theoretical results presented here in analytical form allow photoacoustic sound amplitudes and other pulse characteristics to be scaled for future experimental and engineering applications that exploit the photoacoustic effect.

ACKNOWLEDGMENTS

This research project was supported by Ford Motor Company, Advanced Manufacturing Technology Development. The authors wish to thankfully acknowledge the careful work of the reviewers in preparing the final version of this manuscript.

Bell, A. G. (**1880**). "On the production and reproduction of sound by light," Am. J. Sci. **20**, 305–324.

Bell, A. G. (**1881**). "Upon the Production of Sound by Radiant Energy," Philos. Mag. **11**, 510–528.

Braker, W., and Mossman, A. L. (**1980**). *Matheson Gas Data Book*, 6th ed. (Matheson, NJ), pp. 649–654.

Brassington, D. J. (**1982**). "Photo-acoustic Detection and Ranging—A New Technique for the Remote Detection of Gases," J. Phys. D **15**, 219–228.

Bresse, L. F., and Hutchins, D. A. (**1989**). ''Transient generation by a wide thermoelastic source at a solid surface,'' J. Appl. Phys. **65**, 1441–1446.
Claspy, P. C. (**1977**). ''Infrared Optoacoustic Spectroscopy and Detection,'' in *Optoacoustic Spectroscopy and Detection*, edited by Y.-H. Pao (Academic, New York), pp. 133–166.
Cummings, E. B., Leyva, I. A., and Hornung, H. G. (**1995**). ''Laser-induced thermal acoustics (LITA) signals from finite beams,'' Appl. Opt. **34**(18), 3290–3302.
Dewey, Jr., C. F. (**1974**). ''Opto-Acoustic Spectroscopy,'' Opt. Eng. **13**, 483–488.
Dewey, Jr., C. F., Kamm, R. D., and Hackett, C. E. (**1973**). ''Acoustic Amplifier for Detection of Atmospheric Pollutants,'' Appl. Phys. Lett. **23**, 633–635.
Diebold, G. J., Beveridge, A. C., and Hamilton, T. J. (**2002**). ''The photoacoustic effect generated by an incompressible sphere,'' J. Acoust. Soc. Am. **112**, 1780–1786.
Goldan, P. D., and Goto, K. (**1974**). ''An Acoustically Resonant System for Detection of Low-Level Infrared Absorption in Atmospheric Pollutants,'' J. Appl. Phys. **45**, 4350–4355.
Hanna, M. W. (**1981**). *Quantum Mechanics in Chemistry* (Benjamin/Cummings, Menlo Park, CA).
Hecht, J. (**1992**). *The Laser Guidebook* (McGraw–Hill, Blue Ridge Summit, PA), Chap. 10.
Hoelen, C. G. A., and de Mul, F. F. M. (**1999**). ''A new theoretical approach to photoacoustic signal generation,'' J. Acoust. Soc. Am. **105**, 2685–2694.
Huang, E., Dowling, D. R., Whelan, T., and Spiesberger, J. L. (**2003**). ''High sensitivity photoacoustic leak testing,'' J. Acoust. Soc. Am. **114**, 1926–1933.
Incropera, F. P., and DeWitt, D. P. (**1996**). *Introduction to Heat Transfer* (Wiley, New York), pp. 757–761.
Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**2000**). *Fundamentals of Acoustics*, 4th ed. (Wiley, New York), pp. 177, 181, 191, 224.
Kreuzer, L. B. (**1977**). ''The Physics of Signal Generation and Detection,'' in *Optoacoustic Spectroscopy and Detection*, edited by Y.-H. Pao (Academic, New York), pp. 1–25.
Kreuzer, L. B., and Patel, C. K. N. (**1971**). ''Nitric Oxide Air Pollution: Detection by Optoacoustic Spectroscopy,'' Science **173**, 45–47.
Landau, L. D., and Lifshitz, E. M. (**1959**). *Fluid Mechanics* (Pergamon, Oxford), pp. 86–88.
Liepmann, H. W., and Roshko, A. (**1957**). *Elements of Gas Dynamics* (Wiley, New York), pp. 162–163.
Lück, K. C., and Müller, F. J. (**1977**). ''Simultaneous Determination of Temperature and OH-Concentration in Flames Using High-Resolution Laser-Absorption Spectroscopy,'' J. Quant. Spectrosc. Radiat. Transf. **17**(3), 403–409.
McQuarrie, D. A. (**1983**). *Quantum Chemistry* (University Science Books, Sausalito, CA).
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.
Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (Princeton U.P., Princeton, NJ), Chap. 12.
Okabe, H. (**1978**). *Photochemistry of Small Molecules* (Wiley, New York), pp. 41–43.
Paul, P. H., Farrow, R. L., and Danehy, P. M. (**1995**). ''Gas-phase thermal-grating contributions to four wave mixing,'' J. Opt. Soc. Am. **12**(3), 384–392.
Pierce, A. (**1989**). *Acoustics* (ASA, AIP, New York), p. 417.
Rogers, J. A., Maznev, A. A., Banet, M. J., and Nelson, K. A. (**2000**). ''Optical generation and characterization of acoustic waves in thin films: Fundamentals and Applications,'' Annu. Rev. Mater. Sci. **30**, 117–157.
Reid, R. C., Prausnitz, J. M., and Sherwood, T. K. (**1977**). *The Properties of Gases and Liquids* (McGraw–Hill, New York), pp. 553–560.
Rosencwaig, A. (**1980**). *Photoacoustics and Photoacoustic Spectroscopy* (Wiley, New York).
Schlamp, S., Cummings, E. B., and Hornung, H. G. (**1999**). ''Beam misalignments and fluid velocities in laser-induced thermal acoustics,'' Appl. Opt. **38**(27), 5724–5733.
Squire, H. B. (**1951**). ''The Round Laminar Jet,'' Q. J. Mech. Appl. Math. **4**, 321–329.
Van Wylen, G. J., Sonntag, R. E., and Borgnakke, C. (**1994**). *Fundamentals of Classical Thermodynamics* (Wiley, New York), pp. 47, 753–754.
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.
Zharov, V. P., and Symanovskii, Y. O. (**1990**). ''Laser Photoacoustic Measurements in Plane Gas Jets,'' in *Photoacoustic and Photothermal Phenomena II*, edited by J. C. Murphy, J. W. Maclachlan Spicer, L. C. Aamodt, and B. S. H. Royce (Springer-Verlag, Heidelberg), pp. 372–375.

A Wiener filter approach to the binaural reproduction of stereo sound

Sang-Myeong Kim and Semyung Wang
Mechatronics, Kwang-Ju Institute of Science and Technology, 1 Oryong-dong, Puk-gu, Kwangju 500-712, Korea

(Received 23 December 2002; accepted for publication 15 September 2003)

This paper considers authentic binaural reproduction with loudspeakers using *cross-talk cancellation*. A systematic time domain deconvolution method is presented using both stochastic and deterministic Wiener filters. This method has advantages in that the Wiener filter is inherently causal and stable so that no additional stabilization process, as generally required in the frequency domain methods, is needed. Errors involved in the authentic reproduction are classified into modeling, inverse, and repeatability errors. In addition, the reproduction error is used as the performance measure of a reproduction system. Finally, a series of monaural and binaural reproduction tests were conducted in an anechoic chamber using a PC based reproduction system with a soundcard by changing the length of the inverse filter used. Test results demonstrated that the performance was improved in proportion to the length of the filter used. The reproduction performance error and its test procedure presented here can be used for objectively assessing the performance of a subjective auditory perception test system. [DOI: 10.1121/1.1624070]

PACS numbers: 43.38.Vk, 43.60.Pt [AJZ] Pages: 3179–3188

I. INTRODUCTION

In recent years, the authentic reproduction of three-dimensional virtual sound over loudspeakers has become a reality. The whole collection of techniques in virtual sound may be referred to as *binaural technology*, which can be broadly classified into *binaural synthesis* and *binaural reproduction*. Binaural synthesis is a technique to synthesize a pair of binaural sound signals from an original monaural source signal with the provision of three-dimensional acoustic information: the distance and direction of the source with respect to the listener, and the space of the listening room.[1,2] Specifically, the sense of direction can be rendered by using the head-related acoustic information, such as HRIRs (Head Related Impulse Response) that can be obtained by either experimental[3,4] or theoretical[5] means.

Binaural reproduction (inverse filtering or deconvolution) is a signal processing technique to deliver the synthesized signal pair securely to the listener's eardrums. An electric inverse filter is generally required that can equalize the influence of the electroacoustic path through which the binaurally synthesized sound signals are being reproduced; cross-talk cancellation.[6,7] Theoretically, convolving the inverse filter with the electroacoustic path yields a perfect reproduction of the binaural sound signals at the listener's eardrums. Of interest in the work presented here is the binaural reproduction process that is applied to the stereo reproduction system using two loudspeakers assuming that a pair of perfectly synthesized binaural sound signals has been already provided.

Since the principle of cross-talk cancellation was first introduced to the audio society by Bauer,[8] there has been significant effort to develop more effective algorithms. Various algorithms are available, which can be broadly classified into the time and frequency domain design methods. The time domain design method is accurate,[9] and the frequency domain method is known to be faster, but often inaccurate.[10] The former may be suitable for offline reproduction systems in a fixed listening condition while the latter is for online systems in a variant listening condition whose change is tracked in real time by, for example, a vision-based head tracking system.[11] Offline reproduction is of interest in this research so that only the time domain design method is implemented.

The work presented here describes a theoretical and experimental investigation into authentic binaural reproduction using loudspeakers; cross-talk cancellation. After a brief review of the conventional frequency domain deconvolution method in Sec. II, an efficient time domain deconvolution method for a TITO (two-input-two-output) stereo reproduction system is presented in Sec. III using both stochastic and deterministic Wiener filters. Of particular concern is the limitation of the filter length on the reproduction performance. When FIR (finite impulse response) filters are used to implement the inverse filter, the length given is often not long enough to model its ideal version, since the filter length is strictly constrained by the processing time. The advantages of the time domain method compared with its frequency domain counterpart are discussed using a simple plant example in Sec. IV. Also, errors affecting the reproduction performance are clearly identified and an objective measure to assess the performance is given. Section V presents some monaural and binaural reproduction test results that validate the time domain method presented.

II. FREQUENCY DOMAIN DECONVOLUTION

A. Perfect binaural reproduction

A complete implementation procedure of the binaural technology is illustrated in block diagram form in Fig. 1 for

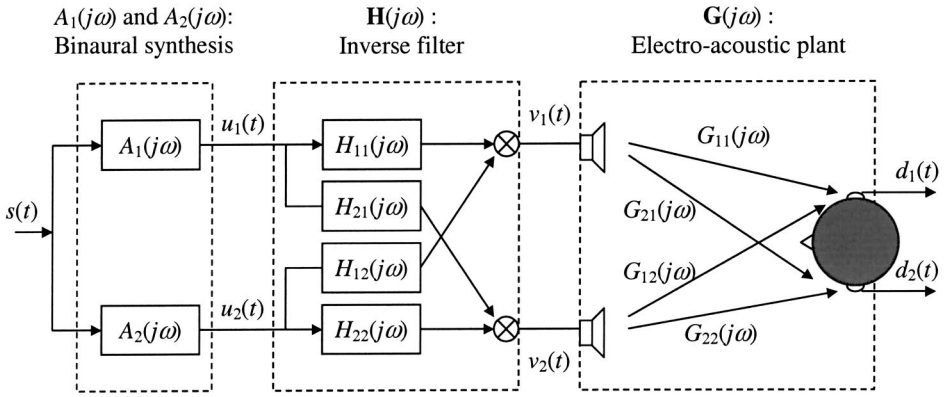


FIG. 1. Binaural technology using two loudspeakers. The monaural source signal $s(t)$ is first synthesized to produce a pair of binaural sounds $\mathbf{u}(t)=\{u_1(t)\ u_2(t)\}^T$ which then passes through the inverse filter matrix to generate a pair of loudspeaker signals $\mathbf{v}(t)=\{v_1(t)\ v_2(t)\}^T$. The signal vector $\mathbf{d}(t)=\{d_1(t)\ d_2(t)\}^T$ denotes the sound pressures at the listener's eardrums.

a stereo reproduction system that uses two loudspeakers.[1,2] A given monaural source signal $s(t)$ is first processed using the two filters, $A_1(j\omega)$ and $A_2(j\omega)$, to yield a pair of the binaurally synthesized sound signals, $u_1(t)$ and $u_2(t)$. The synthesis filters, which can be designed using HRIRs, contain the spatial acoustic information with respect to the listener's head and ears. More source signals can be added using the principle of superposition with the corresponding sets of synthesis filters, if required. It is assumed in this paper that the binaurally synthesized signals, $\mathbf{u}(t)=\{u_1(t)\ u_2(t)\}^T$ in vector form, are already given. Henceforth, we only consider the problem of binaural reproduction (cross-talk cancellation), that is to perfectly reproduce the binaural signals at the listener's eardrums $\mathbf{d}(t)=\{d_1(t)\ d_2(t)\}^T$ via the cancellation of the electroacoustic plant $\mathbf{G}(j\omega)$ using the inverse filter $\mathbf{H}(j\omega)$. The eardrum responses in the frequency domain can be written as

$$\mathbf{d}=\mathbf{GHu}, \tag{1}$$

where $\mathbf{d}=\{d_1\ d_2\}^T$, $\mathbf{u}=\{u_1\ u_2\}^T$, $\mathbf{G}=\left[\begin{smallmatrix}G_{11} & G_{12}\\ G_{21} & G_{22}\end{smallmatrix}\right]$, and $\mathbf{H}=\left[\begin{smallmatrix}H_{11} & H_{12}\\ H_{21} & H_{22}\end{smallmatrix}\right]$ in which frequency dependency of the variables is suppressed for brevity. For time domain descriptions, however, we explicitly specify the time dependency for distinction, for example, as $\mathbf{d}(t)$. Theoretically, the physically realizable (causal stable) perfect inverse filter for a given linear plant $\mathbf{G}$ can be achieved by introducing a time delay term to give[12]

$$\mathbf{H}=\mathbf{G}^{-1}e^{-j\omega\tau}, \tag{2}$$

where the time delay τ should be set large enough for the impulse responses of the inverse filters to be causal stable. Introducing the time delay term is unavoidable since the electroacoustic plants are generally nonminimum phase systems. Substituting Eq. (2) into (1) gives perfect reproduction of the binaural signals with some time delay,

$$\mathbf{d}=\mathbf{u}e^{-j\omega\tau} \quad \text{or} \quad \mathbf{d}(t)=\mathbf{u}(t-\tau), \tag{3}$$

where $\mathbf{d}(t)$ is the time domain representation of $\mathbf{d}$.

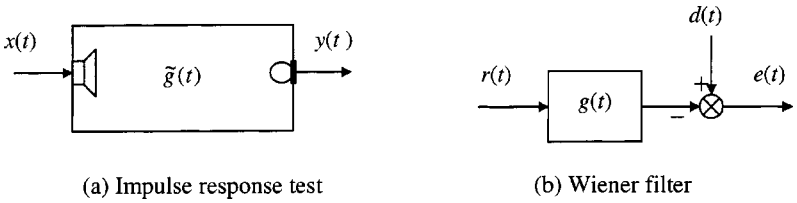
However, such a perfect reproduction is only possible in an ideal situation with some unrealistic assumptions. First, the plant $\mathbf{G}$ should be perfectly linear and noise free, and be perfectly modeled by, for example, a set of FIR filters. Second, the inversion in Eq. (2) should be nonsingular at all frequencies, and the inverse filters should be in causal stable FIR filter form. Last, the plant should be time invariant so that a perfect repetition of the binaural reproduction is guaranteed in every event. In real applications, none of them are perfectly satisfied even in an anechoic chamber so that there are some errors induced by violating each assumption. The deviation between the real plant and its linearlized model can be termed *modeling error*, the error involved in inversion is *inverse error*, and lastly the error induced by a plant change can be defined as *repeatability error*. Each of them will be discussed in detail later.

B. Robust binaural reproduction

If the inverse filters are implemented in FIR filter form, there is no guarantee that the frequency domain inverse method given in Eq. (2) yields a causal stable filter.[10] This is because of the constraint of causality as well as the limitation of the finite filter length imposed on the filters. For example, if the plant matrix $\mathbf{G}$ is singular at some frequencies, its inverse does not exist. Another problematic typical situation may be when the time domain form of the filter $\mathbf{H}$ exists but not in causal and FIR filter form within the given filter length. Thus, a robust deconvolution method is generally employed in practice as presented by Kirkeby *et al.*,[10] and is given by

$$\mathbf{H}=(\mathbf{G}^{\mathbf{H}}\mathbf{G}+\beta\mathbf{I})^{-1}\mathbf{G}^{\mathbf{H}}e^{-j\omega\tau}, \tag{4}$$

where $\mathbf{I}$ is the identity matrix and the positive number β is a robustness coefficient. The role of the coefficient β is twofold; *robustness* to the inverse error, and *robustness* to plant changes. The former is to obtain a causal stable $\mathbf{H}$ in FIR filter form since, when $\beta=0$, Eq. (4) becomes the noncausal inversion given by Eq. (2). Suppose, for example, the given filter length is too short to implement its true inverse, which is the case when $\mathbf{G}^{\mathbf{H}}\mathbf{G}$ in Eq. (4) is very small at some frequencies. It is the coefficient β that limits the amplitude of $(\mathbf{G}^{\mathbf{H}}\mathbf{G}+\beta\mathbf{I})^{-1}$ so as to prohibit its time response from exceeding the given filter length.[10] The latter is to ensure robustness to a subtle change in the plant that is actually the original role of β widely used in robust control.[13] Since β should not be too small or too large for a given plant, a time-consuming searching process for the optimal β is necessary as will be described in detail in Sec. IV. The following section presents a time domain method using the Wiener fil-



(a) Impulse response test (b) Wiener filter

FIG. 2. Impulse response identification in a SISO system; (a) impulse response test, (b) Wiener's problem for finding an optimal filter $g(t)$ that minimizes the error signal in mean square sense when the received $r(t)$ and desired $d(t)$ signals are ergodic and stationary random.

ter that is guaranteed to be causal stable; inherently robust to the inverse error.

III. TIME DOMAIN DECONVOLUTION

Deconvolution in a single channel monaural reproduction system is first considered, and then its theory is extended to a two-channel binaural reproduction system. A monaural reproduction system can be illustrated by simplifying Fig. 1, and is omitted here for brevity. The system can be regarded as an authentic sound reproduction system targeted at a single point in a listening room, which can de-reverberate the room acoustic response as well as the response of the electric equipment used.

As the first step for deconvolution, a linear model of the physical plant is required. Figure 2(a) shows the input–output relationship of the real physical plant $\tilde{g}(t)$, where $x(t)$ and $y(t)$ are the test input and the measured output signals, respectively. The input–output relationship can be written as

$$y(t)=g(t)*x(t)+m(t), \tag{5}$$

where the sign $*$ denotes the linear convolution operator, $g(t)$ denotes a linearlized model of the real physical plant $\tilde{g}(t)$, and $m(t)$ is the modeling error. To obtain the optimal model for $g(t)$ that minimizes the mean square of $m(t)$, the Wiener filter can be used. The Wiener filter $g(t)$ illustrated in block diagram form in Fig. 2(b) is the physically realizable (*causal stable*) optimal filter that minimizes the error $e(t)$ in the mean square sense when the desired $d(t)$ and received $r(t)$ signals are ergodic and stationary random.[6,14,15] Thus, by setting $r(t)=x(t)$ and $d(t)=y(t)$ in Fig. 2(b), the Wiener filter can be represented by

$$x(t)*g(t)\approx y(t) \tag{6}$$

and the filtering error here will be the plant *modeling error*, i.e., $e(t)=m(t)$. Discrete time domain analysis is necessary when the Wiener filter is implemented digitally with sampled signals, and thus a brief solution procedure is given in the Appendix.

In general, some modeling error is unavoidable due to noise present in the real physical electroacoustic plant: the acoustic background noise and the electric noise in the measurement transducers and the data acquisition front end. Furthermore, there is also an additional but serious constraint imposed on the modeling filter of the plant: the *filter length*. The filter length is strictly constrained by computation time, but there are often situations when the electroacoustic plant is too strongly reverberant to be modeled completely by a FIR filter.

The linearized model $g(t)$ can now be used for designing its inverse filter $h(t)$. As illustrated in Fig. 3, this design task is a deconvolution problem that is to perfectly reproduce the original input signal at the output by deconvolving $g(t)$ using the inverse filter $h(t)$.[6] In the diagram, a time delay in $\delta(t-\tau)$ is introduced in order to have a causal inverse filter $h(t)$,[12] and unlike Fig. 1 the locations of the plant model $g(t)$ and the inverse filter $h(t)$ are exchanged since they are linear. It now turns out again to be a Wiener filtering problem, as indicated by a dashed box in Fig. 3. Note here the *deterministic* approach using an impulse $\delta(t)$ instead of white noise $w(t)$ is used as the input since they are equivalent as far as the Wiener filter is concerned as discussed in detail by Brennan and Kim.[15] Advantages of this deterministic approach in the discrete time implementation are its fast calculation time and high accuracy. The relationship between the desired and input signals can be written as

$$\delta(t-\tau)=g(t)*h(t)*\delta(t)+n(t), \tag{7}$$

where $n(t)$ can be defined as the *inverse error* which is generally caused by both the limitation of the filter length used and the constraint of causality imposed on the inverse filter so as to be physically realizable. Once the time delay τ is large enough for the inversion to be causal, it is the filter length which limits the deconvolution performance.

Rewriting Eq. (7) in Wiener filter form gives

$$g(t)*h(t)\approx\delta(t-\tau). \tag{8}$$

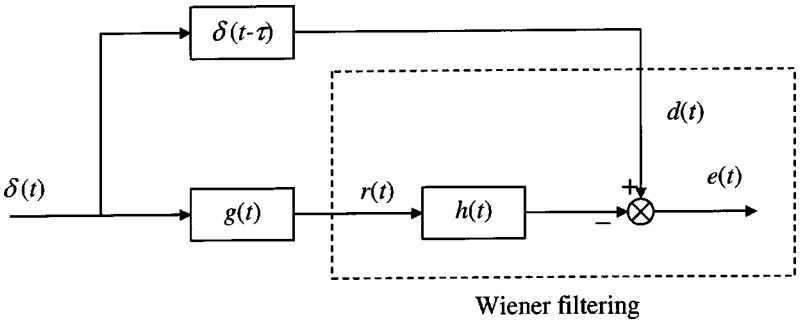
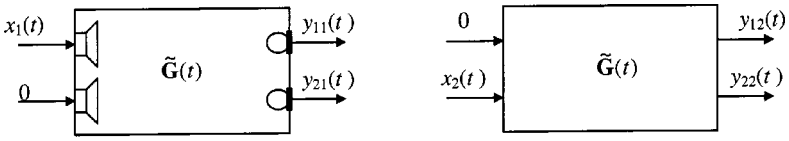


FIG. 3. Deconvolution of $g(t)$ using the inverse filter $h(t)$.



(a) Channel 1 input (b) Channel 2 input

FIG. 4. Input–output relationships of the TITO system $\tilde{\mathbf{G}}(t)$; (a) input at channel 1 and (b) input at channel 2.

The optimal inverse filter $h(t)$ can be obtained using the deterministic Wiener filter approach by setting $r(t)=g(t)$ and $d(t)=\delta(t-\tau)$.[15]

Now, consider the binaural reproduction system using two microphones and two loudspeakers as shown in Fig. 4. The plant is now a TITO (two-input-two-output) system, but its model can be obtained simply by extending its single channel case given in Eq. (5). Suppose a test signal $x_1(t)$ is sent to the loudspeaker 1 only, as shown in Fig. 4(a), then the corresponding outputs at the microphones are $y_{11}(t)$ and $y_{21}(t)$, where the first and second subscripts correspond to the output and input channel numbers, respectively. The input–output relationship can be written as

$$\mathbf{y}_1(t)=\mathbf{G}(t)*\mathbf{x}_1(t)+\mathbf{m}_1(t), \tag{9a}$$

where the optimal model of the real physical (2×2) plant is $\mathbf{G}(t)=\left[\begin{smallmatrix} g_{11}(t) & g_{12}(t) \\ g_{21}(t) & g_{22}(t) \end{smallmatrix}\right]$, the input vector is $\mathbf{x}_1(t)=\{x_1(t)\ 0\}^T$, and its corresponding output and plant modeling error vectors are, respectively, $\mathbf{y}_1(t)=\{y_{11}(t)\ y_{21}(t)\}^T$ and $\mathbf{m}_1(t)=\{m_{11}(t)\ m_{21}(t)\}^T$. Likewise, when a test signal is sent to the channel 2 loudspeaker as shown in Fig. 4(b), it gives

$$\mathbf{y}_2(t)=\mathbf{G}(t)*\mathbf{x}_2(t)+\mathbf{m}_2(t), \tag{9b}$$

where the input signal vector is now $\mathbf{x}_2(t)=\{0\ x_2(t)\}^T$, and the others can be similarly defined. If the same random noise $w(t)$ is used for both of the test input signals, i.e., $x_1(t)=x_2(t)=w(t)$, combining Eqs. (9a) and (9b) gives

$$\mathbf{Y}(t)=\mathbf{G}(t)*w(t)+\mathbf{M}(t), \tag{10}$$

where $\mathbf{Y}(t)=\left[\begin{smallmatrix} y_{11}(t) & y_{12}(t) \\ y_{21}(t) & y_{22}(t) \end{smallmatrix}\right]$, $\mathbf{G}(t)=\left[\begin{smallmatrix} g_{11}(t) & g_{12}(t) \\ g_{21}(t) & g_{22}(t) \end{smallmatrix}\right]$, and $\mathbf{M}(t)=\left[\begin{smallmatrix} m_{11}(t) & m_{12}(t) \\ m_{21}(t) & m_{22}(t) \end{smallmatrix}\right]$ is the modeling error matrix. Rewriting Eq. (10) in Wiener filter form gives

$$w(t)*\mathbf{G}(t)\approx\mathbf{Y}(t). \tag{11}$$

Equation (11) is a compact description of four separate SISO (single-input-single-output) Wiener filter problems. For example, the Wiener filter $g_{11}(t)$ can be obtained by setting the received and desired signals as $r(t)=w(t)$ and $d(t)=y_{11}(t)$, and the corresponding error becomes $e(t)=m_{11}(t)$. The others can be similarly calculated.

Deconvolution of this TITO system is very similar to that of the SISO system given in Eq. (7), and the corresponding extension can be written as

$$\mathbf{I}\delta(t-\tau)=\mathbf{H}(t)*\mathbf{G}(t)*\mathbf{I}\delta(t)+\mathbf{N}(t), \tag{12}$$

where the impulse response matrices of the inverse filter and plant are, respectively, $\mathbf{H}(t)=\left[\begin{smallmatrix} h_{11}(t) & h_{12}(t) \\ h_{21}(t) & h_{22}(t) \end{smallmatrix}\right]$ and $\mathbf{G}(t)=\left[\begin{smallmatrix} g_{11}(t) & g_{12}(t) \\ g_{21}(t) & g_{22}(t) \end{smallmatrix}\right]$, $\mathbf{I}=\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$, and $\mathrm{N}(t)=\left[\begin{smallmatrix} n_{11}(t) & n_{12}(t) \\ n_{21}(t) & n_{22}(t) \end{smallmatrix}\right]$ is the inverse error matrix. It can be seen that, when $\mathbf{H}(t)*\mathbf{G}(t)=\mathbf{I}\delta(t-\tau)$, a perfect deconvolution is achieved. Since $\mathbf{A}\cdot\mathrm{adj}(\mathbf{A})=\Delta\mathbf{I}$ where $\mathbf{I}$ is an identity matrix, and adj($\mathbf{A}$) and Δ are, respectively, the adjoint (the transposed matrix of co-factors) and determinant of a square invertible matrix $\mathbf{A}$,[16] post-convolving the adjoint form of $\mathbf{G}(t)$ to both sides with ignoring $\mathbf{N}(t)$ gives

$$\Delta(t)*\mathbf{H}(t)\approx\begin{bmatrix} g_{22}(t) & -g_{12}(t) \\ -g_{21}(t) & g_{11}(t) \end{bmatrix}*\delta(t-\tau), \tag{13}$$

where $\Delta(t)=(g_{11}(t)*g_{22}(t)-g_{12}(t)*g_{21}(t))$ and the (2×2) matrix on the right correspond to the determinant and the adjoint of $\mathbf{G}(t)$ in a conceptual sense, respectively. One should note that post-convolving the adjoint matrix form to both sides does not raise a causality issue since it is composed of causal plant responses. One should also notice that Eq. (13) is the same form as Eq. (11) that is accepted to be the general Wiener filtering process for multiple-channel systems. Systems with more than two channels can be similarly formulated, but its practicality may be limited in some applications due to computational complexity of the determinant and the adjoint matrix form. Note also the sign $*$ is the linear convolution operator, not circular, so that care should be taken in calculating the convolution in the digital domain. Thus, if each plant $g_{ij}(t)$ is of length I, then $\Delta(t)$ is of length $2I-1$. Each of the elements in $\mathbf{H}(t)$ can be optimally identified by using the SISO Wiener filter, and thus the deterministic Wiener filtering is required to be done four times. For example, the Wiener filter $h_{11}(t)$ is obtained by setting $r(t)=\Delta(t)$ and $d(t)=g_{22}(t)*\delta(t-\tau)$. The others can be obtained separately but similarly. Note also that the $\Delta(t)$ serves as the common received signal for all four Wiener filters so that its zeros can act as the common poles of the four Wiener filters. It should be emphasized that, although it deals with a TITO system, the formulations for modeling and deconvolution, given in Eqs. (11) and (13), respectively, require a multiple number of the simple SISO Wiener filtering. This can be very advantageous in terms of clarity and calculation time compared with others, for example, see Refs. 9 and 17 that uses only a single SISO Wiener filter but with an autocorrelation matrix that is four times larger.

IV. ERROR ANALYSIS

A. Reproduction performance error

The authentic reproduction procedure may be divided into three steps: modeling, inversion, and reproduction. We may define the error of each step using the normalized mean square error as is generally used in the Wiener filter. First, consider the mono (1 channel) system in Fig. 2 whose mathematical model is given by Eq. (5). The normalized mean square of the modeling error (the modeling performance error), is given by[6,14]

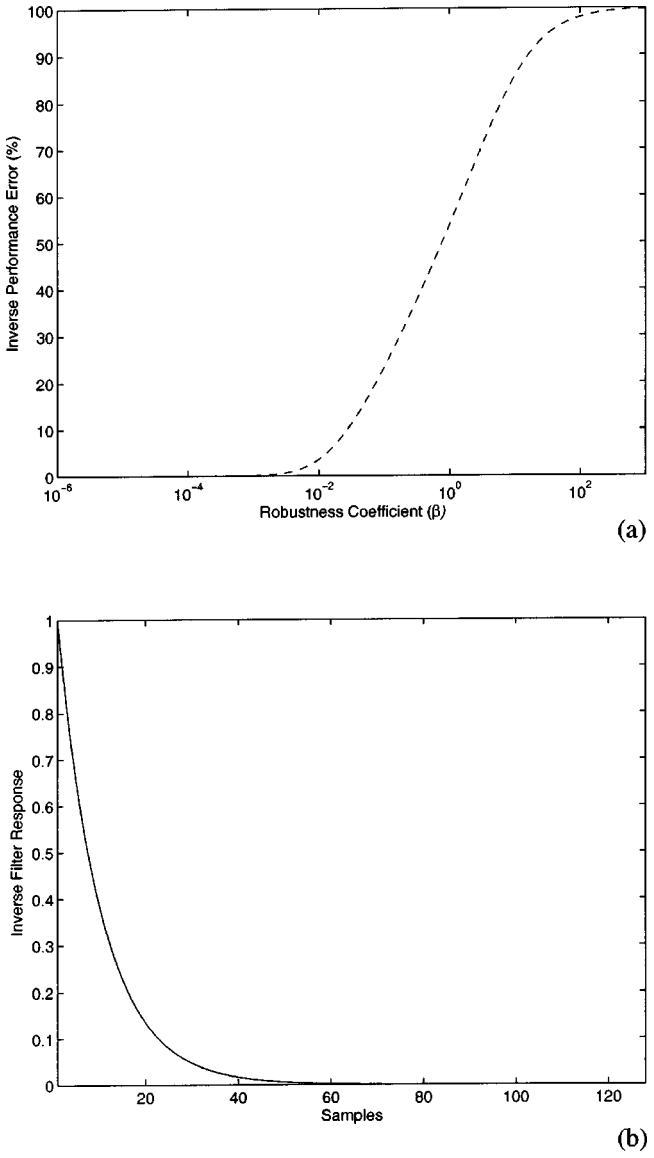


FIG. 5. The plant with $a=0.9$; (a) the inverse performance errors of the time (solid line) and frequency (dashed) domain methods according to the robustness coefficient, (b) the optimal inverse filters of the time (solid) and frequency (dashed) domain methods are compared with the true inverse (dotted) given in Eq. (21).

$$J_m(\%)=\frac{E[m^2(t)]}{E[y^2(t)]}\times 100, \tag{14}$$

where $E[o^2]$ denotes the variance, $m(t)$ is the modeling error, and $y(t)$ is the output signal at the microphone. The inverse performance error can be written as

$$J_n(\%)=\frac{E[n^2(t)]}{E[\delta^2(t-\tau)]}\times 100 \tag{15a}$$

or

$$J_n(\%)=\sum_{i=0}^{I-1} n^2[i]\times 100, \tag{15b}$$

where $n(t)$ is the inverse error and $\delta(t-\tau)$ is a delayed impulse, and $n[i]$ of length I is the inverse error in the discrete time domain. Finally, the reproduction performance error can be defined as

$$J(\%)=\frac{E[e^2(t)]}{E[u^2(t-\tau)]}\times 100, \tag{16}$$

where the reproduction error is the difference between the measured and desired signals i.e., $e(t)=d(t)-u(t-\tau)$.

The modeling, inverse, and reproduction performance errors for the stereo (2 channel) case illustrated in Fig. 4 can be similarly defined by extending those for the single channel case, and they can be written as

$$J_m(\%)=\frac{E[m_{11}^2(t)]+E[m_{12}^2(t)]+E[m_{21}^2(t)]+E[m_{22}^2(t)]}{E[y_{11}^2(t)]+E[y_{12}^2(t)]+E[y_{21}^2(t)]+E[y_{22}^2(t)]}\times 100, \tag{17}$$

$$J_n(\%)=\frac{E[n_{11}^2(t)]+E[n_{12}^2(t)]+E[n_{21}^2(t)]+E[n_{22}^2(t)]}{E[\delta_1^2(t-\tau)]+E[\delta_2^2(t-\tau)]}\times 100, \tag{18}$$

$$J(\%)=\frac{E[e_1^2(t)]+E[e_2^2(t)]}{E[u_1^2(t-\tau)]+E[u_2^2(t-\tau)]}\times 100, \tag{19}$$

where the desired signals and the corresponding errors are given in the denominator and numerator of each equation, respectively. The reproduction error is comprised of modeling, inverse, and repeatability errors, and thus the reproduction performance error can be used as the final objective performance measure of the authentic reproduction system.

B. Example study

The frequency and time domain methods for inverse filter design are compared in terms of the inverse error using a simple example. The example plant considered is

$$\tilde{G}(z)=1-az^{-1}, \text{ where } |a|<1, \tag{20}$$

which is a minimum phase system whose pole and zero are inside a unit circle, and thus its causal stable inverse is simply the inverse of Eq. (20) as given by

$$\tilde{H}(z)=\frac{1}{1-az^{-1}}. \tag{21}$$

Note the plant $\tilde{G}(z)$ has nonzero values at its first two samples while the length of the inverse filter $\tilde{H}(z)$ is variant depending on the constant a.

The filter lengths for both the plant and the inverse filter were set to be 128, and inverse filters were designed for $a=0.9$ and $a=0.999$, respectively. Both the frequency and time domain deconvolution methods were applied. Since the plant has nonzero values for the first two samples only, no modeling error is involved so that the inverse performance error is of particular concern. The cyclic shift algorithm was used for the frequency domain deconvolution method given by Eq. (4) by changing the robustness coefficient.[10] However, although it is possible,[13,17] no plant robustness was considered in the time domain method as given by Eq. (8). In addition, no time delay was needed in both methods, i.e., $\tau=0$ since the plant is a minimum phase system.

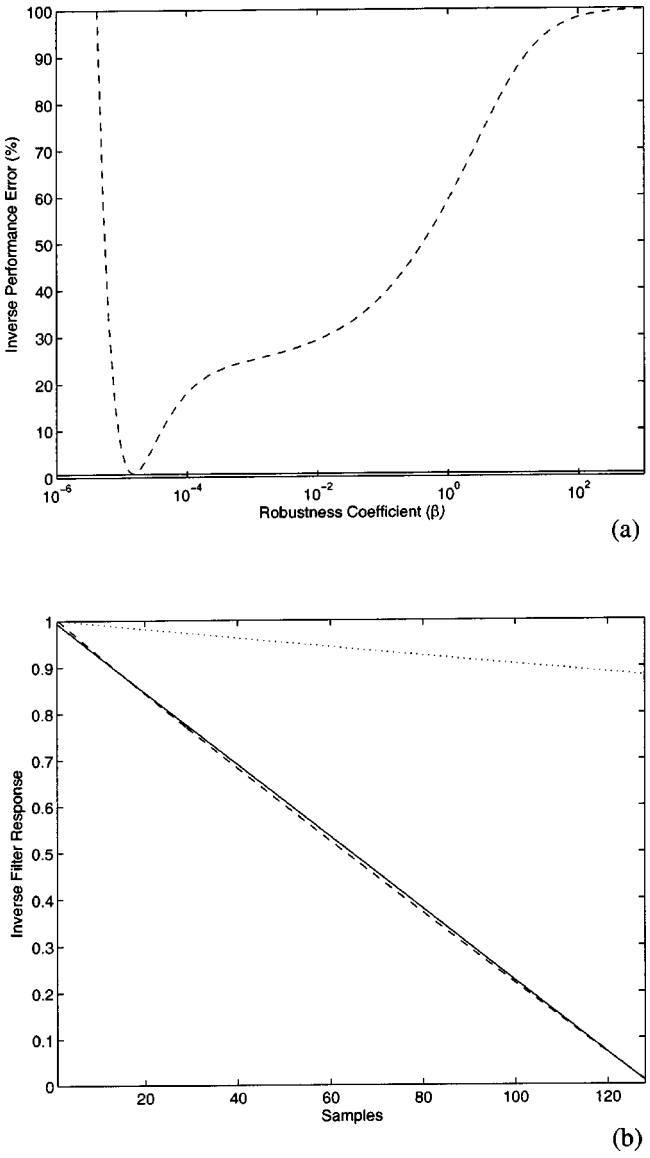


FIG. 6. The plant with $a=0.999$; refer to Fig. 5 for the captions of (a) and (b).

When $a=0.9$, the normalized mean square error of the frequency domain method (dashed lines) is shown in Fig. 5(a) against the robustness coefficient β used. A perfect deconvolution is achieved when β is very small, and as β increases, the performance is reduced and finally approaches to $J_n=100\%$ at about $\beta=10^3$. The time domain method was also applied without considering robustness since the Wiener filter is guaranteed to be causal stable. This method also offered a perfect deconvolution, and its normalized mean square error (solid line) actually plotted in Fig. 5(a) coincides with the abscissa, i.e., $J_n=0$. The optimal filter responses of the frequency and time domain methods that yield perfect deconvolutions are compared with the theoretical one. They turned out to be identical as plotted in Fig. 5(b). This example illustrates that, when the inverse filter length given as $I=128$ is long enough to accommodate the true inverse filter, both frequency and time domain methods are equivalent.

Now consider the case when $a=0.999$. In this case the given filter length of 128 is short for modeling the true inverse filter in Eq. (21). The corresponding results are shown in Fig. 6, which is in the same form as Fig. 5. The error curve of the frequency domain method (dashed lines) shown in Fig. 6(a) no longer monotonously grows in proportion to β, but has a dip at about $\beta=1.5e^{-5}$. This robustness coefficient that yields the minimum error is the optimal robustness coefficient as far as the (numerical) inverse robustness is considered. It is also shown that the frequency domain method is basically unstable when $\beta\approx 0$. The minimum error of the frequency domain method approaches to that of the time domain method (solid lines) shown in Fig. 6(a). The optimal filters of the time (solid lines) and frequency (dashed) domain methods are compared with the true inverse filter (dotted) in Fig. 6(b) where the optimal filters of each method are very similar, but not identical.

This example demonstrates that, when the frequency domain method is employed with a short filter, the optimal β for the inverse robustness should be known *a priori* to ensure plant robustness and to avoid the hazardous unstable region, i.e., ensure inverse robustness. The optimal β is plant dependent and the searching process is often very laborious. The advantage of the fast design speed of this method is true only when this optimal β has already been determined. It also demonstrates that the performance of the frequency domain method with the optimal β is similar to that of the time domain method. If an offline (nonadaptive) realization is the main concern as considered in this work, it may be more systematic and accurate to use the Wiener filter. The following section describes some authentic reproduction tests using the time domain deconvolution method only.

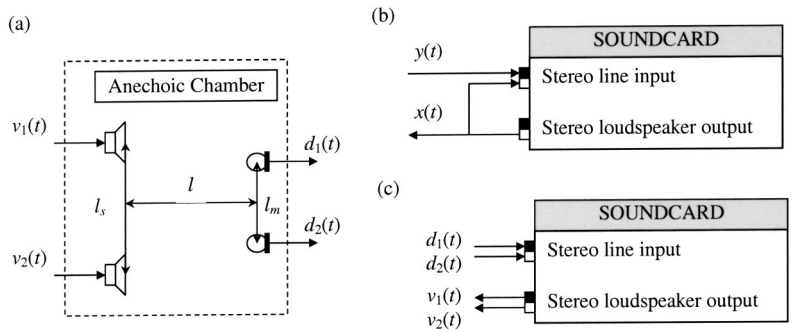


FIG. 7. Experimental setup: (a) geometric setup, (b) wire connection for impulse response identification, and (c) wire connection for binaural reproduction.

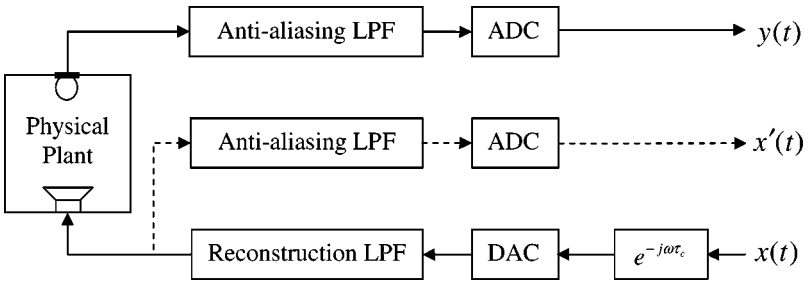


FIG. 8. The electroacoustic plant is represented by the solid line flow where $x(t)$ and $y(t)$ are the input and output signals, respectively. The computer processing time delay term $e^{-j\omega\tau_c}$ should be excluded from the plant by way of triggering.

V. BINAURAL SOUND REPRODUCTION TEST

A. Experimental setup

Binaural as well as monaural reproduction tests were conducted in an anechoic chamber, and the geometric setting for the binaural case is shown in Fig. 7(a), where the distance l between the centers of the loudspeakers and microphones was 0.5 m, and the distances between loudspeakers and microphones were 0.6 m and 18 cm, respectively. For the monaural case, the loudspeaker and the microphone used were $v_1(t)$ and $d_1(t)$, respectively. All the whole experiments including impulse response identification and binaural reproduction were conducted using a PC with a soundcard (LINX TWO®) as a front-end device for data acquisition and transmission. The loudspeakers used were ordinary two-way loudspeakers (Yamaha MSP-5), and the microphones were B&K 1/2-inch condenser microphones (Type 4189). Matlab® signal processing and data acquisition toolboxes were used for signal processing and data acquisition and transmission.

Each of the four impulse responses of the (2×2) plant matrix were first identified with the wire connection setting as illustrated in Fig. 7(b). The loudspeaker input signal $x(t)$ and the microphone output signal $y(t)$ were both monitored at the line input ports: Channel 1 for $x(t)$ and channel 2 for $y(t)$. The electroacoustic plant to be measured is more precisely represented in Fig. 8 where, as indicated as a solid line flow, the plant included the physical plant (acoustic plant, the microphone, and the loudspeaker) as well as the data transmission/acquisition module: the DAC (digital-to-analog converter) and ADC (analog-to-digital converter), and the antialiasing and reconstruction LPFs (low pass filters). In order to identify the proper plant response, it is necessary to exclude the influence of the data transmission delay $e^{-j\omega\tau_c}$ where τ_c is the computing delay time. Since the delay time inside the computer was different depending on the computing load in each event, a known sequence of random noise was transmitted to the loudspeaker and the input channel 1 measuring $x'(t)$ was used only for triggering the sequence. Once the trigger was activated, the two input sequences could be measured synchronously, and then for the impulse response identification the signals $y(t)$ and $x(t)$ were used instead of $x'(t)$. The random signal was of length 128 000 and the sampling frequency was 44.1 kHz. The identified plant was later used for designing its inverse filter.

Finally, Fig. 7(c) illustrates the wire connection for the binaural reproduction test, where two different random signals were properly processed for cross-talk cancellation and sent to the two loudspeakers through the two output channels and the input channels measure the two microphone signals. The reproduction errors were calculated by comparing the two original random signals with the measured signals in each channel. For monaural reproduction, only a single channel was used in both the input and output ports.

B. Monaural reproduction test

Authentic reproduction of a monaural sound signal was first tested using a loudspeaker and a microphone. The impulse response of the plant was first identified using the

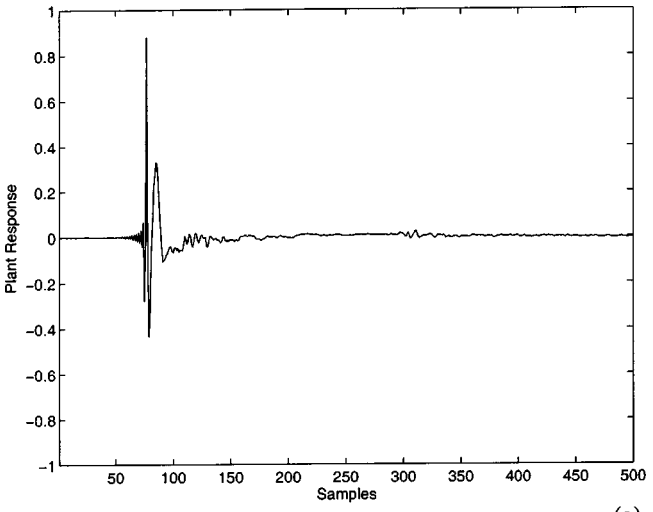
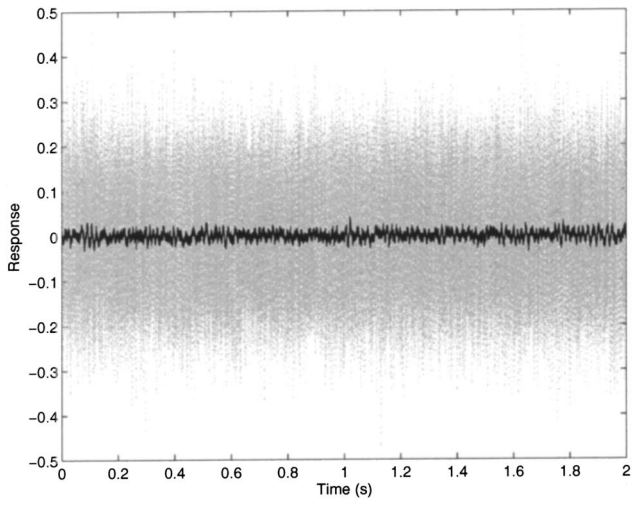
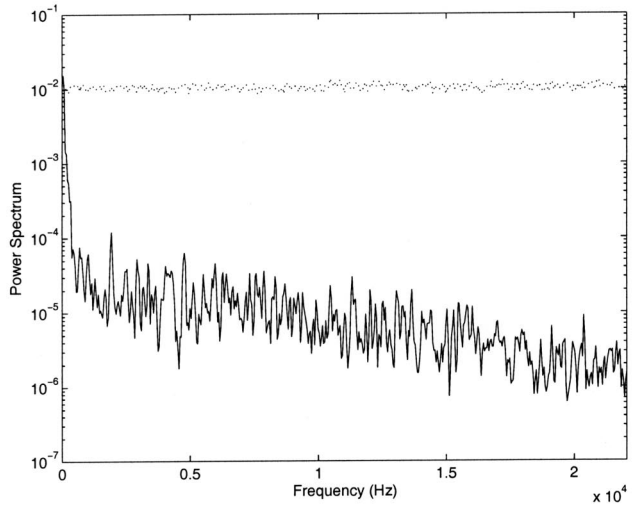


FIG. 9. The electroacoustic plant (a) and its inverse filter (b).



(a)



(b)

FIG. 10. Monaural reproduction performance error of 0.75%. A time delayed version of the original random sequence (dotted lines) and the reproduction error (solid lines) are shown together: (a) time responses, (b) power spectra.

Wiener filter in Eq. (6) with the original random signal $x(t)$ and the measured microphone signal $y(t)$ indicated in Fig. 8. The plant model was then used for the deconvolution process in Eq. (8) to obtain its inverse filter. The plant and its inverse filter are, respectively, shown in Figs. 9(a) and (b), where the impulse response has its maximum at the sample number of 77. The filter lengths of the plant and its inverse models used were both $I=500$, and the delay for τ was $I/2$. The plant shows a sinc functionlike characteristic near the maximum value at the sample number of 77, since linear phase low pass filters were installed in the soundcard. Finally, the inverse filter shown in Fig. 9(b) was implemented in a similar way as illustrated in Fig. 1, and the reproduction result in the time and frequency domains are, respectively, shown in Figs. 10(a) and (b), where dotted lines denote the time delayed version of the original random sequence while the solid lines denote the reproduction error $e(t)$ in Eq. (16). Figure 10(b) demonstrates that the reproduction performance is excellent all over the frequency range of interest except the very low frequency region where the loudspeaker used was very inactive. The reproduction performance error was 0.75%, and the modeling and inverse performance errors were $J_m=0.32\%$ and $J_n=0.36\%$, respectively. Several reproduction tests conducted in the anechoic chamber with a fixed microphone position revealed that the system using the Wiener filter was very stable. This was the reason why the plant robustness was not included in the system considered here.

A series of tests were also conducted by changing the filter length used: $I=200,300,...,800$. Figure 11 shows the modeling (J_m), inverse (J_n), and reproduction (J) performance errors according to the length of the filter used. A dramatic reduction of all three errors can be seen in the step from $I=200$ to $I=300$, and afterward a gradual decrease of the reproduction performance error (J) is achieved mainly due to a gradually decreasing modeling performance error (J_m). It is interesting to note in Fig. 11 that the inverse performance error (dotted line) does not decrease any further after $I=300$. This is because a longer inverse filter, say I

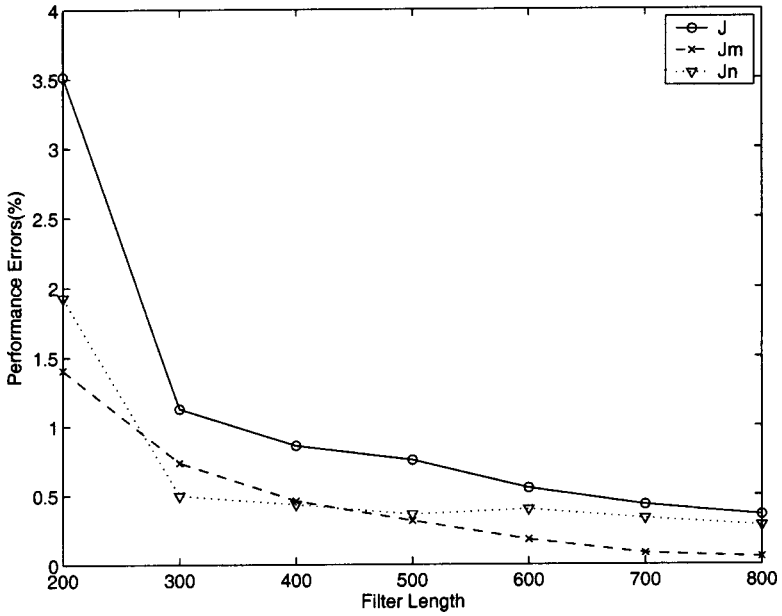


FIG. 11. Modeling (J_m), inverse (J_n), and reproduction (J) performance errors of the monaural test.

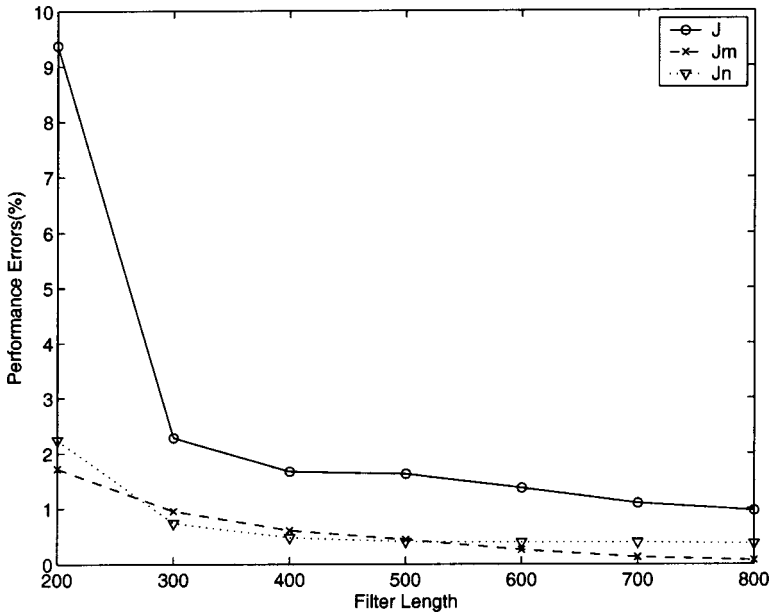


FIG. 12. Modeling (J_m), inverse (J_n), and reproduction (J) performance errors of the binaural test.

$=600$, was based on a longer plant filter of $I=600$ whose tail was nonzero due to slight acoustic reflections.

C. Binaural reproduction test

Binaural reproduction was also tested with two different random noises as the original sound signals, $u_1(t)$ and $u_2(t)$ in Fig. 1. Each elementary plant in the (2×2) electroacoustic plant matrix was identified separately by using the Wiener filter similarly as that described in the preceding section. The same known sequence of random noise was used for the test input signal to the left and right loudspeakers in order to facilitate the measurement with triggering. The (2×2) impulse responses identified showed a symmetric structure with their maximum value at the sample of 77 for the diagonal elements and 89 for the nondiagonal elements. The identified plant $\mathbf{G}(t)$ was then used to obtain the (2×2) inverse filter matrix $\mathbf{H}(t)$.

The inverse filter matrix was implemented as illustrated in Fig. 1, and a series of binaural reproduction tests were conducted by changing the filter length used: $I=200,300,...,800$. Figure 12 shows the modeling (J_m), inverse (J_n), reproduction (J) performance errors according to the length of the filters used. A dramatic reduction of all three errors was found in the step from $I=200$ to $I=300$, and then a gradual decrease of the modeling (J_m) and reproduction (J) performance errors was achieved afterward. Due to the same reason described in the preceding section, the inverse error does not decrease any further after $I=300$. A comparison between Figs. 11 and 12 shows that the reproduction performance error is increased more than double when the binaural system is used.

VI. CONCLUSIONS

This paper has investigated deconvolution algorithms for the authentic binaural reproduction of stereo (2 channel) sound over two loudspeakers. A systematic time domain deconvolution method was formulated using the Wiener filter approach. The main advantage is that the Wiener filter is inherently causal stable so that no laborious process for searching the robustness optimal coefficient for inversion is required.

The errors involved in authentic reproduction have been defined, and they are; the *modeling error* in the plant modeling process, the *inverse error* in the deconvolution process, and the *repeatability error* due to the time-varying plant characteristics in the reproduction process. Since the reproduction error can include all of these errors, the normalized mean square of this was used to assess the performance of a reproduction system. Authentic reproduction tests conducted demonstrated that the reproduction performance was improved in proportion to the length of the filter used, and interestingly this was not because of a smaller inverse error but because of a smaller modeling error for the system considered here. The reproduction performance error and its test procedure presented here can be used for objectively assessing the performance of a subjective auditory perception test system.[18]

ACKNOWLEDGMENT

This work was supported by Grant No. 2000-1-30400-006-3 from the Korea Science and Engineering Foundation.

APPENDIX: DISCRETE TIME DOMAIN REPRESENTATION OF WIENER FILTER

When the Wiener filter is implemented by a FIR (finite impulse response) filter in the discrete time domain, the Wiener filter $g[k]$ is the solution of the discrete form of the Wiener–Hopf equation given by[6]

$$\sum_{k=0}^{\infty} R_{rr}[n-k]g[k]=R_{rd}[n], \quad n\geqslant 0, \tag{A1}$$

where $R_{rr}[n]$ and $R_{rd}[n]$ are, respectively, the autocorrelation of the received signal $r[k]$ and the cross-correlation between the received and desired signal $d[k]$ in Fig. 2(b). Note here that, when $r[k]$ and $d[k]$ are deterministic as is the case

for the deterministic Wiener filter,[6] no time consuming averaging processes are required for calculating the correlations. If the filter has I coefficients, then Eq. (A1) can be expressed in matrix form as

$$\mathbf{Ag}=\mathbf{b}, \tag{A2}$$

where the $(I\times I)$ autocorrelation matrix $\mathbf{A}$ is a Toeplitz matrix, and is given by

$$\mathbf{A}=\begin{bmatrix} R_{rr}[0] & R_{rr}[1] & \cdots & R_{rr}[I-1] \\ R_{rr}[-1] & R_{rr}[0] & \cdots & R_{rr}[I-2] \\ \vdots & \vdots & & \vdots \\ R_{rr}[-I+1] & R_{rr}[-I+2] & \cdots & R_{rr}[0] \end{bmatrix}$$

and the I length Wiener filter coefficient vector $\mathbf{g}$ and the I length cross-correlation vector $\mathbf{b}$ are given by

$$\mathbf{g}=\{g[0]\ \ g[1]\ \ \cdots\ \ g[I-1]\}^T,$$

$$\mathbf{b}=\{R_{rd}[0]\ \ R_{rd}[1]\ \ \cdots\ \ R_{rd}[I-1]\}^T.$$

The matrix $\mathbf{A}$ is real valued and symmetric since $R_{rr}[n]=R_{rr}[-n]$. When the received signal $r[t]$ is spectrally rich as in random processes, the matrix $\mathbf{A}$ is guaranteed to be positive definite.[17] The Wiener filter can be obtained by

$$\mathbf{g}=\mathbf{A}^{-1}\mathbf{b}. \tag{A3}$$

[1] J. P. Blauert, *Spatial Hearing* (MIT Press, Cambridge, MA, 1996).
[2] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia* (Academic, New York, 1994).
[3] B. Gardner and K. Martin, KEMAR HRTF Measurements, MIT's Media Lab, 1994. http://sound.media.mit.edu/kemar.html
[4] C. P. Brown and R. O. Duna, "A structural model for binaural sound synthesis," IEEE Trans. Speech Audio Process. **6**, 476–488 (1998).
[5] P. A. Nelson and Y. Kahana, "Spherical harmonics, singular-value decomposition and the head-related transfer function," J. Sound Vib. **239**, 607–637 (2001).
[6] S. Haykin, *Adaptive Filter Theory*, 3rd ed. (Prentice-Hall, New York, 1996).
[7] B. Widrow and E. Walach, *Adaptive Inverse Control* (Prentice-Hall, New York, 1996).
[8] B. B. Bauer, "Stereophonic earphones and binaural loudspeakers," J. Audio Eng. Soc. **9**, 148–151 (1961).
[9] P. A. Nelson, H. Hamada, and S. J. Elliott, "Adaptive inverse filters for stereophonic sound reproduction," IEEE Trans. Signal Process. **40**, 1621–1632 (1992).
[10] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization," IEEE Trans. Speech Audio Process. **6**, 189–194 (1998).
[11] W. G. Gardner, "Head-tracked 3-D audio using loudspeakers," Proceedings of WASPAA, New Paltz, NY, 1997.
[12] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing* (Prentice Hall, New York, 1989).
[13] S. Skogestard and I. Postlethwaite, *Multivariable Feedback Control: Analysis and Design* (Wiley, New York, 1997).
[14] A. Papoulis, *Signal Analysis* (McGraw-Hill, New York, 1984).
[15] M. J. Brennan and S. M. Kim, "Feedforward and feedback control of sound and vibration—a Wiener filter approach," J. Sound Vib. **246**, 281–296 (2001).
[16] B. Noble and J. W. Daniel, *Applied Linear Algebra*, 3rd ed. (Prentice-Hall, New York, 1988).
[17] S. Elliott, *Signal Processing for Active Control* (Academic, New York, 2001).
[18] S. M. Kim and W. J. Choi, "Objective and subjective evaluations of 3D sound reproduction in a headphone," Inter-noise 2003, Jeju, Korea.

Characterizing computer cooling fan noise

Lixi Huang[a)]
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong

(Received 23 April 2003; revised 9 August 2003; accepted 4 September 2003)

Computer cooling fan noise is studied theoretically, focusing on the radiation from the interaction between rotor blades and motor struts. The source is decomposed into axial thrust, circumferential drag, and radial force. There is no sound-power coupling among the three components. The index of spatial spinning pressure mode plays the key role in noise radiation. The leading modes are the zeroth, or coincident, mode for thrust and the first mode for the drag and radial force. The effect of source noncompactness is quantified and found to be substantial only for higher-order radiation modes. The sound powers of the leading modes follow a sixth-power law, while the next high-order modes follow an eighth-power law. Quantitative analysis shows that the drag force can be equally noisy as the coincident thrust force. Based on an empirical aerodynamic model of rotor–strut interaction, it is found that the total sound power is more sensitive to the number of struts than rotor blades. Numerical examples are given to demonstrate how the struts can be optimized for typical cooling fan conditions. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1624074]

PACS numbers: 43.50.Ed, 43.20.Rz, 43.28.Ra [MSH]

Pages: 3189–3200

I. INTRODUCTION

With the increasing maturity of aerodynamic design, acoustic performance is becoming one of the major indices differentiating one manufacturer from another for a broad range of consumer products in which fans are used. The noise radiated by the computer cooling fans is receiving increasing attention as the CPU power increases rapidly and the trend of slim packaging continues. Due to the stringent low budget imposed on cooling fans, technology developed for large fans may not be suitable for computer cooling fans. For example, the technique of active noise control is apparently too costly. A cooling fan is normally manufactured by plastic molding and the tip clearance tolerance in such a small product is usually as large as aircraft engine fans. The much higher clearance-to-fan-radius ratio means a fertile source of noise mechanisms. To the best of the author's knowledge, a comprehensive theoretical study has not been performed for small axial-flow cooling fans.

Research on the general fan aeroacoustics can be said to begin with Gutin's (1936) effort in quantifying the propeller noise caused by the rotation of steady loading, later called Gutin noise. Lighthill's (1952) acoustic analogy provided a formal platform of investigating aerodynamic sound, which was soon extended by Curle (1955) to include the effect of solid boundaries by replacing it with distributed dipoles and monopoles while the explicit effect of acoustic scattering are neglected. It was not until 1969 that Ffowcs Williams and Hawkings (1969) formally extended Lighthill's acoustic analogy by using the generalized functions to account for the effect of solid boundaries in arbitrary motion. The use of the so-called Ffowcs Williams and Hawkings equation became a dominant feature from the late 1970's in the acoustics for rotary machines. Meanwhile, Tyler and Sofrin (1962) revealed the important role played by the numbers of rotor and stator blades, and the phenomenon of modal cutoff consistent with duct acoustics. This analysis represented a major contribution towards turbomachinery acoustics. The link between this cutoff behavior for the ducted fan was related to that of unducted fan by Lowson (1970), whose work also demonstrated that the radiation of mismatched spinning modes cannot be cut off but is rather reduced in acoustic efficiency. Kaji and Okazaki (1970), using the wake models of Kemp and Sears (1953, 1955), modeled the rotor–stator interaction with careful doublet arrays while preserving the blade effects on sound propagation in the duct and imposing Kutta conditions on trailing edges. A comprehensive review was given by Morfey (1973). Meanwhile, Sharland (1964), and Mugridge and Morfey (1972) focused on broadband noise sources for which order of magnitudes estimates were made. In fact, accurate computation of broadband noise is still a challenge, but numerical schemes using elementary experimental data as input can provide useful predictions (Glegg and Jochault, 1998). Longhouse (1976) found that design features good for reducing broadband noise could contradict those for discrete tones. Ffowcs Williams and Hall (1970) showed that trailing edge scattering of turbulent boundary layer flows is a powerful noise source, and the work on this aspect continues today (Howe, 1999) as aerodynamicists are yet to come up with a confident understanding of boundary layer behavior.

In the area of noise abatement for small axial-flow fans, considerable efforts have also been made. A selected few of them are mentioned here. Fitzgerald and Lauchle (1984) demonstrated the effect of a comprehensive range of corrective measures that can be taken to modify a cooling fan design for better acoustic quality. These include the equalization of downstream strut size, the reduction or elimination of the potential flow interaction with nearby objects, use of bellmouth inlet to minimize inlet flow distortions, and the prevention of possible large-scale flow separation on the suc-

a)Electronic mail: mmlhuang@polyu.edu.hk

tion side of blades. Quinlan and Bent (1998) focused on broadband, high-frequency noise generated by the tip leakage flow with extensive acoustic and aerodynamic measurements. Attempts were also made to reduce tonal noise by uneven blades for propellers (Lewy, 1992) and small radial fans (Boltezar *et al.*, 1998). It was concluded that only spectral shifts can be made while the total sound power remains more or less the same. The optimization in terms of stator/vane blade sweep and lean was conducted on advanced ducted propeller (Envia and Nallasamy, 1999) and turbofan engine (Schulten, 1997). The general conclusion is that a stator blade should interact with as many rotor wakes as possible. Similar study has not been found for unducted, small axial-flow fans with few blades. The acoustic feasibility of active noise control for small axial-flow fans was tested using ordinary loudspeakers (Quinlan, 1992) and the fan itself as the secondary source through axial shaking (Lauchle *et al.*, 1997). In both cases a baffle was used, which may have altered the acoustic directivity. The study on propeller noise by Subramanian and Mueller (1995) paid more attention to acoustic directivity. By simulating the acoustic directivity of sound radiated by the fundamental and the first higher-order spinning pressure modes, Gerhold (1997) was able to cancel low-frequency noise of a ducted fan using a ring of orchestrated loudspeakers.

The present study focuses entirely on the tonal noise of isolated computer cooling fans. This represents a first step towards understanding the noise radiation of such cooling fans installed inside a computer chassis. A chassis is a rather confining environment and it would not be surprising that reflection and scattering of sound by confining walls may alter the noise radiation characteristics significantly. However, it might be expected that the chassis environment would not have much scattering job to perform if the sounds radiated by different parts of the fan cancel themselves out in the first place. In other words, efforts in silencing an isolated cooling fan would by no means be wasted. Having said that, the aerodynamic environment inside a computer does present an extra source of noise which is not encountered in free space. For example, the inlet flow is most likely to be nonuniform due to the presence of nearby stationary objects. Apart from the nonuniformity of the potential flow part of the inlet condition, turbulence and deformation of turbulent eddies by a contracting inlet stream tube may also give rise to noise radiation, cf. Trunzo *et al.* (1981) and Majumdar and Peake (1998). Such radiation may contain both broadband (random) and tonal (deterministic) components. For example, the inlet bellmouth of a sample fan shown in Fig. 1(a) has four sharp edges caused by the intersection of the bellmouth and the square outer frame. The inlet flow is expected to have a four-lobe distortion which may or may not contain discrete vortices. The tonal noise induced is expected to be equivalent to the interaction between the rotor blades and a set of four inlet vanes. Most likely, the noise induced by this inlet feature dominates over interaction with other inlet features which are further away from the rotor. The kinematic characteristics of such interaction noise would be similar to that caused by the rotor blades and the downstream struts, the latter being the focus of the present study. In this sense,

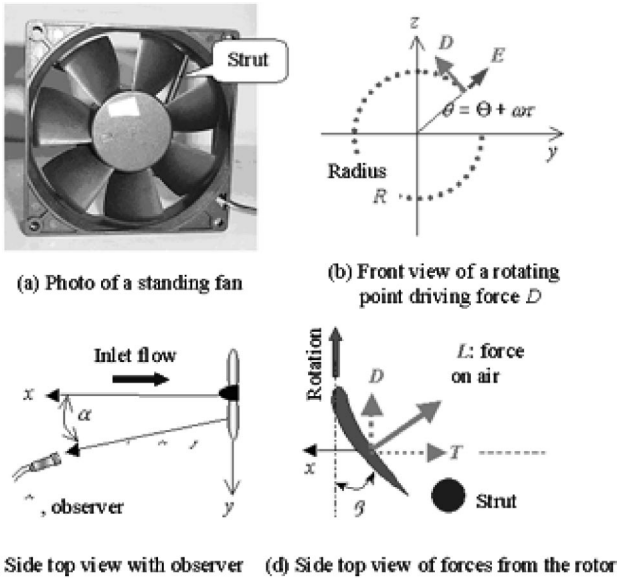


FIG. 1. Sound generated by a rotating force.

the current study deals with the interaction noise inherent in an isolated cooling fan. The interaction is considered to be an unavoidable feature of this category of cooling fans, unlike the freedom Brungart *et al.* (1999) has in their handling of the so-called installation effects for the whole set of vacuum cleaner. In terms of broadband noise, possible sources in the present cooling fan includes the ingestion of inflow turbulence, tip leakage vortex (Fukano *et al.*, 1986), vortex shedding from the rotor blades (Longhouse, 1977), as well as scattering of Tollmien Schlichting waves by the blade trailing edge. The identification and separation of all these possible noise sources are beyond the scope of the current study.

For the typical computer cooling fans, the dominant noise has a typical blade passing frequency (BPF) of 350 Hz for which the fan is physically compact. The effect of blade surface scattering can be ignored and the work of Lowson (1970) is referred to in Sec. II. The dominant noise source is the aerodynamic interaction between the impeller blades and the downstream struts which hold the fan motor. Typical struts are circular or triangular cylinders of small size, and the unsteady forces experienced by the struts are normally much smaller than that on the rotor blades which are, after all, so profiled to generate large lift. Anticipating a possible trend in designing swept and leaned fan, the contribution of the unsteady radial force is, for the first time, included in the formulation. Section III examines the effects of source noncompactness, which is mainly caused by the angular phase distributions instead of physical size. The relative importance of each unsteady force component, thrust, drag, and radial force, is derived analytically for dominant spinning pressure modes. Further progress rests with the spectral knowledge of the rotor–strut interaction force. An empirical model is proposed in Sec. IV with the purpose of investigating the general characteristics of all noise components for various rotor–strut combinations. The current design practices are scrutinized in a quantitative manner with some comparison with experimental results. Numerical examples are then given for a possible optimization of strut design.

II. ELEMENTARY ANALYSIS

Although the details of the point force formulation are available in Lowson (1965, 1970), the noise radiated by the radial component of unsteady force has not been derived. The essential derivation steps are given in this section focusing on the physical interpretation of results and the characterizations appropriate for the computer cooling fan noise. The effect of source noncompactness is investigated in the next section.

A. Point force formulation

As shown in Fig. 1, a typical computer cooling fan consists of a rotor with B blades and S motor struts in its downstream. The struts are typically circular cylinders with a certain leaning angle with respect to the radius to reduce noise, e.g., the fan shown in Fig. 1(a), for which $B=7$, $S=4$.

Using a Cartesian coordinate system (x,y,z) with its origin fixed to the center of a vertically standing fan, the front and side views are shown in the three subfigures of Fig. 1. A cylindrical coordinate system, (x,R,θ), is also introduced and the polar coordinates are shown in Fig. 1(b). For an impeller rotating anticlockwise, the angular coordinate for a point fixed on rotor is given as

$$\theta=\Theta+\omega\tau, \tag{1}$$

where τ is time used for the source, Θ is the initial position of the source point on rotor relative to the stationary reference frame, and ω is the angular speed. No generality is lost if the observer is placed on the horizontal plane, $z=0$, while Θ is allowed to vary. The observer position is denoted by vector $\mathbf{x}=(x_1,x_2,x_3)=(x,y,0)$, and source point by $\mathbf{y}=(y_1,y_2,y_3)$. The distance between them is r and the distance to the origin is defined as r_0. The source is located at $x=x_s$ from the origin and $|x_s|$ is at most comparable to the source radius R. The angular position of the observer, α, is measured from the upstream, as shown in Fig. 1(c). For an observer in the far field, $|x_s|,R\ll r_0$, r can be approximated as

$$r\approx r_0-x_s(x/r_0)-y(R/r_0)\cos\theta. \tag{2}$$

Figure 1(d) shows a blade cross section. The force $\mathbf{F}$ acting on air from a rotating blade is split into three components, thrust force T in the flow direction, drag D in the direction of rotation (the more appropriate term here should be "driving force," but the term "drag" has been used extensively in literature), and radial force E, which can be substantial when the blade leans. The noise radiated by T,D,E is called, respectively, thrust noise, drag noise, and radial noise. The effect of force distribution on the blade surface can be accounted for by the summation of many isolated point forces each with a different initial source position of (x_s,R,Θ) in cylindrical coordinates. For the point force approximation, drag and thrust are related by the pitch angle β

$$T=L\cos\beta,\quad D=L\sin\beta, \tag{3}$$

where L is the airfoil lift. For a fan operating in a condition where the inlet flow is uniform, the unsteady lift on the blade, $L(\tau)$, is mainly produced by the interaction between the moving airfoil and the stationary struts, while the unsteady forces on the struts are produced by a combination of potential and viscous wakes of the rotor. The radial force on a leaning strut can be particularly large. The projection of the total force on the source–receiver direction r is

$$F_r=r^{-1}[-(xT+(ER-x_sT))-yD\sin\theta+Ey\cos\theta], \tag{4}$$
$$\text{(small)}$$

where the contribution from the terms of $(ER-x_sT)$ is small in the far field, $r_0\gg R,|x_s|$, and shall be ignored.

Following Lowson (1965), the sound generated by a rotating point force is

$$p(\mathbf{x},t)=\frac{x_i-y_i}{(1-M_r)c_0r}\frac{\partial}{\partial\tau}\left[\frac{F_i}{4\pi r(1-M_r)}\right], \tag{5}$$

where the right-hand side is evaluated at the radiation time τ, or retarded time, c_0 is the speed of sound, and $M_r=-c_0^{-1}dr/d\tau$ is the Mach number of the source approaching the observer. The Fourier transform of the far-field sound is

$$\begin{aligned}c_n&=\frac{\omega}{\pi}\int_0^{2\pi/\omega}p(\mathbf{x},t)e^{in\omega t}dt\\&=\frac{-\omega}{4\pi^2}\int_0^{2\pi/\omega}\frac{in\omega F_r}{c_0r}e^{in\omega(\tau+r/c_0)}d\tau.\end{aligned} \tag{6}$$

In carrying out the above integration, the leading term of the far-field sound would be retained by substituting r in the denominator by r_0, the same applies to r in F_r given in Eq. (4). But, that in the exponent cannot as it involves the phase angle variation. The phase angle $n\omega r/c_0$ is approximated by

$$\begin{aligned}n\omega r/c_0&\approx nk_0(r_0-x_s\cos\alpha)-nM\sin\alpha\cos\theta,\\M&=\omega R/c_0,\end{aligned}$$

where M is the Mach number of the source rotational speed. Disregarding the mean phase angle corresponding to nk_0r_0, and noting $\omega d\tau=d\theta$, Eq. (6) can be rewritten as

$$\begin{aligned}c_n\approx\frac{-in\omega e^{-ink_0x_s\cos\alpha}}{4\pi^2c_0r_0}\int_\Theta^{2\pi+\Theta}F_r\exp[in(\theta-\Theta\\-M\sin\alpha\cos\theta)]d\theta.\end{aligned} \tag{7}$$

Note that k_0x_s is negligible for small fans, and the term of $e^{-ink_0x_s\cos\alpha}$ is dropped henceforth, although some comments shall be made on this point in the next section. This integration can be evaluated if all force components, denoted as F_j, $j=1,2,3$ for T, D, E, respectively, are Fourier transformed

$$\begin{aligned}F_j&=\sum_{\lambda=-\infty}^{\infty}F_{j\lambda}e^{-i\lambda\omega\tau}\overset{\omega t=\theta-\Theta}{\Rightarrow}\sum_{\lambda=-\infty}^{\infty}e^{i\lambda\Theta}F_{j\lambda}e^{-i\lambda\theta},\\F_{j\lambda}&=\frac{\omega}{2\pi}\int_0^{2\pi/\omega}F_j(\tau)\,e^{i\lambda\omega\tau}d\tau,\end{aligned} \tag{8}$$

in which both positive and negative frequencies are used. The substitution of Eqs. (4) and (8) into Eq. (7) gives

$$c_n \approx \frac{in\omega}{4\pi^2 c_0 r_0} \sum_{\lambda=-\infty}^{\infty} \int_{\Theta}^{2\pi+\Theta} e^{i(\lambda-n)\Theta} \times [T_\lambda \cos\alpha + (D_\lambda \sin\alpha)\sin\theta - (E_\lambda \sin\alpha)\cos\theta] \times \exp[i(n-\lambda)\theta - inM \sin\alpha \cos\theta] d\theta.$$

These integrations lead to the Bessel function of the first kind. Hence

$$c_n \approx \frac{in\omega}{2\pi c_0 r_0} \sum_{\lambda=-\infty}^{\infty} i^{\lambda-n} e^{i(\lambda-n)\Theta} \left[\left(T_\lambda \cos\alpha - \frac{n-\lambda}{nM} D_\lambda \right) \times J_{n-\lambda}(nM\sin\alpha) - (\sin\alpha) i E_\lambda \frac{J_{n-\lambda-1}(nM\sin\alpha) - J_{n-\lambda+1}(nM\sin\alpha)}{2} \right], \tag{9}$$

where the thrust and drag noise terms are identical to those in Lowson (1970). Note that when the rotor does not move, the above formula is no longer valid since the angular position θ used in the expression of F_r in Eq. (4) is no longer related to τ in the time integration of Fourier transform. In other words, one cannot simply substitute $M=0$ into Eq. (9) to find sound generated by stationary struts on which unsteady forces are exerted. Sound generated by one point force on strut can, however, be found through the time-domain formula of Eq. (5), where $M_r=0$ and r is a constant. This procedure is useful when considering sound generated by vortex shedding from struts. When sound is generated by strut experiencing the wake sweeping from the rotor, there is fixed phase relationship among different struts, and the result is similar, though not identical, to Eq. (9) since the source is effectively synchronized with and attached to the rotating blades. The exact formula will be given in the next section.

Since M is very small in computer cooling fans, usually $BM<0.2$, the argument of the Bessel function, $z=nM\sin\alpha$, is very small except for very high frequencies where the source strength has decayed. The following Taylor expansions can be shown to be satisfactory for the range of argument z encountered:

$$J_0(z) \approx 1 - \frac{z^2}{4}, \quad J_\nu(z)|_{|z|\ll|\nu|} \approx \frac{1}{\sqrt{2\pi\nu}} \left(\frac{ze}{2\nu} \right)^\nu. \tag{10}$$

Notice that the second expression works better than the usual approximation of $J_\nu(z) \approx (z/2)^\nu/\Gamma(\nu+1)$ normally adopted for $z\to 0$ for all ν.

It will be shown later that, when the noise made by the interaction between B rotor blades and S struts is taken into account, various acoustic interferences take place. The only possible sound radiated is that given by $n=mB$, $\lambda=kV$, where both m and k are integers. The index of frequency differential, $\nu=n-\lambda$, is the most important parameter controlling how strong a sound component is radiated as it registers mathematically as the order of Bessel functions. As $\nu\to\infty$, the Bessel functions vanish rapidly. This suggests that noise generated by the source component of very different frequencies (Doppler effect) is very low when M is small. For Gutin noise, the source frequency is zero, $\lambda=0$, and the sound-pressure amplitude is $|c_B| \propto J_B(BM) \propto M^B$, which is normally negligible for computer cooling fans. The dominant sound is interaction noise for which ν is small. The following subsection discusses the most typical situations of $|\nu|=0,1,2$ with emphasis on physics related to two factors: one is the change of unsteady force direction as blade rotates; see F_r in Eq. (4), and the other is the change of r during the source motion; see Eq. (2).

B. Modal directivity

The acoustic directivity of the three noise components can be drawn from Eq. (9) by ignoring all constants unrelated to α and by making use of the Bessel function approximations in Eq. (10). The results are shown in Fig. 2 in polar plots of sound intensity with phase angle distribution of sound pressure marked in the lobes, and the simplified expressions given in the vertical labels. Three levels of Bessel function orders are considered, $\nu=n-\lambda=0,1,2$, the results for $\nu=-1,-2$ being similar to $\nu=1,2$. The subfigures in the first row are for thrust noise, the second for the drag noise, and the third for the radial noise. The second row has only two subfigures as the drag noise vanishes for $\nu=0$. First, it is pointed out that thrust noise differs from both drag noise and radial noise. The difference is that the thrust noise has an antiphase relationship when measured in front of and behind the rotational plane, while the drag and radial noises have the in-phase relation across the rotational plane as these two forces lie on the rotational plane. The details of the directivity patterns are analyzed below, while the magnitude of sound power from each noise component is studied in the next section.

For thrust noise, $\nu=0$ can be described as the coincident mode, which means the sound generation at the source frequency itself. The result is a simple dipole in the axial direction, as shown in Fig. 2(a). For $\nu=1$, such as caused by $n=2$, $\lambda=1$, the pattern is shown in Fig. 2(b), similar to that of a lateral quadrupole. This pattern is produced by two factors. One is the dipole nature of the source, which contributes $\cos\alpha$ in the directivity distribution due to the alternating signs of the force projection to the upstream and downstream. The second factor is the motion of the source (Doppler effect). When a force leaves the observer, its effect is equivalent to a decreasing force, and vice versa. The source motion changes r, characterized by $\cos\theta$, and its final effect on sound generation is characterized by $\sin\alpha$. Right on the axis, $\alpha=0,\pi$, r hardly changes when the force circulates, while the change is most dramatic when the observer is on the rotational plane. The combination of $\cos\alpha\sin\alpha$ gives a four-lobe pattern of Fig. 2(b). The second-order radiation ($\nu=2$) is shown in Fig. 2(c) and the pattern can be described as a split dipole.

It can be seen from all subfigures of Fig. 2 that the only possible sound heard on the rotational axis is the coincident thrust noise. The time-domain sound pressure can be found from Eq. (5). In the far field, $r_0 \gg |R|$, F_r does not contain any drag and radial force, and $M_r \approx 0$. The sound becomes

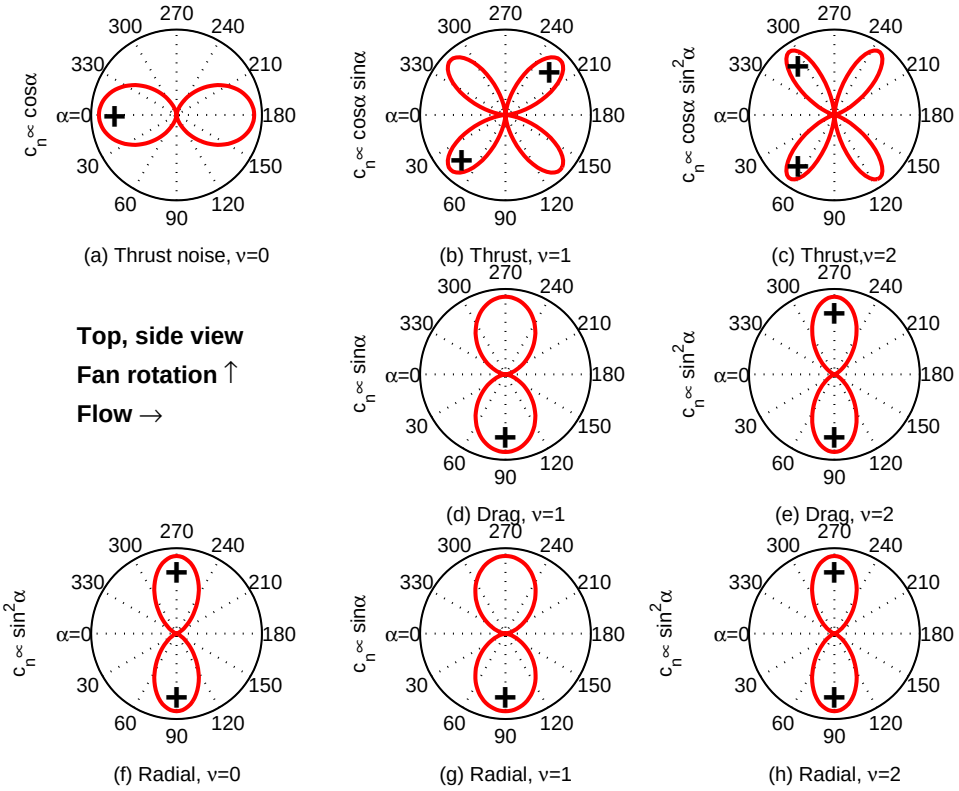


FIG. 2. Acoustic directivity pattern for thrust (top row), drag (middle row), and radial (bottom row) force components. The amplitude is sound intensity, and the symbols of "+" and "−" denote phase angle distribution.

$$p(x,t)\big|_{\alpha=0,\pi}=(\pm 1)\,\frac{\partial T/\partial \tau}{4\pi r_0 c_0}, \tag{11}$$

where the positive sign is for the downstream and negative for the upstream, and thrust T is the total integrated thrust force on all blades at time τ. By measuring the sound on the rotational axis, one can obtain the time-domain characteristics of the lift, $L=T/\cos\beta$, caused by the rotor–strut interaction.

The drag noise is shown in the second row of Fig. 2. The reason why there is no sound at the coincident mode $\nu=0$ is explained as follows. For source frequency $\lambda=1$, D changes its sign once between the two halves of a cycle. In other words, D changes direction once when viewed from the rotor. When the blade rotates, the direction of D also changes once relative to the stationary observer. The result is that there is no change of drag force for this particular frequency component; hence, no sound is generated. More explanation is given below when the radial noise is studied. In short, drag noise is generated at all other frequencies due to both force direction change and distance change. The pattern of sound at $\nu=1$ is given in Fig. 2(d), which is in fact a circulating dipole. The physics is easily appreciated by analyzing what happens for $\lambda=0$ and $n=1$. A steady force appears to be an oscillating one when it circulates around. For $\nu=2$, the pattern shown in Fig. 2(e) differs in phase angle distribution from that of Fig. 2(d). This can be explained easily by looking at $\lambda=0$, $n=2$ with the help of Figs. 1(b) and (c). The source trajectory can be divided into two hemispheres, one near the observer and another far from the observer on $z=0$. When an oscillating D rises from $z<0$ towards $z=0$, the force points towards the observer and makes, for instance, a positive sound. The sound made is weak at $\theta=-\pi/2$ due to its distance, but it is also weak at $\theta=0$ due to the fact that the force is not pointing to the observer despite their proximity. When it moves further up towards $z>0$, it points away from the observer. The half cycle is converted to one full cycle of sound radiation as $\sin 2\theta$. While D makes positive sound during $\theta=-\pi/2\rightarrow 0$ to the near-side observer at $y>0$, it actually makes positive sound to the far-side observer at $y<0$ as well. This is so because of two factors; one is that drag is pointing away from the far-side observer, hence negative sound, and the other is that it also leaves the observer, hence another negative pressure effect. The combination of the two effects makes a positive pressure oscillation. The result is a pattern shown in Fig. 2(e), which is similar to the far-field directivity of a longitudinal quadrupole.

The sound generated by a moving radial force is similar in pattern as the rotating drag force, as shown in the second and third rows of Fig. 2. However, there are differences in details. The reason why there is radial sound heard at the coincident mode of $\nu=0$ is explained for the easiest case of $\lambda=n=1$ and the configuration is shown in Fig. 3. A point source of radial force is circulating with angular position θ. Considering the force signature of $E(\tau)\propto\cos(\theta)$, $\theta=\lambda\omega\tau$, $\lambda=1$, the force changes direction from outward to inward from $\theta=0$ to $\theta=\pi$. As a result, the force appears to be in the same direction to a fixed observer when the source is at $\theta=\pi$ as it is at $\theta=0$. From this perspective, there is little sound generated at the source frequency of $\lambda=1$. However, due to the difference between the two points to the observer at angular position α, $r_1<r_2$, there still appears to be pressure variation felt in the far field once per cycle, $n=1$. The strength is characterized mainly by the Bessel function of

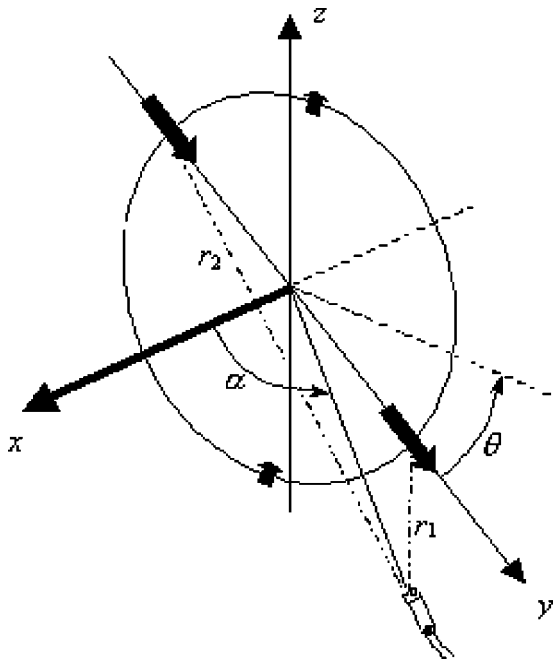


FIG. 3. Sound generation by rotating radial force of the first rotational frequency mode, $\lambda=1$.

$J_{\nu=n-\lambda\pm1}$, which is essentially J_1 when $\nu=0$. This contrasts with the drag noise, which has zero noise generation when $\nu=0$. In the latter case, the highest sound pressure is generated when a drag force passes the point of $\theta=\pi/2$ when D more or less points to the observer located on the plane of $\theta=0$. The sound generated at this point is exactly canceled out by that at $\theta=3\pi/2$ as far as the frequency component $n=\lambda$ is concerned. Of course, there are spectral leaks to other frequencies and the strength of such leak is signified by the Bessel function of order $\nu=n-\lambda$. The difference from the drag noise is that, when $\nu=1$, noise is generated at a magnitude characterized by a Bessel function of zeroth order, for which $J_0(0)=1$. This is explained by using the graph of Fig. 3 and assuming $\lambda=1$, $n=2$. Due to the change of source value once per cycle, i.e., $\lambda=1$, the source force turned inward at $\theta=\pi$, the half cycle of $\theta=0\rightarrow\pi$ actually constitutes one complete cycle of sound radiation. One cycle of drag rotation makes sound at the second harmonic of $n=2$, and the magnitude is characterized by J_0, which implies a high radiation efficiency. In other words, a rotating radial force is actually equivalent to a stationary source of frequency $n=\lambda\pm1$ due to the rotation.

III. ACOUSTIC INTERFERENCE

This section analyzes the total sound power generated by multiple blades and the interaction of sound generated by different force components on blades. Before these are quantified in closed-form formulas using point-force approximation, it is necessary to examine the effect of source noncompactness for different force component and different angular modes.

A. Effect of source noncompactness

The effect of distributed source can be examined by investigating the sum of a series of point forces. The question is how much error is caused by using the instantaneous average of the force components, $F_j(\tau)$, in computing the Fourier transform of Eq. (8), instead of summing the acoustic contribution from each point force given in Eq. (9). A proper integration over the distributed source involves putting right two factors. One is the source position, and the other is the phase angle relationship. The effect of the axial source position x_s is clearly seen in the term of $\exp(-ink_0x_s\cos\alpha)$ in Eq. (7). The error involved in approximating $x_s=0$ for all source points is proportional to nk_0x_s, which is essentially the ratio of the axial extent of the source to the wavelength of frequency $n\omega$. For compact cooling fans, this can be ignored.

The effect of varying the angular position of the source, Θ, manifests itself as $e^{i(\lambda-n)\Theta}$ in Eq. (9) for all three force components in cylindrical coordinates, denoted as $F_j(\tau), j=1,2,3$ for $T(\tau),D(\tau),E(\tau)$, respectively. In time domain, this means that the distributed surface pressure, $p(\Theta,\tau)$, should be integrated over the blade surface S as a lumped component force, denoted as $\tilde{F}_j(\tau)$, with a time delay of $\Theta(\lambda-n)/(\lambda\omega)$. The pair of time-delayed integration and its Fourier transform, denoted as $\hat{F}_{j\lambda}$, are given below

$$\tilde{F}_j(\tau)=\int_S p\left(\Theta,\tau-\frac{\Theta(\lambda-n)}{\omega\lambda}\right)n_j\,ds,$$

$$\hat{F}_{j\lambda}=\frac{\omega}{2\pi}\int_0^{2\pi/\omega}\tilde{F}(\tau)e^{i\lambda\omega\tau}d\tau=\int_S e^{i(\lambda-n)\Theta}\hat{p}_\lambda n_j\,ds, \quad (12)$$

$$\hat{p}_\lambda=\frac{\omega}{2\pi}\int_0^{2\pi/\omega}p(\Theta,\tau)e^{i\lambda\omega\tau}d\tau,$$

where n_j is the component of unit normal vector in the cylindrical coordinate system.

When $n=0$, the time delay is exactly what is required to rewind the source back to $\Theta=0$. In other words, $\tilde{F}_j$ is simply the sum of forces at the different times when they pass through a fixed point in space. For coincident noise, $n=\lambda$, no time delay is needed and the instantaneous force integration gives the correct prediction. For other n, the amount of time delay varies. For $n>\lambda$, time advance occurs. There is no universal time-delay-integration scheme that can capture all frequencies correctly using a lumped source in the circumferential direction.

However, it is normally sufficient to focus on one dominant noise source index for which $|n-\lambda|=0,1,2$. For the example of $B=7$, $S=4$, it can be shown that $\lambda=2S=8$ is the dominant source frequency index, so that a time delay of $\Theta/8\omega$ is suitable for the BPF noise. In this case, the effect of approximating $\tilde{F}_j(\tau)$ by an instantaneous spatial integration is expected to lead to some errors. Two typical cases are considered for error analysis with reference to the blade diagram shown in Fig. 1. In the first case, unsteady flow pressure p on the rotor blade is assumed to have the same phase angle but with a linearly decaying magnitude with respect to

the chordwise coordinate of $\xi \in [0,C]$, where $\xi=0$ means the trailing edge. Denoting the instantaneous integration of force F_j by G_j, one has

$$p(\xi,\tau)n_j = 2A_\lambda e^{-i\lambda\omega\tau}(1-\xi/C),$$

$$G_j(\tau) = \int_0^C p(\xi,\tau)\, n_j\, d\xi = A_\lambda C e^{-i\lambda\omega\tau},$$

$$\hat{G}_{j\lambda} = \frac{\omega}{2\pi}\int_0^{2\pi/\omega} G_j(\tau) e^{i\lambda\omega\tau} d\tau = A_\lambda C.$$

The approximate complex sound amplitude is then $c_n' \propto A_\lambda C$, while the accurate sound is found by integrating $e^{i(\lambda-n)\Theta}\hat{F}_\lambda$ in Eq. (9) over the chord where Θ varies from 0 to ϕ

$$\hat{F}_\lambda(\xi) = 2A_\lambda(1-\xi/C),$$

$$c_n \propto \int_0^C \hat{F}_\lambda(\xi) e^{i(\lambda-n)\Theta}\, d\xi = A_\lambda C\left[\frac{1+i(\lambda-n)\phi - e^{i(\lambda-n)\phi}}{[(\lambda-n)\phi]^2/2}\right].$$

The square bracket represents the deviation of the true amplitude c_n from the approximate one, c_n'. Taking the following parameters as an example, $\lambda=8$, $\phi=2\pi/10$, the ratio of the two amplitudes is 0.99, which is very satisfactory.

In the second case, the unsteady forces on the rotor are still assumed to decay linearly from the trailing edge where interaction takes place, but the pressures are related by time delays according to the rotation

$$p(\xi,\tau)n_j = 2A_\lambda(1-\xi/C)e^{-i\lambda(\omega\tau+\Theta)},$$

$$G_j(\tau) = \int_0^C p(\xi,\tau)n_j\, d\xi = A_\lambda C\,\frac{1-i\lambda\phi-e^{-i\lambda\phi}}{(\lambda\phi)^2/2} e^{-i\lambda\omega\tau},$$

$$G_{j\lambda} = A_\lambda C\,\frac{1-i\lambda\phi-e^{-i\lambda\phi}}{(\lambda\phi)^2/2},$$

$$c_n' \propto A_\lambda C\,\frac{1-i\lambda\phi-e^{-i\lambda\phi}}{(\lambda\phi)^2/2}.$$

The accurate sound is

$$\hat{F}_\lambda(\xi) = 2A_\lambda(1-\xi/C)e^{-i\lambda\Theta},$$

$$c_n \propto \int_0^C \hat{F}_\lambda(\xi)\, e^{i(\lambda-n)\Theta} d\xi = A_\lambda C\left[\frac{1-in\phi-e^{-in\phi}}{(n\phi)^2/2}\right].$$

The ratio of approximate to accurate sound for the example is

$$\frac{c_n'}{c_n} = \left[\frac{1-i\lambda\phi-e^{-i\lambda\phi}}{1-in\phi-e^{-in\phi}} \times \frac{n^2}{\lambda^2}\right],$$

which is 0.8366 or 1.55-dB error. The error grows fast for higher order modes, e.g., $|\nu|=2$.

The radial noncompactness manifests itself through the rotational Mach number, $M=\omega R/c_0$. For the thrust noise, the effect depends on the order $\nu=n-\lambda$. For the coincident mode, there is essentially little influence since $J_0(z)\approx 1-z^2/4\to 1$ for small $z=nM\sin\alpha$. For other modes, $J_\nu(z)\propto z^\nu$. The error caused by placing the lumped source at a radius where the load peaks is easily estimated. The drag noise has different characteristics. The dominant noise comes from when $\nu=\pm 1$ when the Mach number actually does not matter as $J_1(nM\sin\alpha)/M$ is almost constant when M is small. The second-order drag noise features $D_\lambda J_2(nM\sin\alpha)/M \propto DR$, so that the noise radiated is proportional to the integrated torque. The dominant radial noise also comes from $\nu=1$ and the Bessel function involved is $J_0(nM\sin\alpha)$ which is also immune from the variations of the Mach number. In summary, the error caused by lumping forces at an inappropriate radius depends on the dominant noise-making mode. The error will be negligible for all the leading radiation modes.

The summation of sound made by multiple blades is done by assuming that pressure fluctuation caused by the rotor–strut interaction occurs when a blade passes through an arbitrary fixed strut position; hence, the fixed phase angle distribution among all forces on all blades. As was shown by Lowson (1970), the summation of sound from B evenly spaced blades interacting with S evenly spaced struts is

$$c_{n=mB}^{(\text{rotor})} = \frac{imB^2S}{2\pi c_0 r_0}\sum_{k=-\infty}^{\infty} i^{-\nu}\left\{\left(T_{kS}\cos\alpha - \frac{\nu}{nM}D_{kS}\right)\right.$$
$$\times J_\nu(nM\sin\alpha) - \frac{i}{2}(\sin\alpha)$$
$$\left.\times E_{kS}[J_{\nu-1}(nM\sin\alpha) - J_{\nu+1}(nM\sin\alpha)]\right\}, \quad \nu = mB-kS, \tag{13}$$

where T_{kS}, D_{kS}, E_{kS} are derived from the encounter of a single blade with a single strut. Similarly, the sound from struts is found to be of the same form as Eq. (13) except that the subscripts of kS in the forces are changed to mB as there is no Doppler effect from stationary sources, although a pattern of acoustic interference exists

$$c_{n=mB}^{(\text{strut})} = \frac{imB^2S}{2\pi c_0 r_0}\sum_{k=-\infty}^{\infty} i^{-\nu}\left\{\left(T_n\cos\alpha - \frac{\nu}{nM}D_n\right)\right.$$
$$\times J_\nu(nM\sin\alpha) - \frac{i}{2}(\sin\alpha)$$
$$\left.\times E_n[J_{\nu-1}(nM\sin\alpha) - J_{\nu+1}(nM\sin\alpha)]\right\},$$
$$\nu = mB-kS. \tag{14}$$

B. Sound-power integrations

The sound-power integration for each frequency index $n=mB$ is

$$W_n = \int_0^\pi \frac{|c_n|^2}{2\rho_0 c_0} 2\pi r_0^2 \sin\alpha d\alpha, \tag{15}$$

where complex $c_{n=mB}$ is given in Eqs. (13) and (14). For the same sound frequency of index $n=mB$, there are acoustic interferences among sounds generated from different source frequencies (due to the Doppler effect) and among sounds

from different force components at the same frequency. Analytical progress is only possible by considering the dominant order of $\nu = n - \lambda$ for each force component, and by assuming that the three force components of the same frequency share the same phase angle as they are all related to the same unsteady pressure normal to the local blade surface.

As shown in Eq. (10), the Bessel functions decrease rapidly with the order ν, $J_\nu(z) \propto z^\nu$, the order with lowest $|\nu|$ is dominant, and only the dominant order is necessary as the typical value of argument $z = nM \sin\alpha$ is much less than unity for typical cooling fans. For a moderately high value of $z = 0.5$, the ratio of $J_{\nu+1}(z)/J_\nu(z)$ is 0.2582,0.1263, 0.0838,0.0627 for $\nu = 0,1,2,3$, respectively. The highest ratio of 0.2582 means a 11.8-dB difference, which is considered to be large enough. When considering the drag noise, however, the acoustic amplitude is proportional to $\nu J_\nu(z)$ and the ratio $(\nu+1)J_{\nu+1}(z)/[\nu J_\nu(z)]$ is 0.2526,0.1257,0.0836 for $\nu = 1,2,3$, respectively. Having said that, the error caused by neglecting the acoustic interference between a sound of amplitude 1 and another of amplitude 0.25 can be as big as $20\log_{10} 1/0.75 = 2.5$ dB. The maximum possible error will be reduced to about 1.1 dB for a smaller value of $z = 0.25$.

The presence of i in the radial noise term iE_{mB} in Eqs. (13) and (14) means that the radial noise is 90 deg out of phase with the thrust and drag noise in time domain, hence no interference. The interference between the thrust and drag noise for the dominant order ν is illustrated as

$$c_n = (aT_\lambda \cos\alpha + bD_\lambda)\sin^\nu \alpha,$$

where a and b are both real quantities, and the asymptotic formula of Eq. (10) has been employed. The sound power integration, Eq. (15), gives zero coupling between $T_\lambda D_\lambda$

$$W_n \propto \int_0^\pi (aT_\lambda \cos\alpha + bD_\lambda)^2 \sin^{2\nu+1}\alpha d\alpha$$

$$= C_T(\nu)a^2T_\lambda^2 + C_D(\nu)b^2D_\lambda^2,$$

$$C_T(\nu) = 2\int_0^1 x^2(1-x^2)^\nu dx,$$

$$C_D(\nu) = 2\int_0^1 (1-x^2)^\nu dx.$$

To conclude, the sound power contributed by $T_\lambda, D_\lambda, E_\lambda$ can be calculated individually and simply added up, the results being

$$W_n = \frac{(nBS\omega)^2}{4\pi\rho_0 c_0^3}[T_\lambda^2 W_{n\lambda}^{(T)} + D_\lambda^2 W_{n\lambda}^{(D)} + E_\lambda^2 W_{n\lambda}^{(E)}], \quad \nu = n - \lambda,$$

$$\mu = \frac{nM}{4},$$

Mode\Power	$W_{n\lambda}^{(T)}$	$W_{n\lambda}^{(D)}$	$W_{n\lambda}^{(E)}$
$\nu = 0$:	2/3	0	$5.02\mu^2$
$\nu = \pm 1$:	$1.25\mu^2$	0.392	1/3
$\nu = \pm 2$:	$0.667\mu^4$	$1.16\mu^2$	$1.25\mu^2$

(16)

where the new parameter $\mu = nM/4$ is introduced purely for the convenience of comparative study as most coefficients are then close to unity. Several observations are made.

(i) For coincident condition of $\nu = mB - kS = 0$, noise is mainly made by the unsteady thrust.

(ii) For the first rotational mode of $\nu = \pm 1$, the drag noise dominates over thrust noise in terms of noise-making ability. If one assumes a pitch angle of $\beta = 30°$, as shown in Fig. 1, and $nM = 0.5$, $T_\lambda = \sqrt{3}D_\lambda$, but still the drag noise is 14.3 and 8.3 dB higher than the thrust noise for $nM = 0.25,0.5$, respectively.

(iii) For the higher order $\nu = \pm 2$, the noise-making ability of all force components is reduced. The dominance of drag noise over thrust noise is even more pronounced. The physics of such dominance originates from the fact that rotating drag radiates noise by its rapid change of direction, while rotating thrust through its changing distance to the observer.

(iv) The radial noise is about the same as the drag noise for $\nu = \pm 1$ if $E_\lambda = O(D_\lambda)$, and its radiation ability at $\nu = 0$ is exactly 4 times higher than that at $\nu = 2$.

(v) Source radius R enters the expression through the rotation Mach number M. When the sound-power dependency does not contain M, the location of the source does not matter and the acoustic radiation is considered to be of the leading mode. This occurs for $\nu = 0$ in the thrust noise, $|\nu| = 1$ for the torque and radial noise, and the term ''leading'' is meant to differentiate from the ''coincident'' condition of $\nu = 0$. In these cases, the radial location of the source does not matter as far as the distributed forces are lumped properly in time domain if it occupies a significant portion of the circumference.

(vi) When the power dependency contains $(nM)^2$, the radius of the source point is coupled with the fluctuating forces. For drag D at $|\nu| = 2$, this makes a perfect expression for torque. For thrust noise at $|\nu| = 1$, it constitutes a bending moment, while that for the radial force does not have an obvious physical attribute. The drag noise should be renamed torque noise, but it is of secondary importance in terms of sound power. These modes are collectively called secondary modes, or simply higher-order modes.

(vii) If the modal coefficients of the unsteady forces, $T_\lambda, D_\lambda, E_\lambda$ are assumed to be proportional to the mean flow speed, the power dependency for a given geometrical shape can be expressed as $W \propto M^6 R^2 W_{n\lambda}^{(T,D,E)}$, where $W_{n\lambda}^{(T,D,E)}$ represents the dimensionless power from the three force components. For the leading mode radiation, $W_{n\lambda}^{(T,D,E)}$ is constant, so that the power exponent is 6. For secondary mode radiation, it becomes at least 8, depending on the exact circumferential order ν and force component concerned. In reality, the tonal sound is often superimposed on a broadband noise which could originate from the trailing edge scattering, which was shown by Ffowcs Williams and Hall (1970) to have a power dependency exponent of 5. The exponent for the total sound power is often between 5 and 6, which is well known empirically.

(viii) The exact parametric relationship for each component noise cannot be made explicit until theoretical models

are established to relate these force components with design parameters of B,S,ω,R as well as other features of the fan assembly. Mechanisms of lift fluctuation is better understood than those for fluctuating radial forces. However, it may be speculated that significant fluctuations of radial forces could arise from the blade tip and casing due to the motion of tip vortices. Another possible significant source is the leaning design of stator/struts for the purpose of causing destructive acoustic interference of thrust and drag noise.

IV. AERODYNAMIC INTERACTION MODEL

If the lift experienced by a rotor blade passing over S struts is S triangular pulses of height C_L, including rising and falling legs of equal time duration, and each occupies an angular width of $2\pi/S_L$, $S_L>S$, the Fourier component of the lift, L_λ, is found as follows:

$$L_{\lambda=kS}=\frac{S_L C_L}{S\pi^2 k^2}\left(1-\cos\frac{k\pi S}{S_L}\right),$$

where high S_L means short pulses typical of radial struts aligned with the trailing edges of the rotor blades. Although $S_L>S$ would normally be the case, S_L should be independent of S as the former describes the interaction between one rotor blade and one stator blade. Note that the oscillation term of $1-\cos(k\pi S/S_L)$ is associated with the precise pulse shape of rising and falling edges and their slopes, and the term disappears when the pulse has a parabolic shape. In general, it is difficult to determine the precise pulse shape; it would therefore be of more general value if the following relationship is assumed:

$$L_{\lambda=kS}=C_L/(Sk^2). \tag{17}$$

Here, C_L depends purely on the kinematics of how a strut interacts with blade rows, and it has also contained the effect of the pulse duration S_L. It is anticipated that a strut with considerable lean would have lower C_L in general, but surely C_L is not related to B or S. It is also noted that the model is only of limited value when it comes to the comparison of sounds generated at different BPFs since the potentially important oscillation term has been ignored. Note also that, when the individual rotor blade and strut encounter produces a very sharp pulse, the spectrum of unsteady lift approaches that of white noise, and the factor of k^{-2} in Eq. (17) would approach a constant, k^0.

Theoretically, the exact lift waveform can be measured by listening to the sound on the rotational axis, according to Eq. (11). A special configuration has to be made so that the rotor–strut interaction occurs simultaneously for all blades as $T(\tau)$ is the summation of thrust on all blades. Whether the result for such a special configuration can be applied to different combinations of B and S is uncertain.

In most cases, estimates of the unsteady radial forces are not available, radial noise is temporarily set aside. This leaves the thrust and drag noise. The thrust and drag forces are related to the lift by the pitch angle β. When k is replaced by $k=(n-\nu)/S$, and Eq. (17) is substituted into Eq. (16) for thrust and drag noise, the sound power of the mth BPF harmonic is rewritten as

$$W_m^{(TD)}=\frac{W_{n=mB}}{C_l}=\left(\frac{S}{1-\nu/n}\right)^4\left[\frac{W_{n\lambda}^{(D)}+W_{n\lambda}^{(T)}\tan^{-2}\beta}{m^2}\right],$$

$$C_l=\frac{(\omega C_L\sin\beta)^2}{4\pi\rho_0 c_0^3}, \tag{18}$$

where C_l is related to the unsteady lift coefficient and can be treated as a constant in the following discussions.

For the leading noise radiating modes, $\nu=0$ for thrust and $|\nu|=1$ for radial and drag noise, $W_{n\lambda}^{(T)}$ and $W_{n\lambda}^{(D)}$ are constants. This means that the strut number S plays a dominant role as $W_n\propto S^4$, while the rotor blade number does not feature in the power expression at all. For secondary modes, the power coefficients, $W_{n\lambda}^{(T,D)}$, contain μ^2, where $\mu=mBM$, so that B enters the sound-power expression as B^2, which is still less important than S^4. Note that B^2 would also appear in the noise power expression for the leading noise radiation modes if k^{-2} in the scaling law of Eq. (17) is replaced by k^{-1}.

The number of struts is determined partly by structural considerations. Given the power dependence of S^4, focus is therefore put on designs with small S. $S=2,3,4,5$ are all possible designs.

A. Analysis of typical cooling fan

The most typical design shown in Fig. 1(a) has $B=7$ rotor blades and $S=4$ struts behind the impeller. The particular fan in the photo has a hub and tip diameters of 37 and 86 mm, respectively. When it is operated at 3000 rpm in the free-delivery condition, the volume flow rate is around 0.03 m^3/s, while the maximum total pressure rise is about 50 Pa when it is fully blocked. The fan itself has one strut larger than the other three for carrying electric wiring. When this extra size is trimmed, the sound power is found to be 45 dB re: 10^{-12} W. The first 3 BPFs have a sound-power distribution of 44.8, 23.8, 31.0 dB. Details of the experimental setup are described in Wong and Huang (2003).

For this combination of B and S, $\nu=B-2S=-1$ is the dominating mode for the fundamental BPF. For the purpose of analysis, point force is assigned to be where the maximum steady flow loading normally occurs, such as 80% of the tip radius, $R=3.4$ cm, for which $BM\approx0.22$, $\mu=nBM/4=0.055,0.11,0.16$ for $m=1,2,3$, respectively. The pitch angle is assumed to be $\beta=30°$, $\tan^{-2}\beta=3$. Using Eq. (18), the comparison of sound radiating ability of drag and thrust noise is found as follows:

$$W_{n\lambda}^{(T)}=1.25\mu^2=0.0038,\quad W_{n\lambda}^{(D)}=0.392,$$

$$\frac{W_{n\lambda}^{(D)}}{W_{n\lambda}^{(T)}\tan^{-2}\beta}=34.4,$$

which means a difference of 15.4 dB. Despite the fact that thrust force component is 1.73 times larger than the drag force, the dominant noise source is still the drag noise as this is the leading radiation mode. Most radial noise would be

radiated from the strut and it is simply assumed here that it is also much less than the drag noise.

Despite the uncertainties associated with the Fourier transform of the rotor–strut interaction force in Eq. (17), the second and third harmonics are still estimated as follows for reference. The leading modes are $\nu = mB - kS = -1$ for the first BPF as $k=2$. For the second BPF, both $k=3$ gives $\nu = 2$ and $k=4$ gives $\nu = -2$. Both are secondary radiating modes, and their coupling effect cannot be determined unless the interaction force is measured. If their amplitudes are similar and they interfere constructively, the outcome could be a sound power 4 times higher than the estimate based on one of them. Bearing this uncertainty in mind, the estimate now proceeds for $k=3$ as it is likely to have a higher amplitude according to Eq. (17). For the third BPF, $\nu = 3B - 5S = 1$ is the only contributing mode. For all three frequencies, drag noise dominates. By inserting the various values of ν and ignoring the term $W_{n\lambda}^{(T)}$ in Eq. (18), one has

$$W_{m=1}^{(TD)} = 58.8, \quad W_{m=2}^{(TD)} = 1.7, \quad W_{m=3}^{(TD)} = 13.6,$$

$$\Sigma_{m=1}^{3} W_m^{(TD)} = 74.0;$$

$$\frac{W_{m=1}^{(TD)}}{W_{m=2}^{(TD)}} = 35.3, \quad (15.4 \text{ dB}); \quad \frac{W_{m=1}^{(TD)}}{W_{m=3}^{(TD)}} = 4.3, \quad (6.4 \text{ dB}).$$

Compared with experimental results, both the second and the third BPF are overpredicted, but the prediction that the third BPF is higher than the second agrees with the prediction. The numerical discrepancy can arise from many factors assumed in the interaction model of Eq. (17), which is only expected to give a general trend of force spectrum.

B. Effect of reduced number of struts

Using the above result as the reference case, it would be interesting to see how such sound power compares with the coincident thrust noise for $B=6$, $S=2$. In this case, $\nu = B - 3S = 0$, and $W_{n\lambda}^{(T)} = 2/3$. It is found that $W_{m=1}^{(TD)}|_{B=6,S=2} = 32.0$, which is still lower than the drag noise of the reference case. This example serves to highlight the equal importance of the leading drag noise and thrust noise at the coincident mode of $\nu = 0$. Having said that, the problem with the design of $B=6$, $S=2$ is that there is coincidence thrust noise at all higher harmonics.

The second case to be considered is $B=7$, $S=2$. Since coincidence occurs at the second harmonic, another case $B=7$, $S=3$ is also considered, which brings a third harmonic coincidence. As comparison varies from one harmonic to another, the first three harmonics are considered together for the total sound power

	$m=1$	2	3	$\Sigma_{m=1}^{3}$
$W_m^{(TD)}\|_{B=7,S=2}$	11.6	8.0	0.8	20.5
$W_m^{(TD)}\|_{B=7,S=3}$	58.8	0.2	18.0	77.0.

The better design seems to be $B=7$, $S=2$, which is predicted to radiate $10\log_{10}(74.0/20.5) = 5.6$ dB less noise than the reference case. However, there is a likelihood that structural weakness may require a large strut, and the predicted noise-power reduction may not be fully realized. It must be emphasized that, with such few struts, most harmonics are radiated by the leading modes of $\nu = 0, \pm 1$.

All the cases considered so far have a BPF noise radiated by the leading mode. The last case to be considered is $B=7$, $S=5$, for which the BPF noise is given by $\nu = 2$. The dominating modes are considered for six harmonics as follows:

$$\nu_{m=1} = B - S = 2, \quad \nu_{m=2} = 2B - 3S = -1,$$

$$\nu_{m=3} = 3B - 4S = 1,$$

$$\nu_{m=4} = 4B - 6S = -2, \quad \nu_{m=5} = 5B - 7S = 0,$$

$$\nu_{m=6} = 6B - 8S = 2,$$

$$W_{m=1\to 6}^{(TD)} = 8.4, \quad 46.5, \quad 33.1, \quad 1.7, \quad 50.0, \quad 2.7;$$

$$\Sigma_{m=1}^{3}: 88.0; \quad \Sigma_{m=1}^{6}: 142.3.$$

Here, the fifth harmonic dominates despite the natural decay with m. This prediction may not materialize in practice as the phase angle of forces may not be so well synchronized on all blades at very high harmonics due to uncertain aerodynamics of vortex shedding, etc. From the point of view of prediction, the sum of the first three harmonics amounts to the similar level of power as the reference design, which represents a trade-off between the increased S and increased ν for the fundamental BPF. As a summary, the optimal design for $B=7$ seems to be $S=2$ according to the interaction model of Eq. (17).

V. CONCLUSIONS

(1) Unsteady forces can be conveniently separated into axial thrust, circumferential drag, and radial force. Thrust noise radiates efficiently when the number of rotor blade coincides with the struts as $\nu = mB - kS = 0$, but the drag noise and radial noise peaks at the first mode of $\nu = \pm 1$ with very similar radiation efficiency. The traditional design of $\nu = \pm 1$ is thus questionable. The leading radiation mode of drag/radial forces changes from $\nu = 0$ to $\nu = \pm 1$ due to the constant change of force direction. The modes of $|\nu| = 1$ for thrust and $|\nu| = 2$ for drag/radial forces are regarded as secondary as they radiate much less noise. In terms of noise-power law with regard to the fan rotational speed, the leading modes have a sixth-power law, while the secondary modes have an eigth-power law.

(2) Radial noise is formulated for the first time, and is shown to be 90° out of phase with the thrust and drag noise if the force derived from one physical source. The natural radiation mode is also $\nu = \pm 1$. Unlike drag noise, it also radiates at $\nu = 0$. Little has been done or known about this noise component. It is speculated that significant radial noise can be present where tip vortices are strong and structural irregularities exist near the tip regions of the struts and rotor blades.

(3) It is shown analytically that there is no sound radiation power coupling among the three force components as they are orthogonal to each other. Thrust noise and drag noise are separated by directivity differences, while ra-

dial noise and thrust/drag noise can be separated by phase differences. The total sound power at the same frequency is the sum of three component sounds.

(4) The effects of source noncompactness have been quantified. The axial noncompactness is normally negligible for cooling fans operating at low tip speed. In the radial direction, it is also possible to lump the forces at one radius without losing any accuracy. For the leading mode of $\nu=0$ for thrust and $\nu=\pm 1$ for drag/radial noise, the radial position of the forces does not matter. For the secondary modes of $|\nu|=1$ for thrust and $|\nu|=2$ for drag/radial forces, the first-order moment of these forces should be found. For the drag force, the moment is exactly torque. For the thrust force, it becomes the bending moment. An effective mean radius can thus be defined.

(5) In the circumferential direction, no fan is too small to be noncompact. The noncompactness should be overcome by lumping distributed force with a time-delayed integration of distributed force on the circumference, and the time delay required takes the form of $p(\Theta,\tau-\Theta(\lambda-n)/(\lambda\omega))$, where Θ is the angular coordinate of the distributed force. For $\nu=n-\lambda=0$, no time delay is needed. For $n=0$, the time delay means rewinding the blade back to a fixed point in space. For intermediate modes of $\nu=-1,-2,\ldots$, a small amount of backwinding is required, while for $\nu>0$, time advance occurs. As different time delay is required, no universal spatial-averaging scheme can be devised for point-force representation. However, it is generally sufficient to tackle the leading radiation modes, for which the required time delay is small and the simultaneous force integration suffices.

(6) Acoustic directivity patterns are identified for thrust, drag, and radial noise. Thrust noise has the distinct four-lobe pattern when radiating at the noncoincident mode, while a unique axial two-lobe pattern is formed when it operates at the coincident mode. This noise component can be easily separated from the rest by taking the upstream–downstream differential measurement. Drag noise and radial noise cannot be distinguished from the directivity pattern, but their separation could be devised in time-domain operations provided that the radial force and circumferential force originate from the same surface pressure fluctuation on blade surfaces. The interaction of thrust noise with drag/radial noise would create a variety of acoustic directivity.

(7) Acoustic optimization has to rely on detailed knowledge of the fluctuation pressure signature caused by the encounter between one rotor blade and one strut. When the encounter produces a sharp pulse, the radiated noise is expected to be rich in higher harmonics, and the number of rotor blades is expected to be important for the total sound power. If, however, the encounter produces a broader waveform signature with rapid spectral roll-off, the number of struts is expected to be the dominant factor. For the particular example of seven-blade rotor, a design of two strut is predicted to radiate much less noise than a four-strut design.

ACKNOWLEDGMENT

The project is funded by a Competitive Earmarked Research Grant from the Hong Kong SAR Government (PolyU 5162/01E).

Boltezar, M., Mesaric, M., and Kuhelj, A. (**1998**). ''The influence of uneven blade spacing on the SPL and noise spectra radiated from radial fans,'' J. Sound Vib. **216**, 697–711.

Brungart, T. A., Lauchle, G. C., and Ramanujam, R. K. (**1999**). ''Installation effects on fan acoustic and aerodynamic performance,'' Noise Control Eng. J. **47**, 3–7.

Curle, N. (**1955**). ''The influences of solid boundaries upon aerodynamic sound,'' Proc. R. Soc. London, Ser. A **231**, 505–514.

Envia, E., and Nallasamy, M. (**1999**). ''Design selection and analysis of a swept and leaned stator concept,'' J. Sound Vib. **228**, 793–836.

Ffowcs Williams, J. E., and Hall, L. H. (**1970**). ''Aerodynamic sound generation by turbulent flow in the vicinity of a scattering half-plane,'' J. Fluid Mech. **40**, 657–670.

Ffowcs Williams, J. E., and Hawkings, D. L. (**1969**). ''Sound generation by turbulence and by surfaces in arbitrary motion,'' Philos. Trans. R. Soc. London, Ser. A **264**, 321–342.

Fitzgerald, J. M., and Lauchle, G. C. (**1984**). ''Reduction of discrete frequency noise in small, subsonic axial-flow fans,'' J. Acoust. Soc. Am. **76**, 158–166.

Fukano, T., Takamatsu, Y., and Kodama, Y. (**1986**). ''The effects of tip clearance on the noise of low pressure axial and mixed flow fans,'' J. Sound Vib. **105**, 291–308.

Gerhold, C. H. (**1997**). ''Active control of fan-generated tone noise,'' AIAA J. **35**, 17–22.

Glegg, S. A. L., and Jochault, C. (**1998**). ''Broadband self-noise from a ducted fan,'' AIAA J. **36**, 1387–1395.

Gutin, L. (**1936**). ''On the sound field of a rotating propeller,'' *NACA TM 1195* (Translated in 1948 from Zh. Tekh. Fiz. **6**, 899–909).

Howe, M. S. (**1999**). ''Trailing edge noise at low Mach numbers,'' J. Sound Vib. **225**, 211–238.

Kaji, S., and Okazaki, T. (**1970**). ''Generation of sound by rotor–stator interaction,'' J. Sound Vib. **13**, 281–307.

Kemp, N. H., and Sears, W. R. (**1953**). ''Aerodynamic interference between moving blade rows,'' J. Aeronaut. Sci. **20**, 583–598.

Kemp, N. H., and Sears, W. R. (**1955**). ''Unsteady forces due to viscous wakes in turbomachines,'' J. Aeronaut. Sci. **22**, 478–483.

Lauchle, G. C., MacGillivray, J. R., and Swanson, D. C. (**1997**). ''Active control of axial-flow fan noise,'' J. Acoust. Soc. Am. **101**, 341–349.

Lewy, S. (**1992**). ''Theoretical study of the acoustic benefit of an open rotor with uneven blade spacings,'' J. Acoust. Soc. Am. **92**, 2181–2185.

Lighthill, M. J. (**1952**). ''On sound generated aerodynamically. I. General theory,'' Proc. R. Soc. London, Ser. A **211**, 564–587.

Longhouse, R. E. (**1976**). ''Noise mechanism separation and design considerations for low tip-speed, axial-flow fans,'' J. Sound Vib. **48**, 461–474.

Longhouse, R. E. (**1977**). ''Vortex shedding noise of low tip speed, axial flow fans,'' J. Sound Vib. **53**, 25–46.

Lowson, M. V. (**1965**). ''The sound field for singularities in motion,'' Proc. R. Soc. London, Ser. A **286**, 559–572.

Lowson, M. V. (**1970**). ''Theoretical analysis of compressor noise,'' J. Acoust. Soc. Am. **47**, 371–385.

Majumdar, S. J., and Peak, N. (**1998**). ''Noise generation by the interaction between ingested turbulence and a rotating fan,'' J. Fluid Mech. **359**, 181–216.

Morfey, C. L. (**1973**). ''Rotating blades and aerodynamic sound,'' J. Sound Vib. **28**, 587–617.

Mugridge, B. D., and Morfey, C. L. (**1972**). ''Sources of noise in axial flow fans,'' J. Acoust. Soc. Am. **51**, 1411–1426.

Quinlan, D. A. (**1992**). ''Application of active control to axial flow fans,'' Noise Control Eng. J. **39**, 95–101.

Quinlan, Q. A., and Bent, P. H. (**1998**). ''High frequency noise generation in small axial flow fans,'' J. Sound Vib. **218**, 177–204.

Schulten, J. B. H. M. (**1997**). ''Vane sweep effects on rotor/stator interaction noise,'' AIAA J. **35**, 945–951.

Sharland, I. J. (**1964**). ''Sources of noise in axial flow fans,'' J. Sound Vib. **1**, 302–322.

Subramanian, S., and Mueller, T. J. (**1995**). ''An experimental study of propeller noise due to cyclic flow distorsion,'' J. Sound Vib. **183**, 907–923.
Trunzo, R., Lakshminarayana, B., and Thompson, D. E. (**1981**). ''Nature of inlet turbulence and strut flow disturbances and their effect on turbomachinery rotor noise,'' J. Sound Vib. **76**, 233–259.
Tyler, J. M., and Sofrin, T. G. (**1962**). ''Axial flow compressor noise studies,'' SAE Trans. **70**, 309–332.
Wong, J., and Huang, L. (**2003**). ''Identification and control of noise sources in a small axial-flow cooling fan,'' Symposium on Fan Noise 2003, Senlis, France, 23–25 September.

Active local control of propeller-aircraft run-up noise

Murray Hodgson,[a] Jingnan Guo,[b] and Pierre Germain[c]
School of Occupational and Environmental Hygiene and Department of Mechanical Engineering, University of British Columbia, 3rd Floor, 2206 East Mall, Vancouver, British Columbia V6T 1Z3, Canada

(Received 6 September 2003; accepted for publication 30 September 2003)

Engine run-ups are part of the regular maintenance schedule at Vancouver International Airport. The noise generated by the run-ups propagates into neighboring communities, disturbing the residents. Active noise control is a potentially cost-effective alternative to passive methods, such as enclosures. Propeller aircraft generate low-frequency tonal noise that is highly compatible with active control. This paper presents a preliminary investigation of the feasibility and effectiveness of controlling run-up noise from propeller aircraft using local active control. Computer simulations for different configurations of multi-channel active-noise-control systems, aimed at reducing run-up noise in adjacent residential areas using a local-control strategy, were performed. These were based on an optimal configuration of a single-channel control system studied previously. The variations of the attenuation and amplification zones with the number of control channels, and with source/control-system geometry, were studied. Here, the aircraft was modeled using one or two sources, with monopole or multipole radiation patterns. Both free-field and half-space conditions were considered; for the configurations studied, results were similar in the two cases. In both cases, large triangular quiet zones, with local attenuations of 10 dB or more, were obtained when nine or more control channels were used. Increases of noise were predicted outside of these areas, but these were minimized as more control channels were employed. By combining predicted attenuations with measured noise spectra, noise levels after implementation of an active control system were estimated. [DOI: 10.1121/1.1628248]

PACS numbers: 43.50.Ki [KAC] Pages: 3201–3210

I. INTRODUCTION

Nighttime aircraft run-ups at the Vancouver International Airport (YVR) often cause noise-pollution problems in the neighboring communities to the north and south of the airport,[1] even though the affected communities are several kilometers away from the run-up areas. These run-ups consist of revving-up the engine, sometimes to full power, as part of the normal maintenance of the aircraft. Measurements of the run-up noise indicate that it is dominated by low-frequency, pure-tone components starting around 100 Hz. This low-frequency characteristic is, in particular, evident in the affected residential areas.[2] Traditional run-up noise-control measures, such as enclosures or pens, can be very costly, and ineffective at controlling low-frequency noise.

Active noise control (ANC), on the other hand, is particularly effective in the case of low-frequency, pure-tone noise, and is also potentially cost-effective.[3,4] Therefore, applying this novel noise-control technology to reducing aircraft run-up noise at the airport may be a practical option. Previous work has demonstrated that an ANC system works with a run-up hush house, is able to significantly attenuate low-frequency jet-engine noise, and can reduce the overall jet-engine run-up noise on the ground by at least 5 dB in a small area around the hush house. The potential of obtaining run-up noise reduction by ANC over a large area has also been demonstrated.[5] The study presented in this paper investigated the potential performance of ANC systems alone in reducing aircraft run-up noise. The ANC systems involved were multi-channel, local active-noise-control systems, which are shown to have the potential to attenuate run-up noise over a very large area, and to create quiet zones covering the affected residential areas to the north and south of the airport, as shown in Fig. 1. Here, a quiet zone is defined as an area in which the noise has been attenuated by at least 10 dB.

The basic idea behind implementing a local ANC system at a run-up site at YVR is that the multiple control sources cancel the noise in the residential areas by moving the noise energy to the west of the airport—i.e., over the sea. This involves the mechanism of active local noise control in an open (i.e., nonenclosed) environment. Previous studies of ANC in open environments found that the performance of a local ANC system depends on the exact configuration of the control system, the frequency band of the noise, and the number of control channels.[6,7] They also found that the prevailing physical environment affecting the sound field—for example, the presence of a reflective ground—also affects the performance of the ANC system.[8,9] In the present study, how the size of the quiet zone created by an ANC system is affected by the number of control channels, and the configuration of the ANC system, were investigated quantitatively assuming free-field conditions (simulating the effect of very absorptive ground). The effectiveness of ANC for various run-up-noise harmonics was investigated. Also studied was

[a]Electronic mail: hodgson@mech.ubc.ca
[b]Current address: WorkSafe, Department of Consumer and Employment Protection, 1260 Hay Street, West Perth, WA 6005, Australia.
[c]Current address: Valcoustics Canada Ltd., 30 Wertheim Court, Unit 25, Richmond Hill, ON L4S 1B9, Canada.

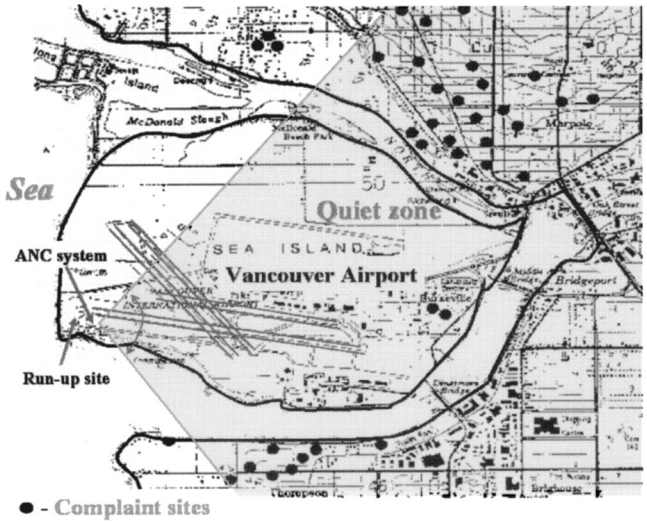


FIG. 1. Predicted quiet zone created by a local ANC system at YVR Airport.

the effect of reflective ground on ANC-system performance, relative to that in the free field since, at low frequency, real grounds represent boundaries that are between these two extremes.

The effectiveness of a local ANC system also depends on the primary noise field and, therefore, on the source's radiation characteristics. Previous studies of ANC in open environments always assumed a point, monopole primary source[6,7] or a noncompact primary source (a source with non-negligible dimensions).[10] However, previous studies have shown that the noise radiated by aircraft run-up is very complex.[11] The radiation pattern is neither that of a monopole, nor of a simple compact source, but is that of a combination of multipoles. The performance of an ANC system was, therefore, investigated in the case of complex primary noise sources, for different system configurations and orientations, leading to new information on the optimal arrangements of the ANC system for controlling propeller-aircraft run-up noise. In summary, this study is a preliminary investigation of the feasibility and effectiveness of applying ANC to run-up noise at YVR, and of how to optimize the ANC system.

II. THEORETICAL BASIS OF ACTIVE LOCAL CONTROL IN OPEN ENVIRONMENTS

There are two major control strategies available when active noise control is applied in open environments—global control and local control. Although global control can result in noise attenuation at all locations in space, it requires that the control sources be placed very close—less than one-quarter of a wavelength—to the primary source to be effective.[3] This condition is very often difficult to fulfill in practice—for example, in the case of aircraft run-up noise control. The local-control strategy, on the other hand, aims to reduce the noise in certain designated areas or directions by creating quiet zones; however, it may cause noise increases in other places. It has been found that the quiet zone created by a local control system can be very large, and the sound-

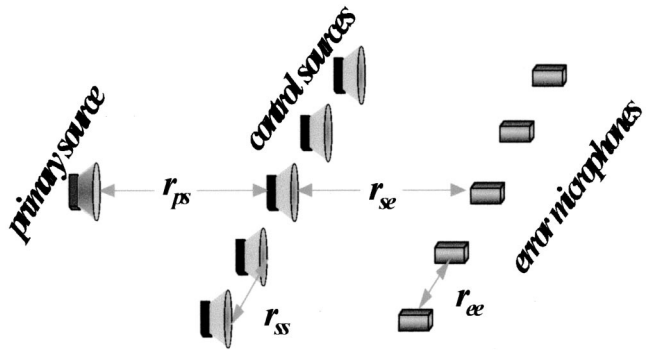


FIG. 2. Multiple-channel active-noise-control system.

pressure increase in the regions outside the quiet zone can be minimized, if the control system is optimally designed.[6]

For the multi-channel active-noise-control system shown in Fig. 2, where equal numbers M of control sources and error microphones are equi-spaced in two parallel lines, the sound pressures at the M error microphones are

$$\mathbf{P}=\mathbf{Z}_{pe}q_p+\mathbf{Z}_{\mathbf{se}}\mathbf{q}_s, \tag{1}$$

in which $\mathbf{Z}_{pe}$ is the vector of acoustical transfer impedances from the primary source to the M error sensors, and $\mathbf{Z}_{\mathbf{se}}$ is the $M\times M$ matrix of acoustical transfer impedances from the secondary sources to the error sensors. q_p and $\mathbf{q}_\mathbf{s}$ are the strengths of the primary source and the M control sources, respectively. Adopting a local-control strategy, the sum of the squared sound pressures at the microphone positions is selected as the ''cost function'' J, with

$$J=\mathbf{P}^H\mathbf{P}=(\mathbf{Z}_{pe}q_p+\mathbf{Z}_{\mathbf{se}}\mathbf{q}_s)^H(\mathbf{Z}_{pe}q_p+\mathbf{Z}_{\mathbf{se}}\mathbf{q}_s), \tag{2}$$

where the superscript H denotes the Hermitian transpose. Then the sound pressures at the microphones can be minimized, and a quiet zone can be created around the microphones. Theoretically, the sound pressures at the microphones can be minimized to zero. The strengths of the secondary sources in the array, minimizing the cost function, are

$$\mathbf{q}_{s0}=-\mathbf{Z}_{\mathbf{se}}^{-1}\mathbf{Z}_{pe}q_p. \tag{3}$$

The sound pressure at any position in the space after control is

$$p=Z_pq_p-\mathbf{Z}_s\mathbf{Z}_{\mathbf{se}}^{-1}\mathbf{Z}_{pe}q_p, \tag{4}$$

and the sound attenuation can be calculated by

$$\Delta L=20\log(|Z_pq_p-\mathbf{Z}_s\mathbf{Z}_{\mathbf{se}}^{-1}\mathbf{Z}_{pe}q_p|/|Z_pq_p|), \tag{5}$$

where Z_p is the acoustic transfer impedance from the primary source to the observation position, and $\mathbf{Z}_s$ is the row vector of acoustical transfer impedances from the M control sources to the observation location.

It has been found that, for the control system shown in Fig. 2, when the number of control channels, the noise frequency to be controlled, the separation between the primary source and the control sources r_{ps}, and the separation between the control-source and the error-microphone arrays r_{se} are all determined, there exists an optimal range of the control-source separation r_{ss} defined as[6]

$$r_{ss-\max} \cong \begin{cases} \frac{\lambda}{2}\sqrt{1+\frac{4r_{se}}{N\lambda}}, & N=2n(n=1,2,3,...,10), \\ \frac{\lambda}{2}\sqrt{1+\frac{N+1}{N-1}\frac{4r_{se}}{N\lambda}}, & N=2n+1(n=1,2,3,...,10); \end{cases} \tag{6}$$

$$r_{ss-\min} \cong \begin{cases} \sqrt{1+\frac{4}{N}}e^{-[2\lambda/r_{se}+\sqrt{\lambda/r_{ps}}]}, & N=2n(n=2,3,...,10), \\ \sqrt{1+\frac{6(N+1)}{N(N-1)}}e^{-[2\lambda/r_{se}+\sqrt{\lambda/r_{ps}}]}, & N=2n+1(n=1,2,3...,10), \end{cases} \tag{7}$$

in which λ is the sound wavelength at the frequency of interest. If the system is designed to respect these optimal ranges, the control system is able to create the largest quiet zone at the cost of the least sound-pressure increase at other locations.

III. PROPELLER-AIRCRAFT NOISE

Noise generated by a propeller is typical of fan noise, which is tonal in nature and comprised of spectral peaks at the blade-passing frequency (BPF) and its harmonics. In order to study the characteristics of propeller-aircraft noise during an engine run-up, measurements were conducted on a typical propeller aircraft—a Beechcraft 1900D twin-engined turboprop—at Vancouver International Airport. The detailed analysis of the propeller aircraft noise was presented elsewhere.[2] The A-weighted noise spectra at near-field (76 m) and far-field (e.g., in the residential area about 3 km away) locations are presented in Figs. 3(a) and (b), respectively; the low-frequency pure-tones are very obvious. Many harmonics appear as sharp peaks at multiples of the fundamental frequency of 112 Hz, i.e., at 224, 336, 448 Hz, and so on. The fundamental and lower harmonics have lower levels due to the A-weighting. At far-field locations the fundamental and first two harmonics are much higher, due to the faster spatial decay of the higher-frequency components with distance during noise propagation. Thus, successful reduction of the first few low-frequency peaks would significantly reduce the impact of the run-up noise on the nearby residents.

The radiation patterns generated by aircraft propellers vary with rotational speed, which affects the various aerodynamic mechanisms; the resulting patterns can be monopoles, dipoles, quadrupoles, or their combinations.[11] Since run-up propeller speeds vary from one aircraft to another, it is difficult to generalize as to which mechanism will govern the run-up noise generated. Figure 4 shows a total A-weighted noise contour measured around the Beechcraft 1900D twin-propeller aircraft. It is very clear that the propeller-aircraft noise has very strong directivity, and cannot simply be treated as generated by a monopole source. In the following ANC simulations, the effectiveness of active-noise-control systems will be predicted for monopole, dipole, and quadrupole sources.

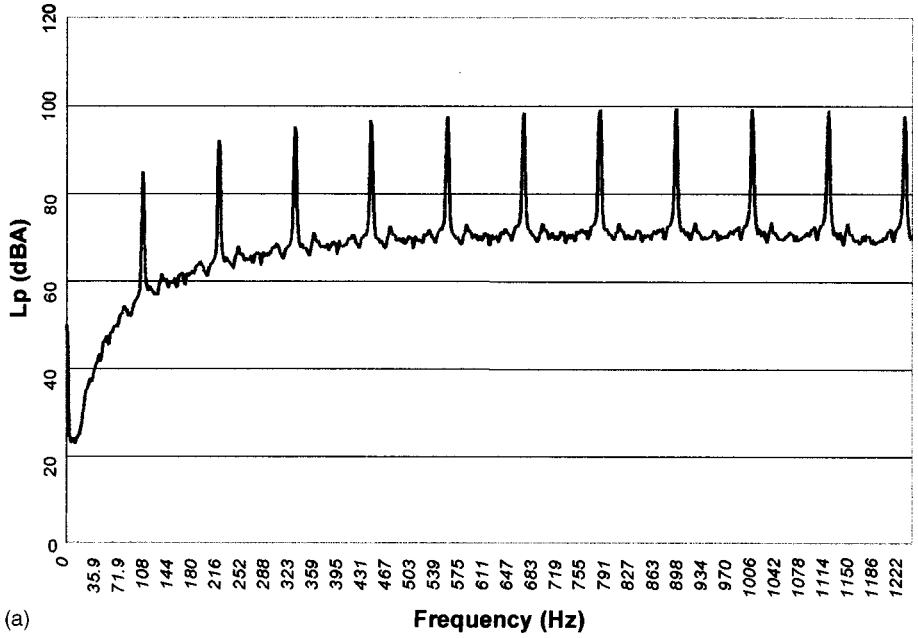
IV. EFFECTIVENESS OF A LOCAL ACTIVE-CONTROL SYSTEM (MONOPOLE SOURCE)

The control effectiveness of a local active-control system is indicated by the size of the quiet zones it creates. Previous studies have demonstrated that an optimally arranged multi-channel control system is able to create a large quiet zone, which is in the shape of a wedge.[6] Its cross-section in the plane containing the control system is shown in Fig. 5. It is clear that, to increase local control effectiveness, the quiet zone angle θ_{qz} must be increased. It has been verified in this study that, even when designed in the optimal range, the size of the quiet zone created by a multi-channel, active-noise-control system for aircraft run-up noise depends on the number of control channels, as well as the configuration of the control system—for example, on r_{ps} and r_{se}. The control efficiency is also affected by the frequency components of the run-up noise, as well as ground around the run-up site. In the following discussion, although the control efficiency is only demonstrated for certain specific system configurations, the conclusions are applicable to other practical system configurations for controlling aircraft run-up noise. In figures showing the attenuation provided by an active-control system, since the attenuation in the quiet-zone can fluctuate strongly, making the extent of the quiet zone difficult to appreciate, attenuations of greater than 10 dB are plotted as attenuations of 10 dB, for clarity.

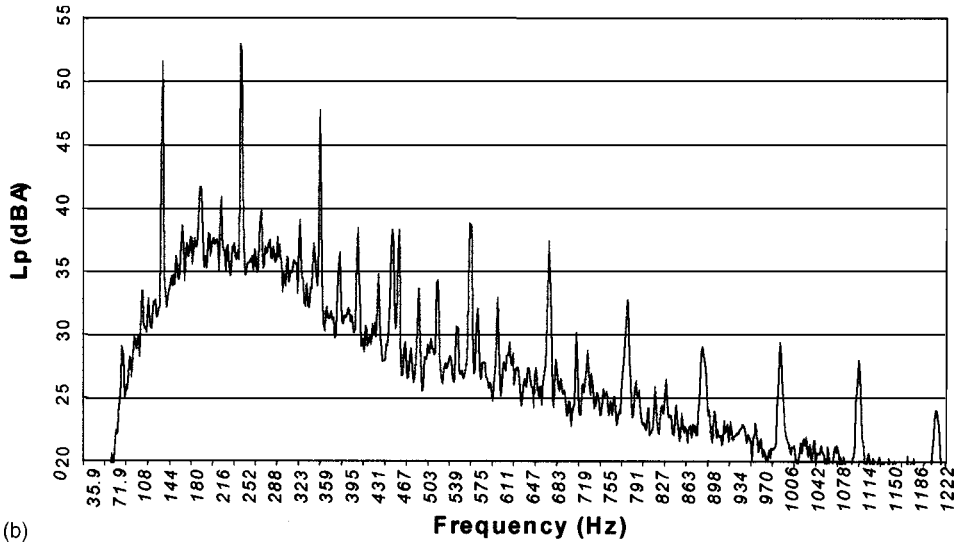
A. Effect of number of control channels on quiet-zone size

It is reasonable to assume that more control channels will result in a larger quiet zone. Therefore, predictions of the quiet-zone angle were performed for multiple-channel control systems with 3, 9, 21, 41, and 61 channels. In this analysis, a possible practical system configuration for real aircraft run-up noise control—with $r_{ps}=20$ m and $r_{se}=40$ m—was chosen as a demonstration. The optimal range of r_{ss} for each configuration was calculated using Eqs. (6) and (7). The test frequency was the Beechcraft blade-passing frequency of 112 Hz. The sound-pressure attenuations generated by control systems with 3 and 41 channels are shown in Figs. 6(a) and (b), respectively.

Figure 6 indicates that the area over which sound pressure is attenuated increases dramatically with the number of control channels. Again, in this and other figures showing the attenuation provided by an active-control system, since the



(a)



(b)

FIG. 3. A-weighted spectra of a Beechcraft 1900D measured at distances of (a) 76 m and (b) 3 km from the aircraft.

attenuation in the quiet-zone can fluctuate strongly, making the extent of the quiet zone difficult to appreciate, attenuations of greater than 10 dB are plotted as attenuations of 10 dB, for clarity. Figure 7 shows the horizontal extent of the quiet zones generated by control systems with 3, 9, 21, 41, and 61 control channels. The configuration of the control system remained unchanged in these simulations, except that the separation of the control sources r_{ss} was selected within the optimal range given by Eqs. (6) and (7) for each control system. They are $r_{ss}=5.2$, 3.7, 2.5, 1.85, and 1.7 m, respectively.

It is very clear that the quiet-zone size, as described by the quiet-zone angle, increases significantly with an increasing number of control sources. Figure 8 plots the relationship between the quiet-zone angles obtained and the number of control channels. The results show clearly that the quiet-zone size does, in fact, increase as more channels are added to the ANC system. The quiet-zone angle is about 20° when there are six control channels. However, it increases to over 80° when the number of control channels increases to 56. The best-fit curve of the relationship between quiet-zone angle and the number of control channels for this control system follows a power law as follows:

$$\theta_{qz}=8.1N^{0.56}, \tag{8}$$

where N is the number of channels. Though it is specific to the configuration studied, this law is useful in deciding the number of control sources to use to achieve a desired quiet-zone angle.

B. Effect of primary-source to control-source distance on quiet-zone size

As mentioned previously, active global control can be achieved in open environments only when the control sources can be located very close to the primary source. For a local-control strategy, the distances between the control sources and the primary source also play very important

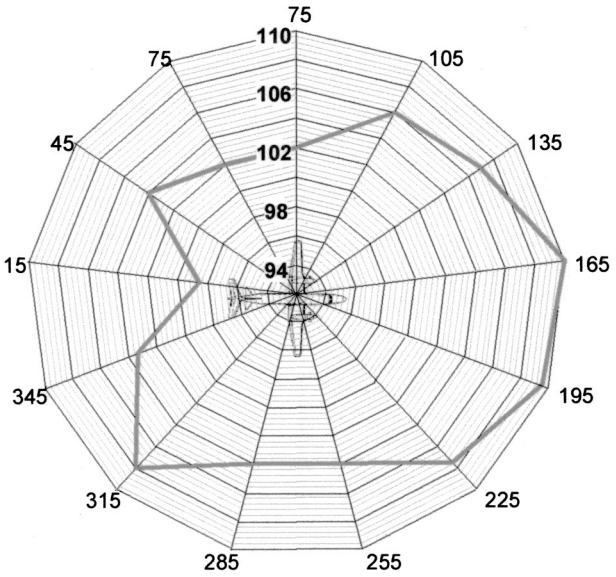


FIG. 4. Total A-weighted sound-pressure-level contour around the Beechcraft 1900D aircraft.

roles in determining control effectiveness. It has been concluded that, for a one-channel active-noise-control system, the size of the quiet zone becomes larger with shorter distance between the primary and control sources.[6] The conclusion of the present study is that, for a multi-channel control system, the size of quiet zone also increases as the control sources are positioned closer to the primary source. It was, therefore, of interest to study the relationship between the quiet-zone angle θ_{qz} and r_{ps}.

Figure 9 illustrates this relationship for a 21-channel system. Here $r_{ps}=1$, 5, 10, 20, and 40 m were examined; r_{se} was varied accordingly, maintaining $r_{se}/r_{ps}=2.0$. Figure 9 shows that a clear trend emerges, along with the associated best-fit regression curve. The best-fit curve again follows a power law, in this case as follows:

$$\theta_{qz}=258.1r_{ps}^{-0.56}. \tag{9}$$

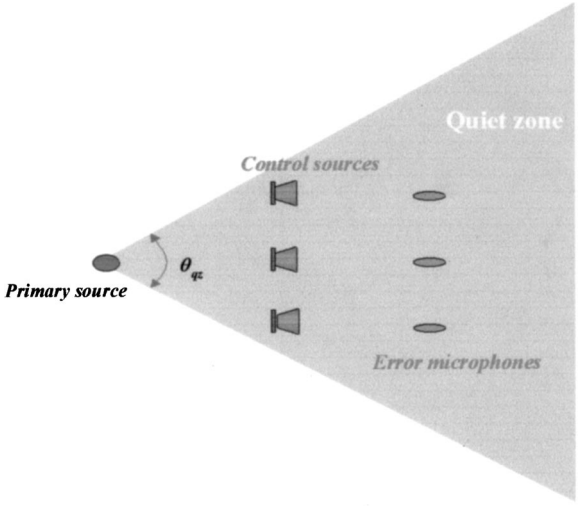


FIG. 5. Quiet zone created by the multi-channel active-control system shown in Fig. 2.

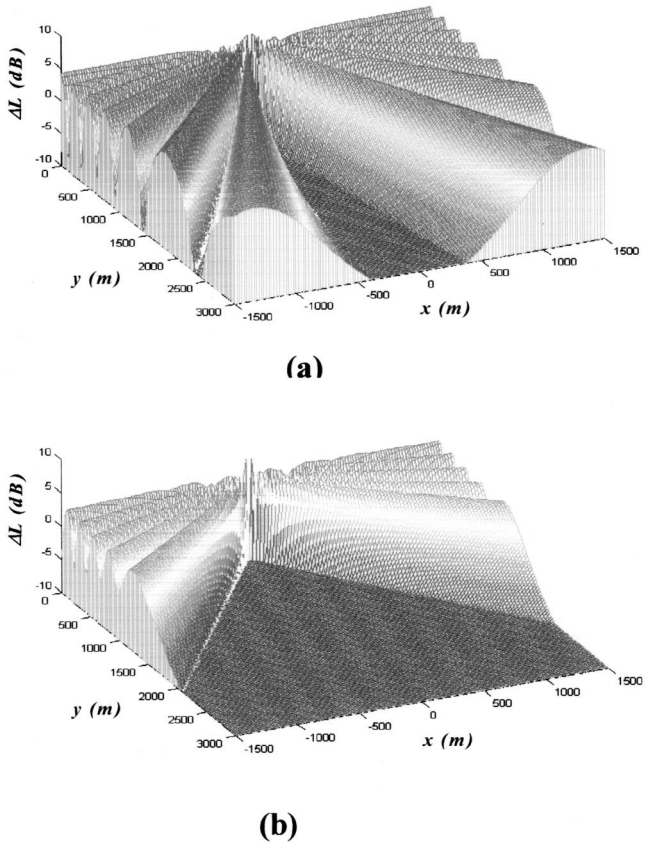


FIG. 6. Sound-pressure attenuations by an ANC system with (a) 3 and (b) 41 control channels.

Note, when using this curve, that the control system becomes global when $r_{ps}<\lambda/4$; therefore, it is only valid for $r_{ps}>0.77$ m, for the 112-Hz frequency. The maximum quiet-zone angle that can be achieved with this 21-channel local-control system is 169°. Though it is specific to the configuration studied, this equation is very useful when a particular quiet-zone angle is sought, and the primary-source to secondary-source distance r_{ps} must be determined.

C. Effect on harmonics

So far, systems targeting only the 112-Hz BPF have been studied, and only results for that frequency presented.

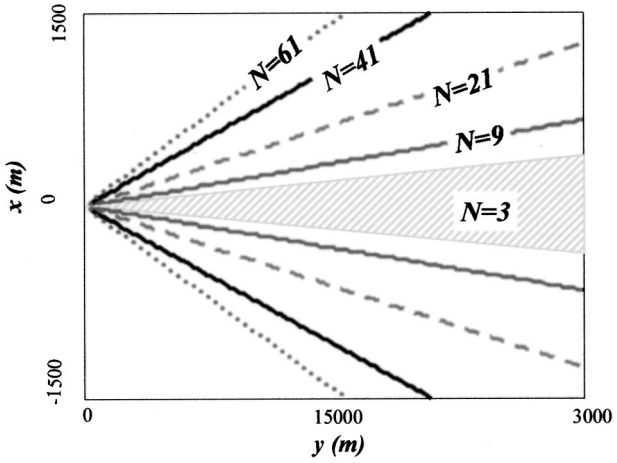


FIG. 7. Quiet zones created by the ANC system with 3, 9, 21, 41, and 61 channels.

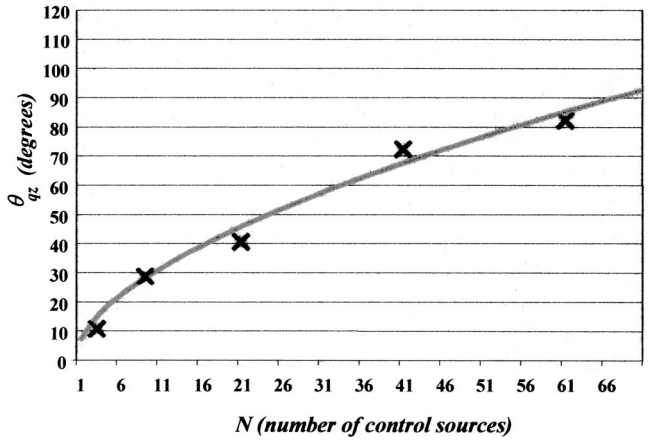


FIG. 8. Variation of quiet-zone angle θ_{qz} with the number of control sources N: (×) prediction; (———) best-fit regression curve.

Since the ultimate goal of this study includes minimizing the total A-weighted sound pressure in certain residential areas, the active-noise-control system needs to be designed not only for the fundamental BPF noise, but also for other low-frequency harmonics, such as the first four BPF harmonics of 224, 336, 448, and 560 Hz. The active-noise-control system needs to be designed to cover as many harmonics as possible, and minimize the noise pressure increase in the areas outside the desired quiet zones.

As mentioned previously, the optimal configuration of a multi-channel control system depends on the noise frequency to be cancelled. A configuration of the system, optimally arranged for the fundamental frequency, may not remain effective or optimal for other harmonic frequencies. However, as the optimal configuration is given in a range, the effective noise frequency also lies in a range, which may cover higher harmonics of interest for control.[12] Calculated using Eqs. (6) and (7), Table I gives the optimal range for the BPF (f_0) of the Beechcraft 1900D, and for its first four harmonics (f_1, f_2, f_3, and f_4), using a nine-channel system configured with $r_{ps}=20$ m and $r_{se}=40$ m.

As seen in Table I, if a system is designed for the 112-Hz BPF, the optimal range only overlaps that for the first harmonic of 224 Hz; the optimal ranges for the other three

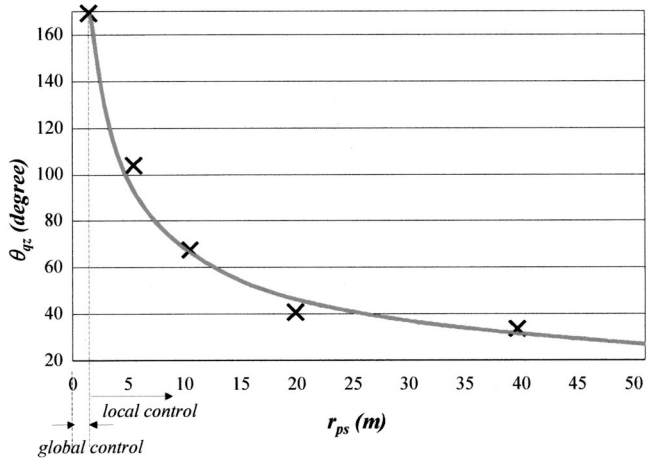


FIG. 9. Variation of predicted quiet-zone angle θ_{qz} with $r_{ps}=r_{se}$: (×) prediction; (———) best-fit regression curve.

TABLE I. Optimal ranges for a nine-channel active-noise-control system.

Frequency (Hz)	$r_{ss\text{-max}}$ (m)	$r_{ss\text{-min}}$ (m)
$f_0=112$	4.40	2.86
$f_1=224$	3.02	1.95
$f_2=336$	2.43	1.55
$f_3=448$	2.10	1.31
$f_4=560$	1.87	1.14

harmonics are far below this range. Therefore, the control system designed for the fundamental frequency is not able to create a quiet zone for harmonics other than the first one. To study the effects on the higher harmonics, simulations using $r_{ss}=2.9$ m, close to the lower limit for f_0, were performed. Since this value is also in the optimal range for f_1, favorable results would be expected for this frequency as well. As expected, it was found that, using this configuration, quiet zones are present only for f_0 and f_1. The other harmonics (f_2, f_3, and f_4) experience some attenuation, but in irregular shapes and patterns in "strips" on either side of the $x=0$ m centerline. Increases of as much as 15 dB outside of the quiet zones were seen for the f_2 and f_4 harmonics.

Another approach was, therefore, taken, this time using $r_{ss}=2.0$ m, a value close to the minimum for f_1. This optimal separation of control sources overlaps those for f_2 and f_3, though it is below the range for f_0 and above the range for f_4. Sound-attenuation simulations done using this value are shown in Fig. 10. As expected, better results are found for f_1, f_2, and f_3. A large quiet zone is also found for f_0, as shown in Fig. 10(a). However, since the r_{ss} value chosen is below the optimal range for this frequency, the sound-pressure increase outside the quiet zone is much higher—up to 10 dB. This might cause problems in the ANC application. Also as expected, the attenuation for f_4 can hardly be seen in Fig. 10(e).

It has been found that, in the area of residential complaints north of the run-up site, the problematic noise is mostly comprised of the fundamental and first harmonic components. Thus, for the example control system discussed here, the optimal separation can be designed for the fundamental frequency—i.e., $r_{ss}=2.9$ m. However, in some other areas, the noise components from higher harmonics may also contribute to the problem. In that case, the design of the optimal range for higher harmonic frequencies may be necessary, e.g., by choosing $r_{ss}=2.0$ m in this example.

D. Effect of reflective ground

Aircraft run-up occurs on a concrete ground. However, at YVR the ground between the run-up area and surrounding communities is, in general, grassland. It has been found that a sound-reflective ground changes the effectiveness of ANC relative to the case of absorptive ground.[8,9] In the present study, as well as free-field conditions simulating very absorptive ground, the effect of reflective ground on the active control of aircraft run-up noise was also investigated. This study of the half-space condition was done for a primary source fixed at the height of the propellers ($h_p=2$ m); the effect of varying the heights of the control-source and error-

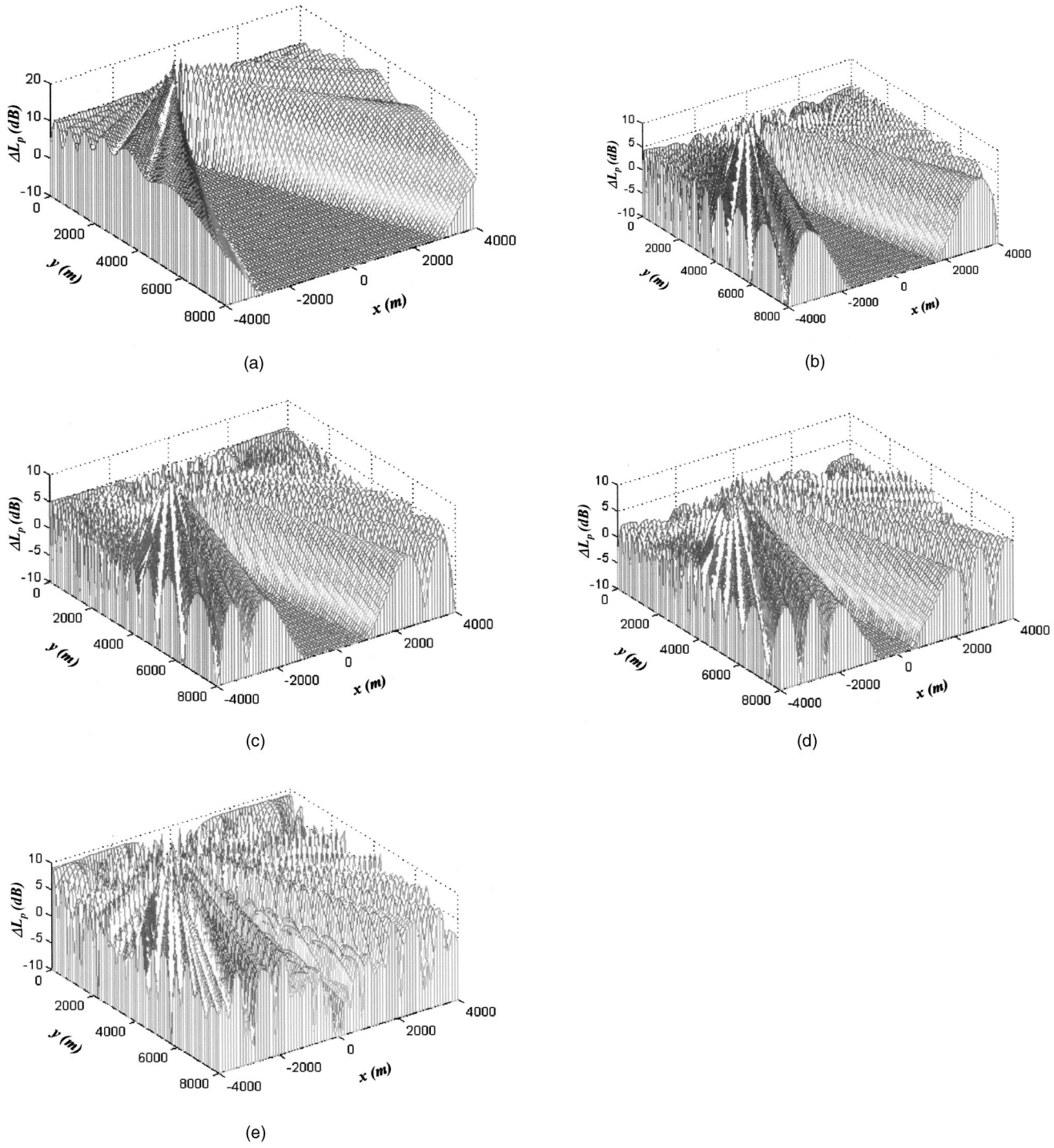


FIG. 10. Sound attenuation by a nine-channel active-control system with $r_{ss}=2.0$ m: (a) $f_0=112$ Hz; (b) $f_1=224$ Hz; (c) $f_2=336$ Hz; (d) $f_3=448$ Hz; and (e) $f_4=560$ Hz.

microphone arrays relative to the primary-source height was predicted. The previous optimal control-system configuration—with $r_{ps}=20$ m and $r_{se}=40\,m$—was again used for these simulations. The result of placing the control sources and error microphones on the ground ($h_c=h_e=0$ m) was compared to that with $h_c=h_e=2$ m. The latter configuration appeared to be slightly better, providing slightly more attenuation. It has previously been shown that, if the control source is at a different height from the primary source, the primary source, control source and error microphone should be aligned (see Fig. 11).[9] For instance, if the primary-source height is 2 m, and the control-source height is 4 m at a distance of 20 m, then an error microphone positioned 40 m from the control source should be at a height of 8 m. However, due to the fact that the complainant area is at about the same height as the aircraft propeller, the configuration of $h_c=h_e=2$ m was selected. These reflective-ground simulations showed that results for the free-field and half-space conditions are almost identical when $r_{se}/r_{ps}>1.0$, and that the optimal heights of the control-source and

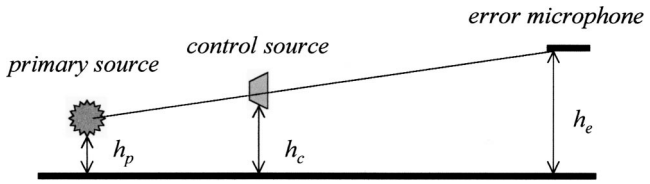


FIG. 11. Optimal alignment of an active-noise-control system.

error-microphone arrays are equal to the primary-source height.

E. Total noise-level reduction

The result of applying the attenuations provided by the nine-channel system optimized for f_1 (i.e., with $r_{ss}=2$ m) to the spectrum of the run-up noise measured in the community located approximately 3 km from Vancouver International Airport was studied. The attenuations of the main spectral peaks, predicted at a position in the center of the quiet zone at a distance 3 km from the primary source, are shown in Table II for both free-field and half-space conditions. Attenuations can be seen for the fundamental and first three harmonics. For the fourth harmonic (f_4), there is an increase in the half-space condition. This problematic increase can be avoided by filtering the reference signal through a low-pass filter with cut-off frequency set at about 500 Hz, so that signals at frequencies higher than 500 Hz are not reproduced by the control source.

Figure 12 shows how the attenuations in Table II affect the harmonics of the Beechcraft 1900D spectrum measured in the community, for one particular aircraft heading. To obtain these results, the predicted attenuations were applied to the measured noise spectrum. A low-pass filter at 500 Hz would have to be used to eliminate the possibility of an increase of sound pressure for harmonics higher than f_3. It can be seen that a total attenuation of 15 dB is provided by this system.

V. MULTIPOLE SOURCES

The simulations presented in the previous sections assumed a single monopole source radiating a spherical wavefront as the primary noise source. The resulting radiation is the same in all directions (omnidirectional). However, it has previously been shown that the radiation pattern generated by aircraft propellers varies with rotational speed, which affects the various aerodynamic mechanisms.[11] The resulting patterns can be monopoles, dipoles, or quadrupoles, or their combinations. Since run-up propeller speeds vary from one

TABLE II. Attenuations provided by an active noise control system optimized for f_1 ($r_{ss}=2.0$ m).

Frequency (Hz)	ΔL_p (dB)	
	Free-space	Half-space
$f_0=112$	−14.9	−14.8
$f_1=224$	−15.0	−14.7
$f_2=336$	−15.0	−14.4
$f_3=448$	−11.1	−20.1
$f_4=560$	−6.5	2.5

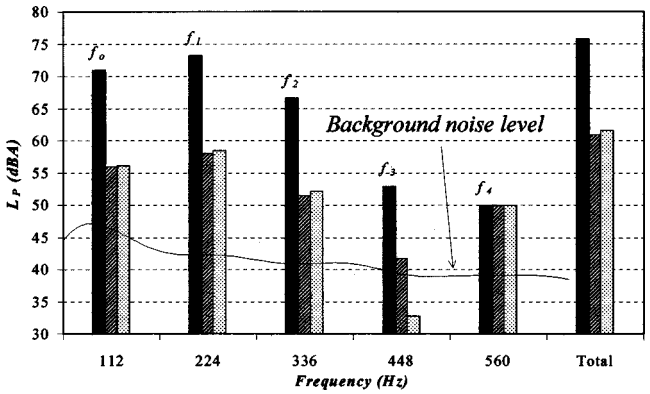
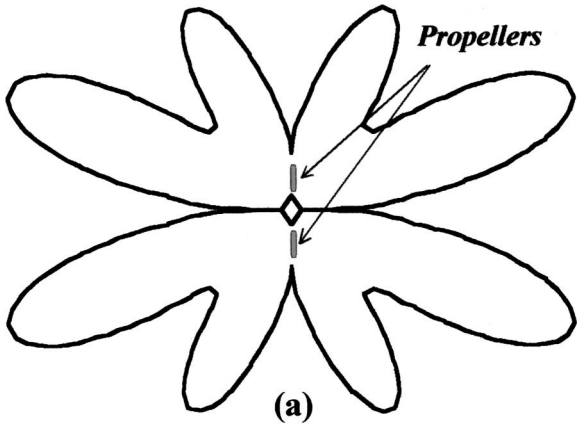


FIG. 12. Noise levels at a nearby community without and with a nine-channel active-control system with $r_{ss}=2.0$ m and a LPF at 500 Hz: *Without* *With, free-field* *With, half-space.*

aircraft to another, it is difficult to generalize as to which mechanism will govern the run-up noise generated. Simulations of ANC effectiveness for all three directivity patterns were performed for both the free-field and half-space conditions; free-field results are presented here. The results demonstrate that a multi-channel, local ANC system is still able to create large quiet zones if the system is designed with optimal arrangements and orientations.

The Beechcraft 1900D is a twin-propeller aircraft, which can realistically be modeled as two sources spaced 5.23 m apart (the same spacing as that of the Beechcraft 1900D propellers) in the simulations. Since their speeds are controlled individually, it is likely that the two propellers do not spin in perfect synchrony; the resulting noise signals from the two sources can likely have any relative phase relationships. Figure 13 shows the directivity patterns created when the two propellers are treated as either two out-of-phase dipoles or two out-of-phase quadrupoles. It is clear that the radiation patterns of the two dipoles and two quadrupoles are very complicated and very different. Thus, the effectiveness of a local ANC system in the case of non-omnidirectional noise sources will be discussed next.

Simulations were conducted for a nine-channel system—again with $r_{ps}=20$ m, $r_{se}=40$ m, and $r_{ss}=2.0$ m, in the free-field condition. The sound attenuation for the fundamental frequency of 112 Hz was calculated. The first simulation involved two propellers treated as two out-of-phase monopoles separated by 5.23 m. For this type of source, the noise has its minimum at the front and rear (0° and 180°), and is a maximum to the sides (90° and 270°) of the aircraft. Figure 14 shows the quiet zones created by the control system for this primary source. Figure 14(a) is the quiet zone created when the system is at 0°—in front of the aircraft—while Fig. 14(b) is the quiet zone created when the system is at 90°—to the right side of the aircraft. It indicates that, although the ANC system is able to create quiet zones when located in both directions, some large areas of sound increase are found both inside and outside the quiet zones for the 0° orientation. Clearly, a larger quiet zone is seen when the system is at 90°—the maximum direction of the primary source. This suggests that the control system is more efficient



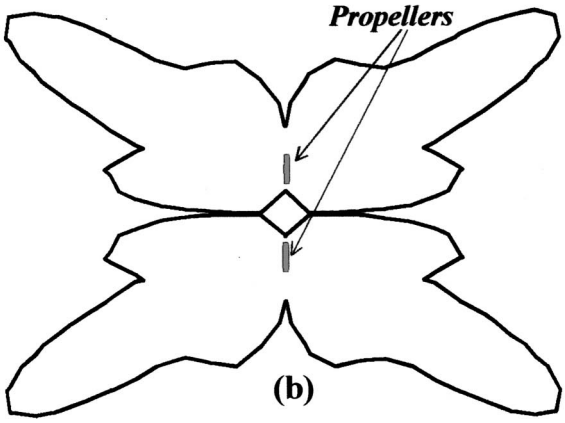


FIG. 13. Predicted sound-radiation patterns of (a) two out-of-phase dipoles and (b) two out-of-phase quadrupoles.

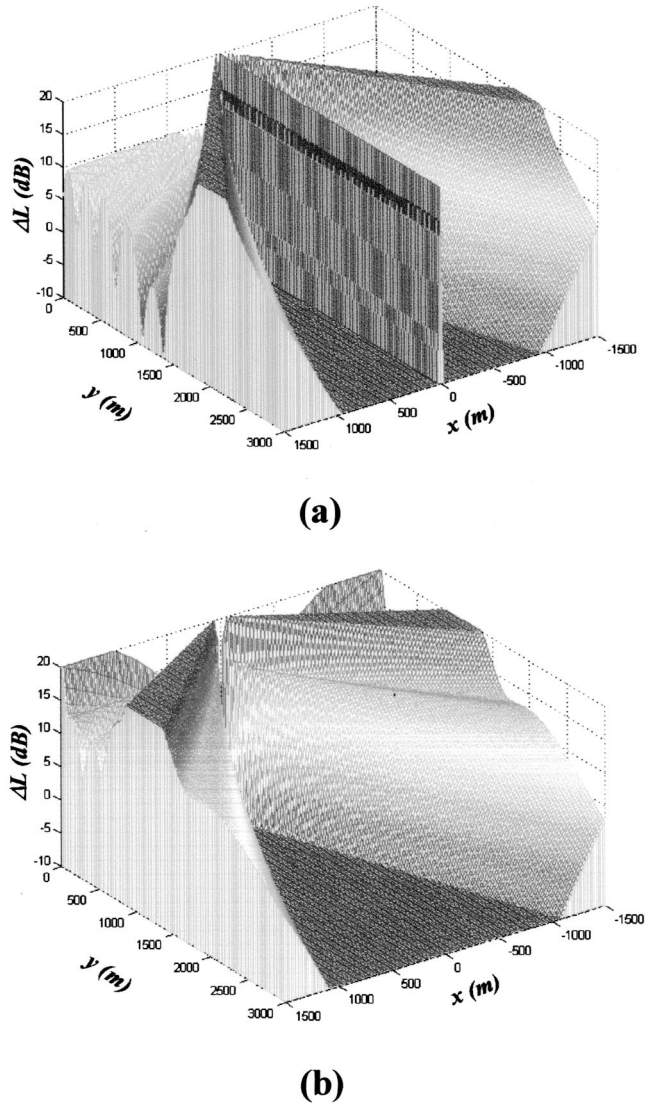


FIG. 14. Quiet zones for two out-of-phase monopoles and an ANC system at (a) 0° and (b) 90°.

when it is located so as to cancel the noise in directions of higher radiation.

Figure 15 shows the large quiet zones created by the ANC system when the two propellers are modeled as two out-of-phase dipoles separated by 5.23 m. Similarly, areas with significant increases of sound pressure are found outside the quiet zone, as well as some small areas inside the quiet zone. Large quiet zones are also found when the two propellers are modeled as two out-of-phase quadrupoles separated by 5.23 m as shown in Fig. 16. However, in this case, the quiet-zone size becomes smaller, and the sound reduction in the center of the quiet zone is decreased.

Observing Figs. 14–16, it can be seen that the quiet zones achieved using the different radiation patterns are similar to those obtained using one monopole source. The main differences are in the significant increases in the hot zones outside of the quiet zones, and some inside the quiet zones. These increases can be very high in some cases, and would need to be investigated and minimized in the practical application of ANC to aircraft run-up noise.

From these simulations, it can be observed how the effectiveness of an ANC system deteriorates as the radiation pattern becomes more complex. This is the result of poorer wavefront matching between the primary and control sources. When moving from monopole to dipole to quadrupole sources, more lobes, with increasingly smaller radii of curvature, are present; these do not match properly the larger

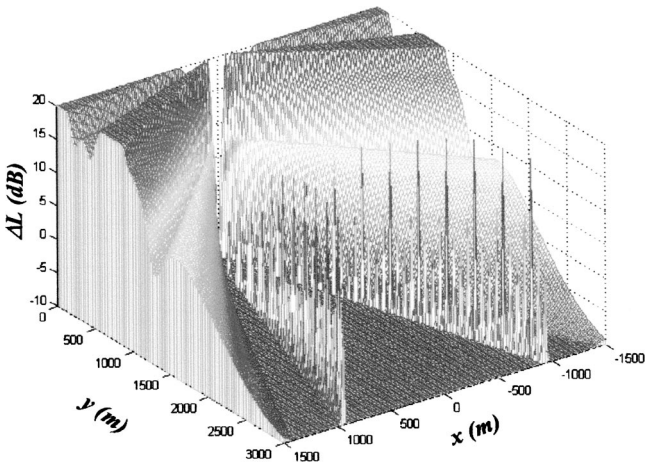


FIG. 15. Quiet zones for two out-of-phase dipoles.

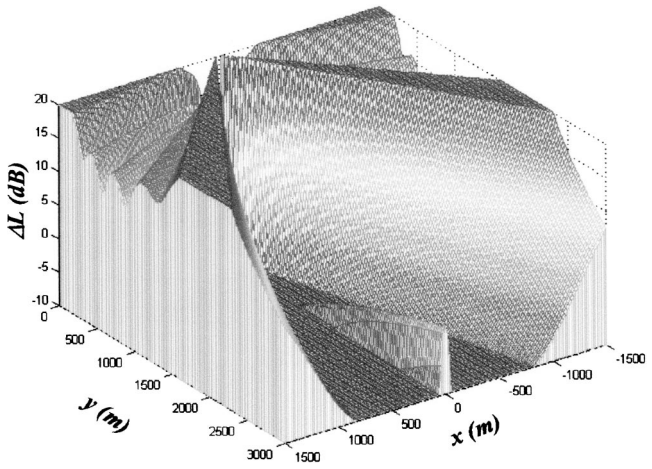


FIG. 16. Quiet zones for two out-of-phase quadrupoles.

radii of curvature of the control-source wavefronts. This is particularly obvious in the case of the quadrupole simulations, for which a ''ridge'' of sound-level increase occurs along the centerline of the sound field. Further simulations showed that this situation can be improved if more channels are introduced into the ANC system.

VI. CONCLUSION

Preliminary simulations were performed to predict the attenuation provided by multi-channel ANC systems for possible implementation at Vancouver International Airport. It was found that 10 dB or more of sound-level attenuation can be obtained in areas (called ''quiet zones'') large enough to cover the complainant receiver locations. The quiet zones, however, are created at the expense of sound-level increases in areas (called ''hot zones''), outside of the quiet zones. However, it was found that a properly designed system limits the sound-level increases to localized areas at the locations of the control sources.

When using a multi-channel system, the spacings of the control sources must lie within an optimal range; this results in the minimum increase of sound energy due to the ANC system. The size of the quiet zone increases as more control channels are added to the ANC system. It was shown that, for a given optimal configuration of the control system, as many as four peaks (the fundamental and the first three harmonics) of the run-up noise spectrum can be attenuated, and that this would result in a total attenuation of 15 dB in a community 3 km from the airport.

The ANC system is also effective at creating quiet zones when the propeller-aircraft noise is modeled as two monopoles, two dipoles, or two quadrupoles. However, the effectiveness is reduced with these multipole noise sources. Real propeller-aircraft noise is likely a combination of these multipole sources. Further work is required to improve the control effectiveness. There are a number of other ways in which further work is required—for example, to better understand propeller-aircraft noise radiation and how to model it, on how to improve the control effectiveness of an ANC system for multipole sources, and on how to deal with changes in the excitation frequency due to propeller-speed changes.

ACKNOWLEDGMENTS

The research was funded by the Vancouver International Airport Authority and the Natural Sciences and Engineering Research Council of Canada.

[1] Vancouver International Airport Authority, ''YVR Aeronautical Noise Management—Annual Report,'' 1997.
[2] P. Germain, J. Guo, and M. R. Hodgson, ''Measurement of propeller-aircraft run-up noise,'' Can. Acoust. **29**(2), 21–27 (2001).
[3] P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic, San Diego, 1992).
[4] C. H. Hansen and S. D. Snyder, *Active Control of Noise and Vibration* (Spon, London, 1997).
[5] J. P. Smith, R. A. Burdisso, C. R. Fuller, and R. G. Gibson, ''Active control of low-frequency broadband jet engine exhaust noise,'' Noise Control Eng. J. **44**(1), 45–52 (1996).
[6] J. Guo, J. Pan, and C. Bao, ''Actively created quiet zones by multiple control sources in free space,'' J. Acoust. Soc. Am. **101**(3), 1492–1501 (1997).
[7] S. E. Wright and B. Wuksanovic, ''Active control of environmental noise,'' J. Sound Vib. **190**(3), 565–585 (1996).
[8] K. A. Cunefare and S. Shepard, ''The active control of point acoustic sources in a half-space,'' J. Acoust. Soc. Am. **93**, 2732–2739 (1993).
[9] J. Guo and J. Pan, ''Effects of reflective ground on the actively created quiet zones,'' J. Acoust. Soc. Am. **103**, 944–952 (1998).
[10] S. Wright and B. Vuksanovic, ''Active control of environmental noise, II: non-compact acoustical sources,'' J. Sound Vib. **202**(3), 313–359 (1996).
[11] H. H. Hubbard, *Aeroacoustics of Flight Vehicles—Theory and Practice, Vol. 1* (Acoustical Society of America, Woodbury, NY, 1995).
[12] J. Guo, M. Hodgson, and J. Pan, ''Investigation of the effective frequency range of active noise control in free space,'' in *Proceedings of Euro-Noise*, Munich (1998), pp. 1023–1028.

Low-frequency absorption using a two-layer system with active control of input impedance

Pedro Cobo[a] and Alejandro Fernández
Instituto de Acústica, CSIC. Serrano 144, 28006 Madrid, Spain

Olivier Doutres
ENSIM, Université du Maine, Rue Aristote, 72085 Le Mans, France

(Received 25 February 2003; revised 24 September 2003; accepted 3 October 2003)

Broadband noise absorption, including low frequencies, may be obtained by a hybrid passive–active two-layer system. A porous layer in front of an air layer provides passive absorption, at medium and high frequencies. Active control of the input impedance of the two-layer system yields absorption at low frequencies. The active control system can implement either pressure-release or impedance-matching conditions. A simple analytical model based upon plane waves propagating in a tube permits the comparison of both control strategies. The results of this simple model show that the pressure-release condition affords higher absorption than the impedance-matching condition for some combinations of geometrical and material parameters. Experimental results corroborate the good performance of the pressure-release condition under the prescribed geometrical setup. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1629306]

PACS numbers: 43.55.Ev, 43.50.Ki [KAC] Pages: 3211–3216

I. INTRODUCTION

The demand for noise reduction, especially at low frequencies, motivates investigation using active control. Passive control provides noise reduction at medium and high frequencies with moderate performance-to-cost ratio, whereas active control is capable of canceling low-frequency noise. Therefore, active control is a complementary, rather than alternative, technique to passive control of noise. A broadband noise problem requires a hybrid passive–active noise control solution.

Noise in halls is an example that requires a broadband solution. Absorbers are typically used to decrease reverberation, this in turn reduces the reflected component of the acoustic field in rooms. Manufacturers provide materials with moderate thicknesses that have absorption coefficients close to 1 for frequencies above 1 kHz. The first peak in the frequency response of the absorption coefficient can be moved to lower frequencies by including an air gap between the absorbing layer and the wall. Increasing the thickness of the air layer will decrease the frequency of the first peak. Thus, this offers an option of designing fully passive low-frequency absorbers. This technique, however, has two severe shortcomings: (1) the two-layer absorber can be too bulky; and (2) the first minimum of the absorption frequency response is moved towards frequencies where higher absorption is required. A solution to this problem is to design the two-layer system to yield high absorption above a few hundred hertz and complement the low-frequency range with an active system. Lueg already proposed to construct walls as vibrated membranes to silence rooms in 1933 (as discussed by Guicking, 1990).

In an active system, the controller drives the actuator with a signal synthesized from the disturbance signal according to some performance objective and control strategy (Elliott, 2001). Previous papers have discussed active control of absorption with different disturbance signals and control strategies. Guicking *et al.* (1985) applied adjustable electronic to afford active impedance or reflection control at the end of a standing wave tube. A standing wave separator calculated the disturbance signal (the incident wave) using measurements from two microphones spaced along the tube, which was processed by the controller. The controller yielded purely active absorption below 1000 Hz. Orduña-Bustamante and Nelson (1992) also used a two-microphone system to calculate a disturbance signal using a feedforward active control system for impedance matching in front of the actuator. Furstoss *et al.* (1997) demonstrated the superior performance of the active absorption system when this is combined with the passive absorption yielded by a layer of absorbing material spaced in front of the actuator. A microphone in the air layer provided the disturbance signal for both feedback and feedforward controllers. Minimization of this disturbance signal (pressure-release condition) gave absorption coefficients above 0.9 at frequencies between 200 and 900 Hz.

Beyene and Burdisso (1997) carried out an experimental comparison between two control conditions on a two-layer hybrid passive–active absorbing system: (1) pressure release, which drove the pressure in the air layer to zero, and (2) impedance matching, which canceled the upstream wave in the air layer. To implement the impedance-matching condition a wave deconvolution circuit was used, similar to the standing wave separator used by Guicking *et al.* (1985), together with two microphones in the air gap. The experimental results of Beyene and Burdisso, further corroborated by Smith *et al.* (1999), showed the superiority of the impedance-matching condition as compared to the pressure-release condition on the hybrid passive–active absorption for broadband noise. Nevertheless, the authors recognize that a

[a]Electronic mail: iacpc24@ia.cetef.csic.es

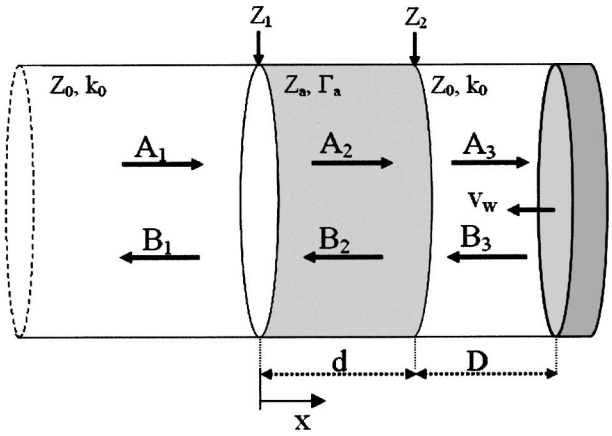


FIG. 1. Schematic representation of the hybrid two-layer passive–active system.

confirmation of this experimental result would require an analytical model that includes the physical properties of the absorption material. The main goal of this paper is to derive such an analytical model. A simple problem of plane waves propagating in a tube is considered. The passive–active absorbing system consists of a layer of porous material separated from an active wall by an air cavity. The semiempirical equations of Allard and Champoux (1992), instead of those of Delany and Bazley (1970), are used to model the propagation through the absorbing material due to the low-frequency contents of the noise. The model yields the absorption coefficient of the system as a function of the geometrical parameters of the system (thicknesses of porous and air layers), a constitutive parameter of the absorbing material (flow resistivity), and the frequency. Following Mechel (1988), expressing these parameters in nondimensional variables offers a clearer comparison of results. When results are presented in such a way it is evident that the pressure-release condition can yield an absorption coefficient higher than the impedance-matching condition for certain combinations of parameters. Experimental confirmation is reported.

II. ONE-DIMENSIONAL PLANE-WAVE MODEL

Figure 1 shows a schematic representation of a hybrid passive–active two-layer system. A plane wave propagates downstream in a one-dimensional medium (a tube, for instance) from a source somewhere on the left. The absorption system at the end of the tube consists of two layers in front of an *active wall* that moves as a piston. The first layer, of thickness d, is made of a porous material. An air layer, of thickness D, exists between the porous layer and the active wall. Let $Z_0=\rho_0\, c_0$ and k_o be the characteristic acoustic impedance and acoustic wave number of the air, respectively. The porous layer is characterized by a complex acoustic impedance Z_a, and a complex propagation constant $\Gamma_a=\alpha_a+ik_a$, where α_a is the attenuation constant and k_a is the phase constant. After reflection and transmission at the interfaces $x=0$, $x=d$, and $x=d+D$, downstream and upstream plane-wave fields are generated. The downstream and upstream waves have pressure amplitudes A_n and B_n, $n=1$, 2, and 3, respectively, as shown in Fig. 1. The absorption coefficient at the input of this system is

$$\alpha=1-|r|^2, \tag{1}$$

where

$$r=\frac{z-1}{z+1} \tag{2}$$

is the reflection coefficient, and

$$z=\frac{Z_1}{Z_0} \tag{3}$$

is the input impedance of the system normalized to that of the air. The input impedance of this two-layer system is (Beyene and Burdisso, 1997)

$$Z_1=Z_a\frac{Z_2\cosh(\Gamma_a d)+Z_a\sinh(\Gamma_a d)}{Z_a\cosh(\Gamma_a d)+Z_2\sinh(\Gamma_a d)}, \tag{4}$$

where Z_2, the input impedance to the air layer, assuming a time dependence $e^{i\omega t}$, is (Beyene and Burdisso, 1997)

$$Z_2=Z_0\frac{V_wZ_0+2B_3e^{ik_0d}\cos(k_0D)}{V_wZ_0+2iB_3e^{ik_0d}\sin(k_0D)}. \tag{5}$$

V_w is the particle velocity at the active termination. Equation (5) reduces to $Z_2=-iZ_0\cot(k_0D)$ when $V_w=0$, the input impedance of the air layer for a rigid termination at $x=d+D$, in the following called the *passive case* (Mechel, 1988).

In order to complete the propagation model for this two-layer system, the acoustical properties of the fibrous layer must be specified. Empirical and analytical models exist for propagation through such materials. For the purpose of this work, an empirical model affords enough precision and will be adopted. Most of the empirical models of propagation in absorbing materials give the acoustic impedance, Z_a, and the propagation constant, Γ_a, as functions of the nondimensional parameter $E=\rho_0 f/R_1$, where ρ_0 is the density of the air, f is the frequency, and R_1 is the flow resistivity of the material. Much of the literature on passive absorbers refers to the work of Delany and Bazley (1970). Nevertheless, these equations are restricted to the range $0.01<E<1$, which is insufficient for the low-frequency contents of the noise involved in active control. The empirical equations of Allard and Champoux (1992) can be used for any $E<1$ providing results similar to those of Delany and Bazley in the valid frequency range. Therefore, the absorbing layer will be characterized by the equations of Allard and Champoux (1992) given by

$$\Gamma_a=i2\pi f\sqrt{\rho(f)/K(f)}, \tag{6a}$$

$$Z_\alpha=\sqrt{\rho(f)K(f)}, \tag{6b}$$

where

$$\rho(f)=1.2+[-0.0364E^{-2}-i0.1144E^{-1}]^{1/2}, \tag{7a}$$

$$K(f)=101\,320\frac{i29.64+[2.82E^{-2}+i24.9E^{-1}]^{1/2}}{i21.17+[2.82E^{-2}+i24.9E^{-1}]^{1/2}}. \tag{7b}$$

Equations (1)–(7) allow the absorption coefficient of the hybrid passive–active system to be obtained as a function of geometrical parameters (the thicknesses of the two layers), material parameters (the flow resistivity), acoustical param-

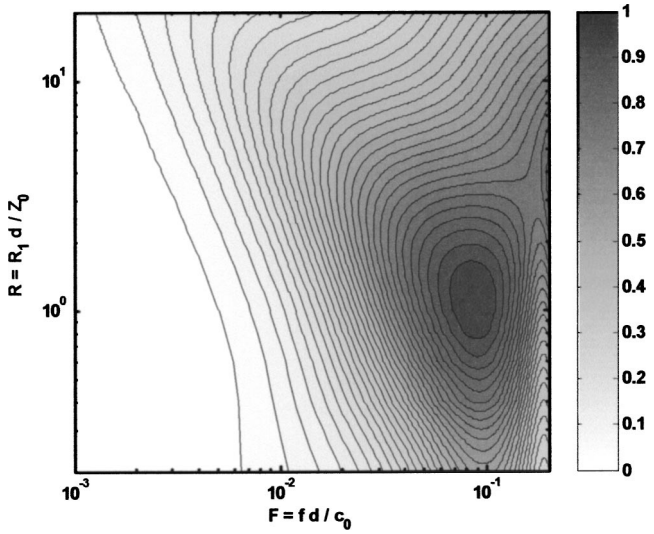


FIG. 2. Contour plot of constant passive absorption against nondimensional variables R and F, for a two-layer system with $D=2d$.

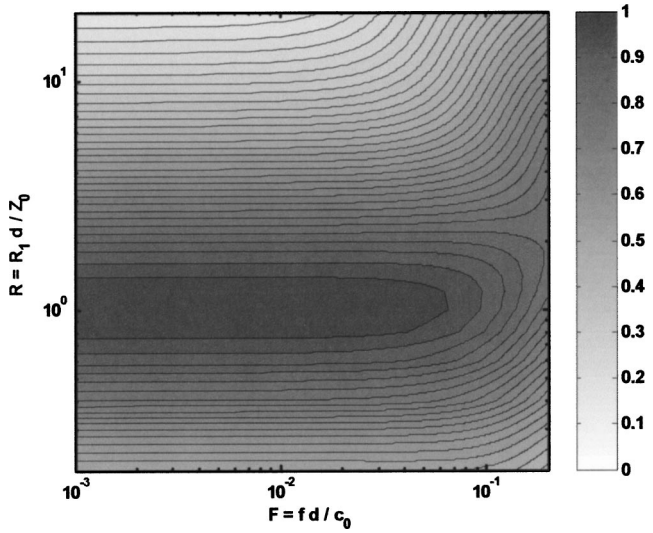


FIG. 3. Contour plot of constant passive–active absorption against nondimensional variables R and F, for a two-layer system with $D=2d$ using a pressure-release condition.

eters (the frequency), and the control condition [V_w in Eq. (5)]. Mechel (1988) showed that a better understanding of the system can be obtained when the absorption coefficient is plotted as a contour map of the nondimensional variables F (defined as $f\ d/c_0$) and R (defined as $R_1\ d/Z_0$), where c_0 is the sound speed in air. Such contour maps will be used in the following to plot the results of the model.

Concerning the control condition, several error sensors have been used, both in front of the absorbing layer and in the air layer. Beyene and Burdisso (1997) and Smith *et al.* (1999) carried out an extensive experimental comparison of two control conditions with the error sensor in the air layer: pressure-release and impedance-matching. In the first case, an active controller drives the total acoustic pressure in the air cavity, measured by a single microphone at the interface between the porous layer and the air cavity, to zero. This implies a zero input impedance of the air layer, $Z_2=0$. In the second condition, the active wall might alter the input impedance to the two-layer system, Z_1, to match that of the air, Z_0. Beyene and Burdisso (1997) demonstrate that when the acoustic thickness of the absorbing layer, $\Gamma_a\ d$, tends to zero, this condition can be achieved by driving the reflected wave to zero in the air layer, $B_3=0$. This happens in most absorbing materials in the low-frequency range. From Eq. (5), this condition implies that $Z_2=Z_0$. A wave deconvolution circuit converted the signals measured by two closely spaced microphones into separate incident and reflected waves in the air layer.

Figures 2–4 show contour maps of the absorption coefficient of a two-layer system with $D=2d$, for the passive (α_p), pressure-release (α_{pr}), and impedance-matching (α_{im}) conditions, respectively. Ranges for the nondimensional variables F and R were chosen to meet the validity margin of Allard and Champoux, $E=F/R<1$. The control of the input impedance enhances the low frequency absorption for both active conditions. In the pressure-release case, the contour lines are parallel to the the F axis, for $F<0.05$, with maximum absorption concentrated around $R=1$. When the acoustic pressure in the air layer vanishes, the input impedance seen by the incident wave can be approximated by the flow resistance, $R_1\ d$ (Furstoss *et al.*, 1997). Therefore, maximum absorption can be expected when this flow resistance matches the acoustic impedance of air, as seen in Fig. 3. The impedance-matching condition, approximated as $B_3=0$ in Eq. (5), requires that the argument of the hyperbolic functions in Eq. (4), $\Gamma_a\ d$, tends to zero (Beyene and Burdisso, 1997). The thinner the porous layer and the lower the flow resistivity, the closer the approximation. Thus, active control by impedance-matching condition performs better for lower R, Fig. 4. For a better comparison between both conditions, Fig. 5 shows contour maps of the difference of absorption coefficients, $\alpha_{\mathrm{pr}}-\alpha_{\mathrm{im}}$. The solid line represents the zero-value contour ($\alpha_{\mathrm{pr}}=\alpha_{\mathrm{im}}$). Above and to the left of this line, α_{pr} exceeds α_{im}, and vice versa, below and right to this line, α_{im} exceeds α_{pr}. The value ($\alpha_{\mathrm{im}}-\alpha_{\mathrm{pr}}$) for lower R is much higher than the value ($\alpha_{\mathrm{pr}}-\alpha_{\mathrm{im}}$) for medium R.

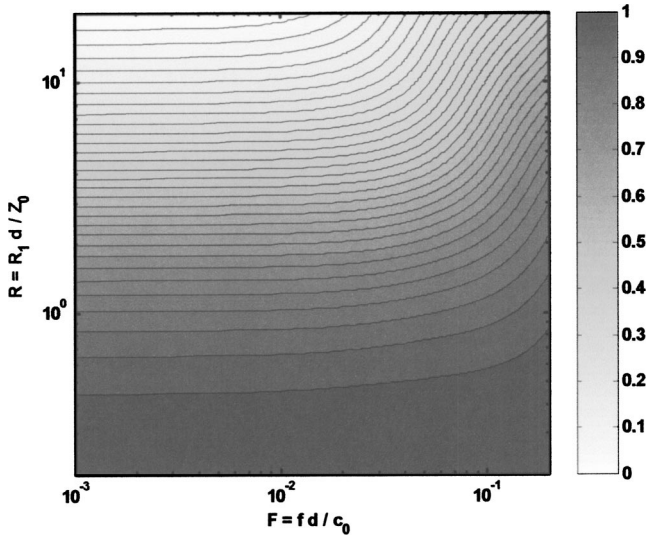


FIG. 4. Contour plot of constant passive–active absorption against nondimensional variables R and F, for a two-layer system with $D=2d$ using an impedance matching condition.

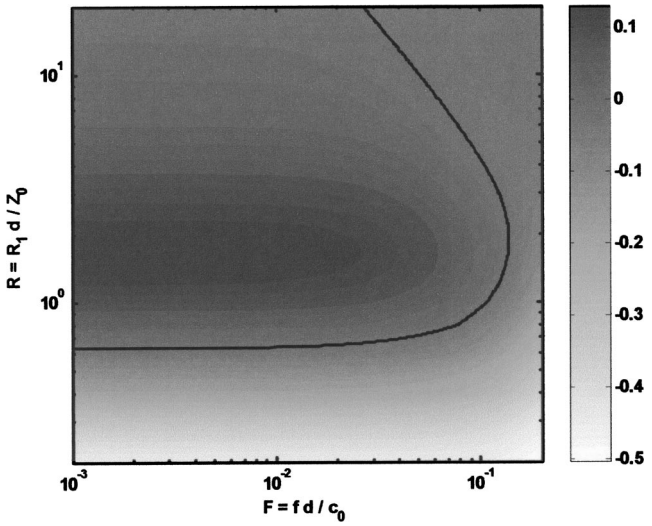


FIG. 5. Absorption difference ($\alpha_{pr}-\alpha_{im}$) of the pressure-release condition against the impedance-matching condition, for a two-layer system with $D=2d$. The thick line corresponds to $\alpha_{pr}=\alpha_{im}$.

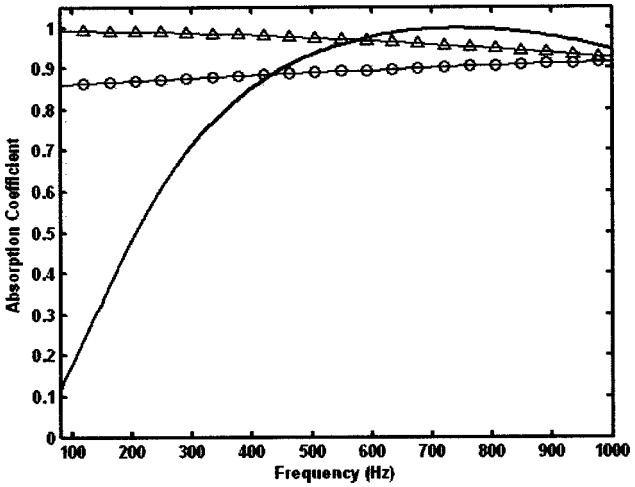


FIG. 6. Absorption coefficient against frequency of a two-layer system with $d=4$ cm, $D=8$ cm, and $R_1=12\,450$ N s/m^4, using passive (———), pressure-release (△———△), and impedance-matching (○———○) conditions.

Therefore, the choice of the better active control condition depends on the particular noise problem to be solved. For a given absorbing material, impedance matching gives higher absorption with a thinner panel. But, the passive absorption should also be taken into account. Active absorption should be restricted to low frequency by low-pass filtering the controller. The cutoff frequency should be determined by the first peak of the passive absorption against frequency curve, which depends on the thickness of the porous layer. The thinner the porous layer, the higher the frequency that requires active control. Thus, a better choice could be a two-layer system with the flow resistance, $R_1\,d$, of the porous layer similar to the acoustic impedance of the air, and with the first peak of the passive absorption at a sufficiently low frequency. Figure 6 shows the absorption coefficient against frequency of a two-layer system with $d=4$ cm, $D=8$ cm, and $R_1=12\,450$ N s/m^4, for the passive, pressure-release, and impedance-matching conditions, respectively. This is an example where active pressure-release control, low-pass filtered at 700 Hz combined with passive absorption offers a high-performance hybrid two-layer system.

III. EXPERIMENTAL SETUP

The model described above predicts higher active absorption with the pressure-release condition than with the impedance-matching condition when the passive layer has a flow resistance close to the air acoustic impedance. Beyene and Burdisso (1997) and Smith *et al.* (1999) reported extensive experimental evidence of the higher performance of the impedance-matching condition for low flow resistance materials. An experiment was therefore devised to measure the absorption coefficient using the pressure-release condition for comparison with the one predicted by the above model.

The experiments were carried out in a standard impedance tube with a diameter of 10 cm and a length of 1 m, Fig. 7. A loudspeaker at one end generates the primary noise. The passive system, at the opposite end, consists of a porous layer in front of an air layer. The active system is comprised of a loudspeaker, a microphone (m_1), and a controller (not shown in Fig. 7). Two microphones, m_2 and m_3 (spaced either 9 or 18 cm apart), are used to calculate the absorption coefficient by the transfer function method. According to the corresponding international standard (ISO 10534-2), both the dimensions of the tube and the distance between the measuring microphones define an operating frequency range of 190 to 1715 Hz, for a microphone separation of 9 cm, and 95 to 857 Hz, for a microphone separation of 18 cm.

The sequence of experiments is as follows. First, a sample of absorbing material having a flow resistance close to Z_0 is chosen. Since the speed of sound and density of air are assumed to be $c_0=343$ m/s and $\rho_0=1.21$ kg/m^3, respectively, the air acoustic impedance is $Z_0=415$ N s/m^3. A layer of glass wool 3 cm thick with a flow resistivity of 14 000 N s/m^4 closely fits this requirement. Then, the thickness of the air layer is selected so that the first maximum of the absorption-frequency response of the passive two-layer system coincides with the cutoff frequency of the low-pass filter of the active system, f_c. For example, for $f_c=800$ Hz, this corresponds to a thickness $D=7$ cm. The passive absorption coefficient predicted by the model is compared with that measured in the tube. For passive measurements, a standard test sample holder (a separate unit tightly fixed to the end of the tube), replaces the secondary loudspeaker. And finally, once the passive system is optimized, the controller is configured to provide maximum active absorption.

A commercial configurable multichannel controller, using the Texas Instruments TMS320 C40 DSP, was the kernel of the active control system. It uses a feedforward strategy to produce the antinoise signal (Nelson and Elliott, 1992; Hansen and Snyder, 1997). It implements both the filtered-x LMS (adaptive FIR) and filtered-u LMS (adaptive IIR) algorithms (Elliott, 2001), and includes also a signal generator and bandpass filters for the reference, error, and control signals. The user can configure the sampling rate, the number of error inputs and control outputs (maximum eight), the type of control algorithm (filtered-x or filtered-u), the length of

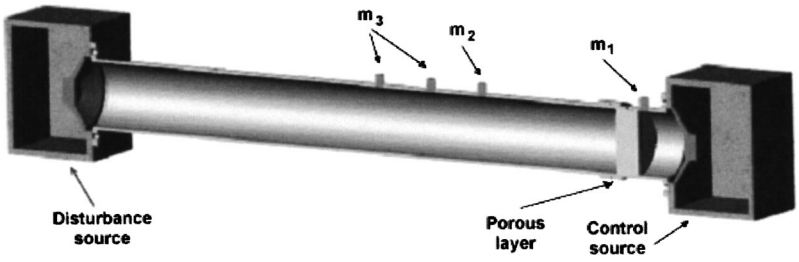


FIG. 7. Scheme of the experimental tube showing the positions of the loudspeakers and the microphones.

the control and identification filters, the convergence factor, and the low- and high-cutoff frequency of the bandpass filters. For the one-dimensional problem considered in this paper, a single channel controller was used.

IV. RESULTS AND DISCUSSION

Several experiments were carried out using the system described above to determine the effectiveness of the pressure-release condition for the nondimensional variable $R=R_1 d/Z_0 \approx 1$. The passive system consists of a 3-cm glass wool layer with $R_1=14\,000$ N s/m^4, in front of a 7-cm air layer. White noise was generated by the primary source at one end of the tube. First, the passive absorption curve was measured with the test sample holder at the opposite end using the signals measured by the microphones m_2 and m_3, shown in Fig. 7. Figure 8 shows that the measured passive curve matches closely with the theoretical one predicted by the model, except for small differences (less than 0.05) between 100 and 300 Hz. The passive system affords high absorption above 800 Hz; thus, the active system should be configured to perform below this frequency.

Next, the passive test sample holder was replaced by the secondary loudspeaker, and the active controller was optimized for maximum absorption. The same white-noise signal used to drive the primary loudspeaker was also utilized as reference signal for the controller. Microphone m_1 was positioned in the air cavity 0.5 cm from the absorbing layer to provide the error signal. Both the reference and the error inputs were low-pass filtered at 800 Hz. The sampling rate was set at 4 kHz. Due to the broadband nature of the disturbance, IIR filters were used for both the control and identification adaptive algorithms of the controller. In this context, optimizing the performance of the controller means finding the best number of forward and backward filter taps, and the best convergence factors that minimize the error signal. Once the controller was optimized and maximum cancellation was obtained, the absorption coefficient was measured again using the two microphones m_2 and m_3 spaced 18 cm apart. Figure 8 shows excellent agreement between the absorption coefficient predicted by the model and that measured in the experiment. The active system yields absorption values above 0.95 in the range 90–850 Hz.

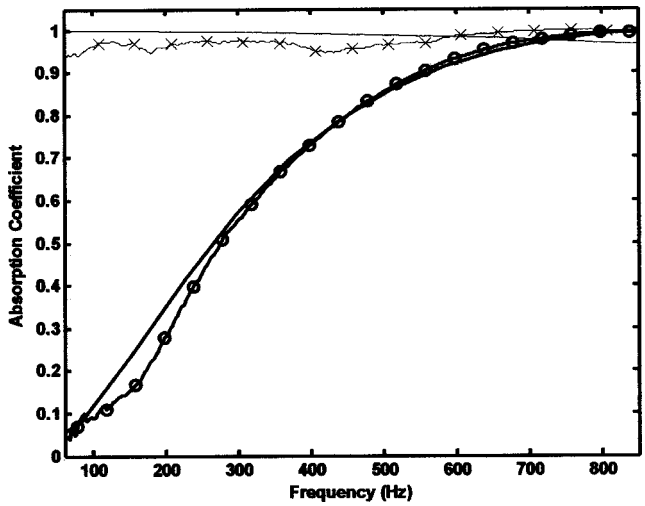


FIG. 8. Absorption coefficient against frequency of a passive–active two-layer system using the pressure-release condition: (———) passive theoretical, (○———○) passive experimental, (—) active theoretical, and (×———×) active experimental.

Additional experiments with other absorbing materials (for example, melamine foam) gave similar results. Therefore, these results demonstrate the feasibility of designing hybrid passive–active systems with high absorption using the pressure-release condition. Moreover, from an electronic hardware point of view, it is much simpler to implement than the impedance-matching condition.

V. CONCLUSIONS

A simple two-layer hybrid passive–active model for plane waves considering both the geometrical and the material characteristics of the passive system, allows a better comparison between the pressure-release and the impedance-matching control conditions. While the impedance-matching condition requires additional electronic hardware to decompose the standing waves into forward and backward plane-wave components, the pressure-release condition needs to drive the pressure measured to zero using only one microphone. Moreover, the pressure-release condition affords higher absorption when the flow resistance of the porous layer is similar to the acoustic impedance of air. Experimental data measured in a standard impedance tube under the pressure-release condition validate this prediction. The absorption coefficient afforded by a 3-cm glass wool layer spaced 7 cm from the secondary loudspeaker is higher than 0.95 at frequencies between 90 and 850 Hz. The model predicts better performance of the impedance-matching condition for absorbing materials with flow resistance less than half of the acoustic impedance of air.

ACKNOWLEDGMENTS

The authors are grateful to the Spanish Ministry of Science and Technology for funding this research under Grant No. DPI2001-1613-C02-01. We acknowledge fruitful discussions with Dr. David K. Anthony.

Allard, J. F., and Champoux, Y. (**1992**). ''New empirical equations for sound propagation in rigid frame fibrous materials,'' J. Acoust. Soc. Am. **91**, 3346–3353.
Beyene, S., and Burdisso, R. A. (**1997**). ''A new hybrid passive/active noise absorption system,'' J. Acoust. Soc. Am. **101**, 1512–1515.
Delany, M. E., and Bazley, E. N. (**1970**). ''Acoustical properties of fibrous absorbent materials,'' Appl. Acoust. **3**, 105–116.
Elliott, S. J. (**2001**). *Signal Processing for Active Control* (Academic, London), pp. 132–166.
Furstoss, M., Thenail, D., and Galland, M. A. (**1997**). ''Surface impedance control for sound absorption: Direct and hybrid passive/active strategies,'' J. Sound Vib. **203**, 219–236.
Guicking, D., Karcher, K., and Rollwage, M. (**1985**). ''Coherent active methods for application in room acoustics,'' J. Acoust. Soc. Am. **78**, 1426–1434.
Guicking, D. (**1990**). ''On the invention of active noise control by Paul Lueg,'' J. Acoust. Soc. Am. **87**, 2251–2254.
Hansen, C. H., and Snyder, S. D. (**1997**). *Active Control of Noise and Vibration* (Spon, London), pp. 559–576.
Mechel, F. P. (**1988**). ''Design charts for sound absorber layers,'' J. Acoust. Soc. Am. **83**, 1002–1013.
Nelson, P. A., and Elliott, S. J. (**1992**). *Active Control of Sound* (Academic, London), pp. 161–179.
Orduña-Bustamante, F., and Nelson, P. A. (**1992**). ''An adaptive controller for the active absorption of sound,'' J. Acoust. Soc. Am. **91**, 2740–2747.
Smith, J. P., Johnson, B. D., and Burdisso, R. A. (**1999**). ''A broadband passive–active sound absorption system,'' J. Acoust. Soc. Am. **106**, 2646–2652.

Energy transmittance predicts conductive hearing loss in older children and adults

Douglas H. Keefe[a] and Jeffrey L. Simmons
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 8 June 2003; revised 12 September 2003; accepted 17 September 2003)

The test performance of a wideband acoustic transfer function (ATF) test and 226-Hz tympanometry was assessed in predicting the presence of conductive hearing loss, based on an air–bone gap of 20 dB or more. Two ATF tests were designed using an improved calibration method over a frequency range (0.25–8 kHz): an ambient-pressure test and a tympanometric test using an excess static pressure in the ear canal. Wideband responses were objectively classified using moment analyses of energy transmittance, which was a more appropriate test variable than energy reflectance. Subjects included adults and children of age 10 years and up, with 42 normal-functioning ears and 18 ears with a conductive hearing loss. Predictors were based on the magnitudes of the moment deviations from the 10th to 90th percentiles of the normal group. Comparing tests at a fixed specificity of 0.90, the sensitivities were 0.28 for peak-compensated static acoustic admittance at 226 Hz, 0.72 for ambient-pressure ATF, and 0.94 for pressurized ATF. Pressurized ATF was accurate at predicting conductive hearing loss with an area under the receiver operating characteristic curve of 0.95. Ambient-pressure ATF may have sufficient accuracy to use in some hearing-screening applications, whereas pressurized ATF has additional accuracy that may be appropriate for hearing-diagnostic applications. [DOI: 10.1121/1.1625931]

PACS numbers: 43.64.Ha, 43.64.Yp, 43.58.Bh [BLM] Pages: 3217–3238

I. INTRODUCTION

Measurements of wideband acoustic transfer functions (ATF) such as pressure and energy reflectance, acoustic admittance, and acoustic impedance have proven useful for studying middle-ear function. Measurements of wideband reflectance at ambient static pressure in the ear canal characterize the acoustic functioning of the normal middle ear in adults (Keefe *et al.*, 1993; Voss and Allen, 1994), in infants and older children (Keefe *et al.*, 1993), and in neonates (Keefe *et al.*, 2000). Ambient-pressure ATF responses predict the presence of a conductive hearing loss in a population of children at risk for otitis media with effusion more accurately than 226-Hz tympanometry (Piskorski *et al.*, 1999). In studies of the efficacy of newborn hearing screening programs, the ambient-pressure reflectance and admittance responses accounted for as much as 28% of the interear variation in the sound-pressure level (SPL) of evoked otoacoustic emissions (OAE) (Keefe *et al.*, 2003a), and predicted middle-ear dysfunction in a population of neonates (Keefe *et al.*, 2003b).

A wideband, tympanometric class of energy reflectance and admittance responses was introduced by measuring these transfer functions in adult and infant ears as joint functions of frequency and static pressure in the ear canal (Keefe and Levi, 1996). A baseline study of wideband energy reflectance and admittance tympanometry was performed on 20 normal-functioning adult ears along with a case study of recurrent otitis media in a 10-year-old (Margolis *et al.*, 1999). Admittance tympanograms were interpretable only up to approximately 2 kHz, but energy reflectance tympanograms were interpretable up to 11.3 kHz. This distinction is thought to arise because an admittance tympanogram requires a compensation procedure to remove the acoustic effect of the volume of air in the ear canal between the probe and the tympanic membrane, whereas an energy reflectance tympanogram does not due to the relative insensitivity of energy reflectance to probe position in the ear canal (Stinson *et al.*, 1982). There is no evidence to suggest that the use of a single frequency, as in clinical tympanometry, is optimal for assessing middle-ear function at all frequencies important in auditory communication systems, no more than would a single frequency suffice for assessing cochlear, behavioral, or neural function. Wideband measurements of middle-ear functioning appear to have promise as a clinical diagnostic test. Feeney *et al.* (2003) reported systematic shifts in ambient-pressure energy reflectance in adults with middle-ear disorders including otitis media with effusion, otosclerosis, ossicular discontinuity, and perforation of the tympanic membrane; they also measured shifts in adults with hypermobile tympanic membranes in the absence of other middle-ear pathology. On the other hand, studies have shown that wideband, pressurized reflectance responses contain information that is absent in wideband reflectance responses measured at ambient static pressure (Margolis *et al.*, 1999, 2001). There have been no studies comparing the wideband test performance of ambient-pressure and pressurized ATFs in the prediction of conductive hearing loss. The purpose of the present study is to make such a comparison.

Based on a time-domain approach to measure reflectance (Keefe, 1996), an ATF technique was devised that was calibrated using a single cylindrical tube closed at the opposite end from that in which the probe was inserted. Isolevel and multilevel reflectance functions were measured in adult ears using this technique (Keefe, 1997). The one-tube reflec-

[a]Electronic mail: keefe@boystown.org

tance calibration procedure in Keefe (1997) used measurements in a long calibration tube, in part, because a design aim was to achieve the simplest possible calibration procedure for clinical use. The calibration was expressed in the form of an incident pressure signal and source reflectance signal, the source reflectance describing the reflection characteristics of the probe assembly. However, the calibration results had insufficient accuracy below 1 kHz and above 5.5 kHz. A more accurate two-tube calibration procedure is described in this report that is a generalization of the one-tube procedure of Keefe (1997). The two-tube procedure provides a measurement range from 0.25 up to 8.0 kHz, which is sufficiently wide to include the frequencies important for speech and music perception.

A limitation in past applications of multifrequency tympanometry has been the complicated nature of the response on the joint frequency–static pressure plane, which is difficult for the nonspecialist to classify into patterns that are clinically meaningful. In contrast to the widely accepted criteria for the interpretation of 226-Hz admittance tympanograms [such as those in ASHA (1990)], no accepted objective, clinical guidelines exist for the interpretation of multifrequency admittance tympanograms, although a preferred method to estimate resonant frequency from such responses has been proposed (Margolis and Goycoolea, 1993). This problem exists also for multifrequency tympanograms of other ATFs such as reflectance. This report describes the application of moment analysis for the interpretation of wideband tympanograms and ambient-pressure responses, and the construction of an objective measure of classifying such moments as predictive of the presence of a conductive hearing loss. Because the energy transmittance has a central maximum on the two-dimensional plane of frequency and static pressure, whereas energy reflectance has a central minimum, the energy transmittance is a more useful input variable with which to calculate moments. The energy-transmittance moments defined below were used in a study to predict a conductive hearing loss in a population of older children and adults with normal hearing, a pure conductive loss, or a mixed hearing loss including both conductive and sensorineural components. Both pressurized and ambient-pressure wideband responses were compared in performance, and each was compared with the performance of 226-Hz tympanometry.

The report is organized as follows. Section II describes the theory of ATF calibration based on the use of two calibration tubes and presents typical calibration results. Section III presents the definitions of the moments of tympanometric functions of frequency and static pressure, and the moments of ambient-pressure functions of frequency. Subsequent sections describe the methods, results, and discussion of the study to predict the presence of conductive hearing loss.

II. CALIBRATION THEORY FOR ATF MEASUREMENTS

This section, which presents the theory and signal processing that underlie the calibration and usage of the ATF measurement system, has more technical material than the other sections of this report. The reader with a specific interest in the test performance of predictors of middle-ear function may omit this section. *Calibration* here means the set of procedures used to measure an ATF in the specified units of the pair of acoustical variables comprising the transfer function. The Appendix contains a list of symbols.

A two-tube calibration method for ATF measurements has been designed, in which a first cylindrical tube is sufficiently long that the incident waveform generated as a short-duration click can be measured in the absence of significant tube reflections [as in the single-tube otoreflectance system described in Keefe (1997)], and a second cylindrical tube is sufficiently short that its response waveform is dominated by multiple reflections between the probe and the opposite, closed end of the tube.

The stimulus is a click with energy from 0.25–8 kHz that is shaped for the frequency responses of the probe loudspeaker and microphone based on measurements in a standard artificial ear simulator (Bruel & Kjaer, IEC 711). This signal shaping is performed in the time domain using a digital-filter deconvolution technique. The electrical stimulus to the miniature loudspeaker of the probe is constructed so that the resulting incident pressure waveform, which is measured by the probe microphone in the long calibration tube, approximates a bandlimited impulse (Agulló *et al.*, 1995). Such a signal shaping has been similarly applied to the design of click stimuli for auditory brain stem response measurements [as in Fig. 1 of Sininger *et al.* (2000)]. The stimulus level is selected to be sufficiently high that the signal-to-noise ratio (SNR) of each response measured in a calibration tube is adequate, but sufficiently low that distortion is minimized. Data are acquired using a signal processor installed in a personal computer. The stimulus is output by a digital to analog converter (DAC) of the processor, and coupled to a miniature loudspeaker in the probe assembly. The pressure response is measured using the probe microphone and digitized by an analog to digital converter (ADC) of the processor. The response is digitally high-pass filtered in real time on the computer to eliminate noise below 0.25 kHz, and a real-time artifact rejection technique is used to eliminate buffers with high peak levels of intermittent noise (Keefe and Ling, 1998).

Tube 1 is a long tube of length approximately 295 cm, and tube 2 is a short tube of length approximately 8.4 cm with the same diameter as tube 1. The tube material is high-density polyethylene (HDPE stock tubing, Hudson Extrusions, Inc.) with a nominal inner diameter of 5/16 in. (or 0.794 cm) and outer diameter of 7/16 in. This material was chosen because of the availability of the desired inner diameter, the rigidity of the tube wall (hardness 65 D), and the ability to coil the longer tube. Each tube is closed at one end by inserting an approximate 2-cm length of steel rod. There are two sets of calibration tubes used in the measurement system (four tubes in all), one set for measurements in adults and older children with a tube diameter of 0.794 cm equal to a typical ear-canal diameter, and another set for measurements in neonates and young children with a tube diameter of 0.476 cm. This is larger than neonatal ear canals but sufficiently large for leak-free insertion of the probe using a child-sized probe tip. Although reflectance is relatively insensitive to discrepancies between ear-canal and calibration-

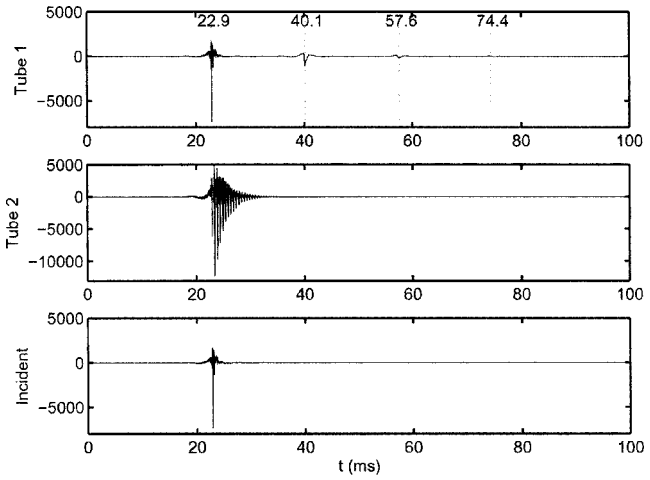


FIG. 1. ADC waveforms recorded for the long tube 1 (top) and the short tube 2 (bottom). The incident ADC waveform (bottom) is defined by separating the incident signal in the tube-1 waveform. The tube-1 waveform shows vertical dotted lines at the peak amplitudes of the incident signal, and for the first-, second-, and third-order round-trip reflections. The time delay (in ms) of each peak is labeled above each dotted line.

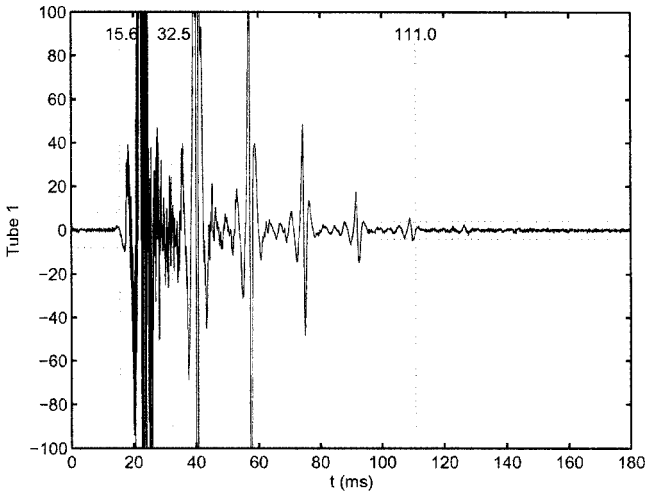


FIG. 2. The ADC waveform for tube 1 is redrawn from Fig. 1 by clipping the amplitude to show the threshold behavior. Vertical dotted lines are drawn and labeled at 15.6 ms for the onset of the incident (and tube-1) signal, 32.5 ms for the end of the incident signal, and 111.0 for the offset of the tube-1 signal.

tube diameters (Keefe *et al.*, 1992), the difference between neonatal and adult ear-canal areas is sufficient to justify the use of two calibration sets with these diameters. The system was calibrated once daily with the large-diameter calibration tubes for the results described in this report.

The probe is inserted into tube 1, the stimulus is repetitively output at a specified level (57 dB SPL), and 16 valid [in the sense of having passed an artifact rejection test of the form described in Keefe and Ling (1998)] response buffers are acquired. Then, it is inserted into tube 2 and the same number of valid buffers are acquired. These buffers are used to calculate the time-domain means and standard deviations, the latter to assess the influence of noise. Noise is less problematical in middle-ear diagnostics compared to noise in OAE measurements because the middle-ear response is on the order of the stimulus level rather than a small, nonlinear residual level as in OAE measurements. As a rough approximation, the stimulus levels of an OAE measurement are similar to those of an ATF measurement of middle-ear response, and their absolute noise levels measured by the method of Schairer *et al.* (2003) are similar. OAE stimulus levels are typically at least 40 dB higher than the measured OAE levels, so that the signal-to-noise ratios of ATF measurements are typically at least 40 dB larger than those of OAE measurements. These signal-to-noise properties are what allow for rapid measurements of middle-ear responses in human ears.

Results are presented in this section for a typical calibration using an implementation of ATF that is further described in Sec. IV C. Figure 1 shows the waveforms recorded by the ADCs in tube 1 and tube 2 in the top and middle panels, respectively. The waveform amplitudes in Fig. 1 are expressed in the discrete steps of the ADC converter prior to the pressure calibration using a frequency-dependent microphone sensitivity. The incident waveform has its peak energy at 22.9 ms; succeeding peaks, which represent multiple internal reflections, occur at 40.1, 57.6, and 74.4 ms, as labeled in Fig. 1 (top). Later reflections are also present, but decay into the noise floor before the end of the 186-ms buffer (the responses in Fig. 1 are plotted in the initial 100 ms of this buffer). The round-trip time in tube 1 between the first-reflected and incident signal is $40.1-22.9=17.2$ ms. The incident signal is an impulse-like signal, as described earlier. A sequence of closely spaced reflections is evident in the tube 2 waveform. The response in this short tube decays more rapidly than that in the longer tube, because the attenuation associated with each reflection from the probe surface is larger than the corresponding attenuation in the long tube associated with distributed viscothermal wall loss. An even more rapidly decaying response occurs in ear recordings because of the additional energy per reflection that is transmitted into the middle ear.

The calibration technique is based upon the fact that the incident click signal in the long tube is separable in time from the first-reflected signal, so that the incident waveform can be extracted from the tube 1 waveform by a simple editing of the waveform buffer (Keefe, 1997). It is also required that the duration of the measurement buffer must be sufficiently long that all late reflections have decayed into the noise floor before the end of the buffer. Otherwise, higher-order reflections would overlap with the incident signal in a time-domain aliasing of the signal. The nominal round-trip time in tube 1 in the absence of viscothermal losses is $\tau=2L_1/c_a$, in which c_a is the phase velocity of sound in air. With a nominal tube length of 295 cm and free-space phase velocity of approximately 34 400 cm/s, the predicted nominal $\tau=17.2$ ms, in accord with the measured round-trip time between the first-order reflection and the incident signal.

The waveform responses $p_1[n]$ in tube 1 and $p_2[n]$ in tube 2 are first edited to calculate to extract the nonzero portion of the signal. An onset threshold parameter ε_{on} defines a minimum waveform amplitude appropriate to the beginning of the buffer (prior to click onset), and an offset threshold parameter ε_{off} defines a minimum waveform amplitude appropriate to the end of the buffer (after click offset). The tube-1 waveform is redrawn in Fig. 2 to show these

threshold parameters and the editing process. The peak amplitudes at 22.9, 40.1, and 57.6 ms, which are evident in Fig. 1, are clipped in Fig. 2 to show the threshold behavior. The additional smaller peak amplitudes at 74.4, 91.7, and 109.0 ms are evident in Fig. 2. The onset threshold parameter has a value of 8 (in the units of the ADC), and horizontal dotted lines are drawn at ± 8 at early times in Fig. 2. At the beginning of the buffer, there is a time delay before the nominal onset time, at which the waveform amplitude first exceeds this onset threshold. The onset time t_0 of the incident signal is defined as the time of the first zero crossing of the tube-1 waveform just preceding this nominal onset time, and is shown in Fig. 2 by the vertical dotted line at 15.6 ms.

The offset threshold parameter has a value of 4 (in the units of the ADC), and horizontal dotted lines are drawn at ± 4 at late times in Fig. 2. At the end of the buffer, there is a nominal offset time beyond which all waveform amplitudes are less than ε_{off}. The offset time is defined as the time of the zero crossing of the waveform just after this nominal onset time, and is shown in Fig. 2 by the vertical dotted line at 111.0 ms. The tube-1 and tube-2 waveform responses are zeroed outside the range between the minimum t_0 of the onset times of tubes 1 and 2 (this turns out equal to the t_0 for tube 2), and the maximum of the offset times of tubes 1 and 2 (this equals the offset time of tube 1 because tube 2 has additional damping due to the greater frequency of reflections from the probe).

The extraction of the incident signal from $p_1[n]$ in Fig. 2 further depends on separating the incident from the first-reflected signal. Because the nominal round-trip time is τ in tube 1, the nominal onset time t_1' of the first-reflected waveform is predicted to occur a time τ after the onset of the incident signal t_0 (15.6 ms in Fig. 2), so that $t_1' = t_0 + \tau = 15.6 + 17.2 = 32.8$ ms. The end of the incident signal t_1 is defined as the time of the first zero crossing of $p_1[n]$ preceding t_1'. The end of the incident signal is $t_1 = 32.5$ ms in Fig. 2. The incident waveform $q[n]$, which is shown in the bottom panel of Fig. 1, is defined equal to $p_1[n]$ for times between t_0 and t_1, and zero otherwise.

It is obvious in Fig. 2 that $p_1[n]$ contains significant energy above the noise at times after the peak of the waveform (22.9 ms) and before the end of the incident signal (32.5 ms). The main pulse of the incident signal extends to approximately 26.2 ms and the maximum amplitude between that time and t_1 is 50 (in ADC units). This error signal should be interpreted with respect to the peak amplitude of 7360 (in ADC units) in Fig. 1, so its relative amplitude of 0.68% compared to this peak. The error signal is most likely due to inhomogeneities δS in the cross-sectional area S of tube 1 that are distributed along its axial length. Such inhomogeneities lead to reflected wavelets as the initial signal propagates down the tube with a relative amplitude of $\delta R \sim (\delta S/S)/2$. Setting $\delta R = 0.68\%$, the relative tolerance in the tube diameter would be approximately $\delta R/4 = 0.17\%$, which appears a reasonable manufacturing tolerance for the extruded polyethylene plastic comprising the tube. In absolute terms, this late-reflected energy between 26.2 and 35.2 ms is a small error in estimating the incident signal. The shorter tube 2 and measurements in an ear canal are much less affected by this source of error because most of the losses in tube 2 are controlled by losses in each of the multiple internal reflections at the probe tip, and, in the ear canal, by losses in each of the multiple internal reflections at the eardrum and the probe tip. This more rapid decay in tube 2 has already been discussed.

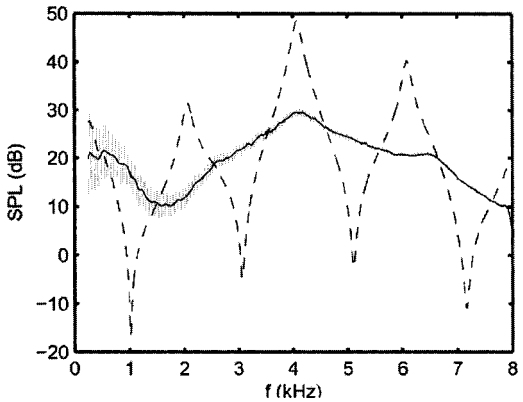


FIG. 3. The SPL spectra are plotted for the long tube (tube 1) as a solid gray curve, the short tube (tube 2) as a dashed black curve, and the incident signal as a solid black curve. The incident SPL traces the midline of the maxima and minima of each of the long- and short-tube responses, a property that is particularly evident when compared to the long-tube SPL with its higher spectral density of maxima and minima.

A reflectance calibration consists of calculating the incident pressure and source reflectance at each frequency (Keefe, 1997), so that only the source reflectance remains to be determined. Responses measured in two known load impedances are sufficient to calculate the pair of Thevenin (or Norton) parameters, and similarly for reflectance parameters. In acoustical circuits, it can be difficult to obtain two ''known'' loads, although the long tube is sufficiently long that it constitutes one known load (i.e., a semi-infinite tube) over sufficiently short measurement times. Another problem is that if the SNR is poor at some frequency, the corresponding estimates of the Thevenin, Norton, or reflectance parameters may be inaccurate at that frequency.

The tube-1 (pressure) spectrum $P_1[k]$, the tube-2 spectrum $P_2[k]$, and the incident spectrum $Q[k]$ are calculated in each frequency bin indexed by the integer k using a discrete Fourier transform (DFT) of the three corresponding waveforms shown in Fig. 1, followed by conversion to pressure using the microphone sensitivity. Their SPLs are shown in Fig. 3 (the noise SPLs are less than -20 dB SPL). The SPL of the incident spectrum varies slowly with frequency. The corresponding SPLs of the tube-1 and tube-2 spectra have a fine structure of maxima and minima due to the presence of acoustic standing waves within the tubes.

An optimization described below is introduced to fit the measured pressure reflectance of each tube to a theoretical model. The theoretical tube model is first described. For a rigid, smooth-walled tube closed at the end of its length L and having a propagation wave number Γ (Keefe, 1984), the model for the (pressure) reflectance for round trip propagation in the tube is [Eq. (2), Keefe, 1997]

$$R^c[k] = e^{-2\Gamma L}. \tag{1}$$

The model reflectances $R_1^c[k]$ and $R_2^c[k]$ are calculated for tubes 1 and 2 (the same subscript convention for tube

number is used throughout), respectively, at the kth frequency f_k. Depending on context, the same spectral variable is equivalently notated as a function of bin index as $Q[k]$, or as a function of frequency as $Q(f)$ for $f=f_k$, or with no explicit argument as Q. The wave number $\Gamma[k]$ is a function of dimensionless numbers r_v^{-1} and r_t^{-1}, which are essentially the ratios of the viscous and thermal boundary layer thicknesses, respectively, to the tube diameter d (Benade, 1968; Keefe, 1984). The viscous boundary layer ratio is expressed in terms of an additional empirically determined loss parameter χ by $r_v^{-1}=(\chi/d)\sqrt{2\eta/(\rho\pi f_k)}$, in which the density of air is ρ and the shear coefficient of viscosity is η. The loss parameter is introduced in an analogous multiplicative manner into the expression described in Keefe (1984) for the thermal boundary layer ratio r_t^{-1}. The parameter $\chi=1$ for an ideal smooth-walled tube, but real tubes have slightly higher losses. The wave number Γ and the characteristic impedance Z_c of the tube, which is used below to calculate impedance, admittance, and absorbed power, are calculated using a third-order asymptotic series in r_v^{-1} and r_t^{-1} (Keefe, 1984).

The optimization procedure iterates to minimize a function MSE (defined below in step 4) by adjusting each of the two tube lengths and the loss parameter χ. The additional losses in real tubes beyond those predicted for an ideal smooth-walled tube are incorporated into the measurement technique by allowing the loss parameter χ to be slightly larger than 1. All three of these variables enter only into step 1 of the optimization via Eq. (1). The calibration procedure begins by measuring the incident spectrum $Q[k]$ in the long tube responses based on the separability criterion above. A single iteration loop of the optimization consists of these four steps.

(1) Using new values of L_1, L_2, and χ, calculate the model reflectances $R_1^c[k]$ and $R_2^c[k]$ using Eq. (1).

(2) Calculate an estimate $R_1^0[k]$ of the source reflectance from the tube-1 data and model, and the incident spectrum by [see Eq. (A7), Keefe, 1996]

$$R_1^0[k]=\frac{P_1[k]-[1+R_1^c[k]]Q[k]}{R_1^c[k]P_1[k]}. \tag{2}$$

In a similar manner, calculate an estimate $R_2^0[k]$ of the source reflectance from the tube-2 data and model, and the incident spectrum. Calculate the estimate of the source reflectance $R^0[k]$ using a weighted average of these two estimates with weighting coefficient $w_2[k]$, which is defined below with the property that $0\leqslant w_2[k]\leqslant 1$, by

$$R^0[k]=(1-w_2[k])R_1^0[k]+w_2[k]R_2^0[k]. \tag{3}$$

(3) Calculate the "data" reflectance $R_1^d[k]$ of tube 1 from the currently estimated source reflectance, the measured tube-1 data, and the incident spectrum by [see Eq. (13), Keefe, 1997]

$$R_1^d[k]=\frac{P_1[k]-Q[k]}{R^0[k]P_1[k]+Q[k]}. \tag{4}$$

This data reflectance is the measurement of an unknown reflectance from its measured pressure spectrum in terms of the calibration parameters (i.e., the source reflectance and incident pressure spectrum). Analytically, Eq. (4) is a rearrangement of the terms of Eq. (2) with the following reversal of "inputs" and "outputs:" $R_1^c[k]$ and $R_1^0[k]$ in Eq. (2) are replaced by $R_1^d[k]$ and $R^0[k]$, respectively, in Eq. (4). In a similar manner, calculate the data reflectance $R_2^d[k]$ of tube 2.

(4) Calculate the mean-squared error (MSE) as the sum of the squared differences between the complex data and model reflectances of each tube, in which the sum extends over all frequency bins k in the measurement bandwidth (0.25–8 kHz)

$$\mathrm{MSE}=\sum_k\left(|R_1^d[k]-R_1^c[k]|^2+|R_2^d[k]-R_2^c[k]|^2\right). \tag{5}$$

This optimization is implemented using the Simplex method (Press *et al.*, 1992).

The optimization procedure is independent of the load characteristics of the two tubes. The load variations of the terminating impedances are described by variations of tube length in the reflectances defined by Eq. (1). The incident signal is independent of load and characterizes the source signal characteristics. For example, the Thevenin impedance $Z^T[k]$ and Thevenin source pressure $P^T[k]$ associated with the probe assembly are related to the incident signal $Q[k]$ and source reflectance $R^0[k]$ by (Keefe, 1996)

$$\begin{aligned} Z^T[k]&=Z_c[k]\left(\frac{1+R^0[k]}{1-R^0[k]}\right),\\ P^T[k]&=\frac{2Q[k]}{1-R^0[k]}, \end{aligned} \tag{6}$$

in which the characteristic impedance of either of the calibration tubes is $Z_c[k]$, which varies inversely with the cross-sectional area of the tube, acoustic frequency, the viscous and thermal boundary layer thicknesses, and other thermodynamic constants (Keefe, 1984). Thus, an acoustic circuit for a waveguide parametrized by characteristic impedance Z_c may be represented by its *reflectance circuit* parameters (Q,R^0) in an equivalent manner to its Thevenin (or Norton) circuit parameters (Keefe, 1996). An ideal pressure source corresponds to $R^0[k]=-1$, and an ideal volume–velocity source corresponds to $R^0[k]=1$.

Returning to the description of the optimization, the weighting coefficient in Eq. (3) is chosen to have the properties that $1-w_2[k]$ is small when $|P_1[k]|$ is small, and $w_2[k]$ is small when $|P_2[k]|$ is small. This rule has the effect that the source reflectance estimate is insensitive to contributions from either tube when its SNR is poor. Otherwise, the bias is to use the short-tube response that is more sensitive to source reflectance. The specific choice of $w_2[k]$ is as follows. Define the minimum pressure amplitude $P_{\min}$ as the minimum of $|Q[k]|$ and $|P_2[k]|$ across the range of measurement frequencies. Next, define the levels normalized by $P_{\min}$ as

$$S_1[k] = 20\log_{10}(|Q[k]|/P_{\min}),$$
$$S_2[k] = 20\log_{10}(|P_2[k]|/P_{\min}). \tag{7}$$

A first weighting variable $w_a[k] = AS_2[k]/(S_1[k] + AS_2[k])$ is defined in terms of a constant A. A second weighting variable $w_b[k]$ is defined in terms of constants B and C such that $w_b[k] = C$ if $(1+B)S_2[k] \geq S_1[k]$, and $w_b[k] = 0$ otherwise. The weighting coefficient $w_2[k]$ is defined as the maximum of $w_a[k]$ and $w_b[k]$. In the implementation, the constants have the values $A = 5$, $B = 0.2$, and $C = 0.975$.

The real and imaginary parts of the final source reflectance are smoothed across frequency after the optimization using a two-stage, median filter (Rabiner and Schafer, 1978), because of the physically reasonable assumption that the source reflectance varies slowly with frequency. A 139-tap median filter (each tap corresponding to a frequency bin) is used in the first stage to eliminate outliers at particular frequencies, followed by a 139-tap Hamming window in the second stage.

The root-mean-squared (rms) reflectance error ΔR is defined as $\Delta R = \sqrt{\text{MSE}/(2K)}$, in which K is the number of frequency bins in the measurement bandwidth ($K = 1441$ in the implementation) and the factor of 2 in the denominator is because Eq. (5) for MSE includes two tubes. The rms error estimates the reflectance measurement variability associated with the calibration. After completion of the optimization stage, the final rms error ΔR should be small compared to the reflectance magnitude differences that one desires to measure, and, for a calibration to be considered accurate, the final optimized χ should be only slightly larger than 1, but never smaller. For the calibration waveforms in Fig. 1, the corresponding ΔR was 0.0065, and the final $\chi = 1.055$. This rms error is relative to the range of reflectance magnitudes from 0 to 1, and should be ≤ 0.01. The final χ typically lies in the range of 1.03–1.08 for an adequate calibration. This calibration is adequate in both respects.

The data and model reflectances are shown in Fig. 4 for tube 1, and in Fig. 5 for tube 2, in which the model reflectances are calculated using Eq. (1). The magnitude in these plots (and in Fig. 6) is the magnitude of the pressure reflectance, such as $|R_1^d[k]|$ for the tube-1 data reflectance, rather than the energy reflectance, which is the squared magnitude of a pressure reflectance. The long-tube reflectance magnitude (Fig. 4) tend towards values close to 0 at moderate and high frequencies, whereas the short-tube reflectance magnitude (Fig. 5) tend towards values close to 1 out to high frequencies, because the integrated wall loss in the short tube is small. Thus, the pair of tubes has quite different reflectance magnitudes at all frequencies. The data and model reflectances have close agreement in magnitude and phase, which is as expected given the small value of ΔR. The data reflectance magnitude sometimes exceeds 1 in Fig. 5 (by amounts on the order of 1%, which is similar in value to ΔR). This behavior is unphysical as the net acoustic losses should reduce the magnitude to values not exceeding 1 according to Eq. (1). These fluctuations represent inaccuracies in the optimization procedure and are corrected in measurements in ears by smoothing the measured tube reflectance across frequency.

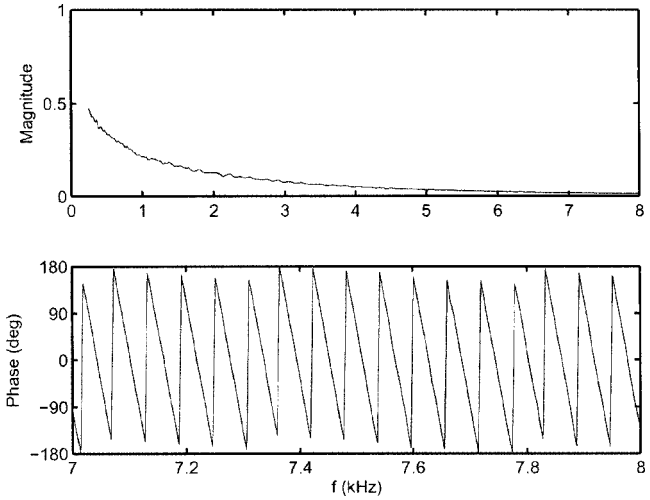


FIG. 4. The top panel shows the relative agreement between the measured ("data") pressure reflectance magnitude $|R_1^d(f)|$ (black curve) and the model pressure reflectance magnitude $|R_1^m(f)|$ (gray curve) for the long tube (tube 1). The bottom panel shows the relative agreement between the measured pressure reflectance phase of $R_1^d(f)$ (black curve) and the model pressure reflectance phase of $R_1^m(f)$ (gray curve). The phase responses are plotted only from 7.0 to 8.0 kHz to facilitate comparison between the data and model reflectances, but the phase responses had similar agreement at lower frequencies. The model-reflectance curves are not visible when they are overlaid by the data-reflectance curves. The data and model reflectances would be identical for a calibration of perfect accuracy.

The source reflectance magnitude and phase (Fig. 6) are slowly varying with frequency: the magnitude remains close to 1 and the phase is close to 0 deg. This is similar to an ideal closed-end reflectance for which, except for a small thermal-boundary-layer effect acting on the cross section of the closed-end termination, the magnitude is equal to 1 and the

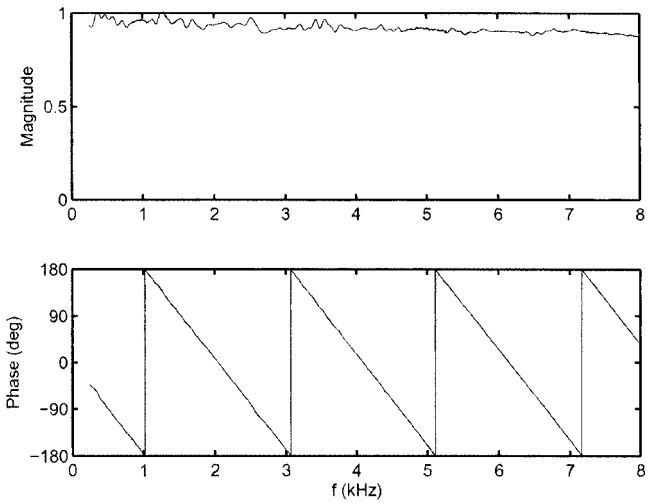


FIG. 5. The top panel shows the relative agreement between the measured ("data") pressure reflectance magnitude $|R_2^d(f)|$ (black curve) and the model pressure reflectance magnitude $|R_2^m(f)|$ (gray curve) for the short tube (tube 2). The bottom panel shows the relative agreement between the measured pressure reflectance phase of $R_2^d(f)$ (black curve) and the model pressure reflectance phase of $R_2^m(f)$ (gray curve). The model-reflectance curves are not visible when they are overlaid by the data-reflectance curves. The data and model reflectances would be identical for a calibration of perfect accuracy.

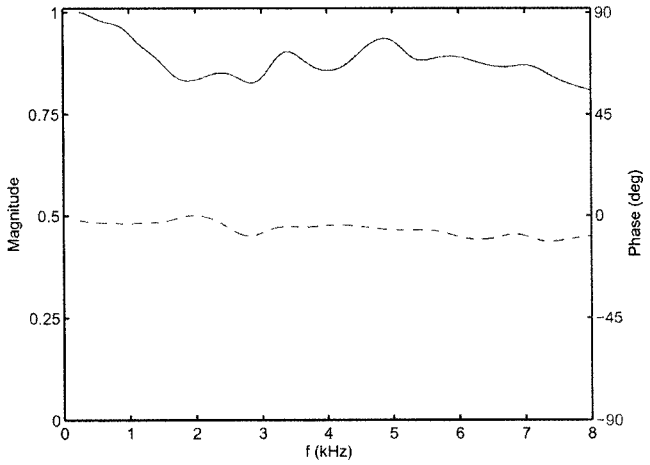


FIG. 6. The source pressure reflectance $R^0(f)$ is plotted for its magnitude (solid curve with values scaled on the left ordinate) and phase in degrees (dashed curve with values scaled on the right ordinate).

phase is equal to 0. The source reflectance varies with changes in probe design; the probe used in these measurements uses hard rubber ear tips (similar to those used in the Grason-Stadler 33) that are more rigid than the disposable, soft foam tips used in other OAE probes. After completing the calibration, the probe is inserted into the subject's ear, and a tip size is selected appropriate to the area of the ear canal. The stimulus is presented and the complex pressure $P[k]$ is measured. The nominal pressure reflectance $R[k]$ of an ear with measured pressure spectrum $P[k]$ is calculated with respect to the characteristic impedance of the calibration-tube set in terms of the incident spectrum and source reflectance by

$$R[k]=\frac{P[k]-Q[k]}{R^0[k]P[k]+Q[k]}. \tag{8}$$

This expression for the nominal ear reflectance is identical in form to that for data reflectance of the tube in Eq. (4), as it is the general relation to calculate a pressure reflectance in terms of the pressure spectrum measured in an unknown system and the reflectance circuit parameters (Q and R^0) of the measurement system.

A correction must be made to the nominal reflectance of the ear to account for the fact that the cross-sectional area of the calibration tube set differs, in general, from that of the subject's ear canal. Each ear tip is precalibrated in terms of the cylindrical-tube area into which it fits snugly, so that the ear-tip size for a snug fit in each subject's ear canal is used to measure the subject's ear-canal area S_e. The correction is based on the fact that at a discontinuity in tube area, the acoustic admittance associated with the plane-wave mode is equal on both sides of the discontinuity whereas the pressure reflectance is not (Morse and Ingard, 1968). The ''discontinuity'' here is that between the characteristic impedance associated with the source reflectance, which varies inversely with the cross-sectional area of the calibration tubes, and the characteristic impedance of the ear canal, which varies inversely with the particular cross-sectional area of the ear canal at the location of the probe assembly within the ear canal.

The acoustic admittance $Y[k]$ at the probe tip is calculated in terms of $R[k]$ in Eq. (8) by

$$Y[k]=Z_c^{-1}[k](1-R[k])/(1+R[k]), \tag{9}$$

in which $Z_c[k]$ is the characteristic impedance of each of the calibration tubes as in Eq. (6). An equivalent expression for the admittance is expressed in terms of the characteristic impedance $Z_{ce}[k]$ of the ear canal and the pressure reflectance $R_e[k]$ of the ear canal by

$$Y[k]=Z_{ce}^{-1}[k](1-R_e[k])/(1+R_e[k]). \tag{10}$$

The characteristic impedance $Z_{ce}[k]$ of the ear canal of cross-sectional area S_e is calculated as

$$Z_{ce}[k]=\rho c[k]/S_e, \tag{11}$$

in which the phase velocity $c[k]$ is

$$c[k]=2\pi f_k/\mathrm{Im}(\Gamma_e[k]), \tag{12}$$

and $\Gamma_e[k]$ is the propagation wave number calculated from the viscothermal model (Keefe, 1984) based on the subject's ear-canal diameter. It follows from Eq. (10) that the reflectance $R_e[k]$ in the ear canal is

$$R_e[k]=(1-Z_{ce}[k]Y[k])/(1+Z_{ce}[k]Y[k]), \tag{13}$$

in which $Y[k]$ is calculated using Eq. (9) in terms of the nominal ear reflectance $R[k]$. The nominal reflectance $R[k]$ and the ear reflectance $R_e[k]$ would be approximately equal only if the calibration tube and ear-canal areas were equal. This equality for equal areas is only approximate due to small differences between $Z_{ce}[k]$ in Eq. (11) and the characteristic impedance that might be calculated as described above using a third-order asymptotic series. Such errors are negligible above 0.25 kHz compared to the errors due to the residual inaccuracy in measuring the ear-canal area using probe-tip size. In summary, $R_e[k]$ is the appropriate variable to represent ear-canal pressure reflectance, because its definition is based on the measured ear-canal area.

Other methods of adjusting the pressure reflectance according to variations in ear-canal area have been described. Voss and Allen (1994) used an average adult ear-canal area. An acoustic estimator of ear-canal area has been proposed (Keefe *et al.*, 1992) that is based on full-spectrum measurements of acoustic resistance at the probe. The resistance is calculated from the admittance in Eq. (9) as the real part of $1/Y[k]$. This estimator was used in a study in human ears that measured responses down to 0.125 kHz (Keefe *et al.*, 1993) and in a study in domestic cat ears down to 0.1 kHz (Huang *et al.*, 2000). Huang *et al.* found the accuracy of the estimator to be within 10% of the diameter (i.e., 21% of the area) in cat ears relative to the diameter measured from an ear-canal mold, although a procedure was required for some ears in which the diameters of the calibration-tube set were adjusted to the estimated diameter in the ear in an iterative manner. The estimator depends on the acoustic resistance spectrum, and can be inaccurate if resistance values are omitted due to truncation of the measurement bandwidth at low frequencies (high-frequency truncation is less problematical). This resistance includes the low-frequency acoustic resistance of the middle ear, which can be unusually large in

ears with conductive hearing loss or in young infants (Keefe *et al.*, 1993) and neonates (Keefe *et al.*, 2000). Because the measurement bandwidth is restricted in this implementation to frequencies no lower than 0.25 kHz, a large resistance below 0.25 kHz would produce inaccuracies in the acoustical estimate of the area. Thus, the direct estimate of ear-canal area using probe-tip size is used, which is thought to have at least a similar accuracy to that of the acoustic estimator (21% in area). The absence of a measured response between 0.125 and 0.25 kHz in the present ATF implementation is justified because of the lack of data showing that responses in this octave are otherwise informative in ways that responses in the next-higher octave (between 0.25 and 0.5 kHz) are not. Elimination of the frequencies below 0.25 kHz simplifies signal-processing issues related to distortion, noise, and the measurement of the incident signal.

The (''data'') reflectance of a third tube not in the calibration tube set was measured by first calibrating the measurement system according to the above procedures, and compared to a model reflectance for that tube. The model reflectance was calculated for a tube of the same area as the calibration tubes but a length of 9.442 cm. The MSE was calculated in a manner similar to Eq. (5) over the frequency range 0.25–8 kHz except that results for only a single tube contributed to the sum. The above length was varied relative to an initial estimate of length of 9 mm to give the smallest MSE. The uncertainty in insertion distance of the probe tip precluded a more precise geometrical measurement of length. The corresponding $\Delta R = \sqrt{\mathrm{MSE}/K}$ was 0.057. Because the energy reflectance of the ear rather than pressure reflectance is the important output variable, an error function for energy reflectance was defined as the average across all frequencies of the absolute value of the difference of the energy reflectances of the model and data. This error function was equal to 0.037, relative to a mean energy reflectance of 0.853.

This measurement technique provides estimates of ATFs including pressure reflectance and admittance, from which the equivalent volume, conductance, impedance, and absorbed power can be calculated (Keefe *et al.*, 1993). The equivalent volume may also be useful in detecting a perforated tympanic membrane (Voss *et al.*, 2001; Feeney *et al.*, 2003) and to estimate the ear-canal volume including the stiffness contribution of the normal tympanic membrane (Keefe *et al.*, 1993).

The energy reflectance of the ear is $|R_e[k]|^2$ and the energy transmittance of the ear is $1-|R_e[k]|^2$.

Among the limitations of this implementation are the fluctuations in the diameter of the long calibration tube, the need for the weighting coefficient $w_2[k]$, and the extent of across-frequency smoothing. One approach would be to simply set $w_2[k]=1$ for all k, so that Eq. (3) would be replaced by $R^0[k]=R_2^0[k]$. The source reflectance would be determined based on its estimate using the tube-2 response alone. This approach works well except at the minima of $|P_2[k]|$ near 1, 3, 5, 7, and 9 kHz in Fig. 3. It is in the vicinity of these frequencies that across-frequency smoothing of the source reflectance is needed. Near these zeros of the tube-2 pressure spectra, the pressure response at the probe is strongly influenced by the higher-order, evanescent modes of the cylindrical tube. These evanescent modes act on the plane-wave impedance by adding a small inertance in series over this frequency range (Keefe and Benade, 1981). The plane-wave impedance (and reflectance) is insensitive to the junction inertance at all other frequencies, for which the plane-wave mode response dominates. By including this junction inertance in the model of the reflectance, it may be possible to improve the calibration accuracy at the zeros of the tube-2 pressure spectra and reduce the extent of the across-frequency smoothing, which is important only near these spectral zeros. Reducing the length of tube 2 would reduce the number of spectral zeros in the measurement bandwidth and, thus, the number of frequencies at which inaccuracies in measuring the source reflectance occur, but there are important disadvantages: (1) the minima of the remaining spectral zeros would have deeper notches and approach closer to the noise; (2) the maxima would have larger peaks and approach closer to the threshold of distortion; and (3) the responses of the higher-order modes would become more complicated for tubes in the limit that the tube length is not larger than the diameter.

This measurement technique may have potential applications to the measurement of ATFs in other areas of acoustics. In particular, the two-tube calibration technique was based on a one-tube technique used to measure the reflectance in the time domain of musical wind instrument air columns (Keefe, 1996). The two-tube calibration technique may be similarly applicable for time-domain reflectance measurements. Equation (1) would be replaced by a time-domain model of reflectance in a cylindrical tube (Polack *et al.*, 1987; Keefe, 1996). Equations (2) through (4) would require a time-domain formulation. The MSE error across frequencies in Eq. (5) would be replaced by an MSE error across time.

III. THEORY UNDERLYING MOMENT CLASSIFICATION OF MULTIFREQUENCY TYMPANOGRAMS

A. Definitions of tympanometric moments

The pattern classification techniques of Vanhuyse *et al.* (1975) have proven useful for interpretating admittance tympanograms at low frequencies. These techniques, reviewed in Van Camp *et al.* (1986) and Margolis and Hunter (2000), classify admittance tympanograms according to the number of extrema in the conductance and susceptance tympanograms. One approach to classifying multifrequency tympanograms plots the sequence of Vanhuyse types by frequency (Holte *et al.*, 1991; Margolis and Goycoolea, 1993; Margolis *et al.*, 1999). Even so, some multipeaked tympanograms fall outside of the Vanhuyse *et al.* categories. There are few objective assessment techniques for other than single-peaked tympanograms; otherwise, the assessment techniques are subjective. The sensitivity and specificity of this approach on a population of normal and pathological ears are unknown.

An alternative characterization of a tympanogram is proposed that is quantitative and objective. *Tympanogram* here means any pressurized ATF or acoustic-pressure response measured in the ear canal, whether narrow band or wideband

in frequency. This alternative approach is based on a moment analysis of the tympanogram in the two dimensions of static pressure and frequency. Such a moment analysis is often used in other domains of the physical sciences and engineering, e.g., in statistical descriptions of data, or in time-frequency representations of signals (Cohen, 1995). A moment analysis of a tympanogram about frequency and static pressure is of interest because of its potential to represent the essential shape of a two-dimensional distribution in terms of a small set of parameters, i.e., the moments. These moments are formally defined in terms of sums over all test frequencies and static pressures.

Consider a tympanogram on a variable $T(P,f)$ that is measured across a range of static pressures P and frequency f. This variable T can represent the admittance magnitude, or the conductance and susceptance, which are the real and imaginary parts, respectively, of the acoustic admittance, or some other variable such as energy reflectance. As discussed later, the present applications focus on energy transmittance, hence the variable name T.

The moments defined over variations in static pressure are similar to well-known tympanometric variables used to interpret single-peaked patterns, but are here defined for arbitrary patterns. Each moment is defined over a sum of discrete static pressures. In a swept-pressure measurement in tympanometry, each discrete static pressure might be defined as the average pressure over a certain time interval, and the entire sweep might be thus expressed as a sequence of discrete pressures over successive time intervals. In energy transmittance results presented below, the tympanometric (i.e., the pressurized ATF) data were measured at N discrete pressures and the standard tympanometric sweep was modified to a sequence of measurements during brief durations in which the static pressure was held constant. The pressure moments of order 0, 1, 2, and 3 are used, respectively, to define the *pressure-averaged tympanometric variable* $\langle T(f)\rangle$, the *mean pressure* $\langle P(f)\rangle$, the pressure variance $\sigma_P^2(f)$, and the *pressure skewness* $\gamma(f)$. Each of these is defined at each acoustic frequency in a wideband response, or at the frequency used in a single-frequency tympanogram. The discrete set of frequencies in these measurements use the $M=61$ 1/12th-octave frequencies between 0.25 and 8.0 kHz. The pressure moments are defined by

$$\begin{aligned}
\langle T(f)\rangle &= \frac{1}{N}\sum_P T(P,f),\\
\langle P(f)\rangle &= \frac{1}{N\langle T(f)\rangle}\sum_P PT(P,f),\\
\sigma_P^2(f) &= \frac{1}{(N-1)\langle T(f)\rangle}\sum_P [P-\langle P(f)\rangle]^2 T(P,f),\\
\gamma(f) &= \frac{1}{(N-1)\langle T(f)\rangle\sigma_P^3(f)}\sum_P [P-\langle P(f)\rangle]^3 T(P,f).
\end{aligned} \tag{14}$$

In each equation, the summation sign with a P below denotes that the sum is over each discrete pressure in the response with frequency held constant. The *mean pressure* is analogous to the tympanometric peak pressure (TPP) in conventional isofrequency tympanometry. The *pressure breadth* is defined as $\Delta P(f)=\sqrt{\sigma_P^2(f)}$, and equals the standard deviation of the response over pressure, just as in the corresponding use in statistics. When applied to a tympanogram with a symmetrical peak, the average pressure is equal to the TPP, the pressure breadth is a measure closely related to the *tympanometric width* (or gradient), and the skewness is zero. For example, the TPP of a 226-Hz admittance magnitude tympanogram is the static pressure of the maximum magnitude. If the tympanometric peak were symmetrical with respect to static pressure about this maximum, then the mean pressure calculated in Eq. (14) would equal the TPP. For a general tympanometric pattern, the mean pressure at any frequency retains the interpretation that it is the static pressure of the ''center'' of the tympanometric variable across all static pressures. The mean pressure would lead to an inaccurate interpretation if the tympanometric variable were bimodal across static pressure, but such patterns are uncommon. For a single-peaked tympanogram that is asymmetrical, the pressure skewness is positive if the tail of the response is larger over positive pressures than negative pressures, and negative otherwise. Thus, the pressure skewness, which is dimensionless, is a quantitative measure of response asymmetry over static pressure.

A complementary set of moments defined over variations in frequency has no analog in single-frequency tympanometry, but may be useful for interpreting multifrequency tympanograms or wideband, ambient-pressure responses. Each frequency moment is defined by a sum of terms over the $M=61$ discrete frequencies. The frequency moments of order 0, 1, 2, and 3 are used, respectively, to define the *frequency-averaged tympanometric variable* $\langle T(P)\rangle$, the *mean frequency* $\langle f(P)\rangle$, the frequency variance $\sigma_f^2(P)$, and the *frequency skewness* $\gamma(P)$. Each of these is defined at each static pressure P in the tympanometric response, or at a single static pressure if used in ambient, or constant, static pressure. The frequency moments are defined by

$$\begin{aligned}
\langle T(P)\rangle &= \frac{1}{M}\sum_f T(P,f),\\
\langle f(P)\rangle &= \frac{1}{M\langle T(P)\rangle}\sum_f fT(P,f),\\
\sigma_f^2(P) &= \frac{1}{(M-1)\langle T(P)\rangle}\sum_f [f-\langle f(P)\rangle]^2 T(P,f),\\
\gamma(P) &= \frac{1}{(M-1)\langle T(P)\rangle\sigma_f^3(P)}\sum_f [f-\langle f(P)\rangle]^3 T(P,f).
\end{aligned} \tag{15}$$

The *frequency breadth* is $\Delta f(P)=\sqrt{\sigma_f^2(P)}$, which equals the standard deviation of the response over frequency. The summation sign with an f below denotes that the sum is over each of the M discrete frequencies in the measurement bandwidth of the response with static pressure held constant. These frequency moments differ from conventional tympanometric variables because they are defined as sums over frequency, whereas conventional tympanometry is formulated for a single frequency. They appear well-suited for use with multifrequency admittance tympanograms, inasmuch as

the variable $T(P,f)$ can be admittance, or one of its components.

In Eqs. (14) and (15), the angle brackets denote an average over either pressure or frequency, respectively, with the function argument signifying the variable that is independent. Thus, $\langle T(P)\rangle$ is the average of $T(P,f)$ over frequency, whereas $\langle T(f)\rangle$ is the average of $T(P,f)$ over static pressure.

Only the zeroth-order pressure moment in Eq. (14) contains information for an ambient-pressure measurement. In this case the zeroth-order pressure moment is the measured spectrum at the ambient pressure $\langle T(f)\rangle = T(0,f)$ [based on $P=0$ and $N=1$ in the top equation of Eq. (14)]. Such an ambient-pressure spectrum is denoted as $T_0(f)$. Each of the four frequency moments in Eq. (15) is well-defined for a measurement at ambient pressure $(P=0)$. These moments are $\langle T(0)\rangle$, $\langle f(0)\rangle$, $\Delta f(0) = \sqrt{\sigma_f^2(0)}$, and $\gamma(0)$.

B. Definitions of other central measures

Other central measures of general multifrequency tympanograms are defined based on analogies with isofrequency tympanometry. The peak-compensated static acoustic admittance (or static admittance) is defined in clinical tympanometry as the difference in the admittance magnitude at TPP and the admittance magnitude at the most positive static pressure (Van Camp *et al.*, 1986). For older children and adults, this compensation procedure removes the effect of ear-canal volume on the admittance magnitude at frequencies up to and slightly above 226 Hz. If the tympanogram has a complicated shape, it can be difficult or impossible to define the static admittance.

An alternative variable is defined that essentially reduces to the static admittance for the special case that the admittance magnitude is the tympanometric variable and the tympanogram is single-peaked, but which is also well-defined for an arbitrary tympanogram. This variable is the *extremal difference*, which is defined as the difference between the largest and smallest values of the tympanometric variable. When the extremal differences are evaluated across a range of static pressures at each fixed frequency, the pressure-averaged extremal difference $\Delta T(f)$ is defined. When extremal differences are evaluated across a range of frequencies at each fixed static pressure, the frequency-averaged extremal difference $\Delta T(P)$ is defined.

A tympanometric variable may be defined based on a quality factor, i.e., the Q, as is often used in other spectral applications. A spectral $Q(P)$ is defined in terms of a pair of pressure moments by

$$Q(P) = \langle f(P)\rangle / \Delta f(P). \tag{16}$$

IV. METHODS

A. Subjects

The subjects were recruited from a clinical population of persons with normal hearing and with hearing loss. Of the 64 subjects tested overall in the study, 35 subjects met inclusion criteria discussed below. Some subjects met inclusion criteria in both ears while others met criteria in only one ear. Results were analyzed by ear, with 23 normal-hearing subjects (18 female and 5 male) providing 42 normal-ear responses (32 female ears and 10 male ears), and 12 subjects with conductive or mixed conductive and sensorineural hearing loss (9 female and 3 male) providing 18 impaired-ear responses (13 female ears and 5 male ears). The groups of normal and impaired ears were approximately balanced across sex. The normal-hearing subjects ranged in age from 10 to 48 years, with a mean age and standard deviation (s.d.) of 19.2 ± 10.3 years. The conductives ranged in age from 11 to 55 years, with a mean age and s.d. of 30.3 ± 16.0 years. Thus, the normal-hearing subjects were younger overall. Patient records on the types of middle-ear pathologies represented in the conductive population were unavailable and, therefore, the subjects with conductive impairments were classified only in terms of their conductive hearing loss.

B. Clinical measurements

Audiometric thresholds, acoustic immittance, and wideband ATF measurements were obtained during the same test session. The static admittance (SA) was measured using 226-Hz tympanometry (Grason-Stadler middle-ear analyzer, GSI-33), and tests were judged as normal or impaired based on clinical guidelines (ASHA, 1990). In particular, the 90% range of normal middle-ear responses for patients of age 10 years to adults included SA in the range from 0.3 to 1.4 mmho. SA scores outside that range were judged as abnormal tympanograms predictive of a conductive hearing loss. The tympanometric peak pressure (TPP) was also measured using 226-Hz tympanometry or recorded as 0-daPa in cases of a flat tympanogram.

Pure-tone air-conduction (AC) thresholds were measured at 0.25, 0.5, 1, 1.5, 2, 3, 4, 6, and 8 kHz. Thresholds were measured with a standard diagnostic audiometer (Grason-Stadler, GSI-61) linked to insert earphones (Etymotic ER-3A) or headphones (Telephonics TDH-50P). Several audiologists measured AC thresholds and each chose earphones or headphones based on their individual clinical preference. For a subset of subjects on whom thresholds were measured twice with earphones and with headphones, the test condition resulting in the better threshold, if there was a difference, was chosen, creating a best-case scenario for any conductive hearing loss. A study by Voss *et al.* (2000) has indicated that middle-ear pathology can have an effect on the sound pressure generated in the ear canal by earphones and that the possible magnitude of this effect differs by type of earphone (e.g., supra-aural versus insert). However, the subects used in that study presented with reduced impedances from pathologies such as mastoid-bowl cavities, tympanostomy tubes, or tympanic membrane perforations. The subjects in the current study, on the other hand, all had intact tympanic membranes and no mastoid-bowl cavities.

Pure-tone bone-conduction (BC) thresholds were measured at 0.25, 0.5, 1, 2, and 4 kHz, with masked thresholds measured where appropriate. Measurements were completed using the audiometer described above and a bone-conduction oscillator (Radioear B-71). Audiometric equipment was calibrated using ANSI S-3.6, 1996 standards. The air–bone gaps

at the octave frequencies from 0.25 to 4 kHz were each calculated as the difference in the AC and BC thresholds.

The criteria for inclusion in the normal-hearing group were an air–bone gap of 15 dB or less at each octave from 0.25 to 4 kHz, BC thresholds of 15 dB or less at each BC test frequency, and AC thresholds of 15 dB or less at 6 and 8 kHz. An ear with an air–bone gap of 15 dB would be clinically regarded as having a borderline conductive hearing loss, but was nonetheless included in the normal group because the degree of loss would likely not require clinical management, other than a possible retest at some future date. Only 1 of the 42 ears in the normal group had an air–bone gap of 15 dB, and that large of a gap was present in that ear only at 1.0 kHz. An ear was included in the normal-hearing group even if the opposite ear did not meet the criteria for normal hearing.

The criterion for inclusion in the impaired group was an air–bone gap of 20 dB or more at any octave frequency from 0.25 to 4 kHz. No inclusion criterion was used regarding potential sensorineural hearing loss. There were 7 ears with a purely conductive loss and 11 ears with a mixed hearing loss. No subject with an ear included in the impaired group had the opposite ear included in the normal group.

C. ATF measurements

The ATF measurements were performed using a diagnostic middle-ear (DME) prototype system (Welch Allyn).[1] The DME hardware is a computer with a digital signal processor (Signalogic C31-4) that includes 16-bit sigma–delta ADCs and DACs operating at a sample rate of 22.050 kHz per channel. One output channel (DAC1) supplies an input wideband signal to the probe loudspeaker, and one input channel (ADC1) is used to digitize the probe microphone response. A second output channel (DAC2) is used for pressurized measurements to supply a dc voltage to the static pressure pump controller. The static pressure pump in the DME is provided by the pump and internal controller of a specially modified Grason-Stadler middle ear analyzer (GSI 33). The GSI 33 was modified so that the customary manual control of static pressure was controlled instead by a dc voltage input through an additional port on the back of the GSI 33 from DAC2 of the DME. The DME probe was physically similar in shape to probes typically used in otoacoustic emission measurements. The probe assembly contained three ports to couple to a loudspeaker, microphone, and static pressure pump (rather than to a second loudspeaker, as would be the configuration in a typical OAE probe). The probe used a Knowles EK-3024 microphone and a Knowles ED-1925 receiver as sound source. A tip from a set of removable rubber probe tips (Grason-Stadler) was affixed to the probe in order to provide a pressure seal of the probe in the ear canal. A plastic tube coupled the output of the GSI 33 pressure pump to the probe, as well as providing a coupling to the GSI 33 static pressure transducer used by its internal controller to maintain a constant static pressure.

The stimulus had a duration of 4096 samples or 186 ms. The stimulus energy was of short duration, containing spectral energy from 0.25 kHz as further described in Sec. II. Calibration was performed only at ambient pressure as the calibration changed only slightly under pressurized conditions (Margolis *et al.*, 1999), as was also confirmed for the DME system.

In ATF measurements at ambient static pressure, the initial 4 responses were discarded and the next 20 responses found to contain no artifact were time averaged, for a total test time of approximately 4.1 s.

In ATF measurements at varying static pressure, measurements were obtained in unequal step sizes of static pressure from +200-daPa down to −300-daPa with finer step sizes in the vicinity of the TPP. Because the tympanometric mode of the DME lacked an initial determination of the TPP, the TPP from the 226-Hz tympanogram was used to reference the static pressure steps used in the pressurized ATF measurements. The nominal static pressure steps relative to the TPP value were 200, 150, 100, 50, 30, 20, 10, 0, −10, −20, −30, −50, −150, −200, and −300 daPa. Suppose an ear had TPP=−30 daPa. Then, its nominal static pressures ranged from 200−30=170 daPa down to −300−30=−330 daPa. The actual static pressures were equal to these nominal pressures except that actual static pressures were limited to lie within the range 200 to −300 daPa. These endpoint pressures were always included in the test sequence. There were, on average across subjects, 17 steps in static pressure in a pressurized ATF measurement.

Data were acquired at each fixed static pressure, the static pressure was then adjusted to the next desired static pressure step at a rate of 150 daPa/s, and the static pressure was allowed to equalize for 0.3 s before starting the next round of data acquisition. In each round, the initial two responses at each static pressure step were discarded and the next eight responses, which were found to contain no artifact, were time averaged. The duration of the pressurized ATF test was approximately 40 s. The resolution of static pressure measurements was not varied in this study so that the influence of varying resolution on the measured ATF and the resulting energy transmittance moments is unknown. The above choices were determined in order to maintain a test time under a minute and were judged to be suitable to extract informative tympanometric shapes. The effectiveness of the chosen resolution was ultimately assessed by the ability of the system to predict a conductive hearing loss.

The ATF test protocol for each subject consisted of an ambient ATF pretest, a pressurized ATF test, and an ambient ATF post-test. The typical total test time was approximately 50 s. Inclusion of ambient pre- and post-tests allowed assessment of the influence of static pressurization on the ambient measurement, as the pressurized test occurred in between the pretest and post-test. Such changes were always small: across all frequencies and subjects, the means and standard deviations of the absolute value of the difference in energy reflectance between the pre- and post-tests were 0.0489 ±0.0206 for normal-hearing ears and 0.0474±0.0220 for hearing-impaired ears. The pretest and post-test results from ambient measurements are independently assessed in Sec. V as a measure of measurement variability, including variability induced in the subject's ear due to the presence of pressurization just before the post-test measurement.

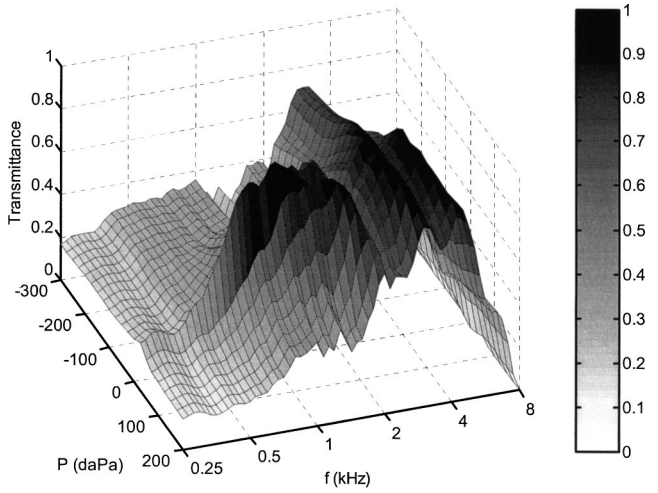


FIG. 7. Energy transmittance tympanogram versus frequency (in kHz) and static pressure (in daPa) in the ear canal for a normal-hearing ear. Energy transmittance values close to 1 are coded in dark gray and values close to 0 are coded in light gray.

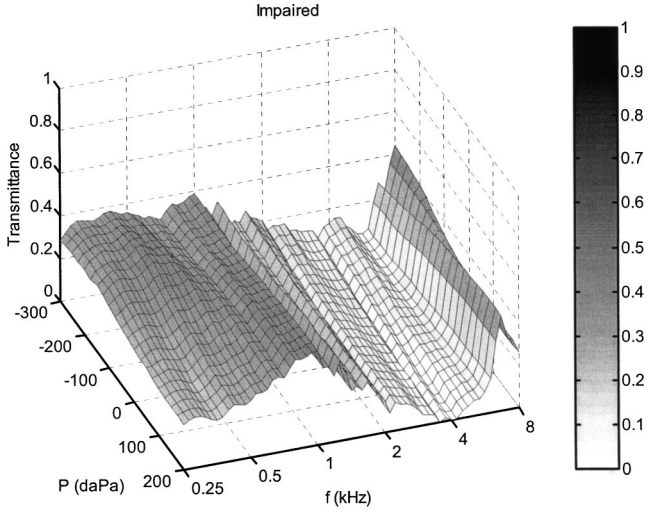


FIG. 8. Energy transmittance tympanogram versus frequency (in kHz) and static pressure (in daPa) in the ear canal for an ear with a conductive middle-ear pathology. Energy transmittance values close to 1 are coded in dark gray and values close to 0 are coded in light gray.

V. RESULTS

A. Static admittance

In the 42 normal-hearing ears, 38 226-Hz tympanograms had static admittance within the normal range (0.3 to 1.4 mmho). In the remaining tympanograms, 3 had SA in the range from 1.5 to 1.7 mmho, while the other was hypercompliant (SA=2.3 mmho). In the 18 ears with conductive hearing loss, 13 ears had SA in the normal range. In the remaining tympanograms, two were flat (SA=0 mmho), two had SA in the range from 1.5 to 1.7 mmhos, while the other was hypercompliant (SA=3.4 mmho). Based on the normal range of 0.3 to 1.4 mmho, the SA specificity was 0.90 and the sensitivity was 0.28. Based on an extended normal range of 0.3 to 1.7 mmho, the SA specificity was 0.98 and the sensitivity was 0.17.

B. Transmittance tympanometry

All results are presented using the (energy) transmittance $T(P,f)$. This is because: (1) the transmittance, like the (energy) reflectance, is insensitive to probe location so it is readily interpreted at high as well as at low frequencies, and (2) the transmittance, unlike the reflectance, has a central maxima across static pressure and frequency, and so is well-suited to the calculation of pressure and frequency moments. The energy reflectance is poorly suited for a moment analysis due to the absence of a central maximum. The transmittance response in Fig. 7 shows a pattern typical of a normal ear. As a function of frequency, the transmittance is approximately 0.2 at 0.25 kHz, increases with increasing frequency to a maximum of approximately 0.9 near 3 kHz, and decreases to lower values at 8 kHz. The corresponding reflectance in this ear has a minimum of approximately 0.1 near 3 kHz, as its pattern is simply inverted. As a function of static pressure, the transmittance at low frequencies (i.e., 0.25 kHz) has a single maximum, in this case near 0 daPa. This pressure maximum is evident as a ridge of high static pressure at increasing frequencies from 0.25 to approximately 1.5–2 kHz, and in more complicated shapes on the static pressure–frequency plane at higher frequencies.

A transmittance response from a conductive-impaired ear (Fig. 8) differed markedly from that of a normal ear. The 226-tympanometry on this ear had an SA of 0.6 cm^3 in the normal range, and a TPP of −165 daPa. The subject had a mixed hearing loss, with a sensorineural loss at 2 and 4 kHz, and a conductive loss at all frequencies tested except 2 kHz. As a function of frequency, the transmittance had lower values with increasing frequency, and rarely exceeded 0.4. As a function of static pressure, the low-frequency transmittance had no clearly defined maximum but had its largest values at low frequencies at the most negative static pressure (−300-daPa). The general tendency is that of a flattened tympanogram as a function of static pressure.

Low transmittance (as in Fig. 8) is associated with low transfer of energy into the middle ear, and would be consistent with a conductive hearing loss. To quantify this observation, it is necessary to establish a relationship between the shape of $T(P,f)$ and the presence or absence of a conductive hearing loss. This is difficult using the double-variable functions plotted in Figs. 7 and 8 because there are too many data for the number of test ears in this study. The calculation of pressure and frequency moments converts each double-variable function into a set of single-variable functions that can be more readily interpreted. As described earlier, the set of static pressures over which the pressure moments were calculated in each ear varied according to the TPP value for that ear. In each case, there were smaller steps in static pressure near the TPP and larger steps distant from the TPP. This led to efficient calculations of pressure moments using a relatively sparse number of static pressure steps. The frequency moments were calculated at each static pressure for a particular ear. Each frequency moment, which is a function of static pressure, was interpolated using a cubic spline function to a grid of static pressures from 200 down to −300 daPa in 25-daPa steps.

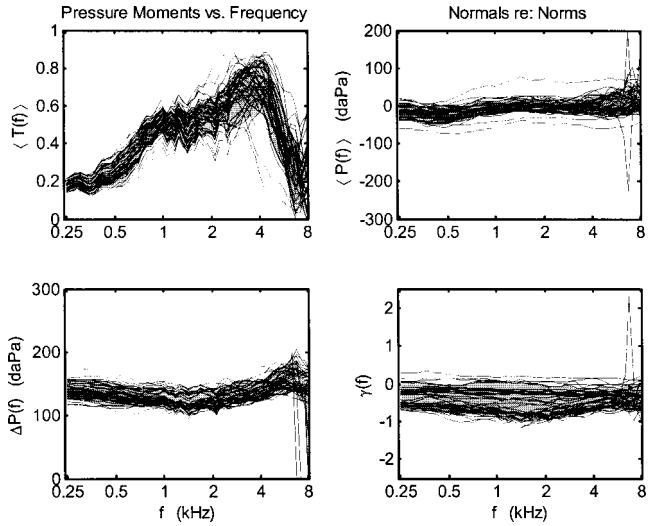


FIG. 9. The pressure moments of transmittance tympanograms versus frequency (in kHz) are calculated at each frequency and summed across static pressure. The highest order moment used is 0 for $\langle T(f)\rangle$ (upper left panel), 1 for $\langle P(f)\rangle$ (upper right panel), 2 for $\Delta P(f)$ (lower left panel), and 3 for $\gamma(f)$ (lower right panel). Each of the 42 normal-hearing responses is plotted as a line, and the 10th to 90th percentiles of the normal-hearing range are indicated by a gray band.

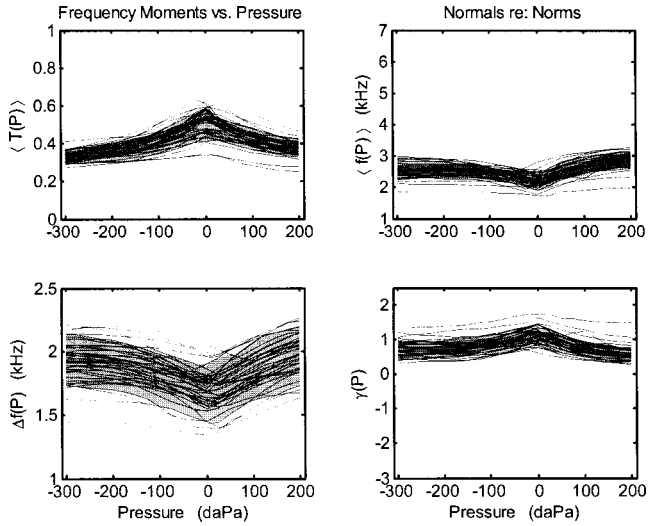


FIG. 10. The frequency moments of transmittance tympanograms versus static pressure (in kHz) are calculated at each static pressure and summed across static pressure. The highest order moment used is 0 for $\langle T(P)\rangle$ (upper left panel), 1 for $\langle f(P)\rangle$ (upper right panel), 2 for $\Delta f(P)$ (lower left panel), and 3 for $\gamma(P)$ (lower right panel). Each of the 42 normal-hearing responses is plotted as a line, and the 10th to 90th percentiles of the normal-hearing range are indicated by a gray band.

The pressure moments calculated from Eqs. (14) for the 42 normal ears are shown in Fig. 9 with gray shading indicative of the 10th–90th percentiles of the normal range. $\langle T(f)\rangle$ resembles the ambient-pressure transmittance—compare to the inverted adult energy reflectance patterns in Keefe *et al.* (1993), and Voss and Allen (1994). However, $\langle T(f)\rangle$ includes the additional averaging over static pressure at each frequency. It was noted that $\langle P(f)\rangle$ is analogous to TPP in tympanometry; $\langle P(f)\rangle$ in Fig. 9 is close to 0 daPa, as would be expected for the TPP in normal-functioning ears. $\Delta P(f)$, which is a measure of tympanometric width, has an approximately constant breadth of 130 daPa. The pressure skewness $\gamma(f)$ has values tending to lie between 0 and -1 (i.e., skewed to lower pressures), but closer to 0. A few ears in each panel of Fig. 9 show departures from the normal range, especially at frequencies above 6 kHz. This is probably due to a small measurement artifact in the implementation such that the energy reflectance sometimes slightly exceeded 1 at large positive and large negative static pressures. This may be due to an influence of static pressure on the probe transducers. These data were analyzed by setting energy reflectance to 1 whenever its value exceeded 1, which means that the energy transmittance was set to 0 for such cases. These artifacts were more often present at frequencies above 6 kHz, but the 10th to 90th percentiles were never influenced by these outliers. Therefore, there was no need to exclude them.[2]

The frequency moments calculated from Eqs. (15) for the 42 normal ears and their corresponding 10th to 90th percentiles are shown in Fig. 10. The (median) $\langle T(P)\rangle$ ranges from approximately 0.33 to 0.48 over various static pressures. The mean frequency $\langle f(P)\rangle$ ranges from 2.5–3 kHz, and indicates the frequency region where the peak transmittance in $T(P,f)$ occurs, as is evident in Fig. 7. The mean frequency is smaller in the vicinity of the mean pressure, or near 0 daPa in normal ears. This indicates the influence of the transmittance peak in the 1–1.5-kHz region, which is reduced at static pressures remote from the mean pressure $\langle P(f)\rangle$ (Margolis *et al.*, 1999), and thus pushes the mean frequency to higher values. The $\Delta f(P)$ varies from approximately 1.7 to 2.0 kHz, indicating that the spectral bandwidth of transmittance is smaller near the mean pressure than at the extremal static pressures. The $Q(P)$ calculated from Eq. (16) ranges from 1.27 to 1.49. The frequency skewness $\gamma(P)$ is positive, and more so near the mean pressure.

The pressure moments for the 18 conductive-impaired ears (Fig. 11) are substantially different from those for the normal ears, but the differences occur at different frequencies for different impaired ears. The 18 individual curves are responses from the 18 ears with conductive loss, while the gray shading indicates the same 10th to 90th percentiles calculated from the normal data of Fig. 9. Almost half of the conductives have reduced $\langle T(f)\rangle$ compared to normals in the 0.7- to 5-kHz range. This parallels the increased energy reflectance observed in impaired ears of children at risk for otitis media with effusion (Piskorski *et al.*, 1999), in neonates with cleft lip or palate, and in neonates tested in the first 24 h after birth, for whom there is more likely to be vernix present in the ear (Keefe *et al.*, 2000). $\langle T(f)\rangle$ for the impaired ears separate into two groups at low frequencies (0.25–0.5 kHz): there are 7 ears in a high-transmittance group in which $\langle T(f)\rangle$ exceeds 0.2 at 0.25 kHz, and 11 ears in a low-transmittance group. At these low frequencies, low transmittance should correspond to high stiffness, and high transmittance to low stiffness. Of the 7 ears in the high-transmittance group (i.e., for the transmittance *re*: 0.25–0.5 kHz) at frequencies higher than 0.5 kHz, 2 ears remain relative low in transmittance (never exceeding 0.4) while the other 5 ears have higher transmittance, often in the normal range. Of the 11 ears in the low-transmittance group (i.e., for

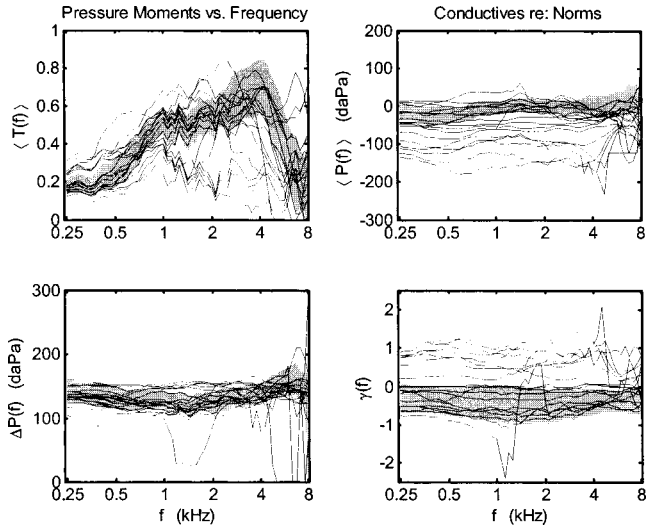


FIG. 11. The pressure moments of transmittance tympanogram versus frequency (in kHz) are calculated at each frequency and summed across static pressure. The highest order moment used is 0 for $\langle T(f)\rangle$ (upper left panel), 1 for $\langle P(f)\rangle$ (upper right panel), 2 for $\Delta P(f)$ (lower left panel), and 3 for $\gamma(f)$ (lower right panel). Each of the 18 conductive-impaired responses is plotted as a line, while the 10th to 90th percentiles of the normal-hearing range are indicated by a gray band.

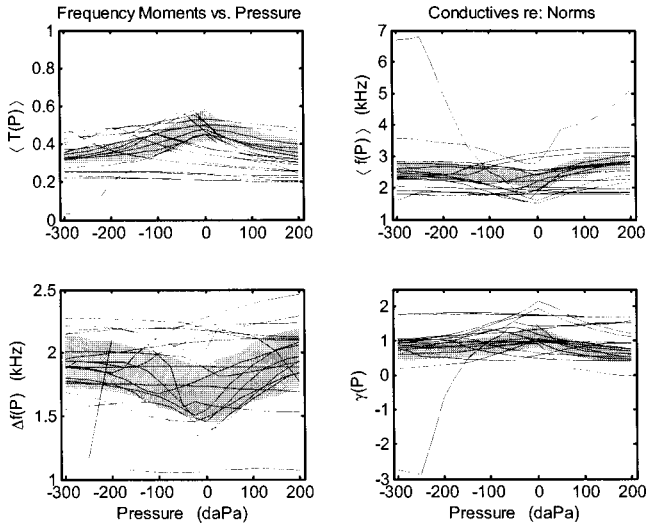


FIG. 12. The frequency moments of transmittance tympanograms versus static pressure (in kHz) are calculated at each static pressure and summed across static pressure. The highest order moment used is 0 for $\langle T(P)\rangle$ (upper left panel), 1 for $\langle f(P)\rangle$ (upper right panel), 2 for $\Delta f(P)$ (lower left panel), and 3 for $\gamma(P)$ (lower right panel). Each of the 18 conductive-impaired responses is plotted as a line, while the 10th to 90th percentiles of the normal-hearing range are indicated by a gray band.

the transmittance *re*: 0.25–0.5 kHz) at frequencies higher than 0.5 kHz, a more detailed analysis shows there is no particular pattern except that responses usually remain outside the normal range. This may indicate different structure–function relations in the two groups of ears. However, these two groups of ears do not have clearly separable responses on the other moments. The mean pressure $\langle P(f)\rangle$ is often negative in impaired ears, akin to negative TPP in 226-Hz tympanograms in impaired ears; the normal- and impaired-ear distributions are the most widely separated on mean pressure near 4 kHz. $\Delta P(f)$ in conductives is often in the normal range, although some conductives are far outside that range. As a general rule, the pressure skewness $\gamma(f)$ tends to be positive when the mean pressure is negative, although there is considerable individual variability.

The frequency moments for the 18 conductive-impaired ears (Fig. 12) are also substantially different from those for the normal ears (Fig. 10), but the differences occur at different static pressures for different impaired ears. Across the four frequency moments, the conductives are approximately as likely to have higher as lower values than the normal baselines. The $\langle T(P)\rangle$ responses tend to fall into two groups: one group with low transmittance and often flat across static pressure, and the other group with a maximum at a negative static pressure.

The extremal difference transmittance functions, which were defined by analogy with the static admittance in 226-Hz tympanometry, are shown in Fig. 13 for the 18 conductives relative to the normal baselines. As with the moments, many of the extremal difference transmittance functions in the conductive group lie outside the normal range, although there is also considerable overlap within the normal baseline.

These moments appeared to be useful in separating normal and impaired ears, by simply plotting the moments measured in a given ear in comparison to the normal baselines, as in Figs. 11 and 12. This observation led to the creation of an objective test for a conductive hearing loss.

For each ear, the moments were calculated with reference to the normal baselines. Consider the four pressure moments defined over frequency (Figs. 9 and 11). At each frequency, an error signal was defined as the distance outside the normal range, measured in units normalized to the difference between the 90th and 10th percentile values (this normalization means that the vertical distance on each plot is one unit between the 10th and 90th percentiles, and the error signals are measured in these units). If the moment was within the normal range, the error signal was defined to be zero. Each error was thus positive or zero at each frequency, and a positive error was coded irrespective of whether the individual moment was above the 90th percentile or below the 10th percentile. The errors were then averaged across the frequency range for each pressure moment, providing for each ear a single error for each of the four pressure moments. A similar procedure was used to define the error for the four frequency moments defined over each static pressure (Figs.

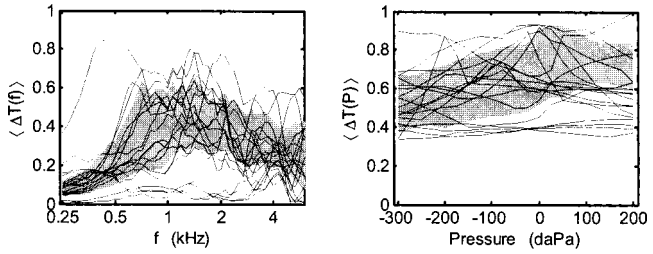


FIG. 13. The extremal range $\langle \Delta T(f)\rangle$ over all static pressures is plotted as a function of frequency (in kHz) in the left panel. The extremal range $\langle \Delta T(P)\rangle$ over all frequencies is plotted as a function of static pressure (in daPa) in the right panel. Each of the 18 conductive-impaired responses is plotted as a line, while the 10th to 90th percentiles of the normal-hearing range are indicated by a gray band.

TABLE I. Areas under the ROC curve to predict conductive hearing loss based on transmittance tympanometry.

Pressure moments	ROC area	Frequency moments	ROC area	Moment order	ROC area
$\langle T(f)\rangle$	0.86	$\langle T(P)\rangle$	0.86	0	0.92
$\langle P(f)\rangle$	0.88	$\langle f(P)\rangle$	0.92	1	0.93
$\Delta P(f)$	0.75	$\Delta f(P)$	0.79	2	0.89
$\gamma(f)$	0.78	$\gamma(P)$	0.85	3	0.84
Mean of moments (0–3)	0.94	Mean of moments (0–3)	0.93	Global mean of moments (0–3)	0.95
Mean of moments (0–2)	0.96	Mean of moments (0–2)	0.94	Global mean of moments (0–2)	0.95
$\langle \Delta T(f)\rangle$	0.83	$\langle \Delta T(P)\rangle$	0.85	Mean extremal differences	0.86

10 and 12), and the error was averaged over all static pressures. There were thus a total of eight moment errors for each ear. An ''ideal'' normal ear would have all errors equal to 0 if all the moments were within the normal baseline. In practice, some normal ears were outside the normal baseline range for some frequencies and/or some pressures, but the expectation from Figs. 9 and 10 is that the errors should remain small. Ears with a pure or mixed conductive loss have moments that are often outside the normal range (Figs. 11 and 12), so their calculated moment errors should be large. Thus, the errors can be used in a clinical signal-detection approach to predict the presence or absence of a conductive hearing loss, with large errors predicting a hearing loss.

As a refinement of this approach, the pressure–moment errors for each ear were averaged over moment orders (0–3) to provide a single pressure–moment error function, and the four frequency–moment errors were averaged over moment orders (0–3) to provide a single frequency–moment error function. Similar functions were defined by restricting the average to moment orders (0–2), as it was found that the skewness moments of order 3 were less accurate in predicting hearing loss than the other moment orders. As an alternative refinement using the moment orders from 0 to 3, pairs of error functions for pressure and frequency moments of the same order were averaged to provide error functions for each order. For example, the errors for $\langle T(f)\rangle$ and $\langle T(P)\rangle$ were averaged to define an error function for order 0. Finally, the eight pressure and frequency moment errors were averaged over moment orders from 0 to 3 to define a global moment error (0–3), and over moment orders 0 to 2 to define a global moment error function (0–2). Consideration of the relative predictive performance of these error functions provides insight on what properties of the transmittance tympanogram are helpful in predicting a conductive loss, and to test whether the composite error functions are better predictors. Error functions were defined in a similar manner for the extremal difference transmittance functions $\langle \Delta T(f)\rangle$ and $\langle \Delta T(P)\rangle$ plotted in Fig. 13.

Each error function is defined for both normal and impaired ears and encodes the ''distance'' of each ear from the normal baseline region of the distribution of moments. If these moments are capable of classifying normal and impaired ears, then the error functions of normal ears should be smaller than those of impaired ears. A hypothesis is that an increasing error function value is monotonically associated with a increasing likelihood of conductive hearing loss. If such a monotonic relationship exists, then error functions would be predictors of whether a conductive hearing loss is present. Framed in this manner, the usual tools of clinical signal detection are applicable to evaluating the extent to which the error functions predict hearing loss. The clinical signal detection literature describes the area under the receiver operating characteristic (ROC) curve as an unbiased measure of test accuracy (Swets and Pickett, 1982). A convenient method to calculate this ROC area uses the nonparametric Mann–Whitney U test (Bamber, 1975; Hanley and NcNeil, 1982; Song, 1997), which is the method used in the present analyses. It is not necessary to plot the ROC curve in order to calculate this ROC area.

The ROC area predicts how well a given test can correctly classify normal versus impaired ears, with a value of 1 representing perfect accuracy and a value of 0.5 representing chance performance. An alternative method to evaluate relative test performance is to compare the test sensitivity (the proportion of impaired ears correctly identified as impaired) at a fixed test specificity (the proportion of normal ears correctly identified as normal). Because a practical test of conductive loss should not fail many normal ears, it is appropriate to compare test sensitivities at a relatively high fixed specificity. A fixed specificity of 0.90 was chosen for this study.

The ROC area was calculated for each error function, along with specificity and sensitivity functions associated with each error function. The transmittance sensitivities were also compared with the sensitivity of 226-Hz tympanometry for the normal range of SA given in the ASHA (1990) guidelines. The measured specificity for the SA test in the present study was 0.90, which is consistent with its specificity reported in AHSA (1990). Its measured sensitivity in this study was 0.28, which means that tympanometry was unable to predict an air–bone gap of 20 dB or more in 72% of the ears in the impaired group.

The ROC areas listed in Table I ranged from 0.75 to 0.92 for the eight individual moment error functions. The mean pressure performed best, while the breadth moments and frequency skewness performed the worst (ROC areas <0.80). When the error functions of the pressure and frequency moments were averaged for each moment order (right pair of columns in Table I), the ROC areas ranged from 0.84 to 0.93 with the zeroth- and first-order moments performing best. The third-order skewness moments per-

TABLE II. Sensitivity for predicting conductive hearing loss based on transmittance tympanometry at a fixed specificity of 0.90.

Pressure moments	Sensitivity	Frequency moments	Sensitivity	Moment order	Sensitivity
$\langle T(f)\rangle$	0.72	$\langle T(P)\rangle$	0.61	0	0.83
$\langle P(f)\rangle$	0.61	$\langle f(P)\rangle$	0.67	1	0.78
$\Delta P(f)$	0.50	$\Delta f(P)$	0.61	2	0.72
$\gamma(f)$	0.56	$\gamma(P)$	0.61	3	0.67
Mean moments (0–3)	0.78	Mean moments (0–3)	0.83	Global mean moments (0–3)	0.89
Mean moments (0–2)	0.78	Mean of moments (0–2)	0.89	Global mean of moments (0–2)	0.94
$\langle \Delta T(f)\rangle$	0.61	$\langle \Delta T(P)\rangle$	0.61	Mean extremal differences	0.50

formed worst (ROC area of 0.84). When averaged over all four moments (row 5 in Table I), the error function for the pressure moments had an ROC area of 0.94 and that for the frequency moments was 0.93. The global error function constructed over all eight moments had a higher ROC area of 0.95 (row 5). Thus, when averaging over moment order or type of moment, the ROC areas tended towards higher values, but it was unknown whether including all moment orders produced the best result. Because the third-order moments were the least accurate, another set of averages was calculated for moment orders 0 through 2 with results listed in row 6 of Table I: the pressure moment ROC area was 0.96, the frequency moment ROC area was 0.94, and the ROC area for the global mean moments (0–2) was 0.95. These were the highest ROC areas that were calculated. A set of averages was calculated for moment orders 0 and 1, but the resulting ROC areas were smaller than the ROC areas listed in rows 5–6 of Table 1. The extremal-difference transmittance functions had ROC areas in the range of 0.83 to 0.85 (row 7 of Table I). Inclusion of their error functions did not improve the performance of the global mean moment error functions.

The corresponding test sensitivities at a fixed specificity of 0.90 are listed in Table II. The sensitivities of the individual moment tests ranged from 0.50 to 0.72, and each of these was larger than the 226-Hz tympanometric sensitivity of 0.28. As with the ROC areas, averaging over moment order or type of moment increased the test sensitivity, with the best test performance for the global error functions, which had the highest sensitivity of 0.94 for the global mean of moments (0–2). The corresponding average for each order resulted in the lowest sensitivity of 0.67 for moment order 3. The extremal-difference transmittance error functions had similar predictive performance to the error functions of the individual moments, and adding their error functions to the global error function did not increase the resulting sensitivity.

C. Ambient-pressure transmittance

The studies reviewed earlier showed that wideband reflectance (and, thereby, wideband transmittance) measured at ambient pressure is effective at predicting a conductive hearing loss associated with OME and middle-ear dysfunction in neonates. The present study compares the performance of energy reflectance measured at ambient static pressure with transmittance tympanometry and 226-Hz admittance tympanometry. As with the pressurized measurements, the energy reflectance was converted to an ambient-pressure energy transmittance $T_0(f)$, as described in Sec. III A. Both pretest and post-test ambient pressure results were measured, and the pretest response was the direct analog of the ambient pressure test.

The pretest responses are plotted in Fig. 14 for the 42 normal ears (left panel), and 18 impaired ears (right panel), with the 10th to 90th percentiles of the normal group shown in gray. The normal responses are similar to those reported as ambient-pressure energy reflectance functions in previous studies (Keefe *et al.*, 1993; Voss and Allen, 1994). Especially at frequencies above 1 kHz, approximately half of the impaired-ear responses are well separated from the normal-ear responses. These responses are most similar to the zeroth-order pressure moment $\langle T(f)\rangle$ plotted in the upper left-hand panels of Figs. 9 and 11, but there is greater variability in the ambient normal-ear responses than the pressurized responses. That is, the 10th to 90th percentiles are broader in Fig. 14 than in Fig. 9.

The error functions of the ambient-pressure responses were calculated for each of the 60 ears in the same manner as for the error functions of the pressurized responses. As described in Sec. III A, all four frequency moments are well defined at the $P=O$ pressure corresponding to ambient pressure. The resulting ROC areas are listed in Table III, which has an otherwise similar format to the pressurized responses in Table I. The ROC area of the error function based on $T_0(f)$ was 0.87 for the pretest and 0.80 for the post-test, and serves as a baseline of comparing the ROC areas for other

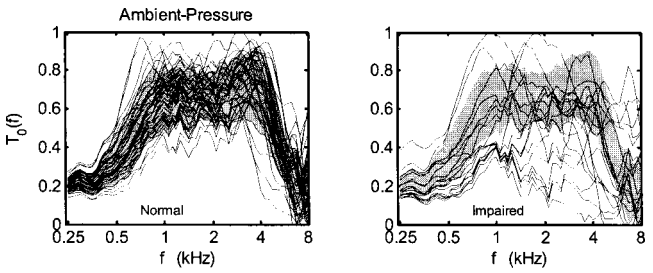


FIG. 14. Energy transmittance measured at ambient static pressure versus frequency for normal-hearing subjects (left panel) and conductive-impaired subjects (right panel). Each individual response is plotted as a line, while the 10th to 90th percentiles of the normal-hearing range are indicated by a gray band.

TABLE III. Areas under the ROC curve to predict conductive hearing loss based on ambient-pressure transmittance. Areas are provided for pretest, and for the post-test within parentheses. Cells for which moments were undefined in the ambient-pressure test are left blank.

Pressure moments	ROC area	Frequency moments	ROC area	Moment order	ROC area
$T_0(f)$	0.87 (0.80)	$\langle T(0)\rangle$	0.70 (0.60)	0	0.85 (0.76)
		$\langle f(0)\rangle$	0.77 (0.77)	1	
		$\Delta f(0)$	0.79 (0.80)	2	
		$\gamma(0)$	0.67 (0.66)	3	
Mean moments (0–3)	0.87 (0.80)	Mean moments (0–3)	0.87 (0.82)	Global mean moments (0–3)	0.89 (0.83)
Mean moments (0–2)	0.87 (0.80)	Mean moments (0–2)	0.88 (0.83)	Global mean moments (0–2)	0.90 (0.84)

ambient-pressure moments. The pretest ROC area was similar to that for $\langle T(f)\rangle$ in the pressurized responses (with its ROC area of 0.86 in Table I). In the case of the post-test, the normal baselines were those calculated in the pretest, so that the post-test had aspects of a replication in the same ears, except that the post-test followed the pressurized measurement and also included any physiological aftereffects of pressurization on the middle ear.

The individual frequency moments had ROC areas ranging from 0.67 to 0.79 in the pretest (and 0.60 to 0.80 in the post-test). The frequency skewness (moment order 3) was the least accurate. The mean error function of the four frequency moments had an ROC area of 0.87 in the pretest (and 0.82 in the post-test), while that averaged only over orders (0–2) had an ROC area of 0.88 in the pretest (and 0.83 in the post-test). Thus, this average over only three values per ear (i.e., the moments calculated for orders 0–2) was perhaps a slightly better predictor, and no worse, than the original $T_0(f)$. The global mean error function, which was calculated as the average of the mean error functions of the pressure and frequency moments, resulted in an ROC area of 0.89 in the pretest (and 0.83 in the post-test) for the average over moments (0–3) (row 5 of Table III), and ROC area of 0.90 in the pretest (and 0.84 in the post-test) for the average over moments (0–2) (row 6 of Table III). Thus, the global mean performed slightly better than $T_0(f)$ in each of the ambient tests, with the best performance obtained for averages over moments 0–2, as in the analyses of the pressurized responses.

The corresponding ambient-pressure test sensitivities at a fixed specificity of 0.90 are listed in Table IV. The sensitivities of the individual moment pretests ranged from 0.44 to 0.72, each of which was larger than the 226-Hz tympanometric sensitivity of 0.28. The frequency skewness (moment order 3) had the lowest sensitivity among the moments. The pretest sensitivity of $T_0(f)$ was nearly equal to that of the global error function, with, however, a decrease in sensitivity of $T_0(f)$ in the post-test (from 0.72 to 0.56), compared to no decrease in sensitivity of either of the global error functions in the pretest and post-test (0.72 in both). Thus, the moment analysis is concluded to slightly improve the generalizability of the ambient-pressure test performance, with the global error function over moments (0–2) as the best-performing ambient-pressure test, or at least no worse than the global error function over moments (0–3).

VI. DISCUSSION

A. Summary of findings

Pressure and frequency moments were calculated for energy transmittance responses measured in normal-hearing ears and ears with a conductive hearing loss. Predictors of hearing loss based on these moments were defined in terms of error functions relative to the distribution of moments measured in ears with normal hearing. The predictors based on the individual pressure and frequency moments of orders 0 through 3 were calculated, and the predictors derived from the third-order moments were the least accurate across moment order. It is possible that the third-order moments representing the skewness of responses were unduly sensitive to boundary effects associated with the finite measurement bandwidth and the finite range of static pressures. The most accurate predictors were based on a global error function, which was calculated as an average across error functions over moment type and moment orders 0, 1, and 2. The results obtained using these predictors are summarized below.

TABLE IV. Sensitivity for predicting conductive hearing loss based on transmittance measured at ambient pressure at a fixed specificity of 0.90. Sensitivities are provided for pretest, and for the post-test within parentheses. Cells for which moments were undefined in the ambient-pressure test are left blank.

Pressure moments	Sensitivity	Frequency moments	Sensitivity	Moment order	Sensitivity
$T_0(f)$	0.72 (0.56)	$\langle T(0)\rangle$	0.50 (0.39)	0	0.72 (0.44)
		$\langle f(0)\rangle$	0.61 (0.50)	1	
		$\Delta f(0)$	0.56 (0.56)	2	
		$\gamma(0)$	0.44 (0.39)	3	
Mean moments (0–3)	0.72 (0.56)	Mean moments (0–3)	0.72 (0.72)	Global mean moments (0–3)	0.72 (0.72)
Mean moments (0–2)	0.72 (0.56)	Mean moments (0–2)	0.72 (0.78)	Global mean moments (0–2)	0.72 (0.72)

In order to predict a conductive hearing loss based on measurements in the same group of ears at a fixed specificity of 0.90, the sensitivity of 226-Hz tympanometry (0.28) was less than the sensitivity (0.72) of ambient-pressure transmittance, and both were less than the sensitivity (0.94) of pressurized energy transmittance. The area under the ROC curve for ambient-pressure energy transmittance (0.90) was less than that for pressurized energy transmittance (0.95). This suggests that energy transmittance is an accurate predictor of conductive hearing loss in human ears of older children and adults, and that it is a much better predictor than 226-Hz tympanometry. For the pressurized energy transmittance test, a slightly higher test specificity of 0.93 was associated with a reduced sensitivity of 0.78. If test specificity and sensitivity are judged to be of equal importance in a clinical test to predict conductive hearing loss, the best test performance obtained was for the pressurized ATF test with a specificity of 0.90 and sensitivity of 0.94, although the relatively small number of impaired ears ($N=18$) is a limitation of the study.

It was noted on several occasions that approximately ''half'' of the impaired-ear moments lie outside the normal-ear moment baselines at particular frequencies in Fig. 11 and at particular static pressures in Fig. 12. Thus, it may appear surprising that the error functions are capable of identifying 90% of normal ears and 94% of impaired ears, as opposed to perhaps 50% (i.e., ''half'') of impaired ears. The solution to this conundrum is that the impaired ears have moments that are more likely to lie outside the normal baselines at some frequencies or at some static pressures than the normal ears. The individual composition of the half of the impaired ears that are outliers varies across frequency (or static pressure). Across all frequencies and static pressures, the moments of nearly all the impaired ears do lie sufficiently outside the normal baselines that their global error function becomes large. This is evident in individual plots of the 18 impaired ears relative to the normal baselines, but is not obviously evident in the summary plots in Figs. 11 and 12. The error functions encode this information as a cumulative distance measure from the normal baseline range of responses. This property also explains why it is helpful to assess the middle-ear response at more than one frequency, as some middle-ear pathologies produce shifts in energy transmittance only at higher frequencies.

The potential use of an ambient-pressure ATF test in clinical measurements is supported by these results, but it may be limited to applications in which ear-canal pressurization is determined to be undesirable. The pressurized ATF test is concluded to be more suitable for a comprehensive hearing diagnostic instrument, as it was the most accurate in predicting a conductive hearing loss. The pressurized ATF test provides other frequency-specific diagnostic information that ambient-pressure ATF does not, such as is encoded in the pressure and frequency moments, and in the full energy transmittance and admittance tympanograms. Some examples in individual ears are described in the following section. The admittance tympanogram may also be useful in assessing low-frequency pathologies in which knowledge of ear-canal volume is important, e.g., by comparing the equivalent-volume, or susceptance, tympanogram across static pressures (Keefe and Levi, 1996).

B. Selected individual studies

The single ear with both a conductive hearing loss and a hypercompliant 226-Hz tympanogram had unusual pressure and frequency moments compared to those of normal ears. For the pressure moments, this ear had the largest $\langle T(f)\rangle$ of any ear tested between 0.35–0.6 kHz (Fig. 11), and the smallest $\Delta P(f)$ (close to 100 daPa in Fig. 11) of any ear tested between 0.25–0.8 kHz. The large $\langle T(f)\rangle$ at low frequencies is likely to be associated not with energy transfer through the middle ear into the cochlea, but rather with excess energy absorption within the middle ear. The small pressure breadth $\Delta P(f)$ at low frequencies, which is akin to tympanometric width, suggests that the hypercompliant tympanic membrane was associated with a narrow-width tympanogram in this ear with hearing loss. The frequency moments of this ear were within the range of moments of the other conductives.

The single normal-hearing ear with a hypercompliant 226-Hz tympanogram had all its moments lying approximately within the 10th to 90th percentiles of the normal group; in particular, its pressure breadth was not unusually small. Thus, the transmittance moments may prove useful in classifying ears with hypercompliant 226-Hz tympanograms as to the presence of conductive hearing loss.

There were two ears with flat 226-Hz tympanograms ($SA=0$) in the group of ears with conductive hearing loss. These ears had moments that were outside the normal range of many of the moments, but their responses stand out as the pair of nearly horizontal curves with $\langle T(P)\rangle \approx 0.22$ in Fig. 12. This variable is the average energy transmittance over all frequencies, and its insensitivity to static pressure is an expression that the entire transmittance tympanogram was approximately flat.

Only one normal ear (male, age 34 years) had a TPP (of 110 daPa) at 226 Hz exceeding 30 daPa. The static admittance was 1.0 mmho. This normal-hearing ear was misclassified by each of the pressurized and ambient-pressure predictors. It might be useful for a middle-ear test to identify such an ear, because a positive TPP$>$50 daPa is a risk factor for acute otitis media (Margolis and Hunter, 2000). As expected from the TPP value, this ear had the largest mean pressure $\langle P(f)\rangle$ across frequency, with a value in the range from 60–70 daPa between 1 and 2 kHz (upper right panel, Fig. 9). The ear also had the largest pressure breadth $\Delta P(f)$ (170 daPa at 0.25 kHz) in the normal group at low frequencies (lower left panel, Fig. 9). The ear response was outside the normal distribution of $\langle T(P)\rangle$, with a local maximum transmittance of 0.55 just above 100 daPa (upper left panel, Fig. 10). A peak near 110 daPa was also evident for the skewness $\gamma(p)$ (lower right panel, Fig. 10). It had the largest frequency breadth $\Delta f(P)$ between -200 to 20 daPa, which is above the normal range, and a minimum close to 110 daPa (lower left panel, Fig. 10). The ear had an abnormally low ambient-pressure transmittance in the 1–2 kHz range, and was the lowest response in Fig. 14 (left panel) at 1.5 kHz.

Only one normal ear (female, age 12 years) had a negative TPP at 226 Hz less than −30 daPa, with TPP=−75 daPa at 226 Hz. Its static admittance was 0.5 mmho. This ear was also one of the four ears misclassified as having a conductive hearing loss by the pressurized test. Aside from the responses with sharp peaks and dips above 6 kHz, the mean pressure $\langle P(f)\rangle$ for this ear was the most negative outlier from the normal range with a value fluctuating between −75 to −40 daPa across frequency (upper right panel, Fig. 9). Its skewness $\gamma(f)$ was above the normal range at many frequencies (lower right panel, Fig. 9). Its frequency moments were entirely within the normal range (Fig. 10).

The normal ear (female, 15 years) that had the highest global error and failed both of the best pressurized and ambient-pressure tests had a normal 226-Hz tympanogram (SA=0.3 mmho, TPP=30 daPa). This ear had the lowest pressurized $\langle T(f)\rangle$ from 3–6 kHz (upper left panel, Fig. 9) and the lowest ambient-pressure $T_0(f)$ from 4–6 kHz (left panel, Fig. 14) of all normal ears tested. Its pressure breadth $\Delta P(f)$ was usually elevated above the normal range from 0.25–4 kHz (lower left panel, Fig. 9), and its mean pressure $\langle P(f)\rangle$ was occasionally elevated (upper right panel, Fig. 9). Its frequency moments were nearly always below the normal range, and were often the lowest across all normal ears for $\langle T(P)\rangle$, $\langle f(P)\rangle$, and $\Delta f(P)$, while its value for $\gamma(P)$ was the highest. This ear, more than any other in the normal group, had an unusual energy transmittance without an adequate explanation.

The fourth ear that failed the best pressurized and ambient-pressure tests had the next-lowest: ambient-pressure tests had the next-lowest transmittance $\langle T(f)\rangle$ (upper left panel, Fig. 9) and next-lowest $T_0(f)$ (left panel, Fig. 14) from 4–6 kHz; these transmittances were also below their normal range from 0.5–1 kHz.

For the impaired group, the pressurized global error function correctly identified 17 of 18 ears as impaired. For the impaired ear that was misclassifed as normal, the subject (female, age 30 years) had a purely conductive air–bone gap of 35 dB at 0.25 kHz, a borderline gap of 15 dB at 0.5 kHz, and no significant gap at higher frequencies. The 226-Hz tympanogram also incorrectly classified the ear as normal (SA=0.4 mmho, $TPP=10$ daPa). Consistent with a low-frequency loss, the pressurized $\langle T(f)\rangle$ (upper left panel, Fig. 11) and ambient-pressure $T_0(f)$ (right panel, Fig. 14) responses were slightly below, or just within the 10th percentile of, the normal group from 0.25–0.5 kHz, and tended to lie within the normal range at higher frequencies. Among the other pressure moments, $\langle P(f)\rangle$ and $\gamma(f)$ were in the normal range at all but at most 1 of the 61 test frequencies. However, $\Delta P(f)$ was below the normal range over approximately 80% of the test frequencies, and was the next to lowest among the conductives from 0.25–0.6 kHz (lower left panel, Fig. 11). In the frequency moments, $\langle T(P)\rangle$ and $\gamma(P)$ were in the normal range at all pressures, but $\langle f(P)\rangle$ tended to be slightly higher than the normal range for $-200\leqslant P\leqslant -50$, and $\Delta f(P)$ had an unusual sigmoidal shape, rather than a single-peaked shape, and was above the normal range for $-70\leqslant P\leqslant -10$ daPa.

C. Implications for future research

The use of static pressure distends the highly compliant ear-canal walls of neonates and young infants (Holte *et al.*, 1991), and the compliant and lossy properties associated with ear-canal wall motion are thought to influence the interpretation of acoustical transfer-function measurements, particularly at frequencies below 1–1.5 kHz (Keefe *et al.*, 1993, 2000). It is hypothesized that the low-frequency information in the energy transmittance tympanogram should be weighted less strongly in constructing predictors of middle-ear dysfunction in infants tested in newborn hearing screening programs. This is consistent with the weightings of the predictors of middle-ear dysfunction in neonates that were derived from ambient-pressure ATF measurements (Keefe *et al.*, 2003b).

More research is needed to examine whether additional information, beyond that in the pressure and frequency moments, is contained in the two-dimensional transmittance function (as plotted in Figs. 7 and 8), which might be useful in the identification of particular middle-ear pathologies. The individual-ear studies described above suggest that there may be instances in which a frequency-specific weighting of moments, or an unequal weighting of the contributions of different moments leads to improved performance. An unequal weighting approach has been used to predict conductive hearing loss associated with otitis media with effusion (Piskorski *et al.*, 1999) and to predict transient middle-ear dysfunction in neonates (Keefe *et al.*, 2003b). A person skilled in interpreting energy reflectance tympanograms may obtain additional diagnostic information, as described in Margolis *et al.* (1999) and Margolis *et al.* (2001). Feeney *et al.* (2003) reported frequency-specific patterns for a range of middle-ear disorders using ambient-pressure energy reflectance. While some disorders appeared to have a unique pattern in the reflectance spectrum, Feeney *et al.* commented on the similarity between the responses measured in an otosclerotic ear and a normal-hearing ear with a negative TPP. Frequency- and pressure-specific patterns in pressurized energy transmittance responses may provide additional information with which to classify middle-ear disorders. Two of the individual-ear responses described above suggest that it may be possible to classify pressurized transmittance responses in ears with hypercompliant tympanic membranes according to the likelihood that a conductive hearing loss is present or absent.

The advantage of the tympanometric moment approach is that it provides an objective means of accurately detecting the presence or absence of a conductive hearing based on a measured ATF response. The potential application of wideband ATF responses to identifying particular middle-ear pathologies, for which a conductive hearing loss may not even be present, is a topic for future research.

One limitation of the pressurized ATF measurement using the DME prototype system was that an initial 226-Hz tympanogram was required to determine the TPP. The static pressures in each ear were selected as steps relative to this TPP, or 0 daPa in the case of a flat tympanogram. A practical clinical instrument would need to integrate this preliminary static pressure sweep into a test protocol, or else adopt a

different procedure of choosing the static pressure steps or sweep parameters.

VII. CONCLUSIONS

A method to measure ambient-pressure and pressurized ATFs over a bandwidth from 0.25 to 8.0 kHz was described based on the use of a calibration procedure in two cylindrical tubes. A moment formulation was devised for wideband tympanometric responses and ambient-pressure spectral responses, and applied to ATF measurements of energy transmittance in the ear canal. Comparing tests at a fixed specificity of 0.90, the sensitivity was 0.28 for static admittance at 226 Hz, 0.72 for ambient-pressure ATF, and 0.94 for pressurized ATF. Pressurized ATF was accurate at predicting conductive hearing loss with an area under the receiver operating characteristic curve of 0.95. While ambient-pressure ATF tests may have sufficient accuracy to use in selected hearing screening applications, pressurized ATF tests may be more suitable for clinical diagnostics inasmuch as they are more accurate than ambient-pressure ATF in predicting a conductive hearing loss in older children and adults.

ACKNOWLEDGMENTS

Michael Gorga at BTNRH and two anonymous journal reviewers provided helpful comments on earlier versions of this report. The data collection, the development of the moment analysis methodology, and their application to analyzing the research data were supported by the NIDCD (DC03784).

APPENDIX: LIST OF SYMBOLS

Any variable such as X that is denoted as $X[k]$ represents the variable value in the kth discrete frequency bin.

c_a	Phase velocity of sound in air
$c[k]$	Phase velocity of sound in a tube
d	Tube diameter
f	Frequency
f_k	Center frequency of kth frequency bin
$p_1[n]$	Pressure waveform in tube 1
$p_2[n]$	Pressure waveform in tube 2
$q[n]$	Incident pressure waveform
r_t^{-1}	Thermal boundary layer thickness ratio (dimensionless)
r_v^{-1}	Viscous boundary layer thickness ratio (dimensionless)
t_0	Onset time of incident signal
t_1	End time of incident signal
t_1'	Nominal onset time of first-reflected signal
$w_a[k]$	Weighting variable defined using A
$w_b[k]$	Weighting variable defined using B and C
$w_2[k]$	Weighting coefficient for source reflectance from tube-2 data
A	Constant for smoothing source reflectance
B	Constant for smoothing source reflectance
C	Constant for smoothing source reflectance
K	Number of discrete frequencies used in calibration
L_1	Length of tube 1
L_2	Length of tube 2
MSE	Mean squared error between reflectance data and model
M	Number of discrete frequencies used in tympanogram
N	Number of static pressure steps used in tympanogram
P	Static pressure (but only when P is used without $[k]$)
$P[k]$	(Acoustic) pressure spectrum measured in an ear
$P_{\min}$	Minimum pressure amplitude
$P^T[k]$	Thevenin source pressure
$P_1[k]$	Pressure spectrum of tube 1
$P_2[k]$	Pressure spectrum of tube 2
$Q[k]$	Incident spectrum of source
$R[k]$	Nominal pressure reflectance in an ear (before the area correction)
$R_e[k]$	Pressure reflectance in an ear (after the area correction)
$1-\lvert R_e[k]\rvert^2$	Energy transmittance in an ear (unspecified static pressure)
$R^c[k]$	Pressure reflectance model for tube of arbitrary length L
$R_1^c[k]$	Pressure reflectance model for tube 1 (''model reflectance of tube 1'')
$R_2^c[k]$	Pressure reflectance model for tube 2 (''model reflectance of tube 2'')
$R_1^d[k]$	Pressure reflectance calculated for tube 1 using current source reflectance (''data reflectance of tube 1'')
$R_2^d[k]$	Pressure reflectance calculated for tube 2 using current source reflectance (''data reflectance of tube 2'')
$R_1^0[k]$	Source reflectance estimated from data measured in tube 1
$R_2^0[k]$	Source reflectance estimated from data measured in tube 2
$R^0[k]$	Source reflectance
SA	Static admittance (from 226-Hz tympanogram)
S_e	Cross section of ear canal at the probe tip
$S_1[k]$	Sound level of $Q[k]/P_{\min}$
$S_2[k]$	Sound level of $P2[k]/P_{\min}$
$T(P,f)$	Tympanometric variable, or (energy) transmittance, at static pressure P and frequency f
TPP	Tympanometric peak pressure (from 226-Hz tympanogram)
$Y[k]$	Acoustic admittance
Z_c	Characteristic impedance in calibration tubes
Z_{ce}	Characteristic impedance in an ear
$Z^T[k]$	Thevenin impedance of measurement system
Γ	Propagation wave number in calibration tubes
Γ_e	Propagation wave number in an ear
ΔR	rms reflectance error
ε_{on}	Onset threshold parameter (in ADC units)
ε_{off}	Offset threshold parameter (in ADC units)
η	Shear coefficient of viscosity of air

ρ	Equilibrium density of air
τ	Round-trip time to closed end of calibration tube
χ	Loss parameter (dimensionless)

Pressure moments:

$\langle T(f)\rangle$	Pressure-averaged transmittance (or general tympanometric variable)
$\langle P(f)\rangle$	Mean pressure
$\sigma_p^2(f)$	Pressure variance
$\gamma(f)$	Pressure skewness
$\Delta P(f)$	Pressure breadth

Frequency moments:

$\langle T(P)\rangle$	Frequency-averaged transmittance (or general tympanometric variable)
$\langle f(P)\rangle$	Mean frequency
$\sigma_f^2(P)$	Frequency variance
$\gamma(P)$	Frequency skewness
$\Delta f(P)$	Frequency breadth

Other variables related to moment variables:

$T_0(f)$	Ambient-pressure transmittance
$\Delta T(f)$	Pressure-averaged extremal difference
$\Delta T(P)$	Frequency-averaged extremal difference
$Q(P)$	Quality factor

[1]Douglas H. Keefe was involved during 1998–2001 in the design and implementation of the Welch Allyn diagnostic middle-ear (DME) prototype system prior to the time that Welch Allyn discontinued commercial development of the system. The research data reported in this study were obtained with a DME system that was commercially obtained by Keefe at Boys Town National Research Hospital. The research data were collected after Welch Allyn discontinued commercial development of the DME system. A goal of this report is to describe the system in sufficient detail that another knowledgeable investigator might replicate the system and use it for research work, and to contribute to improved methods for measuring ATFs in the ear canal.

[2]As a general remark, the effect of static pressure on the calibration (i.e., the source reflectance and incident pressure) is much smaller than the effect of static pressure on responses measured in the ear canal. Thus, to first approximation, it is possible to simply ignore the effect of static pressure on the calibration, as was done in this study. The constancy of the 10th–90th percentiles to whether ears with artifact are included or excluded confirms that this first approximation is a reasonable approach. To a second approximation, greater accuracy would be obtained by performing separate calibrations over the range of static pressures to be used in measurements in ears.

Agulló, J., Cardona, S., and Keefe, D. H. (**1995**). ''Time-domain deconvolution to measure reflection functions from discontinuities in waveguides,'' J. Acoust. Soc. Am. **97**, 1950–1957.

American Speech-Language-Hearing Association (**1990**). ''Guidelines for screening for hearing impairments and middle-ear disorders,'' ASHA **32** (Suppl. 2), 17–24.

Bamber, D. (**1975**). ''The area above the ordinal dominance graph and the area below the receiver operating characteristic graph,'' J. Math. Psychol. **12**, 387–415.

Benade, A. H. (**1968**). ''On the propagation of sound waves in a cylindrical conduit,'' J. Acoust. Soc. Am. **44**, 616–623.

Cohen, L. (**1995**). *Time-Frequency Analysis* (Prentice-Hall, New York).

Feeney, M. P., Grant, I. L., and Marryott, L. P. (**2003**). ''Wideband energy reflectance measurements in adults with middle-ear disorders,'' J. Speech Lang. Hear. Res. **4**, 901–911.

Hanley, J. A., and NcNeil, B. J. (**1982**). ''The meaning and use of the area under a receiver operating characteristic (ROC) curve,'' Radiology **143**, 29–36.

Holte, L., Margolis, R. H., and Cavanaugh, Jr., R. M. (**1991**). ''Developmental changes in multifrequency tympanograms,'' Audiology **30**, 1–24.

Huang, G. T., Rosowski, J. J., Puria, S., and Peake, W. T. (**2000**). ''A noninvasive method for estimating acoustic admittance at the tympanic membrane,'' J. Acoust. Soc. Am. **108**, 1128–1146.

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**, 58–62.

Keefe, D. H. (**1996**). ''Wind-instrument reflection function measurements in the time domain,'' J. Acoust. Soc. Am. **99**, 2370–2381; **100**, 3985(E) (1996).

Keefe, D. H. (**1997**). ''Otoreflectance of the cochlea and middle ear,'' J. Acoust. Soc. Am. **102**, 2849–2859.

Keefe, D. H., and Benade, A. H. (**1981**). ''Impedance measurement source and microphone proximity effects,'' J. Acoust. Soc. Am. **69**, 1489–1495.

Keefe, D. H., Bulen, J. D., Hoberg Arehart, K., and Burns, E. M. (**1993**). ''Ear-canal impedance and reflection coefficient of human infants and adults,'' J. Acoust. Soc. Am. **94**, 2617–2638.

Keefe, D. H., Folsom, R. C., Gorga, M. P., Vohr, B. R., Bulen, J. C., and Norton, S. J. (**2000**). ''Identification of neonatal hearing impairment: Ear-canal measurements of acoustic admittance and reflectance in neonates,'' Ear Hear. **21**, 443–461.

Keefe, D. H., Gorga, M. P., Neely, S. T., Zhao, F., and Vohr, B. R. (**2003b**). ''Ear-canal acoustic admittance and reflectance measurements in human neonates. II. Predictions of middle-ear dysfunction and sensorineural hearing loss,'' J. Acoust. Soc. Am. **113**, 407–444.

Keefe, D. H., and Levi, E. (**1996**). ''Maturation of the middle and external ears: Acoustic power-based responses and reflectance tympanometry,'' Ear Hear. **17**, 361–373.

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., Ling, R., and Bulen, J. C. (**1992**). ''Method to measure acoustic impedance and reflection coefficient,'' J. Acoust. Soc. Am. **91**, 470–485.

Keefe, D. H., Zhao, F., Neely, S. T., Gorga, M. P., and Vohr, B. R. (**2003a**). ''Ear-canal acoustic admittance and reflectance effects in human neonates. I. Predictions of otoacoustic emission and auditory brainstem responses,'' J. Acoust. Soc. Am. **113**, 389–406.

Margolis, R. H., and Goycoolea, H. G. (**1993**). ''Multifrequency tympanometry in normal adults,'' Ear Hear. **14**, 408–413.

Margolis, R. H., and Hunter, L. L. (**2000**). ''Acoustic immittance measurements,'' in *Audiology: Diagnosis*, edited by R. J. Roeser, M. Valente, and H. Hosford-Dunn (Thieme Medical, New York).

Margolis, R. H., Paul, S., Saly, G. L., Schachern, P. A., and Keefe, D. H. (**2001**). ''Wideband reflectance tympanometry in chinchillas and humans,'' J. Acoust. Soc. Am. **110**, 1453–1464.

Margolis, R. H., Saly, G. L., and Keefe, D. H. (**1999**). ''Wideband reflectance tympanometry in normal adults,'' J. Acoust. Soc. Am. **106**, 265–280.

Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (McGraw-Hill, New York).

Piskorski, P., Keefe, D. H., Simmons, J. L., and Gorga, M. P. (**1999**). ''Prediction of conductive hearing loss based on acoustic ear-canal response using a multivariate clinical decision theory,'' J. Acoust. Soc. Am. **105**, 1749–1764.

Polack, J.-D., Meynial, X., Kergomard, J., Cosuard, C., and Bruneau, M. (**1987**). ''Reflection function of a plane sound wave in a cylindrical tube,'' Rev. Phys. Appl. **22**, 331–337.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (**1992**). *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge).

Rabiner, L. R., and Schafer, R. W. (**1978**). *Digital Processing of Speech Signals* (Prentice-Hall, Englewood Cliffs, NJ).

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. **114**, 944–966.

Sininger, Y. S., Cone-Wesson, B., Folsom, R. C., Gorga, M. P., Vohr, B. R., Widen, J. E., Ekelid, M., and Norton, S. J. (**2000**). ''Identification of neonatal hearing impairment: Auditory brain stem responses in the perinatal period,'' Ear Hear. **21**, 383–399.

Song, H. H. (**1997**). ''Analysis of correlated ROC areas in diagnostic testing,'' Biometrics **53**, 370–382.
Stinson, M. R., Shaw, E. A. G., and Lawon, B. W. (**1982**). ''Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal,'' J. Acoust. Soc. Am. **72**, 766–773.
Swets, J. A., and Pickett, R. M. (**1982**). *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory* (Academic, New York).
Van Camp, K. J., Margolis, R. H., Wilson, R. H., Creten, W. L., and Shanks, J. E. (**1986**). *Principles of Tympanometry*, ASHA Monographs Number 24 (American Speech-Language-Hearing Association, Rockville).
Vanhuyse, V. J., Creten, W. L., and Van Camp, K. J. (**1975**). ''On the W-notching of tympanograms,'' Scand. Audiol. **4**, 45–50.
Voss, S. E., and Allen, J. B. (**1994**). ''Measurement of acoustic impedance and reflectance in the human ear canal,'' J. Acoust. Soc. Am. **95**, 372–384.
Voss, S. E., Rosowski, J. J., Merchant, S. N., Thornton, A. R., Shera, C. A., and Peake, W. T. (**2000**). ''Middle-ear pathology can affect the ear-canal sound pressure generated by audiologic earphones,'' Ear Hear. **21**, 265–274.
Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2001**). ''Middle-ear function with tympanic-membrane perforations. I. Measurements and mechanisms,'' J. Acoust. Soc. Am. **110**, 1432–1444.

A longitudinal study of distortion product otoacoustic emission ipsilateral suppression and input/output characteristics in human neonates

Carolina Abdala[a)]
Children's Auditory Research and Evaluation Center, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 4 April 2003; revised 13 September 2003; accepted 17 September 2003)

Past work has shown that distortion product otoacoustic emission (DPOAE) ($2f1-f2$) ipsilateral suppression and input/output (I/O) characteristics are not adult-like in prematurely born neonates [Abdala, J. Acoust. Soc. Am. **110**, 1465–1476 (2001)]. These age differences are most pronounced at $f2=6000$ Hz and have been interpreted to indicate a subtle immaturity in human cochlear function prior to term birth. It is still not clear, however, whether term-born neonates are completely adult-like in cochlear function. To study this question, DPOAE suppression and I/O functions for $f2=6000$ Hz were measured in a group of prematurely born neonates at weekly intervals over a period of 7–8-weeks, a group of normal-hearing adults, and during a one-time test session in a group of term-born neonates. Results show that there was no significant change in suppression tuning, suppression growth, and various I/O characteristics across test session for premature neonates, but there was an age-group effect; even once prematurely born neonates reached the equivalence of term-like status (38–40-weeks postconceptional weeks), they continued to show narrower suppression tuning than adults, shallower suppression growth for low-frequency side suppressor tones, and an elevated amplitude saturation plateau on the I/O function. Term-born neonates showed DPOAE results that were comparable to those measured from premature neonates and unlike adult findings. These results suggest that a subtle immaturity in cochlear function persists into the postnatal period. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1625930]

PACS numbers: 43.64.Jb, 43.64.Kc [BLM] Pages: 3239–3250

I. INTRODUCTION

Several characteristics of the $2f1-f2$ DPOAE are not adult-like in premature neonates. Premature neonates have narrower suppression tuning curves (STC) than adults at 1500 and 6000 Hz and atypical, compressive growth of suppression for suppressor tones lower in frequency than $f2$ (Abdala, 1998, 2001; Abdala and Chatterjee, 2003). The DPOAE input/output function (DPOAE amplitude as a function of primary tone level) has also shown non-adult-like characteristics in neonates born prematurely. Our laboratory has reported that premature neonates have a larger percentage of monotonic functions, lacking high-level amplitude saturation, particularly at $f2$ frequencies of 1500 and 6000 Hz. When amplitude saturation is present, it is observed at relatively high primary tone levels in premature neonates (≈80 dB SPL versus ≈70 dB SPL in adults) (Abdala, 2000). In seeming contradiction, others have described more compressive, saturating DPOAE I/O functions in neonates (Lasky, 1998b); however, this investigator presented primary tone levels up to 65 dB SPL only and was, therefore, unable to observe or describe *high-level* DPOAE saturation. These two studies of the neonatal DPOAE I/O function were, in fact, measuring two different features of the function and therefore, the seemingly discrepant results are not in conflict. Although observed at both 1500 and 6000 Hz, DPOAE age effects are most robust and reliably present in the high-frequency range, and have been interpreted to indicate an immaturity in cochlear function prior to term birth.

In contrast to the clear DPOAE-based immaturities observed in premature neonates, *term-born* neonates have produced somewhat equivocal results. Eggermont and colleagues (1996) reported that estimates of the cochlear traveling wave delay, measured using DPOAE phase data, are basically adult-like across frequency in premature neonates. Although these investigators observed longer travel times for the youngest ages tested (30–33-weeks PCA), they attributed them to an immature conductive system, rather than cochlear immaturity. Thus, they concluded that the human cochlea was completely adult-like in function by 35-weeks postconceptional age.

Brown and colleagues (1995) also studied maturation of peripheral auditory function in humans by measuring DPOAE amplitude and phase as a function of $f2/f1$ ratio in neonates and adults. The typical $f2/f1$ adult functions are known to be bandpass in shape with an amplitude peak approximately one-half octave below $f2$. $f2/f1$ ratio functions measured from neonates had a significantly more shallow low-pass slope at 4000-Hz. Estimates of mean group delay (based on DPOAE phase) in neonates were also non-adult-like at 4000-Hz. Brown and colleagues interpreted these findings to indicate that the mechnical response of the neonatal cochlea is still immature in this frequency range. This result was dissimilar from the results obtained in our laboratory in a separate study with premature and term-born neonates (Abdala, 1996). We found that the $f2/f1$ ratio functions

a)Electronic mail: cabdala@hei.org

TABLE I. Characteristics of premature neonates.

Neonate subject	Gestational age (wks)	Range of PCA[a] during test period (# test sessions)	Ear/Gender	Birthwt (grams)	APGAR (5 min)
002	29	33–38 (7)	L/M	1285	7
004	29	33–38 (6)	L/M	1250	8
009	30	33–40 (8)	L/F	1370	7
011	30	31–37 (7)	L/F	1890	8
015	26	33–45 (8)	L/F	1000	8
016	32	33–43 (8)	R/M	2040	8
017	32	33–39 (7)	L/F	2370	7
019	28	33–38 (6)	R/M	1185	6
024	26	33–45 (7)	R/M	930	5

[a]Postconceptional age.

were completely adult-like at 1500 and 6000 Hz in premature and term-born neonates. Certainly, the different DPOAE paradigms, different $f2$ frequencies used, and the variable age classifications applied for each experiment contribute to these seemingly discrepant results.

Term-born neonates have adult-like DPOAE suppression tuning when recorded with an $f2$ frequency of 1500-Hz. However, at 6000-Hz, both term and premature neonates have narrower tuning curves than adults (Abdala *et al.*, 1996; Abdala, 1998). Term neonates also show excessively compressive suppression growth for low-frequency suppressor tones at both 1500 and 6000 Hz, although their values are moving in the direction of adult values (Abdala and Chatterjee, 2003). Additionally, the DPOAE input/output (I/O) functions recorded in term neonates show amplitude saturation at higher primary tone levels then adult functions, although the configuration of the function itself is becoming adult-like (Abdala, 2000).

As is evident, term-born neonates show some adult-like features of cochlear function as measured by DPOAE suppression and I/O characteristics; however, there are also systematic immaturities that persist in this group of subjects. These immaturities suggest that maturation of the cochlear response may persist beyond the gestational time period associated with term birth. Clearly, the time course for maturation of human cochlear function and the postconceptional age at which this function becomes completely mature is not presently known.

The purpose of this investigation was to examine neonatal immaturities in cochlear function (as measured by DPOAE suppression and I/O characteristics) by testing prematurely born neonates over a 6–8-week period between 33-weeks postconceptional age (PCA) and the equivalence of term status (38–40-weeks PCA). The longitudinal design was used to enhance detection of changes in cochlear function with age. By applying this experimental design, we sought to further elaborate the form and time course of previously documented immaturities in human cochlear function and elucidate age effects that may have been previously missed using only group mean data in a cross-sectional experimental design.

II. METHODS

A. Subjects

Nine neonates born prematurely, ten term-born neonates, and five normal-hearing adults participated as subjects in this experiment. The nine premature neonates had a mean birth weight of 1480 grams and a mean 5-min APGAR score of 7.1. These subjects were initially tested between 31- and 33-weeks postconceptional age (PCA) and then repeatedly at weekly intervals until they reached the equivalent of ''term-like'' status (38–40-weeks). Three of these neonates were tested again at 44–46-weeks, which is equal to approximately 1 month corrected age. Basic information about each premature neonate is presented in Table I. Term-born neonates were born between 38–41-weeks gestational age and tested within 72 h of birth. All neonatal subjects passed a click-evoked ABR hearing screening at 35 dB HL in order to participate in this experiment and, other than prematurity, they had no other risks for hearing loss (Gerkin, 1984). Adult subjects ranged in age between 18 and 35 years. They had audiometric thresholds <15 dB HL for the frequencies of 500–8000-Hz and a negative history of otologic disease. Adult subjects were tested repeatedly at weekly intervals for five sessions each.

B. Instrumentation and signal processing

An Ariel DSP16+ signal processing and acquisition board housed within a Compaq Prolinea 590 personal computer with Pentium processor was used to generate stimuli and acquire data. The Ariel board was connected to an Etymotic Research ER-10C probe system and to an analog high-pass filter (12-dB/oct; 710-Hz high-pass cutoff). The ER-10C probe contains two output transducers and a low-noise microphone. The two primary tones and the suppressor tone were generated by the DSP processor. The primary tone at $f1$ was generated by one D/A converter and delivered via one transducer. The primary tone at $f2$ and the suppressor tone (fs) were produced by the second D/A converter and output through the second transducer. The signal at the probe microphone was high-pass filtered and sampled at a rate of 50 kHz with a sweep length of 4096 samples, giving a frequency resolution of 12.2-Hz. Twenty-five sweeps of the microphone signal were added and comprised one block.

C. Data acceptance criteria

Acceptance criteria were as follows: (1) noise measurements for three frequency bins (12.2-Hz wide) on either side of the $2f1-f2$ frequency had to be <0 dB SPL to assure appropriate subject state; and (2) the measured DPOAE level had to be at least 5-dB above the average noise measured in the same six bins around the distortion product frequency to be accepted into the grand average.

Each subject required either one or two blocks of 25 sweeps to achieve the absolute noise criterion of 0 dB SPL and the S/N ratio of 5-dB. In addition, sweeps were accepted into a block of data only when the estimated rms level in that sweep did not exceed a user-controlled artifact rejection threshold. This level was set for each subject based on observations of baseline activity level determined early in the test session, and modified if necessary during the experiment.

D. Calibration

Intermodulation distortion produced by the recording system at $2f1-f2$ was measured with the probe in a Zwislocki coupler for all test conditions. The mean level of distortion was -21 dB SPL. In no case did the level exceed -17 dB SPL. The recording system noise floor was determined using a similar method with no tones present. The level of system noise floor ranged between -22 and -27 dB SPL depending on frequency.

An *in situ* calibration procedure was conducted on both output transducers before each subject was tested. Tones of fixed voltage were presented to the transducers at 250-Hz intervals from 500 to 15 000 Hz, and the resulting SPL of these tones recorded in each ear canal. Based on this information, an equalization of output levels was performed for each subject to achieve target stimulus levels across test frequencies.

E. Procedure

Following the acquisition of informed parental consent, neonates were tested at Women and Children's Hospital Los Angeles County+USC Medical Center in a quiet room away from the NICU. They were swaddled and fed if necessary, then placed in a hospital isolette for testing. Adult subjects were tested at the House Ear Institute in the Children's Auditory Research Laboratory. They were tested within a sound-treated IAC booth while sitting comfortably in lounge chairs, reading or napping.

During each session, a DP-gram and both a DPOAE STC and a DPOAE I/O function were collected at $f2=6000$ Hz. The most robust age effects for DPOAE suppression and amplitude growth have been previously observed at 6000-Hz. For this reason, 6000-Hz was chosen as the test frequency to enhance the probability of detecting DPOAE changes as the neonates increased in age.

1. DPOAE suppression

DPOAE STCs were recorded at $f2=6000$ Hz with an $f2/f1$ ratio of 1.2. Primary tone levels were presented at 65–55 dB SPL. Custom-designed software for the collection of DPOAE suppression tuning curves was developed at the Children's Auditory Research and Evaluation Center, House Ear Institute. Once the subject was asleep, they were positioned either on their back or slightly angled onto their side with blankets for support at their back. The positioning was conducted to ensure optimal accessibility to the external auditory meatus for placement of the probe. Once the probe was placed at the meatus snugly and the cable fixed to the isolette or the blankets with surgical tape to minimize movement, the test was initiated. If the probe tip slipped out minimally during testing (as noted by a 2–3-dB drop in stimulus and DPOAE levels), it was nudged back into place until appropriate levels were achieved and the test continued. If the probe tip came out of the ear completely during the recording of a DPOAE STC, it was refit and levels were recalibrated before continuing.

At the beginning of the test, an unsuppressed DPOAE was initially measured and then a suppressor tone (fs) was presented simultaneously with the primary tones and its level increased in 5-dB steps over a range of intensities (40 to 85 dB SPL). Between 12 and 15 suppressor tones with frequency ranging from one octave below to 1/4 octave above $f2$ were presented at intervals between 25 and 150 cents (one octave=1200 cents). To generate isosuppression tuning curves, for each suppressor tone, the suppressor level that reduced DPOAE amplitude by 6-dB was determined using linear interpolation (between data points just prior to and following 6-dB of suppression) and then plotted as a function of its frequency.

2. DPOAE input/output functions

Custom-designed software was implemented for the collection of DPOAE I/O functions. The probe was fit at the entrance of the ear canal as described in the preceding section; however, a probe was never replaced or refit during measurement of the I/O function. Each I/O function was recorded with only one fit. If the probe slipped out during recording of an I/O function, the test was restarted, including recalibration. I/O functions were collected at $f2=6000$ Hz ($f2/f1=1.2$) and were generated by presenting primary tones ranging from 35 to 85 dB SPL in 5-dB intervals. In adults, the transducer was not able to reliably produce 85 dB SPL at 6000-Hz in all subjects; therefore, only DPOAEs recorded from 35–80 dB SPL were considered. Because of the small ear-canal size in neonates, it was possible to present primary tones up to 85 dB SPL in this age group. This additional 5-dB range was helpful when establishing amplitude saturation levels for neonates. Because saturation levels for adults were never higher than 75 dB SPL, the lack of data at 85 dB SPL did not bias their saturation level measurements.

The level separation between tones was fixed at 10-dB (L1>L2). At least two growth functions were collected per ear and the mean of these two I/O functions was entered into data analysis. This initial within-subject averaging of DPOAE amplitude was conducted to reduce the influence of run-to-run variability and thus, reduce the impact of spurious or outlier data points.

F. Data analysis

1. DPOAE suppression

DPOAE suppression data were analyzed in the following manner: (1) tuning curve width was quantified with a Q10 value; (2) slope of the low- and high-frequency flank of the tuning curve was determined by fitting a regression line from the tip of the STC to end points on the low and high-frequency boundaries; (3) tuning curve tip frequency and level were measured; (4) suppression growth slope was determined for each suppressor tone by fitting a regression line to the linear portion of the DPOAE amplitude×suppressor level plot.

2. DPOAE input/output functions

DPOAE I/O function data were analyzed in the following manner: (1) configuration of DPOAE I/O function was quantified with a categorical decision re: absence/presence of amplitude saturation in each individual function. A plateau in DPOAE amplitude at mid- to high-stimulus levels has been attributed to saturation of cochlear amplifier function and appears to provide an important landmark. If DPOAE amplitude did not saturate at the highest levels tested, the growth function was considered to be *monotonic*; if an amplitude plateau was present, the function was considered to be *saturating*; (2) DPOAE amplitude *saturation level* was calculated as the primary tone level associated with an amplitude plateau. The criteria used for establishing saturation level are provided in Appendix A; (3) DPOAE *detection threshold* was calculated as the last primary tone level at which a valid DPOAE measurement could be detected.

III. RESULTS

A. DPOAE suppression

Noise measurements remained consistently low for DPOAE measurements in neonatal subjects. The noise floor, averaged across test session and paradigm, ranged from −9 (for I/O functions) to −14 dB SPL (for suppression), thus allowing for sufficient range to suppress the DPOAE by 6-dB and allowing for the recording of threshold level DPOAEs with sufficient SNR. Figure 1 displays a sequential series of DPOAE STCs ($f2=6000$ Hz) recorded from two randomly selected prematurely born subjects who were tested weekly for a total of 7–8 sessions. The basic shape and morphology of the STCs is consistent across session; the tip is sharp and remains centered at a stable frequency around $f2$. Each tuning curve is asymmetrical, with the high-frequency flank steeper than the low-frequency flank. Although some test-to-test jitter is present on the low-frequency slope at the highest suppressor levels, there does not appear to be any systematic change in morphology or width across test session.

Figure 2 shows individual Q10 values from each subject as a function of test session (postconceptional age for neonates). Each string of symbols connected by one continuous line represents the sequential Q10 values measured from any one subject. Q10 values from term-born neonates tested during one session only are also shown. The mean Q10 values for each age group are provided in the small inset graph.

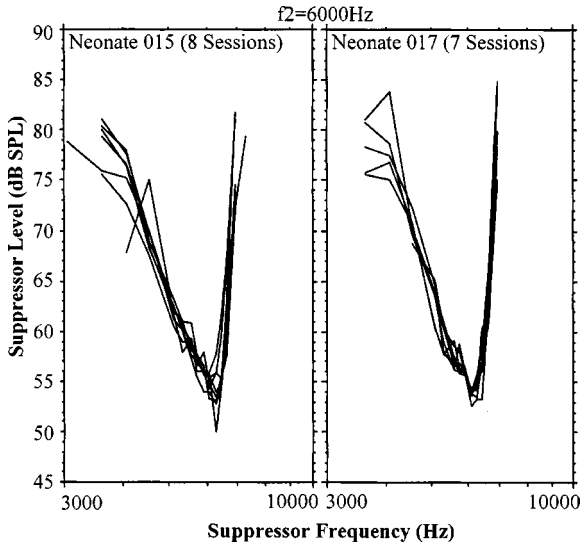


FIG. 1. DPOAE suppression tuning curves (STC) recorded during repeated test sessions with two premature neonates. For both neonates, the first test session was conducted at 33-weeks postconceptional age (PCA) and then weekly thereafter for 7–8 sessions. STCs were recorded with an $f2$ frequency of 6000-Hz and primary tone levels at 65–55 dB SPL.

The Q10 values from neonates are generally larger than values from adults, indicating a narrower DPOAE STC. As is evident from the inset figure, the mean Q value from the premature group was 3.7, whereas the mean value from the adult group was 2.3. Additionally, eight out of the ten Q values recorded from the term-born subjects overlapped with values recorded from premature neonates and not with adult data. Even though Q10 values from adults and neonates fall into two fairly distinct clusters, one adult produces data that overlap with the lower range of neonatal values. Of the 25 Q10 measurements collected from adult subjects, only five of these observations overlapped with neonatal data.

There appears to be more intersubject variability in Q among neonates than adults, suggesting greater flux in the data. Because only 4–5 test sessions were conducted with adults and 7–8 sessions with neonates, it was not possible to conduct statistical analyses using all neonatal sessions. Therefore, an age group×test session (PCA) ANOVA with repeated measures on test session was conducted including all adult data and only five neonatal Q data sessions ranging from 33- to 37-weeks PCA. All nine neonatal subjects were included in each of these five test sessions used in the repeated measures analysis. This analysis was conducted to assess whether there was any change in Q10 across test session (i.e., with increasing PCA) or between adult and premature age groups. Results indicated an effect of age group ($f=24.51$; $p<0.0008$) but no effect of test session. That is, premature neonates as a group had narrower STCs than adults and there was no significant change in the width of the tuning curve with increasing postconceptional age. Neonatal tuning curves did not become more adult-like in width between 33- and 37-weeks PCA.

Figure 3 shows DPOAE STC slope on the low-frequency flank plotted as a function of test session (PCA).

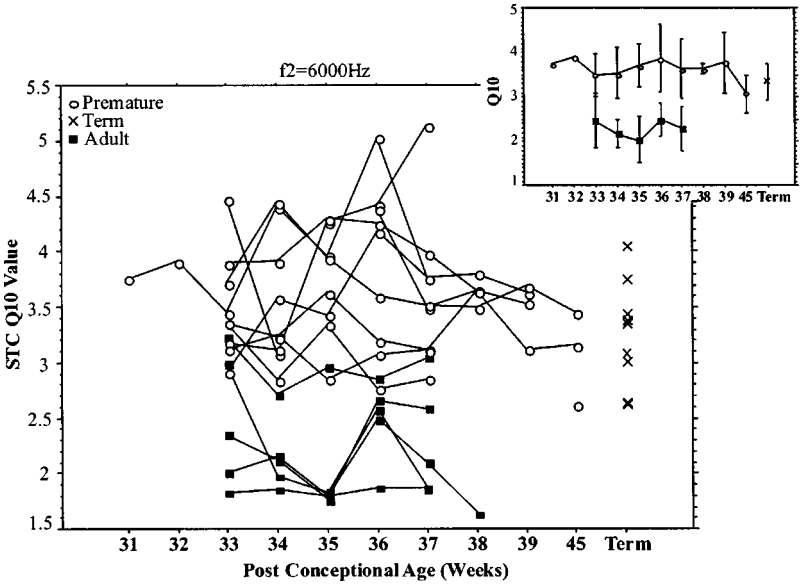


FIG. 2. DPOAE STC Q10 values for nine premature neonates tested repeatedly at weekly intervals for 7–8-weeks (open circles), five normal-hearing adults tested weekly for five sessions (filled squares), and ten term-born neonates tested during one session within 72 h of birth (X). Symbols connected by a continuous line represent the repeated measurements from one subject. The inset panel displays the mean Q10 data from each of the three age groups. Error bars are ±1 s.d.

The high-frequency slope data are not shown because values from adults and neonates show complete overlap. Multiple studies have shown no age effect on the high-frequency flank of DPOAE STCs (Abdala *et al.*, 1996; Abdala, 1998, 2001). The low-frequency slope results shown in Fig. 3 are similar to the Q10 data. There was an effect of age group ($f=21$; $p=0.0001$) but there is no effect of test session. These results indicate that premature neonates had a steeper slope than adults on the low-frequency side of the tuning curve, and this slope did not become more adult-like across test session as the neonates increased in postconceptional age.

ANOVAs with repeated measures on test session were also conducted on both DPOAE STC tip level and tip frequency. As in the previous statistical analyses, only neonatal data from 33- to 37-weeks PCA were included. Results indicated that: (1) the suppressor level associated with the tip of the DPOAE suppression tuning curve did not show age group or test session effects; (2) the tip frequency showed an effect of age group ($f=15.988$; $p=0.003$) with adults showing a higher frequency tip than neonates but there were no test session effects. These analyses indicate that STC tip features did not change systematically during the course of repeated testing, as premature neonates increased in postconceptional age from 33- to 37-weeks.

The pattern of DPOAE suppression growth was analyzed by determining the rate or slope of suppression growth produced by each suppressor tone. For this analysis and discussion, suppressor tones lower than the $f2$ frequency were termed ''low-frequency side'' (LFS) suppressors and those higher than $f2$, ''high-frequency side'' (HFS) suppressor tones. Figure 4 displays the mean slope of DPOAE suppression growth as a function of suppressor frequency for premature neonates at four different ages, term neonates, and adults. The standard deviations for these mean data are presented in Appendix B. The vertical dashed line is placed at the $f2$ frequency for reference. The horizontal solid line indicates linear growth in suppression (1-dB/dB slope). Data points below the horizontal line indicate compressive suppression growth, and those above it show expansive growth of suppression.

The normal and expected pattern of suppression growth in the mature and fully functional cochlea is illustrated by the filled squares. LFS suppressors (to the left of the vertical line) produce steep growth of suppression as expected, due

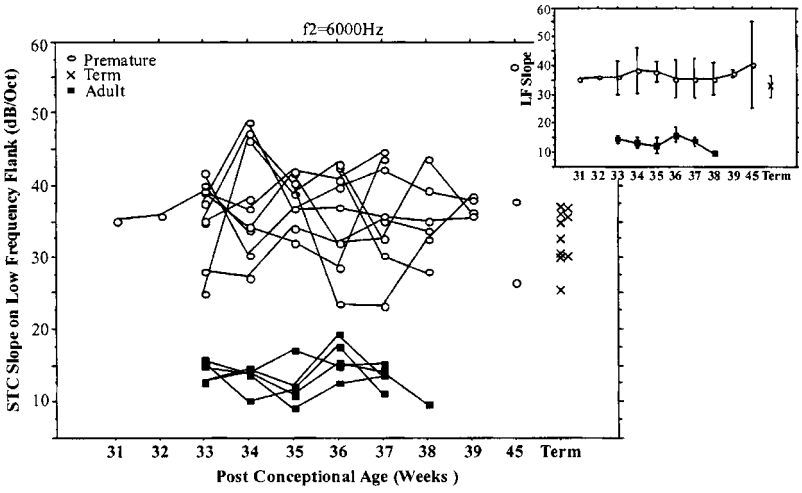


FIG. 3. Slope of DPOAE STC low-frequency flank for nine premature neonates tested repeatedly at weekly intervals for 7–8 sessions (open circles), five normal-hearing adults tested weekly for five sessions (filled squares), and ten term-born neonates tested during one session within 72 h of birth (×). Symbols connected by a continuous line represent the repeated measurements from one subject. The inset panel displays the mean low-frequency slope data from each of the three age groups. Error bars are ±1 s.d.

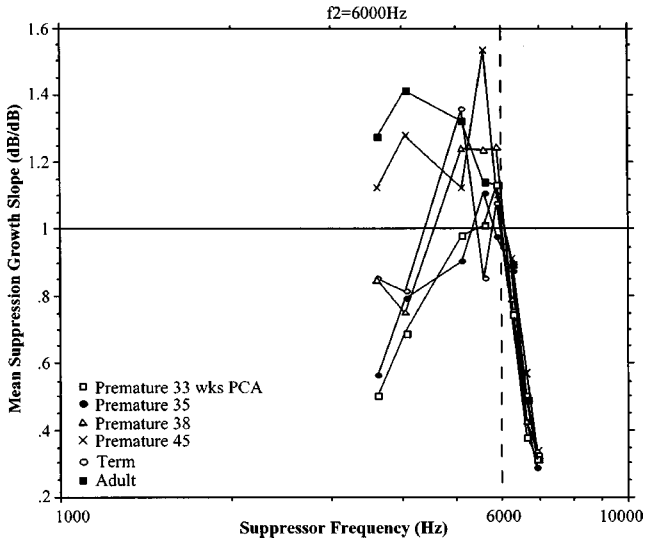


FIG. 4. Mean slope of suppression growth as a function of eight suppressor tones ranging in frequency from 3620 to 6932 Hz. Mean data are shown for adults, premature neonates tested at four different ages: 33-, 35-, 38-, and 45-weeks postconceptional age (PCA) and term-born neonates. The vertical dashed line is at $f2$ and the horizontal solid line denotes linear growth of suppression (1-dB/dB). The s.d. for these mean data is provided in Appendix B.

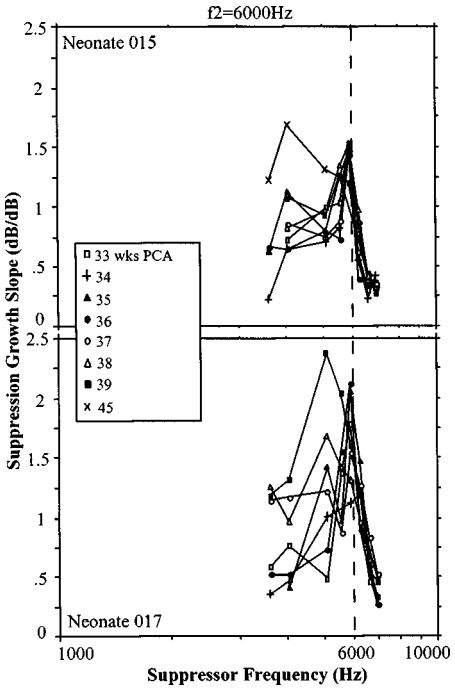


FIG. 5. Slope of suppression growth as a function of suppressor frequency for two randomly selected premature neonates tested at weekly intervals. The PCA at test is designated by symbol.

to an upward spread of masking which causes an overlap with the DPOAE generation site at $f2$ as suppressor level is increased. HFS suppressors do not overlap greatly with the DPOAE generation site even when the suppressor is presented at relatively high levels; thus, they are not effective at reducing DPOAE amplitude. They produce shallow, compressive slopes of DPOAE suppression growth. The nonlinear behavior of the basilar membrane is clearly represented in the asymmetry of suppression growth slope for tones lower than $f2$ versus those higher in frequency than $f2$.

As shown in previous studies, neonates here have a non-adult-like pattern of suppression growth for LFS suppressor tones. When the premature neonates were tested at both 33- and 35-weeks PCA, they showed compressive ($<$1-dB/dB) suppression growth slope for all suppressor tones lower in frequency than $f2$. By 38-weeks PCA, this same group of neonates had adult-like suppression growth for LFS suppressor tones closest to $f2$, but the very lowest suppressor tones continued to produce shallow suppression growth. Term-born neonates show a similar pattern (open circles). By 45-weeks PCA (1 month corrected age), the LFS suppression growth slope from prematurely born neonates was generally adult-like. This mean value, however, was based only on three observations and therefore should be interpreted with caution. In contrast to the age-related changes in LFS suppression growth, HFS suppression growth was adult-like and compressive at all ages tested. Although the variability in suppression growth slope (see the standard deviations presented in the Appendix) suggests overlap among ages, even for the lowest frequency suppressor tones, a developmental change in suppression growth patterns across frequency is clearly evident with increasing subject age. The normal contrast between compressive (HFS) and steep (LFS) suppression in adults is clearly absent in the neonates, with the exception of the 1-month-olds (i.e., 45-weeks PCA). The neonatal pattern, in contrast, shows a peak of suppression growth slope around $f2$ and decreased slope on either side of this. Although the large standard deviations in suppression growth slope attest to the variable nature of this maturational process among neonates, the developmental trend is clear.

Figure 5 presents suppression growth data from two randomly selected neonatal subjects during repeated test sessions. Within each panel, the postconceptional ages at test are designated by symbol. Test sessions during the earliest ages (33–36-weeks PCA) produced non-adult-like, excessively compressive suppression growth for LFS tones and a maximum slope value around the $f2$ frequency, rather than frequencies lower than $f2$. This ''$f2$-peaked'' pattern appears to be unique to neonates. The data from both neonates, displayed in Fig. 5, show a trend toward more adult-like suppression growth as subjects increased in age from 33-weeks to the equivalence of ''term-like'' maturational status. Neonate 015 was tested at 1 month corrected age and appears to show a completely adult-like pattern of suppression growth by this time.

In order to contrast the age effects on suppression growth for low- versus high-frequency side suppressor tones, mean slope values were calculated by averaging suppression growth slope for the two lowest (3620 and 4063 Hz) and two highest frequency suppressor tones (6932 and 7239 Hz). Figure 6 shows these mean low- and high-frequency slope values as a function of test session for adults and premature neonates. An ANOVA of age group$\times$test with repeated measures on session confirmed an age effect for the low-frequency slope estimate ($f=22.57$; $p=0.0005$) and no session effect. There appears to be an upswing in slope of suppression growth for LFS suppressors by 39- to 45-weeks. This may indicate a move toward adult-like patterns of sup-

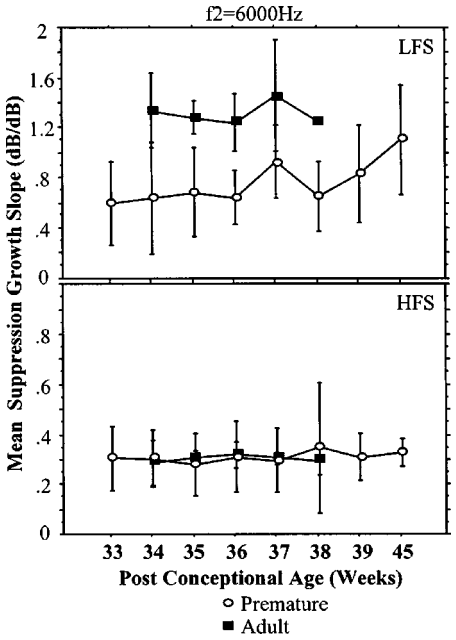


FIG. 6. Mean slope of suppression growth for an index of low-frequency side (LFS) suppression and high-frequency side (HFS) suppression. Data are shown as a function of PCA for premature neonates and test session for adults.

pression; however, the mean value at 45-weeks is based on only three observations. There was neither session nor age effect for high-frequency suppression growth.

These data confirm previously observed immaturities of DPOAE suppression tuning and growth in premature neonates. However, the data extend previous results to show that nine individual neonates tested repeatedly over 7–8 sessions remain non-adult-like in DPOAE suppression tuning and suppression growth for $f2 = 6000$ Hz, even once they reach the equivalence of term status. Additionally, term-born neonates tested during one session are more similar in their suppression tuning and suppression growth to premature neonates than to adults.

B. DPOAE input/output functions

Figure 7 shows mean DPOAE amplitude as a function of primary tone level (L1) for five adult subjects, nine prematurely born neonates tested at 33- and 35-weeks PCA, and ten term-born neonates. In adults, DPOAE amplitude grew monotonically as the stimulus increased to moderate levels, then saturated and formed a plateau at moderate levels (65–70 dB SPL). This saturating pattern was evident in all adults tested. The mean I/O configuration from premature subjects tested at 33-weeks PCA does not show amplitude saturation, even at the highest primary tone levels presented. The mean I/O function from these same neonates recorded 2-weeks later shows an amplitude saturation plateau, but at relatively high primary tone levels (75 dB SPL). Likewise, the mean function from term neonates was relatively monotonic, with an amplitude plateau developing at the highest primary tone level.

The percentage of functions showing DPOAE amplitude saturation increased with postconceptional age. At 33-weeks, one-third of the I/O functions collected were monotonic and

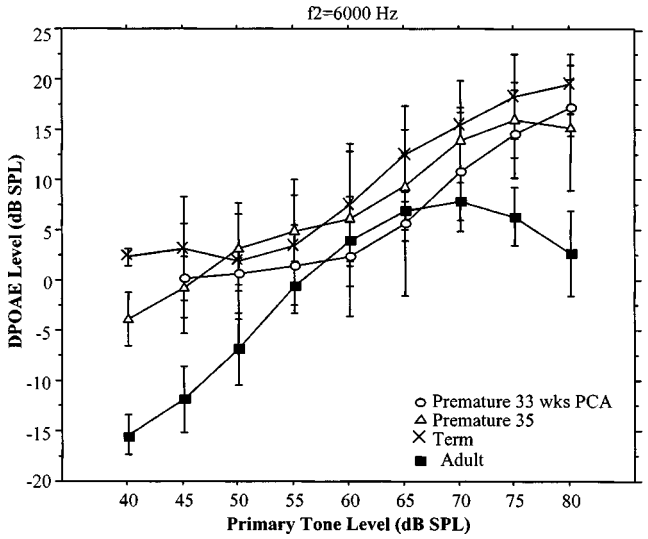


FIG. 7. An input/output function displaying mean DPOAE amplitude as a function of primary tone level for adults and a group of prematurely born neonates tested at two PCA ages: 33- and 35-weeks and term-born neonates. Error bars are ±1 s.d.

nonsaturating. By 38-weeks PCA, only one of eight functions had a nonsaturating configuration. Beyond this age, there are reduced observations; therefore, it is not clear whether prevalence of amplitude saturation in the DPOAE I/O function becomes completely adult-like by the equivalence of term gestation (40-weeks).

The saturation level (stimulus level associated with DPOAE amplitude saturation) remained elevated at least through 39-weeks PCA. Figure 8 shows DPOAE amplitude saturation level as a function of test session (PCA). For the purposes of this analysis, any I/O function that was monotonic and did not show amplitude saturation at the highest level tested (85 dB SPL for neonates; 80 dB SPL for adults) was assigned a value of 90 dB SPL. This occurred in 6 of 63 I/O functions for neonates and 0 of 25 functions in adults. An ANOVA of age×test session with repeated measures on test session was conducted on saturation level data. The neonatal

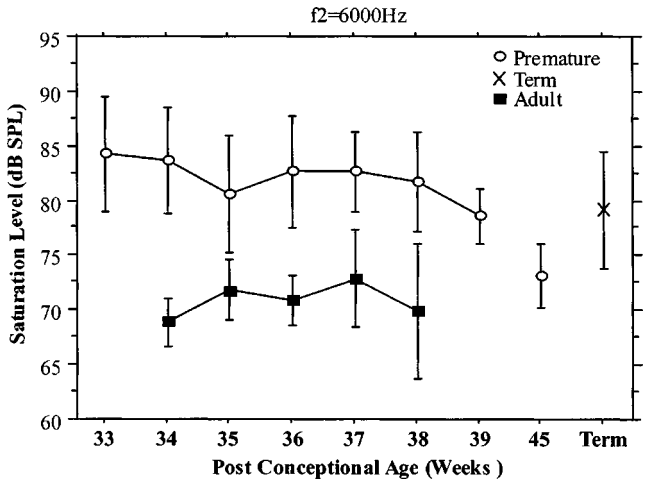


FIG. 8. Mean DPOAE amplitude saturation level as a function of PCA in premature neonates as they increase in age from 33- through 45-weeks PCA. Data are shown for a group of adults, prematurely born neonates, and a one-time test session with term-born neonates. Error bars are ±1 s.d.

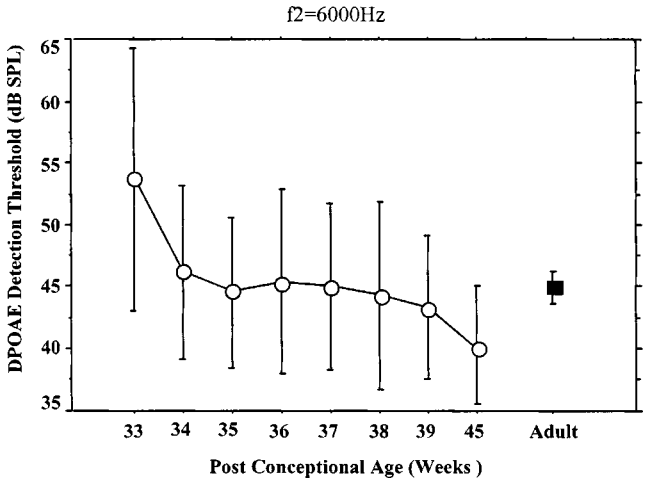


FIG. 9. Mean DPOAE detection threshold (derived from I/O functions) plotted as a function of postconceptional age in premature neonates during weekly test sessions. The results of one test session with a group of adult subjects are also shown for comparison.

data collected between the ages of 33- to 37-weeks PCA were included. This analysis confirmed an age effect ($f = 24.21$; $p = 0.0006$) but no effect of test session, indicating that neonates have significantly elevated saturation thresholds that do not change significantly over this 5-week period. As the neonates increase in age from 39- to 45-weeks, it appears from data in Fig. 8 that the saturation level is moving toward adult-like values. However, both groups at 39- and 45-weeks PCA include only 3–4 observations; therefore, these data should be interpreted with caution.

The DPOAE detection threshold was also determined from the DPOAE I/O function. This threshold is defined as the lowest primary tone level at which a valid DPOAE can be recorded. Figure 9 displays mean DPOAE detection threshold as a function of test session (PCA) for neonates and a one-time measurement for ten adult subjects. Neonates at the youngest ages appear to have an elevated threshold. This threshold decreases slightly with increased PCA but, by 34–35-weeks the mean value is already adult-like at approximately 45 dB SPL. Even though this mean value is adult-like early, the variability in threshold is greater in neonates than adults. This indicates that there are still some neonates at each PCA showing elevated DPOAE detection thresholds.

Figure 10(a) presents DPOAE detection thresholds plotted as a function of PCA for four individual neonates. Although there are fluctuations at any given age, the general trend is for threshold to decrease with increasing age. Six out of the nine neonatal subjects showed a pattern of consistently decreasing DPOAE threshold. As is evident from Fig. 10(a), during the first one or two test sessions (33- and 34-weeks PCA), DPOAEs could only be recorded using moderate to high-level primary tones (>60 dB SPL) in three out of the four subjects shown. As each neonate increased in age, it became possible to record DPOAEs to primary tones presented between 35 and 45 dB SPL. In order to rule out that the age-related decrease in DPOAE detection threshold was produced by an age-related reduction in noise, mean noise-floor data are included in Fig. 10(b). Figure 10(b) shows the

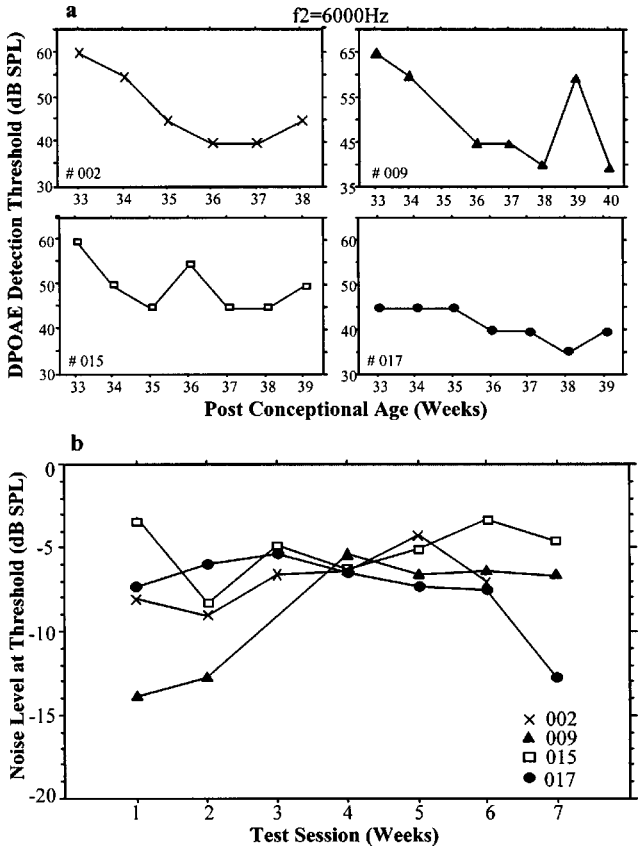


FIG. 10. (a) DPOAE detection threshold for four prematurely born neonates as they increase in age from 33-weeks through 38–40-weeks. (b) The mean noise-floor value measured at DPOAE threshold for each of the four subjects shown in panel (a).

mean noise-floor level at DPOAE detection threshold as a function of test session for the four neonatal subjects shown in Fig. 10(a). The noise floor did not systematically decrease across session. Therefore, this factor cannot easily explain the decreasing DPOAE detection threshold with increasing PCA.

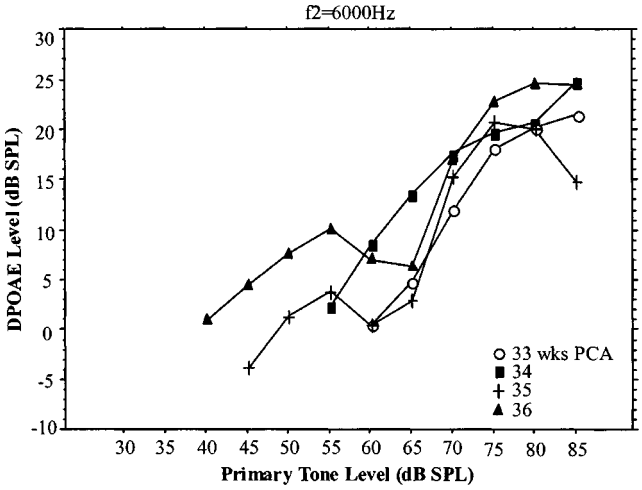


FIG. 11. DPOAE input/output functions for one prematurely born neonate recorded at four postconceptional ages. Age is designated by symbol. Thresholds decreased systematically from 60 to 40 dB SPL during the 3-week testing period. The noise floor did not show an age-related decrease during this same time period.

The changes in DPOAE threshold and I/O configuration are illustrated in Fig. 11 for one neonatal subject as he/she increased in age from 33- to 36-weeks. The initial function recorded at 33-weeks PCA (open circles) is monotonic and there is no amplitude saturation noted. Additionally, DPOAEs are only present for high-level stimuli (>60 dB SPL), leading to an elevated DPOAE detection threshold. By 35–36-weeks, threshold is adult-like and the configuration of the I/O function became nonmonotonic and showed two distinct segments and a clear amplitude plateau.

IV. DISCUSSION

The primary findings in this study are: (1) prematurely born neonates that have achieved the equivalence of term age remain non-adult-like in DPOAE suppression tuning, suppression growth, and DPOAE I/O characteristics; (2) DPOAE suppression and I/O function data from term-born neonates overlap with data from premature neonates and not with adult data; (3) there was no significant change in DPOAE suppression during the 7–8-week test period from 33- to 40-weeks postconceptional age, although there was some change in characteristics of the I/O functions during this same time period. These data suggest that cochlear function in humans is not fully adult-like by term birth and continues to mature into the first months of postnatal life.

A. Term-born neonates

Existing evidence strongly suggests that the cochlea is structurally mature by term birth (Bredberg, 1968; Igarishi, 1980; Lavigne-Rebillard and Pujol, 1986, 1987, 1988; Lavigne-Rebillar and Bagger-Sjoback, 1992). The time course for maturation of human cochlear *function* is less clear, although it is known that onset of human auditory function is precocial. Studies of both the auditory brainstem response and fetal reflexes measured *in utero* have shown that hearing onset occurs between 26 and 28 fetal weeks (Birnholtz and Benecerraf, 1983; Despland and Galambos, 1980; Kuhlman *et al.*, 1988; Starr *et al.*, 1977). It appears that the early morphological maturity of the human cochlea, and early onset of auditory function may have led to the assumption that the human cochlea is completely adult-like in function by term birth. However, this question has not been adequately tested or resolved.

The available studies on human maturation of cochlear function have offered conflicting results. Most of these studies have used a cross-sectional experimental design and analyzed group mean data to address the question. Some results have indicated that certain DPOAE-based assays of cochlear function are mature by term birth or shortly thereafter (Abdala *et al.*, 1996; Bargones and Burns, 1988; Eggermont *et al.*, 1996), and that the optimal parameters for DPOAE generation are the same in neonates and adults (Abdala *et al.*, 1996; Brown *et al.*, 2000, 1994). In contrast, other laboratories have reported non-adultlike DPOAE features in newborns (Abdala, 1998, 2001; Brown *et al.*, 1995; Lasky, 1998a,b). Clearly, the body of literature previously generated in this area has not aptly elucidated the time course for the maturation of human cochlear function.

The longitudinal data presented in the present study unequivocally indicate that DPOAE suppression tuning, suppression growth, and I/O characteristics are not mature in term-born neonates or prematurely born neonates that have reached 38–40-weeks postconceptional age for $f2 = 6000$ Hz. This finding suggests that any further maturation of cochlear function must occur in the postnatal period.

B. Basis of immaturity

The only DPOAE feature that became adult-like early in prematurely born subjects was the DPOAE detection threshold. At the earliest ages tested (33-weeks PCA), DPOAEs could only be recorded using high-level stimuli (>60 dB SPL). However, by the time prematurely born neonates reached 35-weeks, DPOAEs were observed to primary tones as low as 45 dB SPL. Among the many variables measured and quantified in this study, DPOAE detection threshold was the only metric of cochlear function that became completely adult-like prior to the time period associated with term gestation (40-weeks). This measure may simply not provide an accurate indication of prolonged maturation of cochlear responses.

All other DPOAE-based indices of cochlear function measured in this study suggest that there is an immaturity in cochlear function beyond 40-weeks PCA. Based on the very early maturation of human cochlear anatomy and morphology during the fetal period, this immaturity in function is likely to be very subtle. The magnitude of adult–neonatal differences observed in DPOAE suppression and I/O characteristics supports this contention; DPOAE STCs from premature neonates are not grossly abnormal in shape or markedly broadened. Neonatal STCs retain a bandpass characteristic and most of the typical tuning curve features but are slightly more narrow and sharp.

The basis for this subtle immaturity in postnatal cochlear function is not presently known. The excessively shallow DPOAE suppression growth in neonates for LFS suppressor tones has been interpreted as an immaturity in cochlear nonlinearity (Abdala and Chatterjee, 2003). It has been hypothesized that the cochlea of term neonates, as well as prematurely born neonates who have reached adult-like status, is more compressive than the adult cochlea and thus, shows excessive nonlinearity. On the surface, this hypothesis may appear to contradict the more linear appearing DPOAE I/O function observed in neonates. However, the two results are not inconsistent with one another. The lack of high-level saturation in the neonatal DPOAE I/O function indicates that the same stimulus levels presented to neonates and adults do not produce the same effect on DPOAE generation. Adults show DPOAE amplitude that peaks, saturates, and sometimes decreases beyond 70 dB SPL primary tone level, yet these same levels produce continued DPOAE amplitude growth in neonates and saturation, if present, is not noted until stimulus levels reach 80 or 85 dB SPL. Our interpretation of this result is not that the neonatal cochlea is more linear than the adult cochlea, but that the cochlear amplifier may simply have different ''boundaries'' of function in neonates and adults. If DPOAE saturation reflects the upper limit for cochlear amplifier function, it may be that we are

not able to reach this upper limit in neonates due to restrictions in transducer output and concerns about subject comfort. Thus, we may be measuring only the linear portion of the neonatal function and unable to access the saturated portion, although present at higher levels.

Together, these three non-adult-like features of the neonatal DPOAE, highly compressive growth of suppression for LFS suppressor tones, excessively narrow DPOAE suppression tuning, and a wider range of function for active processes, appear consistent with a cochlear amplifier that is not completely mature. Immaturity in the outer hair cell (OHC) may provide one possible mechanism. DPOAEs are integrally linked to OHCs and probably are a result of their unique contractile properties (Brownell, 1990; Mountain, 1980; Siegel and Kim, 1982). Electromotility of OHCs has been unequivocally linked to enhancement of cochlear sensitivity. Absence of electromotility results in the loss of cochlear sensitivity measured *in vivo* for mice with a targeted prestin deletion (Liberman *et al.*, 2002). Thus, it is plausible that an immaturity in OHC function and/or motility in neonates could produce non-adult-like DPOAE suppression as observed in this and other experiments. Although this possibility has not been empirically ruled out, OHCs are morphologically mature very early in fetal life (22nd fetal week) and protracted immaturity in OHC function into the postnatal period does not seem likely (Igarishi, 1980; Lavigne-Rebillard and Pujol, 1986, 1987, 1988).

Another possible source of these age-related effects on DPOAE suppression and I/O characteristics could be the relatively late-developing medial olivocochlear (efferent) system. Although the medial efferent system does not appear to be required for normal auditory function (Scharf *et al.*, 1997, 1994), stimulation of this pathway can alter cochlear output and effectively modulate OAE amplitude (Mountain, 1980; Siegel and Kim, 1982; Warren and Liberman, 1989; Wiederhold and Kiang, 1970). One hypothesized function of the medial efferent system is to provide constant adjustment of cochlear amplifier gain and the maintenance of the cochlea at an optimum mechanical state for efficient function of active processes (Guinan, 1996; Johnstone *et al.*, 1986). The constellation of age effects observed in this study for term-born neonates and prematurely born neonates that have reached the equivalence of term may be based in immaturity of medial olivocochlear system function and the subsequent inefficient regulation of the cochlear amplifier. Consistent with this hypothesis, several previous studies have observed neonatal immaturities in medial efferent function as measured by OAE contralateral suppression (Abdala *et al.*, 1999; Morlet *et al.*, 1993; Ryan and Piron, 1994). It must be noted, however, that the medial efferent system is not maximally effective at high frequencies and DPOAE contralateral suppression is only observed at low- to midfrequency ranges. Thus, although this hypothesis merits further investigation, it is possible that DPOAE immaturities observed in the present study at 6000-Hz are not strongly linked to immaturity in the medial olivocochlear system.

Attenuation of primary tone levels through an immature middle ear could contribute to some of the age effects observed. Similarly, attenuation of the DPOAE during reverse transmission through the middle ear may influence DPOAE results. However, attempts to simulate this attenuation effect suggest that it is not possible to account for excessively narrow suppression tuning or atypically shallow suppression growth based only on this factor (see Abdala and Chatterjee, 2003, for further discussion).

Finally, it should be recognized that the presence of spontaneous otoacoustic emissions (SOAEs) near $2f1-f2$ can potentially impact the DPOAE response measured in the ear canal (Long *et al.*, 1993). This interaction has been well documented and its impact on DPOAE suppression tuning curves has been investigated. SOAEs, although likely to enhance the amplitude of a distortion product close in frequency, do not alter DPOAE suppression in any systematic fashion (Abdala, 1998). The presence of SOAEs could also have influenced the measurement of the DPOAE amplitude on the I/O function. However, these interactions are likely to produce narrow fluctuations in amplitude (i.e., ''notches'') rather than the extended saturation plateau that was measured in the present study. Thus, although they may have impacted DPOAE amplitude to some extent in some of the subjects participating in this experiment, it is not likely that they can explain the age effects observed in either DPOAE suppression or I/O characteristics, nor the lack of session effect as neonates increased in age.

C. Summary

The results of this experiment strongly suggest that a subtle immaturity of human cochlear function in the frequency range of 6000 Hz persists into the first months of postnatal life. This finding prompts many intriguing questions. For example, the perceptual consequences of this physiological immaturity, if any, are not known. Does this type of cochlear immaturity produce perceptional consequences that can be measured in human infants? It is also not known *when* cochlear function becomes completely adult-like in human infants. Measures of frequency tuning based on the auditory brainstem response and psychoacoustic paradigms are adult-like by 6 months (Abdala and Folsom, 1995; Folsom and Wynne, 1987; Olsho, 1985; Spetner and Olsho, 1990), suggesting that the more peripheral sensory organ is also functioning in an adult-like fashion by this age. Recent results from our laboratory suggest that the cochlea becomes functionally mature in humans sometime between 3 and 6 months of age (Abdala and Dumont, 2003); however, systematic research is needed to establish the full time course of this maturational process. Finally, the finding of postnatal immaturity in DPOAE-based indices of cochlear function suggests that future research be conducted to investigate the anatomical/physiological source(s) of the immaturities within the cochlea.

ACKNOWLEDGMENTS

The author would like to acknowledge Dr. Ellen Ma for data collection and Ms. Leslie Visser-Dumont for data collection and management and for assistance in the preparation

of this manuscript. This experiment was funded by a grant from the National Institutes of Health, NIDCD R29 DC03552 and by the House Ear Institute.

APPENDIX A: CRITERIA FOR ESTABLISHING SATURATION LEVEL

(1) ≤1-dB amplitude change among at least three consecutive points, unless they are the two highest levels (75 and 80 dB SPL), then only two consecutive points with ≤1-dB are required. The second point (going from left to right) along a two- or three-sequence plateau is the saturation level.

(2) At least two consecutive points with decreased amplitude. The second point in this sequence is the saturation level.

(3) ≤2-dB amplitude change among at least four consecutive points. The second point in this sequence is the saturation level.

APPENDIX B: STANDARD DEVIATIONS FOR MEAN VALUES IN FIG. 4

	P/33	P/35	P/38	P/45	Term	Adult
3620	0.284	0.248	0.386	0.127	0.530	0.492
4063	0.430	0.455	0.312	0.576	0.342	0.324
5119	0.300	0.346	0.301	0.159	0.482	0.486
5582	0.223	0.413	0.367	0.256	0.286	0.240
5914	0.466	0.519	0.366	0.343	0.275	0.283
6266	0.099	0.303	0.191	0.123	0.233	0.344
6638	0.120	0.129	0.101	0.197	0.174	0.063
6932	0.129	0.125	0.227	0.057	0.077	0.052

Abdala, C., and Folsom, R. (**1995**). ''Frequency contribution to the click-evoked auditory brainstem response in human adults and infants,'' J. Acoust. Soc. Am. **97**, 2394–2404.

Abdala, C. (**1996**). ''Distortion product otoacoustic emission ($2f1-f2$) amplitude as a function of $f2/f1$ frequency ratio and primary tone level separation in human adults and neonates,'' J. Acoust. Soc. Am. **100**, 3726–3740.

Abdala, C., Sininger, Y., Ekelid, M., and Zeng, F.-G. (**1996**). ''Distortion product otoacoustic emission suppression tuning curves in human adults and neonates,'' Hear. Res. **98**, 38–53.

Abdala, C. (**1998**). ''A developmental study of DPOAE ($2f1-f2$) suppression in premature human neonates,'' Hear. Res. **121**, 125–138.

Abdala, C., Ma, E., and Sininger, Y. (**1999**). ''Maturation of medial efferent system function in humans,'' J. Acoust. Soc. Am. **105**, 2392–2402.

Abdala, C. (**2000**). ''Distortion product otoacoustic emission ($2f1-f2$) amplitude growth in human adults and neonates,'' J. Acoust. Soc. Am. **107**, 446–456.

Abdala, C. (**2001**). ''Maturation of the human cochlear amplifier: Distortion product otoacoustic emission suppression tuning curves recorded at low and high primary tone levels,'' J. Acoust. Soc. Am. **110**, 1465–1476.

Abdala, C., and Chatterjee, M. (**2003**). ''Maturation of cochlear nonlinearity as measured by distortion product otoacoustic emission suppression growth in humans,'' J. Acoust. Soc. Am. **114**, 932–943.

Abdala, C., and Visser-Dumont, L. (**2003**). ''Cochlear function in older infants,'' Abstr. Am. Aud. Soc. Bulletin, Scottsdale, AZ, Vol. **28**, pp. 23–24.

Bargones, J., and Burns, E. (**1988**). ''Suppression tuning curves for spontaneous otoacoustic emissions in infants and adults,'' J. Acoust. Soc. Am. **83**, 1809–1816.

Birnholtz, J., and Benacerraf, B. (**1983**). ''The development of human fetal hearing,'' Science **222**, 516–518.

Bredberg, G. (**1968**). ''Cellular pattern and nerve supply of the human organ of Corti,'' Acta Oto-Laryngol., Suppl. **236**, 1–135.

Brown, A., Sheppard, S., and Russell, P. (**1994**). ''Acoustic distortion products (ADP) from the ears of term infants and young adults using low stimulus levels,'' Br. J. Audiol. **28**, 273–280.

Brown, A., Sheppard, S., and Russell, P. (**1995**). ''Differences between neonate and adult cochlear mechanical responses,'' Aud. Neurosci. **1**, 169–181.

Brown, D., Bowman, D., and Kimberley, B. (**2000**). ''The effects of maturation and stimulus parameters on the optimal $f2/f1$ ratio of the $2f1-f2$ distortion product otoacoustic emission in neonates,'' Hear. Res. **145**, 17–24.

Brownell, W. (**1990**). ''Outer hair cell electromotility and otoacoustic emissions,'' Ear Hear. **11**, 82–92.

Despland, P., and Galambos, R. (**1980**). ''Use of the auditory brainstem responses by premature and newborn infants,'' Neuropadiatrie **11**, 99–107.

Eggermont, J., Brown, D., Ponton, C., and Kimberley, B. (**1996**). ''Comparison of DPE and ABR traveling wave delay measurements suggests frequency specific synapse maturation,'' Ear Hear. **17**, 386–394.

Folsom, R., and Wynne, M. (**1987**). ''Auditory brainstem responses from human adults and infants: Wave-V tuning curves,'' J. Acoust. Soc. Am. **81**, 412–417.

Gerkin, K. (**1984**). ''The high-risk register for deafness,'' ASHA **26**, 17–23.

Guinan, J. (**1996**). ''Physiology of olivocochlear efferents,'' in *The Cochlea*, edited by P. Dallos, A. Popper, and R. Fay (Springer, New York), pp. 435–502.

Igarishi, Y. (**1980**). ''Cochlea of the human fetus: A scanning electron microscope study,'' Arch. Histol. Jpn. **43**, 195–209.

Johnstone, B., Patuzzi, R., and Yates, G. (**1986**). ''Basilar membrane measurements and the traveling wave,'' Hear. Res. **22**, 147–153.

Kuhlman, K., Burns, K., Depp, R., and Sabbagha, R. (**1988**). ''Ultrasonic imaging of normal fetal response to external vibratory acoustic stimulation,'' Am. J. Obstet. Gynecol. **158**, 47–51.

Lasky, R. (**1998a**). ''Distortion product otoacoustic emissions in human newborns and adults. I. Frequency effects,'' J. Acoust. Soc. Am. **103**, 981–991.

Lasky, R. (**1998b**). ''Distortion product otoacoustic emissions in human newborns and adults. II. Level effects,'' J. Acoust. Soc. Am. **103**, 992–1000.

Lavigne-Rebillard, M., and Bagger-Sjoback, D. (**1992**). ''Development of the human stria vascularis,'' Hear. Res. **64**, 39–51.

Lavigne-Rebillard, M., and Pujol, R. (**1986**). ''Development of the auditory hair cell surface in human fetuses,'' Anat. Embryol. **174**, 369–377.

Lavigne-Rebillard, M., and Pujol, R. (**1987**). ''Surface aspects of the developing human organ of Corti,'' Acta Oto-Laryngol., Suppl. **436**, 43–50.

Lavigne-Rebillard, M., and Pujol, R. (**1988**). ''Hair cell innervation in the fetal human cochlea,'' Acta Oto-Laryngol. **105**, 398–402.

Liberman, M., Gao, J., He, D., Wu, X., Jia, S., and Zuo, J. (**2002**). ''Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier,'' Nature (London) **19**, 300–304.

Long, G., Sun, C., and Talmadge, C. (**1993**). ''Interactions between spontaneous emissions and external tones: Suppression, frequency shifting, and distortion product generation,'' in *Biophysics of Hair Cell Sensory Systems*, edited by J. Duifuis, J. Horst, P. van Dijk, and S. van Netten (World Scientific, Singapore), pp. 40–47.

Morlet, T., Collet, L., Salle, B., and Morgon, A. (**1993**). ''Functional maturation of cochlear active mechanisms and of the medial olivocochlear system in humans,'' Acta Oto-Laryngol. **113**, 271–277.

Mountain, D. (**1980**). ''Changes in endolymphatic potential and crossed olivocochlear bundle stimulation alter cochlear mechanics,'' Science **210**, 71–72.

Olsho, L. (**1985**). ''Infant auditory perception: Tonal masking,'' Infant Behav. Dev. **8**, 371–384.

Ryan, S., and Piron, J. (**1994**). ''Functional maturation of the medial efferent olivocochlear system in human neonates,'' Acta Oto-Laryngol. **114**, 485–489.

Scharf, B., Magnan, J., Collet, L., Ulmer, E., and Chays, A. (**1994**). ''On the role of olivocochlear bundle in hearing: A case study,'' Hear. Res. **75**, 11–26.

Scharf, B., Magnan, J., and Chays, A. (**1997**). ''On the role of the olivoco-

chlear bundle in hearing: 16 case studies,'' Hear. Res. **103**, 101–122.
Siegel, J., and Kim, D. (**1982**). ''Efferent neural control of cochlear mechanics? Olivocochlear bundle stimulation affects cochlear biomechanical nonlinearity,'' Hear. Res. **6**, 171–182.
Spetner, N., and Olsho, L. (**1990**). ''Auditory frequency resolution in human infancy,'' Child Dev. **61**, 632–652.
Starr, A., Amlie, R., Martin, W., and Sanders, S. (**1977**). ''Development of auditory function in newborn infants revealed by auditory brainstem response,'' Pediatrics **60**, 831–839.
Warren, E., and Liberman, M. (**1989**). ''Effects of contralateral sound on auditory-nerve responses. I. Contributions of cochlear efferents,'' Hear. Res. **37**, 89–104.
Wiederhold, M., and Kiang, N. (**1970**). ''Effects of electrical stimulation of the COCB on single auditory nerve fibers in the cat,'' J. Acoust. Soc. Am. **48**, 950–965.

Influence of hearing sensitivity on mechano-electric transduction

Mark E. Chertoff,[a] Xing Yi, and Jeffery T. Lichtenhan
Department of Hearing and Speech, University of Kansas Medical Center, Kansas City, Kansas 66160

(Received 22 April 2003; revised 10 September 2003; accepted 17 September 2003)

This study examined the relation between the extent of permanent hearing loss and the change in a third-order polynomial transducer function (PTF) representing mechano-electric transduction (MET). Mongolian gerbils were exposed to noise for 1 to 128 h. A control group received no exposure. The cochlear microphonic (CM) was recorded from a round-window electrode and stapes velocity was recorded with a laser Doppler vibrometer in response to Gaussian noise. A nonlinear systems identification procedure provided the frequency-domain coefficients of the PTF and their associated coherence functions. In the control group, the PTF in the high frequencies was dominated by linear and cubic terms. In noise-exposed animals, the magnitude of these terms decreased with increasing threshold, suggesting a progressive decrease in the receptor currents through basal hair cells. Moreover, the linear coherence increased and the cubic coherence decreased, indicating that MET in the cochlear base became linear. In the low frequencies, noise exposure altered the group delay of the CM, demonstrating a redistribution of hair-cell currents. The low-frequency PTF was characterized by an increase in the contribution in the quadratic term. With increasing threshold, the slope of the PTF decreased and the saturation for positive CM was eliminated. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1625932]

PACS numbers: 43.64.Kc, 43.64.Ld, 43.64.Nf [BLM] Pages: 3251–3263

I. INTRODUCTION

Mechano-electric transduction (MET) is the nonlinear process in which cochlear partition motion deflects hair-cell stereocilia and opens and closes ion channels located within or in the vicinity of the stereocilia (Hudspeth and Jacobs, 1979). Opening transduction channels increases the apical conductance of hair cells allowing an influx of current, mainly potassium ions, resulting in cell depolarization. In mammals, the single-channel conductance of outer hair cells is approximately 112 pS and the number of channels per cell about 50 (Geleoc *et al.*, 1997). Upon depolarization, activation of fast-acting potassium channels in the basolateral wall of cochlear hair cells allows for rapid outflow of potassium and a subsequent repolarization of the hair-cell receptor potential (Mammano and Ashmore, 1996). Closing transduction channels decreases the apical conductance of hair cells and results in hyperpolarization.

A nonlinear function which describes MET is a Boltzmann function (Holton and Hudspeth, 1986) given as, $p = 1/(1+e^{-z(x-x_0)/kT})$, where P is the probability of channel opening, Z is the gating sensitivity of the transduction process, and X is the hair bundle's deflection from its resting position X_0. K is the Boltzmann constant and T is temperature. This function and other multistate Boltzmann functions are derived from physical arguments regarding the energy requirements for ion channel mechanics and are used to describe hair-cell current versus stereocilia displacement from isolated hair cells. *In vivo*, the Boltzmann function and other nonlinear transducer functions describing MET can be estimated from the cochlear microphonic (Nieder and Nieder, 1971; Patuzzi and Moleirinho 1998; Chertoff *et al.*, 1996). The cochlear microphonic (CM) is an extracellular electric potential created from the spatial summation of current passing through individual hair cells. Recorded from the round window in animals, the CM is dominated by hair-cell currents from cells in the base of the cochlea, particularly outer hair cells (Patuzzi *et al.*, 1989a; Dallos *et al.*, 1972).

Previously, we applied a nonlinear systems identification procedure (Bendat, 1990) to the CM to obtain the coefficients of a third-order polynomial function estimating MET (Chertoff *et al.*, 1996, 1997; Bian *et al.*, 1998, 2001). The coefficients of the linear and nonlinear terms relate ear-canal pressure to the CM as a function of frequency. These coefficients were computed assuming a nonlinear system followed by a linear system (NL system), which could lead to inaccuracies in the polynomial coefficients and reduce the associated coherence functions. A more anatomically appropriate organization would be a linear system followed by a nonlinear system (LN system), where the middle ear is considered the linear system intervening between the ear-canal pressure and the cochlea. Recently, with the addition of a laser interferometer to measure stapes velocity, the polynomial transducer function (PTF)

$$y = A_1(f)x + A_2(f)x^2 - A_3(f)x^3 + \mathrm{dc}, \tag{1}$$

where x is stapes velocity, y is the CM, and dc is direct current, was used to estimate MET without the influence of the middle ear (Choi *et al.*, 2002). For a sound pressure of 98 dB SPL, coherence functions averaged across frequency were similar to coherence values relating ear-canal pressure and CM, indicating that the middle ear did not influence the accuracy of the nonlinear terms. However, it is still possible that a linear system in the cochlea (e.g., passive basilar-membrane mechanics) precedes a nonlinearity such as those involved in hair-cell transduction. Nevertheless, the difficulty in computing and interpreting the multidimensional spectra

[a]Electronic mail: mchertof@kumc.edu

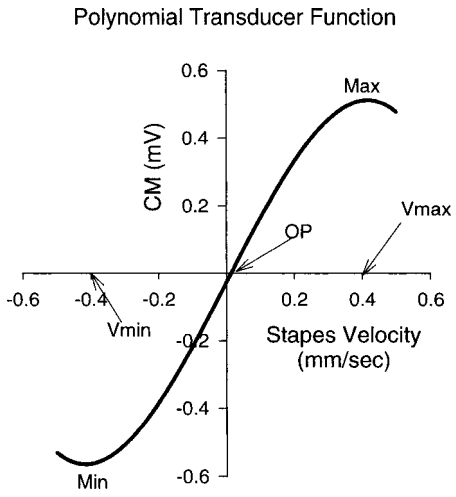


FIG. 1. Example of a polynomial transducer function [Eq. (1)] relating stapes velocity to cochlear microphonic. The coefficients are $A_1 = 1.9$ mV/mm/s, $A_2 = 0$ mV, $A_3 = 3.7$ mV/(mm/s)^3, and dc=0.026 mV. Arrows indicate the physiologic indices derived from the coefficients.

required for LN systems identification (Bendat, 1990) still makes the NL assumption a viable approach.

From the polynomial coefficients and dc component, indices can be derived (see the Appendix) which describe physiologic processes involved in MET. An example of the PTF obtained from normal animals in the present study and associated indices is shown in Fig. 1. Given that basilar-membrane (BM) displacement is proportional to stapes velocity (Dallos and Durrant, 1972; Dallos, 1973), the slope of the PTF reflects the sensitivity of MET or the amount of current passing through hair cells for a given amount of BM displacement. V_{max} and V_{min} are the BM displacements required for the saturation of hair-cell currents when stereocilia ion channels are fully open and closed, respectively. The operating point (OP), defined as the stapes velocity at which the CM=0, reflects the position of the BM for which the current through hair cells is a constant, i.e., there is no change in the silent current (Zidanic and Brownell, 1990).

The magnitude of the terms in the PTF and associated physiologic indices change in damaged cochlea (Bian and Chertoff, 1998, 2001). Moreover, the changes are specific to different cochlear alterations. That is, the coefficients and indices change in different ways when the cochlea is damaged with different noxious agents such as noise versus drugs (Bian and Chertoff, 1998, 2001). These results are important for the long-term goal of this research for developing techniques that characterize and distinguish the underlying pathophysiology associated with sensorineural hearing loss.

These previous studies used damaging agents that created a mild-to-moderate temporary hearing loss. Furthermore, with careful titration of noise and drug dosage, hearing thresholds were equated across the groups of damaged animal. Therefore, simple differences in sensitivity could be ruled out as the variable causing changes in the PTF. In a clinical setting, a more realistic situation is that patients will present with different amounts of permanent hearing loss. Presently, however, it is not yet known how, or to what extent, different degrees of permanent hearing loss will affect MET as described by the PTF. In this study, we expose animals to different durations of noise to create a wide distribution of permanently elevated hearing thresholds. The purpose of this study was to examine the relation between hearing sensitivity and the PTF representing MET.

II. METHODS

A. Subjects

Mongolian gerbils (*Meriones unguiculatus*) weighing between 40 and 60 grams were used as subjects. Hearing was assessed by recording an auditory brainstem response (ABR) to a 16-kHz tone burst (for methods see below). Animals with ABR thresholds at or below 50-dB peak SPL were included in this study. Initially, 57 animals were included in the study; however, only 43 survived to complete the experimental protocol. The experimental procedures were approved by the Institutional Animal Care and Use Committee of the University of Kansas Medical Center.

B. Signal generation and delivery

The signal used for noise exposure was a band of noise (3-kHz bandwidth) centered at 8 kHz, created by bandpass filtering (Khron-Hite, model 3202) a 10-kHz-wide noise (Grason-Stadler noise generator, model 1724). The noise was amplified (Raymer, model 811-100) and sent through a loudspeaker system (Scott, model 166) that was suspended about 0.5 in. above an animal's cage. The average level (117 dB SPL) was determined by placing a probe microphone (Etymotic ER-7C) into small holes drilled in the walls of the cage at 39 locations. The sound level and spectrum of the noise were periodically (about once a month) monitored (Stanford Research Systems Spectrum Analyzer, model SR 760) at one point in the cage to verify that no changes had occurred. The frequency of the noise was chosen so that the hearing loss would occur in the base of the cochlea where the round-window electrode is most sensitive.

Signals used to evoke ABRs and compound action potentials (CAPs) were tones windowed with a $\cos^2$ function yielding a rise/fall time of approximately 1 ms. Frozen Gaussian noise 125 ms in duration and windowed with a $\cos^2$ function ($\sim$5 ms rise/fall time) was used to evoke CM responses. Tone bursts and Gaussian noise were created in an array processor (Tucker-Davis Technology, TDT AP2), downloaded to a 16-bit digital-to-analog converter sampling at 65.536 kHz (TDT DA1), routed to an attenuator, and passed through a headphone buffer (TDT HB6) to a headphone (Etymotic ER-2). The headphone, encased in μ metal, was connected to a tube that was positioned into the entrance of the ear canal for ABR recordings. Signal level was monitored by a calibrated probe microphone (Etymotic ER-7C) attached alongside the headphone tube. For CM and CAP recordings, the headphone tube was attached to a speculum sealed to the bony external ear canal. A probe microphone inserted into the speculum monitored signal level at approximately 5 mm from the umbo of the tympanic membrane. The spectrum of the Gaussian noise was kept at a constant flat spectrum of 0.1–10.24 kHz, using an inverse transfer function routine created by Chertoff and Chen (1996). This frequency band was chosen to be below the best frequency of

the cochlear partition where the round-window electrode is most sensitive to reduce the influence of vector summation from out-of-phase receptor currents. Gaussian noise was presented at 78, 88, and 98 dB SPL; however, only the results from the 98-dB SPL condition will be presented. This level was chosen to characterize the linearity and nonlinearity of MET for signals above hearing threshold.

C. Animal preparation

Animals were sedated with 64 mg/kg of Nembutal and electrodes were placed under the skin at the vertex of head and behind the pinna for ABR recordings. Body temperature was monitored with a rectal thermometer and maintained at 37 °C using a heating pad (Harvard).

Recording CAP, CM, and stapes velocity required the opening of the bulla. In a manner similar to Choi *et al.* (2002), animals were initially anesthetized with 64 mg/kg of Nembutal and 1/3 the initial dose was administered to maintain sedation. The right pinna and surrounding skin and muscle were removed to produce easy access to the bulla. A dental drill (Volvere GX) was used to open the bulla. A silver ball electrode was placed in the round-window niche (active), and a needle electrode was inserted into the neck (ground) to record the electrical activity. The bulla remained open throughout the recordings. The round-window niche and bulla were kept dry by placing a cotton wick into the niche.

D. Physiologic recordings

1. Auditory brainstem response

ABRs were acquired using needle electrodes placed under the skin on the vertex of the head (positive), and behind the right ear (negative) and the left ear (ground). The electrical signal was bandpass filtered between 300 and 3000 Hz, amplified 1000 times by a preamplifier (Stanford SR560), 10 times by a dual variable filter (Steward VB 10 M), and digitized at 65.536 kHz (TDT AD2). The signal was averaged from 1200 stimulus presentations and saved on disk. ABRs were recorded for stimulus levels starting at 100-dB peak SPL and descending in 5-dB steps. Electroencephalic activity was monitored on-line using an oscilloscope throughout the ABR recording.

2. Compound action potential and cochlear microphonic

Compound action potentials were recorded in response to 1-, 2-, 4-, 8-, and 16-kHz tone bursts. The physiologic signal was recorded by a round-window electrode, amplified 500 times, and bandpass filtered (0.03–30 kHz) by a preamplifier (Stanford SR560). Subsequently, the signal was low-pass filtered at 25 kHz amplified 10 times (Steward VBF 10 M), and finally digitized at 65.536 kHz (TDT AD2). Signal level started at 100-dB peak SPL and descended in 5-dB steps. The averaged electrical activity was monitored on a computer screen by two investigators and, when the CAP was identified, the CAP was stored on disk for off-line analysis. At most, a CAP consisted of the average of 500 stimulus presentations

The CM in response to Gaussian noise was recorded with the same electrode configuration, filtering parameters, and sampling frequency as the CAP. The amplification of the electrical signal, however, varied depending on the individual animal. Animals with elevated CAP thresholds required more amplification in order for the electrophysiologic signal to fill the analog-to-digital converter. Amplification ranged from 5000 to 50 000. One response was the average of 20 stimulus presentations of one noise signal. Twenty different noise stimuli were presented to obtain 20 input–output records.

3. Stapes velocity

Stapes velocity was measured with a laser doppler vibrometer (Polytec hearing laser vibrometer, HLV-1000) after the stapes was painted silver to enhance laser-light reflection. A joystick mounted on the microscope positioned the laser beam from the Polytec sensor head to the footplate of the stapes. The reflected interferometer signal was demodulated in the Polytec controller. The electrical signal, proportional to stapes velocity, was low-pass filtered at 30 kHz and amplified 100 times (Stewart VBF 10 M), digitized, and stored on disk.

E. Experimental design

An animal's hearing was assessed and, if normal, the gerbil was randomly assigned to either a control group or an exposure group (1, 1.5, 2, 4, 8, 16, 64, and 128 h). If assigned to an exposure group, the animal was placed in an $11.5\times7.0\times4.5$-in. cage with three other gerbils and allowed access to food and water. The cage was placed on a table inside of a single-person sound-treated booth (Eckel Industries, model AB-200) and noise exposure initiated. An animal in the control group was placed in the sound booth for 16 h with no noise exposure

After exposure, ABRs were recorded once per week for 7 weeks or until permanent hearing loss was obtained. Permanent hearing loss was defined as a threshold that did not fluctuate more than 5 dB on two consecutive recording sessions. Initially two experimenters judged threshold from the printed waveforms; however, because the inter- and intra-judge reliability were high ($r_{\text{inter}}=0.978$, $r_{\text{intra}}=0.982$, computed from 13 animals) only one judge was needed for the remainder of the study. Five dB was chosen as the criterion because this was the average error in estimating threshold from one judge. After permanent hearing loss was achieved CAP, CM, and stapes velocity were recorded. CAP thresholds were determined in the same manner as for the ABR.

F. Signal analysis

The CM, stapes velocity, and the ear-canal acoustic signal were simultaneously recorded and saved to disk. The 8192-point time-history records were segmented into 1024 points, windowed with a Hanning window with 50% overlap (Bendat and Piersol, 1986) and spectral estimates computed. The frequency-domain coefficients of the PTF characterizing MET (A_1, A_2, and A_3) and their associated coherence functions were obtained using the nonlinear systems identifica-

tion procedure, as described previously (Chertoff *et al.*, 1996, 1997; Bian and Chertoff, 1998, 2001). The dc value was obtained by computing the mean of the CM waveform and subtracting the mean of the first 40 points where no response occurred due to a time delay of the headphone tube. This controlled for a dc shift in the A/D converter

G. Statistics

The relation between CAP hearing threshold and a dependent variable (e.g., A_1, A_2, etc.) was determined by computing correlation coefficients between threshold at each frequency and a dependent variable. Linear and nonlinear regression equations were fitted for the dependent variable using the frequency whose threshold yielded the largest and most significant correlation with the dependent variable.

The presence of a polynomial coefficient in the PTF was determined by comparing the coefficient's coherence value to coherence values computed from uncorrelated input–output records. That is, linear and nonlinear coherence functions from nine animals were obtained from response records that were reversed in time. Cumulative probability distributions of these values were constructed and the 95th percentile computed across nine animals was considered the cutoff value for a coefficient to be a valid term in the PTF description of MET. The coefficients were coded as present or absent and the relation between hearing loss and the presence of a coefficient was determined using multiple logistic regression (Neter *et al.*, 1996). This statistic provided the probability of a coefficient being present as a function of hearing loss. In some cases, the best regression equation contained a combination of thresholds at two frequencies. In these cases, only the best predictor frequency is reported. The Nagelkerke R^2 (SPSS version 7.5, 1997) was used as a measure of the goodness of fit of the logistic regression equation.

III. RESULTS

A. Threshold

The goal of the noise exposure was to produce a group of animals with a diverse distribution of permanent hearing threshold levels in the high frequencies. A scatter plot of CAP threshold is shown in Fig. 2 [panel (A)] where a slight ''jitter'' was added to some of the animals' frequency values to visualize the animals with identical thresholds. Our laboratory threshold norms are indicated by the dotted line. CAP thresholds varied as a function of frequency with the greatest elevation at 4, 8, and 16 kHz. Although many different thresholds were represented at each of the high frequencies (4, 8, and 16 kHz), not all threshold values are equally represented.

The relation between exposure duration and threshold is shown in panel (B) (Fig. 2). The mean (±1 s.d.) CAP thresholds at 8 and 16 kHz increased gradually and saturated at approximately 4 h of exposure. At 4 kHz, threshold changed abruptly between 2 and 4 h of exposure. The mean thresholds of the lower frequencies, 1 and 2 kHz, showed small changes with exposure durations starting at about 4 h as compared to the control animals.

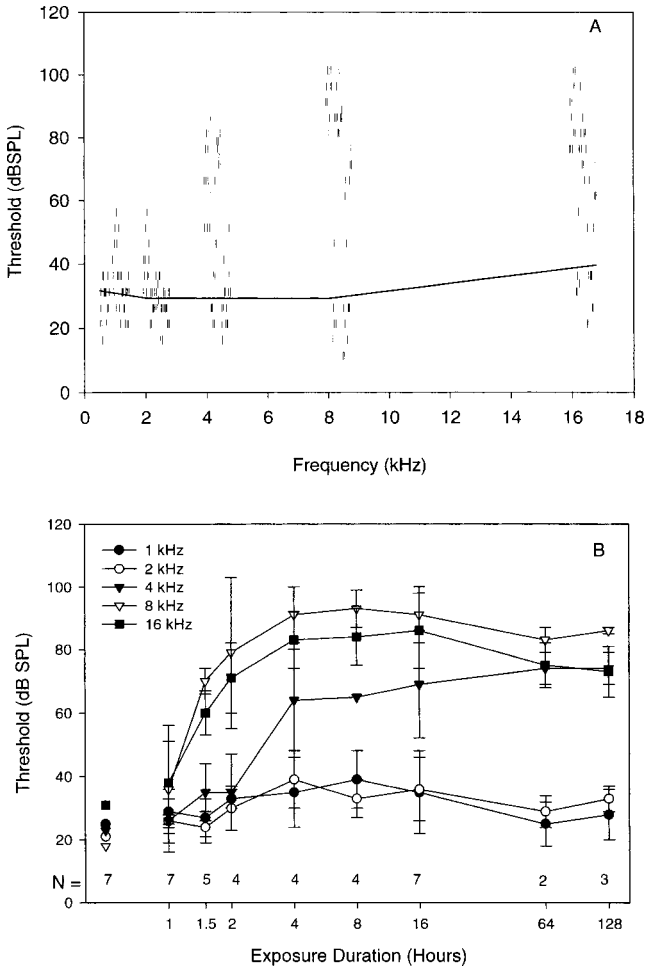


FIG. 2. CAP thresholds for each animal [panel (A)]. Solid line represents the upper limit of normal hearing (mean CAP threshold +2 s.d.) for animals in our laboratory. Panel (B) illustrates the mean (±1 s.d.) CAP thresholds for animals in the control and exposure groups. The number of animals in each group is indicated above exposure duration.

B. Polynomial transducer function: Linear term (A_1)

Figure 3, panel (A) illustrates linear term (A_1) of the PTF [see Eq. (1)] for the control animals (thick lines, ±1 s.d. from the mean) and for each of the noise-exposed animals (thin lines). Because an amplifier setting for one noise-exposed animal was not recorded, one animal was eliminated here and in any other variable that would require the correct amplification. For many of the noise-exposed animals A_1 was reduced, especially at frequencies greater than 2 kHz, whereas in some animals A_1 fell within the variation of the control animals. The relation between A_1 and CAP threshold was determined by dividing the magnitude of A_1 into four 2500-Hz frequency bands and averaging the magnitudes across frequency in each band for each animal. The magnitude of A_1 was directly related to CAP threshold [Fig. 3, panel (B)]. In the low-frequency band (≤2500 Hz), A_1 decreased at a rate of 0.32 dB per dB increase in hearing threshold at 16 kHz ($y=-0.32x-47.19$, $r^2=0.69$), whereas A_1 in the high-frequency band (≥7500 Hz) decreased considerably at 0.87 dB per dB increase in hearing threshold at 4 kHz ($y=-0.87x-35.82$, $r^2=0.79$). Although the middle two frequency bands showed a significant regression func-

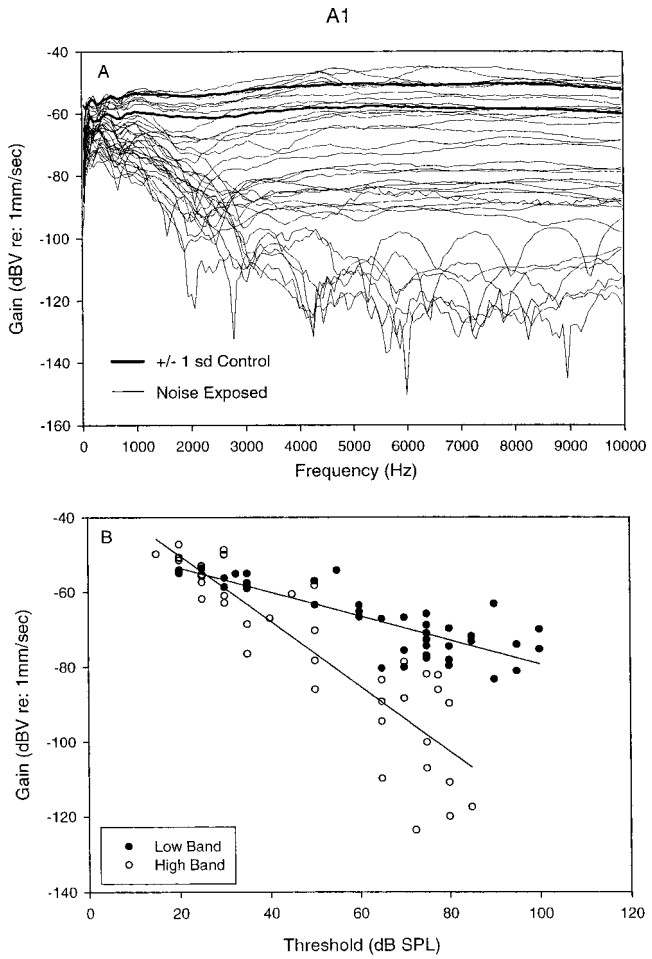


FIG. 3. Panel (A) illustrates the magnitude of the A_1 coefficient for control animals and noise-exposed animals. Thick lines represent ± 1 s.d. from the mean of the control animals. Thin lines are the experimental animals. Panel (B) shows the average A_1 amplitude for each animal computed across a low-frequency band ($\leqslant$2500 Hz, filled symbols) and a high-frequency band ($\geqslant$7500 Hz, open symbols). Solid lines are the best-fitting regression functions relating A_1 magnitude to CAP threshold at 16 kHz (low-frequency band) and 4 kHz (high-frequency band).

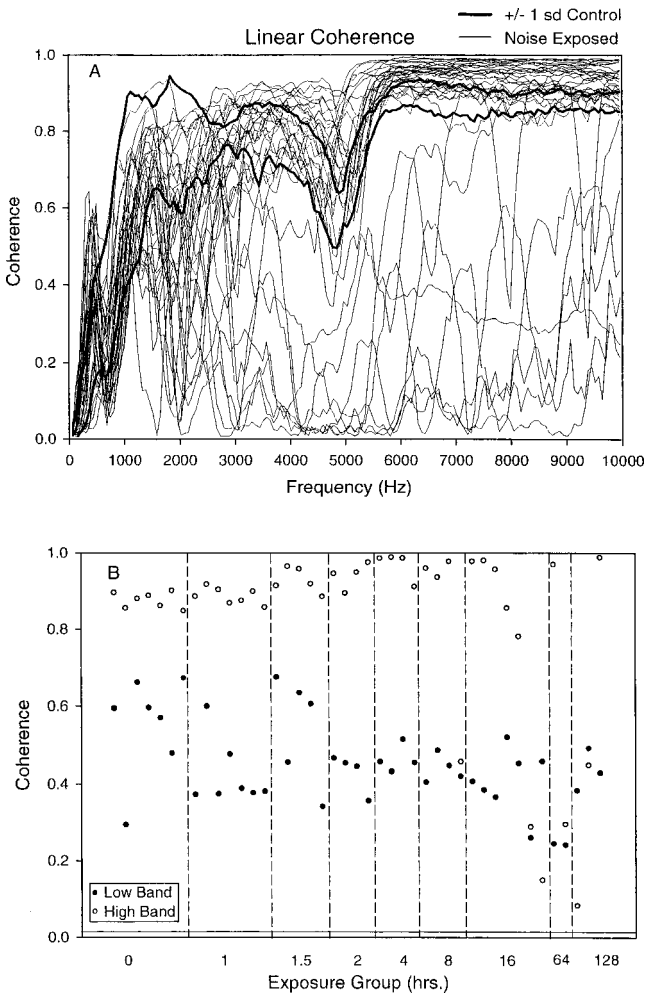


FIG. 4. Linear coherence functions for control animals (± 1 s.d. from the mean, thick lines) and noise-exposed animals (thin lines) are illustrated in panel (A). Panel (B) shows the average linear coherence for each animal in the low-frequency band (filled symbols) and the high-frequency band (open symbols) in the control and the exposure groups. The solid line indicates the criterion for a significant coherence value.

tion with threshold, t-tests between the slopes of their regression equations did not differ from the high-frequency band. Therefore, here, and in the remaining analysis only the low- and high-frequency bands are presented.

The linear coherence function for the control group (± 1 s.d. from the mean) is indicated by the thick lines in Fig. 4, panel (A). The thin lines are the linear coherence functions for the remaining, noise-exposed, animals. In many of these animals the linear coherence was larger than the control animals for frequencies greater than $\sim$5 kHz, whereas a few animals showed coherence values in the high frequencies to be much lower than the control animals. Below $\sim$5 kHz the linear coherence in the exposed animals appears similar to the control animals. Panel (B) illustrates the average linear coherence computed across frequency in the high-frequency band (open symbols) and low-frequency band (closed symbols) for each animal ($n=43$) as a function of noise exposure group. For exposures $\geqslant$1.5 h, the linear coherence for many of the animals was larger than animals in the control group and those exposed to only 1 h of noise. Animals exposed for 16 or more h showed more variable results, some having coherences reaching almost 1.0, whereas others fell below 0.2. In contrast, the linear coherence in the low-frequency band did not show any noticeable trend with increasing noise exposure duration. The solid line is the criterion for a significant coherence and illustrates that all animals still had a significant linear coherence.

The angle of the A_1 coefficient represents the phase between stapes velocity and CM as a function of frequency for 37 animals. Six animals were eliminated because the phase was noisy and not interpretable. The control group (thick lines ± 1 s.d.) and some of the exposed animals (thin lines) showed a slight phase lag as a function of frequency [Fig. 5, panel (A)], whereas others had steep phase lags for frequencies below $\sim$3 kHz followed by shallow lags in the high frequencies. Two straight lines were fit to the phase of each animal's response, one for the low-frequency band and the other for the high-frequency band. The slope of the best-fitting line was used to calculate the group delay (in seconds) for each animal by

$$\text{delay} = \frac{1}{2\pi} \cdot \frac{d\theta}{df}, \tag{2}$$

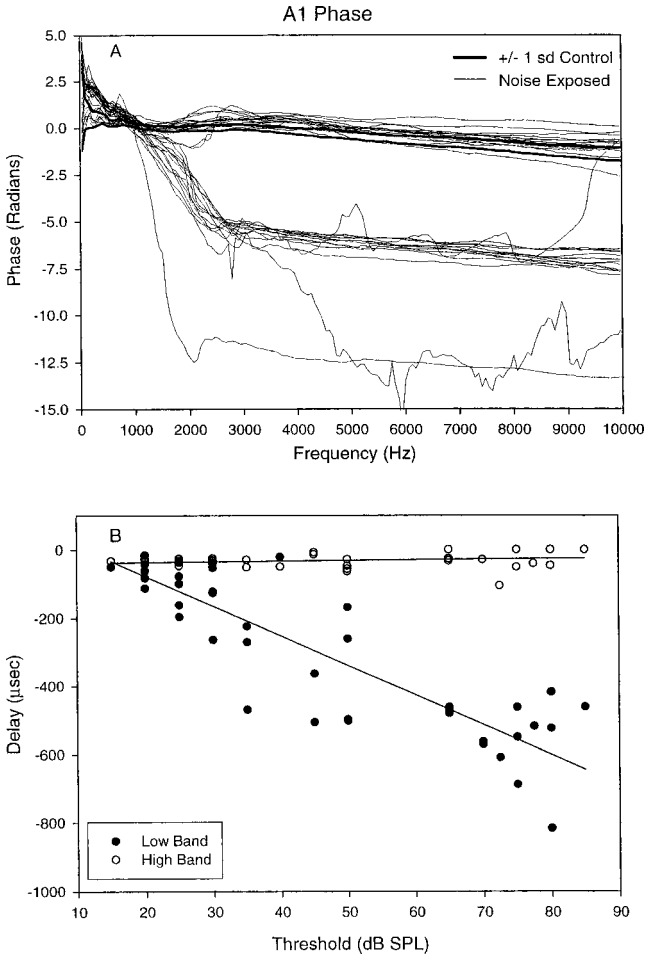


FIG. 5. Panel (A) illustrates the phase of the A_1 coefficient for the control animals (± 1 s.d. from the mean, thick lines) and the noise-exposed animals (thin lines). Panel (B) shows the group delay for each animal in the low-frequency band (filled symbols) and high-frequency band (open symbols). Solid lines are the best-fitting lines relating group delay to CAP threshold at 4 kHz (low- and high-frequency bands).

where θ is the phase and $d\theta/df$ is the slope of the phase. Straight lines fit well to the phase for the low and high bands with the average $r^2=0.85$ (± 0.12 s.d.) and 0.88 (± 0.13 s.d.) for the linear regression in the low and high bands, respectively.

Panel (B) (Fig. 5) illustrates the delay of the high and low band for each animal as a function of hearing threshold. No significant relation was found between hearing threshold at any frequency and delay in the high-frequency band. In contrast, the low-frequency band delay varied significantly with hearing threshold at 4 kHz ($y=-8.72x+97.5$, $r^2=0.76$). The delay of the low-frequency energy in the CM increased at a rate of approximately 9 μs for every 1-dB increase in hearing threshold.

C. Polynomial transducer function: Nonlinear terms (A_2 and A_3)

The magnitude of A_2 and A_3 decreased in some of the exposed animals as compared to the control animals [Fig. 6, panels (A) and (B)]. The quadratic coherence tended to increase [panel (C)] in the low frequencies, and the cubic coherence [panel (D)], in some noise-exposed animals, was similar to the control animals whereas in other animals the coherence was reduced. Figure 7 [panel (A)] shows the quadratic coherence for the high- and low-frequency bands for each animal. In the high-frequency band the quadratic coherence was below the criterion for a significant contribution to the PTF. In the low-frequency band, however, the quadratic coherence was generally above the criterion in many of the noise-exposed animals. The cubic coherence in the high-frequency band showed a noticeable decrease as a function of noise exposure [panel (B)]. Only two animals with noise exposure $\geqslant 2$ h had a significant cubic coherence. The cubic coherence in the low-frequency band showed a similar trend but was more variable, with 12 of the animals still having a significant cubic coherence for noise exposure $\geqslant 2$.

A coherence value below criterion indicated that the associated nonlinear term did not contribute to the PTF description of MET. The relation between the presence of a term and hearing threshold is shown by the logistic regression results in Table I. In the low-frequency band, an animal with a 23-dB SPL threshold at 2 kHz had a 75% chance of having an A_2 term in its PTF. For the high-frequency band, no significant relation existed between hearing threshold and the presence of A_2. In the low-frequency band, A_3 had a 75% chance of being present in the PTF for a threshold at 63 dB SPL at 4 kHz. However, with a threshold increase of 11 dB (74 dB SPL) the chance of having an A_3 term reduced to 25%. A more dramatic effect occurs in the high-frequency band. For a threshold of 75 dB SPL at 8 kHz, there was a 75% chance of having an A_3 coefficient, whereas at a threshold of 80 dB SPL the chance was reduced to 25%.

Considering only the animals with present nonlinear coefficients, the magnitude of the coefficients decreased with increasing hearing threshold levels (Fig. 8). In the low-frequency band [panel (A)], the magnitude of A_2 decreased at 0.36 dB per dB of hearing threshold at 16 kHz ($y=-0.36x-51.04$, $r^2=0.57$). For the A_3 coefficient [panel (B)], the magnitude of low- and high-frequency bands decreased precipitously at a rate of 0.62 and 0.69 dB per dB of hearing threshold, respectively, at 4 kHz ($y_{\text{low band}}=-0.62x-37.45$, $r^2=0.69$; $y_{\text{high band}}=-0.69x-31.54$, $r^2=0.60$).

D. Transducer function indices

The coefficients of the PTF and dc value were used to derive indices (see the Appendix) that represent physiologic features of MET. The slope of the PTF, or sensitivity of MET, was defined at zero stapes velocity. This definition

TABLE I. Values in percentile columns are threshold in dB SPL. Empty cells indicate no significant relation between the presence of a variable and threshold.

	Low band					High band				
Variable	75%	50%	25%	Freq. (kHz)	R^2	75%	50%	25%	Freq. (kHz)	R^2
A_2	23	20	18	2	0.55					
A_3	63	69	74	4	0.77	75	77	80	8	0.91
OP						75	83	89	16	0.63
$V_{\max}$	46	59	72	4	0.54	75	77	80	8	0.91
$V_{\min}$	84	89	95	16	0.65	73	76	79	8	0.91

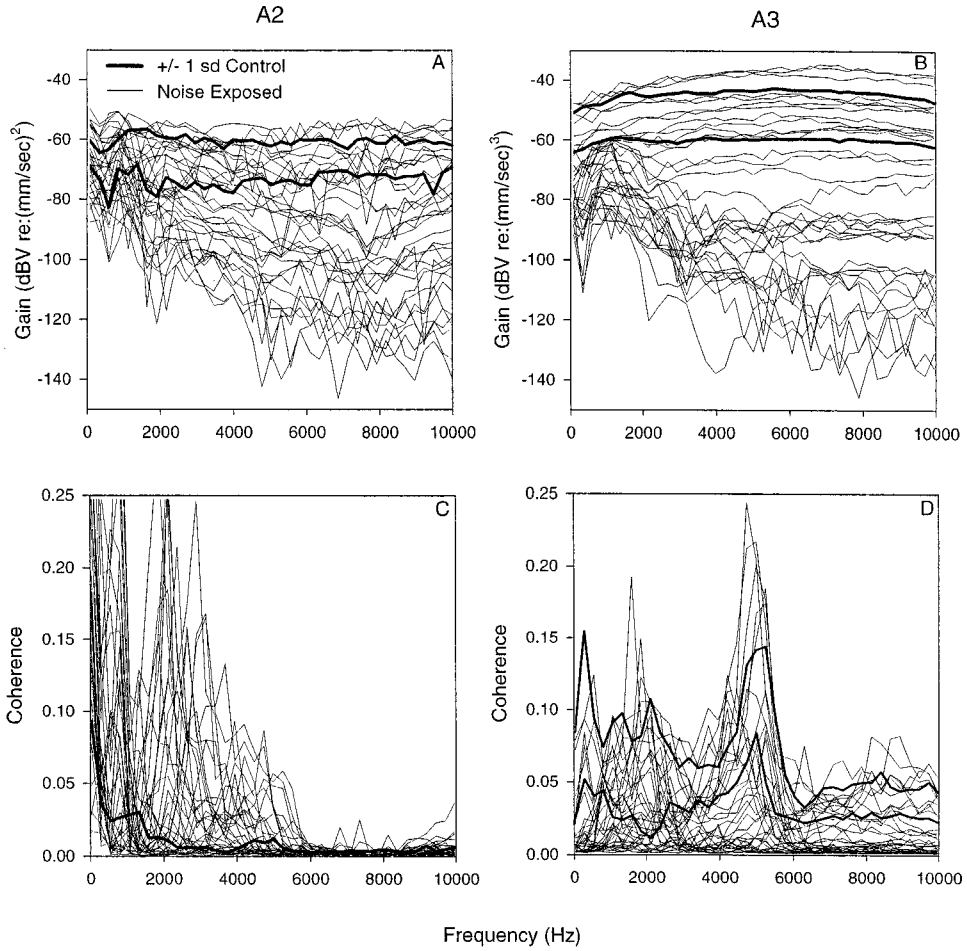


FIG. 6. Magnitude of the nonlinear coefficients A_2 [panel (A)] and A_3 [panel (B)] for the control animals (± 1 s.d. from the mean, thick lines) and the noise-exposed animals (thin lines). Panels (C) and (D) show the nonlinear coherence functions. Control animals and noise-exposed animals are designated as in panels (A) and (B).

differs from our previous studies (Chertoff *et al.*, 1997) that examined the slope at the inflection point. Nonlinear terms are required to compute the inflection point and, in the present study, these terms were not always present. Therefore, we chose to define the slope of the PTF at $x=0$, and thus, the influence of hearing threshold on the presence and magnitude of the slope of the PTF is identical to that reported for A_1.

The presence of the operating point (OP) index was altered in noise-exposed animals. A present OP (a stapes velocity measure, i.e., mm/s) was defined as a stapes velocity in which the OP remained within the domain of the stapes velocities for each animal. In other words, the minimum and maximum stapes velocities were recorded for each animal and the computed OP was required to be within this domain, otherwise the OP was considered absent. In the low-frequency band, there was no significant relation between the presence of an OP and hearing threshold at any frequency. In contrast, the presence of an OP in the high-frequency band was related to hearing threshold (Table I). For a hearing threshold of 75 dB SPL at 16 kHz, the chance of having an OP within the domain of stapes velocities was 75%. The probability was reduced to 25% when hearing threshold reached 89 dB SPL.

For the animals that maintained an OP, the location of the OP was related to hearing threshold [Fig. 9, panel (A)]. In the low- and high-frequency bands, the OP increased with hearing loss. The best-fitting lines were quadratic functions, indicating an increase in the rate at which the OP is shifted with elevation of hearing threshold at 4 kHz ($y_{\text{low band}} = (5.31e-5)x^2-(3.61e-3)x+0.062$, $r^2=0.42$; $y_{\text{high band}} = (1.19e-4)x^2-(7.72e-3)x+0.12$, $r^2=0.61$).

Noise exposure modified the stapes velocities at which the PTF saturates ($V_{\max}$,$V_{\min}$). $V_{\max}$ and $V_{\min}$ were considered present if they were within the domain of the extrema of the stapes velocities for each animal and the coefficients responsible for the calculation were present. For $V_{\max}$, when threshold at 4 kHz reached 72 dB SPL, there was only a 25% chance that $V_{\max}$ was present in the low-frequency band (Table I). In the high-frequency band, the chance was reduced to 25% when hearing threshold at 8 kHz reached 80 dB SPL. In the high-frequency band, $V_{\min}$ showed almost identical results as $V_{\max}$. However, in the low-frequency band, threshold needed to reach 95 dB SPL at 16 kHz for a 25% chance that $V_{\min}$ was present.

Figure 9, panel (B) illustrates the relation between the saturation indices and threshold in animals in which the indices were present. For the low-frequency band, $V_{\max}$ increased at 0.006 mm/s per dB of hearing threshold at 4 kHz ($y=0.006x+0.30$, $r^2=0.42$). $V_{\min}$ showed a smaller change and a much poorer fit ($y=-0.003x-0.29$, $r^2=0.14$). There was no significant relation between $V_{\max}$ or $V_{\min}$ in the high-frequency band and hearing threshold level at any frequency.

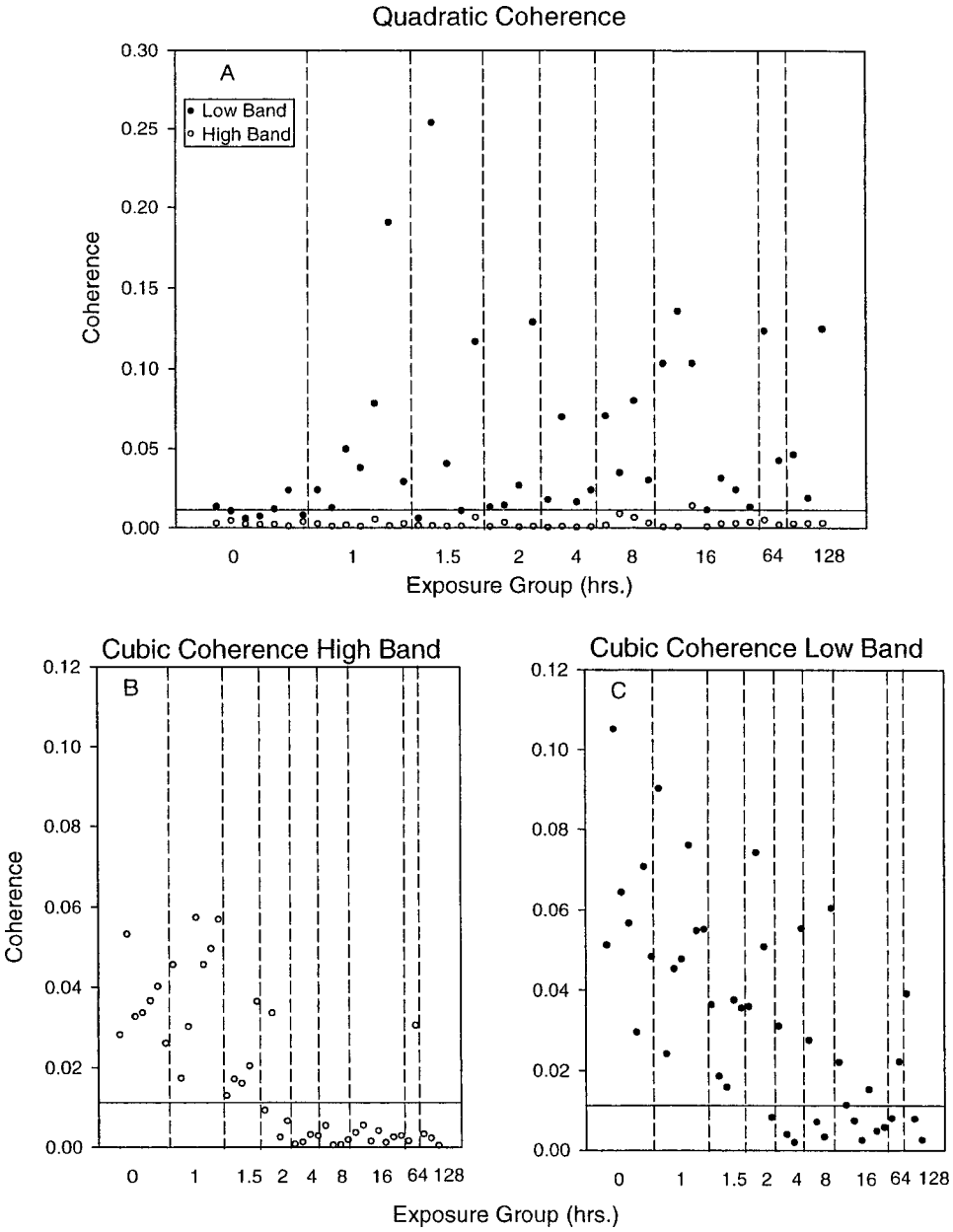


FIG. 7. Average nonlinear coherence values for each animal in the low-frequency band and (solid symbols) and the high-frequency band (open symbols) for A_2 [panel (A)] and A_3 [panels (B) and (C)]. The solid line indicates the criterion for a significant coherence value.

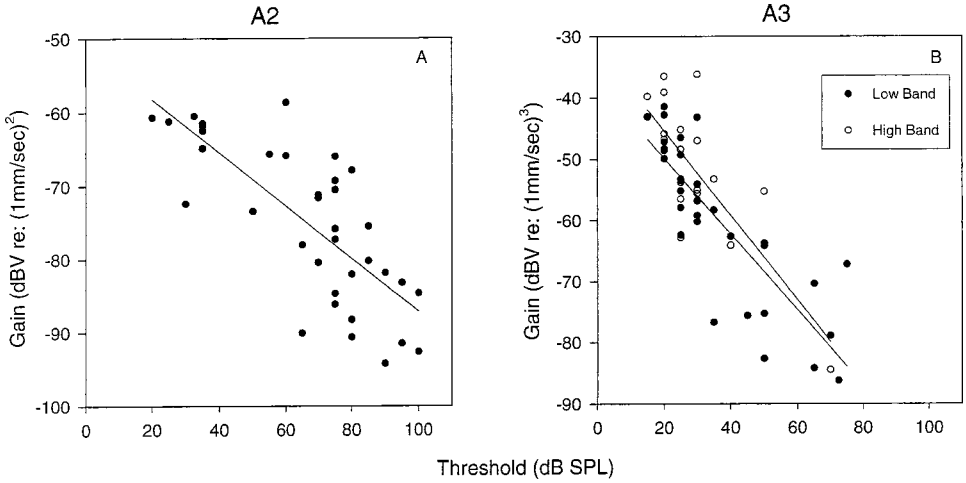


FIG. 8. Magnitude of A_2 [panel (A)] and A_3 [panel (B)] in the low-frequency band (solid symbols) and the high-frequency band (open symbols) for animals with significant nonlinear coherence values. Solid lines are best-fitting regression functions relating the magnitude of the nonlinear coefficients to CAP threshold at 16 kHz (A_2) or 4 kHz (A_3, low- and high-frequency bands).

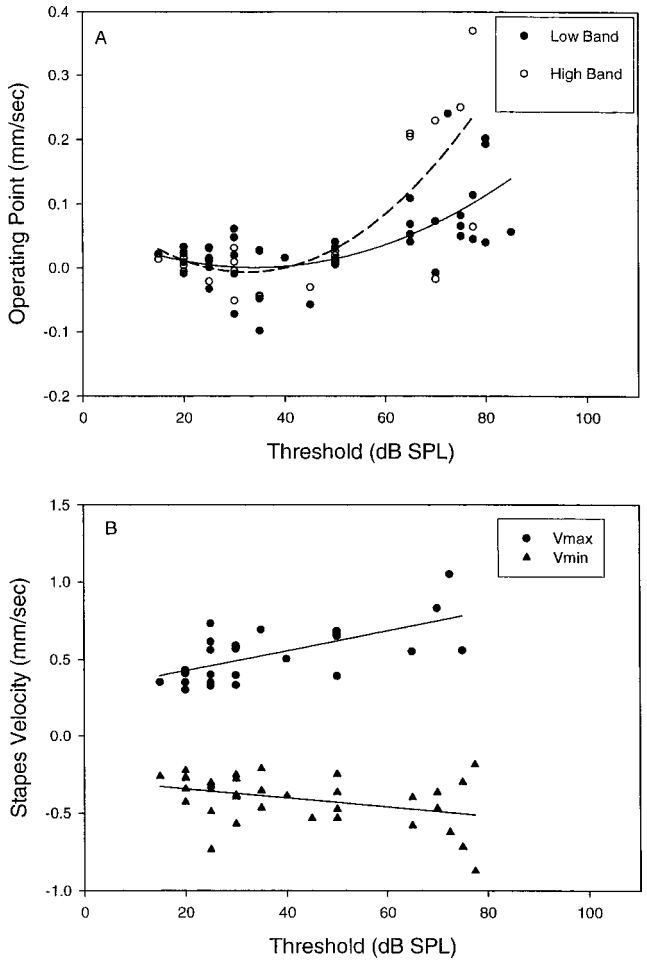


FIG. 9. Panel (A) illustrates the operating point of the PTF for each animal in the low-frequency band (solid symbols) and the high-frequency band (open symbols). The solid and dashed lines are best-fitting quadratic functions relating the operating point to CAP threshold at 4 kHz for the low-frequency band and high-frequency band, respectively. The saturation parameters, V_{max} and V_{min}, in the low-frequency band are shown in panel (B). Solid lines indicate the linear regression functions relating the saturation indices to CAP threshold at 4 kHz.

IV. DISCUSSION

A. Noise exposure and threshold

Exposure to the band of noise, centered at 8 kHz, produced the largest amount of hearing loss at 8 and 16 kHz. This is consistent with the noise-exposure literature indicating that the greatest amount of hearing loss occurs at and above the center frequency of the noise (Clark, 1991). Another consistent finding is that threshold reached an asymptote as a function of noise duration (Melnick, 1991). The maximum threshold was reached in animals with 4 or more hours of exposure. Interestingly, the noise exposure also produced hearing loss in some animals for frequencies lower than the noise center frequency (e.g., 4 kHz). This suggests that with increasing noise duration, excitation of energy on the BM spread to more apical regions. This is consistent with the explanation provided by Syka and Popelar (1980) and Dolan *et al.* (1975), who showed that hearing sensitivity for frequencies below the characteristic frequency of the exposure noise does increase as a function of exposure duration.

B. Effect of noise exposure on the PTF

1. High-frequency band

In the high-frequency band, the linear term of the PTF was present in all control- and noise-exposed animals. The quadratic term did not contribute significantly to the transducer function representing MET, and the cubic coefficient significantly contributed to the PTF of MET for animals with thresholds below ~80 dB SPL. The lack of a quadratic term and present cubic term indicates that MET in the basal turn of these animals consists mostly of odd symmetry. That is, the modulation of hair-cell receptor currents is similar for BM excursions to scala vestibuli and scala tympani. This odd symmetry is consistent with the symmetrical nature of *in vivo* estimates of the OHC transducer function in the base of the cochlea (Russell and Sellick, 1983; Russel *et al.*, 1986).

The contribution of the linear and cubic coefficients (i.e., linear and cubic coherence functions) to the PTF description of MET was modified in noise-exposed animals in the high-frequency band. In some of the animals, the linear coherence was larger than the control animals, and in animals with hearing thresholds $\geqslant$~80 dB SPL the cubic coherence was insignificant, demonstrating that the cubic term of the PTF did not contribute to the description of MET. The increase in linear coherence and the elimination of the cubic term indicates that the transducer function in these animals could be represented simply by the linear (A_1) term. That is, MET was still operational in these animals but now MET was ''linear.'' In other animals (e.g., $\geqslant$16 h), however, the magnitude of the linear coherence was substantially smaller than the control animals and the nonlinear coherence was insignificant. This indicates that the linear term of the PTF accounted for a limited proportion of MET in these animals. Because the nonlinear coherence did not increase to compensate for the decrease in linear coherence, this suggests that either the signal (CM)-to-noise ratio was small, perhaps due to greater structural damage in these animals, or MET was altered in a manner which could not be accounted for by a third-order polynomial equation.

2. Low-frequency band

In the low-frequency band, the linear and cubic terms were present in all of the control animals and in all of the animals exposed to noise for $\leqslant$1.5 h, signifying the symmetry of the PTF. Interestingly, the quadratic term became more relevant to the PTF description of MET in animals exposed to noise. This suggests that MET in the noise-exposed animals was more asymmetric than control animals. One explanation is that the asymmetry reflects the contribution of receptor currents from inner hair cells in the base of the cochlea whose transducer functions are asymmetric (Kros, 1996). Another explanation, however, is that the results reflect the summation of currents from remote apical regions of the cochlea. In noise-exposed animals, group delays were present, suggesting a redistribution of the CM source of the low-frequency receptor currents to apical regions. In these regions, *in vivo* estimates of the transducer functions of both IHCs and OHCs are similarly asymmetric (Dallos, 1986).

C. Physiological implications

1. High-frequency band

The magnitude of the linear term (A_1) of the PTF is the gain or sensitivity of MET. As shown by von Bekesy (1960), the CM, which is produced mainly from OHC receptor currents, is proportional to BM displacement. Furthermore, as indicated by Dallos and Durant (1972) and Dallos *et al.* (1972), BM displacement is proportional to stapes velocity. Therefore, the gain of MET reflects the amount of current passing through OHCs for a given BM displacement.

In the high-frequency band, the gain decreased with increasing hearing threshold and the group delay was not modified. The lack of a change in the group delay indicates that the location of the high-frequency receptor currents recorded by the round-window electrode did not change substantially because of noise exposure. The currents through these cells, however, decreased dramatically with increasing hearing threshold level. This decrease in current and relation with hearing threshold differs from those reported by Patuzzi *et al.* (1989b) and our previous work. Rearranging the equation from Patuzzi (see his Fig. 3) provides the relation between the CM remaining after exposure to pre-exposure ($nOhc$) and is given as

$$nOhc=[Hl+b(Hl-Hl_{\max})-Hl_{\max}]/[b(Hl-Hl_{\max})-Hl_{\max}], \quad (3)$$

where Hl is hearing threshold shift, $Hl_{\max}$ is the maximum shift (55 dB), and b is a constant (0.85). Using the regression equation from the present study (Sec. III B), we can compute the change in CM for exposed animals relative to the control animals. For A_1 in dB

$$A_1=-0.87(x_s+c)-35.82, \quad (4)$$

where, x_s is threshold shift and c is normal threshold defined as 25 dB SPL, then A_1 in V/mm/s is

$$A_1=k10^{-0.044x_s}, \quad (5)$$

where $k=1.3$ mV/mm/s. Assuming the middle-ear transfer function did not differ for the exposed animals compared to the control animals, then the CM of the exposed animals relative to those with no threshold shift [i.e., $x_s=0$ in Eq. (5) is given as

$$\mathrm{CM}_{e/c}=10^{-0.044x_s}. \quad (6)$$

A plot of Eqs. (3) and (6) is shown in Fig. 10 along with the noise-exposure conditions from Chertoff *et al.* (1997) and Bian *et al.* (1998). At a 30-dB threshold shift, our previous results are in good agreement with Patuzzi regarding the percent of CM remaining after damage. By contrast, the results from the present study show that the hair-cell currents are drastically reduced when compared with previous results at equal threshold shifts.

The most likely explanation for the large difference between this study and our previous studies and Patuzzi's study, is that previously animals were exposed for short durations that induced only temporary threshold elevation. Short duration exposures may cause minor morphological changes such as the temporary closing of stereocilia transduction channels (Patuzzi *et al.* 1989b; Patuzzi, 1998), pillar cell buckling, and/or stereocilia–tectorial-membrane uncoupling (Nordmann, Bohne, and Harding, 2000). These effects would decrease hair-cell currents, reduce the gain of the cochlear amplifier, and lead to a given threshold elevation. In the present study, we created permanent hearing loss. In addition to reducing the cochlear amplifier, hair cells may have been lost over a large extent of the cochlear partition. Moreover, changes to metabolic activities as well changes to the cuticular plate (Gao *et al.*, 1992; Clark and Pickles, 1996) could have occurred. Any combination of these morphological changes in the cochlea would lead to a greater reduction in hair-cell current in the permanently damaged ear than in those with just temporary damage.

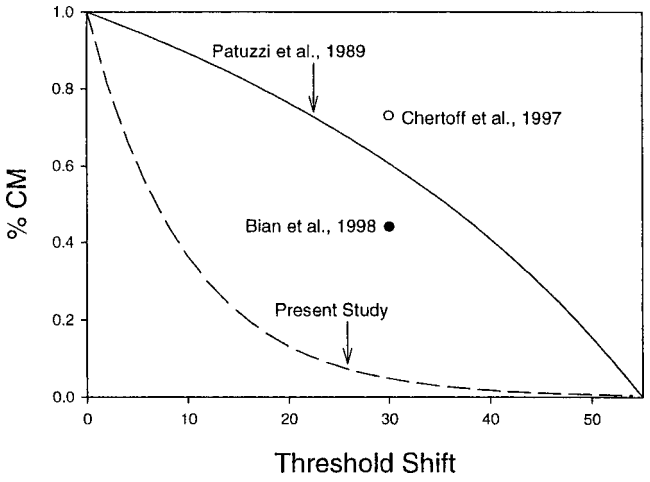


FIG. 10. Percentage of CM remaining after noise exposure as a function of a shift in hearing threshold. Solid line is the solution from Eq. (3) that was rearranged from the equation provided by Patuzzi *et al.*, 1989b. Dashed line is the solution to Eq. (6) relating the percent of CM to permanent threshold shift obtained in the present study. Solid and open symbols indicate the CM remaining from temporary threshold shift obtained from our previous studies.

It is also possible that the difference in the remaining CM for equal temporary and permanent threshold shift reflects alterations in eighth-nerve afferents and not differences in hair-cell receptor currents. Pujol and Puel (1999) report that noise exposure leads to the swelling of type 1 afferents connected to inner hair cells and disrupts other postsynaptic structures. The changes recover within approximately 5 days. Therefore, the remaining CM for both temporary and permanent hearing elevations may be similar but the change occurred in the CAP threshold. That is, the temporary CAP threshold loss could simply be larger than permanent hearing loss because of the initial alteration and subsequent recovery of eighth-nerve afferents.

Another feature of the effect of noise exposure and threshold elevation on the PTF in the high frequencies was the loss of the cubic term and the ''linearization'' of MET. The cubic term, because its sign is opposite to the linear term, causes the PTF to saturate and roll over. This saturation for positive and negative stapes velocities (i.e., $V_{\max}$ and $V_{\min}$) represents the basilar-membrane displacement required for hair-cell ion channels to fully open and close. Because of the missing cubic term, the transducer function does not saturate (i.e., $V_{\max}$ and $V_{\min}$ are absent). This result is an

extension of our previous results where temporary threshold shift was induced in animals exposed to an 8-kHz tone for either 10 min at 110 dB SPL (Bian and Chertoff, 1998) or 20 min at 100 dB SPL (Chertoff *et al.*, 1997). In these studies, the sound pressure required to saturate MET (SPL_{max} and SPL_{min}) increased after noise exposure. This implies that for a given BM displacement, the ion channels in surviving hair cells contributing to the CM no longer open or close in a normal manner. This may be due to an alteration in the channels themselves or due to a reduction in the mechanical drive to the ion channels resulting from a change in the compliance of the basilar membrane or stiffness of the hair-cell stereocilia (Saunders *et al.*, 1991).

The operating point (OP), defined in this study as the stapes velocity at which CM=0, represents the position of the cochlear partition for which a constant current passes through hair cells. For animals with a valid OP, the OP shifted towards larger stapes velocities as hearing threshold increased. This result is consistent with other investigators who show a shift in the OP after noise exposure (Bian and Chertoff, 1998; Kirk *et al.*, 1998). The shift in the OP of the BM may be due to a change in the stiffness of the stereocilia or an uncoupling with the tectorial membrane. Others have reported that OHC volume can change after noise exposure (Dew *et al.*, 1993) that would alter the OP. The consequences of a shift in OP would change the signal processing of OHCs and influence IHCs. For example, because the OHC feedback to the BM is most efficient where the slope of the OHC transducer function is steepest, a change in the OP would decrease the gain provided by OHCs and reduce the efficiency of OHC feedback to the IHCs (Russel and Kossl, 1991).

2. Low-frequency band

The gain of MET in the low-frequency band decreased as hearing threshold increased. However, unlike the high-frequency band the group delay increased, indicating that the location of the receptor currents recorded by the round-window electrode changed to a more apical cochlear location in animals with high thresholds. The actual place of the low-frequency receptor currents can be estimated from the cochlear frequency place and travel time map of gerbil constructed by Schmiedt and Zwislocki (1977). From Fig. 5 an average delay of 400 μs would correspond to a cochlear place of 8 to 10 mm from the cochlear base.

This change in location of the CM source may explain much of the change in gain in the low-frequency band. As explained by Dallos (1973), the voltage recorded by a single electrode at the round window should attenuate at a rate of 2 dB/mm along the cochlear partition away from the stapes footplate. For a round-window electrode placed ~3 mm from the most basal region of the gerbil cochlea and a shift of the CM source to an 8–10-mm cochlear location, a distance of 5–7 mm results. Given a 2-dB/mm attenuation rate, we would expect to see a decrease of roughly 10–14 dB in the CM. Comparing the A_1 gain for animals with normal thresholds at 16 kHz (mean gain −57.40 dB) with those with elevated thresholds (mean gain −69.48 dB) shows approximately a 12-dB decrease in gain [Fig. 3, panel (B)]. This suggests that the reduced low-frequency gain is mainly due to a change in source location and voltage decay. That is, once basal hair cells are damaged the CM produced from the low-frequency part of the stimulus can be ''seen'' from the apical areas.

In addition to changes in gain in the low-frequency band, other changes in MET in noise-exposed animals resemble those in the high-frequency band. For example, in the high-frequency band, the loss of the A_3 coefficient resulted in a loss of the saturation parameters V_{max} and V_{min}. This is comparable to the low-frequency band but, whereas V_{max} and V_{min} were absent at similar thresholds (~80 dB SPL), these indices in the low-frequency band were lost at different thresholds. V_{min} remained until threshold reached ~95 dB SPL, whereas V_{max} was lost at ~72 dB SPL. This reflects the contribution of the A_2 term in the low-frequency band to the V_{min} index. As shown in the Appendix, both A_2 and A_3 contribute to the V_{min} index, whereas only the A_3 term contributes to the V_{max} index. Thus, a present A_2 and a loss of A_3 would influence V_{max} more than V_{min}. Physiologically, this suggests that in the noise-exposed animals the saturation of hair-cell currents and the associated cochlear mechanics differs for cochlear partition displacement towards scala tympani versus scala vestibuli. Perhaps the effects of the noise exposure on cochlear mechanics in the apical regions are more apparent for cochlear partition displacements in one direction versus the other. However, before a more definitive interpretation can be made, it must be ruled out that these effects are not solely due to changes in source location and vector summation of their receptor currents. Studies using intracochlear electrodes would be useful for examining this issue.

D. Summary and conclusions

Figure 11 shows the average PTFs from animals in the control, 1.5-, and 4-h groups. These exposure groups were chosen to embody the results of this study because they represent the least amount of exposure to obtain moderate (65 dB SPL) and moderate-to-severe hearing (87 dB SPL) loss in the high frequencies (average thresholds at 4, 8, and 16 kHz). For normal-hearing animals (25 dB SPL) in both high- and low-frequency bands, MET is characterized by an odd symmetric PTF consisting of linear and cubic terms. With moderate hearing loss, the slope of the PTF decreases (high- and low-frequency bands) and is less saturated. In addition, in the low-frequency band, the quadratic term of the PTF becomes significant which results in an asymmetric PTF. For moderate-to-severe hearing loss, the PTF is reduced to simply a linear function with a shallow slope and a shift in the *x* intercept.

These changes in the PTF suggest alterations in the structures and physiologic events associated with MET resulting in sensory hearing loss. A shallow slope indicates a decrease in the current passing through ion channels, perhaps due to changes in stereocilia and/or a reduction in the number of hair cells. The decrease and eventual loss of saturation suggests an alteration in the mechanical drive to the stereocilia and a shift in the *x* intercept or OP reflects a change in the resting position of the BM. Future research determining

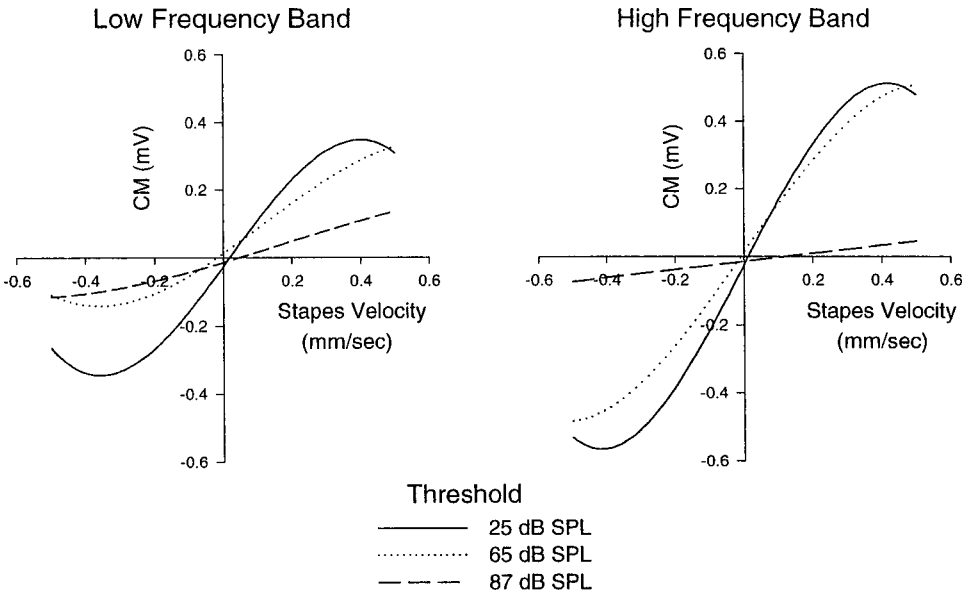


FIG. 11. Average PTFs in the low-frequency band (top) and the high-frequency band (bottom) for normal-hearing and hearing-impaired animals. The solid line represents animals with normal hearing. The dotted line indicates animals with moderate hearing loss, and the dashed line represents animals with moderate-to-severe hearing loss. PTFs were obtained from the average linear and nonlinear coefficients and dc values for animals in the control, 1.5-, and 4-h exposure groups.

the relation between alterations in the PTF and anatomic and physiologic processes involved in MET may be useful for diagnosing and developing new remediation strategies for sensory hearing loss.

ACKNOWLEDGMENTS

The authors would like to thank Lin Bian and Chul-Hee Choi for their valuable comments on this paper. This study was supported by Grant No. 2R01 DC02117 from the National Institute on Deafness and other Communication Disorders of NIH, and USPHS Grant No. HD02528 to the Smith Mental Retardation and Human Development Research Center at the University of Kansas Medical Center.

APPENDIX: DERIVATION OF PHYSIOLOGIC INDICES

The physiologic indices, slope, operating point (OP), stapes velocity at the maximum CM ($V_{\max}$), and minimum CM ($V_{\min}$), are derived from the polynomial transducer function

$$y = A_1 x^1 + A_2 x^2 - A_3 x^3 + \mathrm{dc}, \tag{A1}$$

where x is the stapes velocity, y is the CM, and the coefficients are derived from a nonlinear systems identification procedure (Bendat, 1990). The slope of the PTF is given as

$$\frac{dy}{dx} = A_1 + 2A_2 x - 3A_3 x^2, \tag{A2}$$

and if defined at $x=0$, then the slope is

$$\frac{dy}{dx} = A_1. \tag{A3}$$

The OP is defined as the roots of Eq. (A1), where the root closest to $x=0$ was chosen as the physiologic relevant parameter. If the nonlinear terms are not present in Eq. (A1), then

$$\mathrm{OP} = \frac{\mathrm{dc}}{A_1}. \tag{A4}$$

Four different conditions exist for the saturation indices $V_{\max}$ and $V_{\min}$. If all coefficients are present in Eq. (A1), and $dy/dx = 0$ in Eq. (A2), then

$$x = V_{\max} \quad \text{and} \quad V_{\min} = \frac{(-2A_2 \pm \sqrt{4A_2^2 + 12A_3A_1})}{6A_3}. \tag{A5}$$

If the quadratic coefficient is absent in Eq. (A1), then

$$x = V_{\max} \quad \text{and} \quad V_{\min} = \pm\sqrt{\frac{A_1}{3A_3}}. \tag{A6}$$

When A_3 is not present in Eq. (A1),

$$y = A_1 x + A_2 x^2 + \mathrm{dc}, \tag{A7}$$

$$\frac{dy}{dx} = A_1 + 2A_2 x, \tag{A8}$$

$$0 = A_1 + 2A_2 x, \tag{A9}$$

and

$$x = V_{\min} = \frac{-A_1}{2A_2}, \tag{A10}$$

which is $V_{\min}$ only; $V_{\max}$ is not present. When neither of the nonlinear coefficients exists in Eq. (A1), $y = A_1 x$ and $V_{\max}$ and $V_{\min}$ are not present.

von Bekesy, G. (**1960**). *Experiments in Hearing* (McGraw-Hill, New York).
Bendat, J. S. (**1990**). *Nonlinear System Analysis and Identification from Random Data* (Wiley, New York).
Bendat, J. S., and Piersol, A. G. (**1986**). *Random Data: Analysis and Measurement Procedures*, 2nd ed. (Wiley, New York).
Bian, L., and Chertoff, M. E. (**1998**). "Differentiation of cochlear pathophysiology in ears damaged by salicylate or pure tone using a nonlinear systems identification technique," J. Acoust. Soc. Am. **104**(4), 2261–2271.
Bian, L., and Chertoff, M. E. (**2001**). "Distinguishing cochlear pathophysiology in 4-aminopyridine and furosemide treated ears using a nonlinear systems identification technique," J. Acoust. Soc. Am. **109**(2), 671–685.
Chertoff, M. E., and Chen, J. (**1996**). "An *in situ* calibration procedure for click stimuli," J. Acoust. Soc. Am. **7**, 130–136.

Chertoff, M. E., Steele, T., and Bian, L. (**1997**). ''Characterizing cochlear mechano-electric transduction in ears damaged with pure tones,'' J. Acoust. Soc. Am. **102**(1), 441–450.
Chertoff, M. E., Steele, T. C., Ator, G. A., and Bian, L. (**1996**). ''Characterizing cochlear mechano-electric transduction using a nonlinear systems identification procedure,'' J. Acoust. Soc. Am. **100**(6), 3741–3753.
Choi, C., Chertoff, M. E., and Yi, X. (**2002**). ''Characterizing cochlear mechano-electric transduction with a nonlinear system identification technique: The influence of the middle ear,'' J. Acoust. Soc. Am. **112**(6), 2898–2909.
Clark, W. W. (**1991**). ''Recent studies of temporary threshold shift (TTS) and permanent threshold shift (PTS) in animals,'' J. Acoust. Soc. Am. **90**(1), 155–163.
Clark, J. A., and Pickles, J. O. (**1996**). ''The effects of moderate and low levels of acoustic overstimulation on stereocilia and their tip links in the guinea pig,'' Hear. Res. **99**, 119–128.
Dallos, P. (**1973**). *The Auditory Periphery. Biophysics and Physiology* (Academic, New York).
Dallos, P. (**1986**). ''Neurobiology of cochlear inner and outer hair cells: Intracellular recordings,'' Hear. Res. **22**, 185–198.
Dallos, P., Billone, M. C., Durrant, J. D., and Raynor, W. S. (**1972**). ''Cochlear inner and outer hair cells: Functional differences,'' Science **177**, 356–358.
Dallos, P., and Durrant, J. D. (**1972**). ''On the derivative relationship between stapes movement and cochlear microphonic,'' J. Acoust. Soc. Am. **52**(4), 1263–1265.
Dew, L. A., Owen, R. G., and Mulroy, M. J. (**1993**). ''Changes in size and shape of auditory hair cells *in vivo* during noise-induced temporary threshold shift,'' Hear. Res. **66**, 99–107.
Dolan, T. R., Ades, H. W., Bredberg, G., and Neff, W. D. (**1975**). ''Inner ear damage and hearing loss after exposure to tones of high intensity,'' Acta Oto-Laryngol. **80**, 343–352.
Gao, W., Ding, D., Zheng, X., Ruan, F., and Liu, Y. (**1992**). ''A comparison of changes in the sterocilla between temporary and permanent hearing losses in acoustic trauma,'' Hear. Res. **63**, 27–41.
Geleoc, G. S. G., Lennan, G. W. T., Richardson, G. P., and Kros, C. J. (**1997**). ''A quantitative comparison of mechanoelectrical transduction in vestibular and auditory hair cells of neonatal mice,'' Proc. R. Soc. London, Ser. B **264**, 611–621.
Holton, T., and Hudspeth, A. J. (**1986**). ''The transduction channel of hair cell from the bull-frog characterized by noise analysis,'' J. Physiol. (London) **375**, 195–227.
Hudspeth, A. J., and Jacobs, R. (**1979**). ''Stereocilia mediate transduction in vertebrate hair cells,'' Proc. Natl. Acad. Sci. U.S.A. **76**(3), 1506–1509.
Kirk, D. L., and Yates, G. K. (**1998**). ''4-aminopyridine in scala media reversibly alters the cochlear potentials and suppresses electrically evoked oto-acoustic emissions,'' Audiol. Neuro-Otol. **3**, 21–39.
Kros, C. J. (**1996**). ''Physiology of mammalian cochlear hair cells,'' in *The Cochlea*, edited by P. Dallos, A. Popper, and R. Fay (Springer, New York), pp. 347–350.
Mammano, F., and Ashmore, J. F. (**1996**). ''Differential expression of OHC potassium current in the isolated cochlea of guinea pigs,'' J. Physiol. (London) **496**, 639–646.
Melnick, W. (**1991**). ''Human temporary threhsold shift (TTS) and damage risk,'' J. Acoust. Soc. Am. **90**(1), 147–154.
Netter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. (**1996**). ''Multiple logistic regression,'' in *Applied Linear Statistical Models* (McGraw-Hill, New York), pp. 580–585.
Nieder, P., and Nieder, I. (**1971**). ''Determination of microphonic generator transfer characteristic from modulation data,'' J. Acoust. Soc. Am. **49**(2), 478–492.
Nordmann, A. S., Bohne, B. A., and Harding, G. W. (**2000**). ''Histopathological differences between temporary and permanent threshold shift,'' Hear. Res. **139**, 13–30.
Patuzzi, R. B., Yates, G. K., and Johnstone, B. M. (**1989a**). ''The origin of the low frequency microphonic in the first cochlear turn of guinea-pig,'' Hear. Res. **39**, 177–188.
Patuzzi, R. B., Yates, G. K., and Johnstone, B. M. (**1989b**). ''Changes in cochlear microphonic and neural sensitivity produced by acoustic trauma,'' Hear. Res. **39**, 189–202.
Patuzzi, R. B. (**1998**). ''A four-state kinetic model of the temporary threshold shift after loud sound based on inactivation of hair cell transduction channels,'' Hear. Res. **125**, 39–70.
Patuzzi, R. B., and Moleirinho, A. (**1998**). ''Automatic monitoring of mechano-electrical transduction in the guinea pig cochlear,'' Hear. Res. **125**, 1–16.
Pujol, R., and Puel, J. (**1999**). ''Excitotoxicity, synaptic repair, and functional recovery in the mammalian cochlea: A review of recent findings,'' Ann. N.Y. Acad. Sci. **884**, 249–254.
Russell, I. J., Cody, A. R., and Richardson, G. P. (**1986**). ''The responses of inner and outer hair cells in the basal turn of the guinea-pig cochlea and in the mouse cochlea grown *in vitro*,'' Hear. Res. **22**, 199–216.
Russell, I. J., and Kossl, M. (**1991**). ''The voltage responses of hair cells in the basal turn of the guinea-pig cochlear,'' J. Physiol. (London) **435**, 493–511.
Russell, I. J., and Sellick, P. M. (**1983**). ''Low frequency characteristics of intracellularly recorded receptor potentials in guinea pig cochlear hair cells,'' J. Physiol. (London) **338**, 179–206.
Saunders, J. C., Cohen, Y. E., and Szymko, Y. M. (**1991**). ''The structural and functional consequences of acoustic injury in the cochlea and peripheral auditory system: A five year update,'' J. Acoust. Soc. Am. **90**(1), 136–146.
Schmiedt, R. A., and Zwislocki, J. J. (**1977**). ''Sound transmission and cochlear-microphonic characteristics,'' J. Acoust. Soc. Am. **61**(1), 133–149.
SPSS Professional Statistics 7.5 (**1997**). SPSS Inc., Chicago, pp. 47–48.
Syka, J., and Popelar, J. (**1980**). ''Hearing threshold shifts from prolonged exposure to noise in guinea pig,'' Hear. Res. **3**, 205–213.
Zidanic, M., and Brownell, W. (**1990**). ''Fine structure of the intracochlear potential field I. The silent current,'' Biophys. J. **57**, 1253–1268.

A short latency vestibular evoked potential (VsEP) produced by bone-conducted acoustic stimulation

Neil P. McAngus Todd[a)]
Department of Psychology, University of Manchester, Manchester M13 9PL, United Kingdom

Sally M. Rosengren and James G. Colebatch
Institute of Neurological Sciences and UNSW Clinical School, Prince of Wales Hospital, Randwick, Sydney 2031, Australia

(Received 8 June 2003; revised 22 September 2003; accepted 24 September 2003)

In this paper data are presented from an experiment which provides evidence for the existence of a short latency, acoustically evoked potential of probable vestibular origin. The experiment was conducted in two phases using bone-conducted acoustic stimulation. In the first phase subjects were stimulated with 6-ms, 500-Hz tone bursts in order to obtain the threshold V_T for vestibular evoked myogenic potentials (VEMP). It was confirmed that the difference between bone-conducted auditory and acoustic vestibular thresholds was slightly over 30 dB. The estimated threshold was then used as a reference value in the second part of the experiment to stimulate subjects over a range of intensities from -6 to $+18$ dB ($re{:}V_T$). Averaged EEG recordings were made with eight Ag/AgCl electrodes placed on the scalp at Fpz, F3, F4, F7, F8, Cz, T3, and T4 according to the 10–20 system. Below V_T auditory midlatency responses (MLRs) were observed. Above V_T two additional potentials appeared: a positivity at about 10 ms (P10) which was maximal at Cz, and a negativity at about 15 ms (N15) which was maximal at Fpz. Extrapolation of the growth functions for the P10 and N15 indicated a threshold close to V_T, consistent with a vestibular origin of these potentials. Given the low threshold of vestibular acoustic sensitivity it is possible that this mode may make a contribution to the detection of and affective responses to loud low frequency sounds. The evoked potentials may also have application as a noninvasive and nontraumatic test of vestibular projections to the cortex. [DOI: 10.1121/1.1628249]

PACS numbers: 43.64.Ri, 43.64.Qh, 43.66.Wv, 43.80.Lb [BLM] Pages: 3264–3272

I. INTRODUCTION

In the last decade or so there has been considerable research on the acoustic sensitivity of the vestibular system and sacculus in particular (for review, see Todd, 2001). In lower vertebrates the sacculus is an important auditory structure and, although it is conventionally considered to be primarily vestibular in higher vertebrates, it appears to have conserved an acoustic sensitivity throughout vertebrate phylogeny. Its functional significance has yet to be demonstrated beyond doubt but several hypotheses have been proposed (Todd, 2001).

In the case of humans this work has centered around the use of vestibular evoked myogenic potentials (VEMPs). These are generated by modulation of tonic EMG activity, may be readily obtained from the muscles of the neck, which are the targets for the vestibulocollic reflex system, and consist of a biphasic wave with latencies of approximately 13 and 23 ms (Colebatch *et al.*, 1994). The most likely network is a trisynaptic pathway involving transduction at the otolith, and transmission via the vestibular nucleus, an inhibitory interneurone in the vestibulospinal tract and the motoneurone of the neck muscles. It has been established that, for tone bursts, the VEMP has a threshold of about 90 dB (A), exhibits a tuning with a maximum around 300 Hz (Todd *et al.*, 2000), and shows adaptation characteristic of a normal sensory system (Todd and Cody, 2002).

More recently it has been shown that the sacculus is in fact quite sensitive to bone-conducted (BC) sound (Welgampola *et al.*, 2003). BC hearing is a complex phenomenon since it may be mediated by at least three routes (Tonndorf, 1966): (a) via the external ear (Stenfelt *et al.*, 2003), (b) via inertial effects of middle ear ossicles and inner ear fluid (Stenfelt *et al.*, 2002), and (c) via compressional effects of the skull (Stenfelt *et al.*, 2000). The third route is usually described by means of the skull transfer function which is characterized by a number of resonances and antiresonances, the most prominent being a low-frequency antiresonance at about 500 Hz. Such antiresonances can give rise to the ''lateralization effect'' (Stenfelt *et al.*, 2000) whereby stimulation of one mastoid can result in a greater response in the contralateral ear. The detection of sound conducted from air to bone, after transmission through the external ear canal (the first route) has been blocked by a combination of muffs and plugs, defines the ''bone conduction limit'' (Berger, 1983). This indicates an attenuation minimum of about 40 dB at 2 kHz and a maximum of about 55 dB at 500 Hz. The 500-Hz maximum of attenuation is likely to be due to the low-frequency antiresonance.

The final saccular response to BC stimulation then is likely to be determined by three principal factors. The first factor is the output of the stimulator. Commercial audiometric stimulators fall off rapidly in their output below 500 Hz

a)Electronic mail: todd@psy.man.ac.uk

and therefore it is not possible to deliver significant stimulation for very low frequencies. The second factor is the skull transmission properties, which as noted above is characterized by a 500-Hz antiresonance and complicating lateralization effects. The third factor is the resonance properties of the sacculus itself, which the above previous studies would indicate lies below the skull antiresonance at about 300 Hz.

Investigations of saccular sensitivity to BC sound confirm that the threshold difference between auditory and acoustic vestibular responses is considerably less than the 80–90-dB difference for air conducted (AC) sound. For example, in Welgampola *et al.* (2003) the auditory-vestibular threshold difference was estimated to be about 30 dB at 250 Hz. These experiments also show that the optimal stimulus frequency for obtaining a VEMP by BC is a little less than the 300 Hz for AC. Welgampola *et al.* (2003) provide evidence that the VEMP amplitude is largest at 250 Hz (although it was not possible to investigate lower frequencies in this study) with an apparent dip at 500 Hz consistent with a skull antiresonance. Sheykholeslami *et al.* (2001) report an apparent resonance at about 200 Hz but show no sign of the antiresonance. Nevertheless, the fact that the auditory-vestibular acoustic threshold gap is so small for BC suggests that this mode of stimulation may be useful for investigating vestibular projections additional to the projection which mediates the VEMP. The ratio of saccular to cochlear activation will be much higher, and thus it should in principle be possible to dissociate the two sources.

First-order vestibular afferents send projections to the vestibular nuclear complex in the brain-stem, which consists of four principal nuclei. Second-order afferents project to a number of structures (Benson, 1982) including the cerebellum, the spinal tract, the reticular formation, oculomotor nuclei, and thalamus. Within the thalamus, ventro-posterior and other areas, including the medial division of the medial geniculate body, intralaminar nuclei and pulvinar, have been identified as receiving the second-order vestibular afferents (Blum *et al.*, 1979).

In turn the thalamic areas project to multiple areas in the cortex. In animal studies several cortical targets have been determined including area 2 in the intraparietal sulcus, area 3av in the central sulcus, parieto-insular cortex (PIVC) and area 7 of the inferior parietal lobule (Benson, 1982). Human imaging techniques using caloric and galvanic stimulation (see Paulesu *et al.*, 1997 for review) have found activation in areas consistent with animal work, including the supramarginal gyrus (BA 40, the homologue of area 7). Other areas have also been implicated, including the contralateral temporoparietal cortex, temporal gyrus, putamen, cingulate cortex, and premotor cortex. A cross-comparison with human lesion studies emphasizes the putamen and insular areas (Paulesu *et al.*, 1997). The picture which emerges is that of a network of areas which are projected to in parallel by vestibular afferents. The various projections are likely to be involved in sensory, motor, and affective functions of the vestibular apparatus.

Vestibular projections to the cortex have been investigated using averaged EEG. Most previous studies have used imposed perturbations of the head or head and body, which have the disadvantage of potentially causing artifacts and of activating the vestibular apparatuses of both sides simultaneously (e.g., Hood and Kayan, 1985; Sohmer *et al.*, 1999). Recently, de Waele *et al.* (2001), using direct stimulation of the exposed vestibular nerve, reported a short latency vestibular evoked potential which started at about 6 ms with a peak latency of about 12 ms. Scalp voltage distribution and subsequent dipole source analysis of the potentials implicated five areas, including prefrontal and/or frontal, ipsilateral temporoparietal, contralateral premotor, and contralateral parietal regions. Of these, the prefrontal/frontal source was considered to be the most surprising since it had not been reported in previous studies. A possible ocular source was thought to have been excluded since the response was preserved despite curarization. An aim of the work reported here was to attempt to extend the de Waele *et al.* (2001) study using noninvasive, acoustic stimulation.

II. METHOD

Given that we would be using acoustic stimulation, we expected any vestibular evoked potentials (VsEP) to be mixed with an auditory midlatency response (MLR). The MLR consists of a series of waves—N_0, P_0, N_a, P_a, and N_b—with latencies of approximately 10, 12, 18, 27, and 35 ms (Picton *et al.*, 1974). In order to maximize the chances of observing a VsEP we chose 500-Hz, 6-ms bone-conducted tone bursts, which are close to the optimal stimulus for obtaining a VEMP for the particular stimulator used. As a control stimulus we used 5-kHz, 6-ms bone-conducted tone bursts which are beyond the region of saccular acoustic sensitivity (Sheykholeslami *et al.*, 2001). Stimulus intensities were defined relative to the VEMP threshold (V_T), which was determined in the first part of the experiment.

A. Subjects

Two populations of subjects were employed for the experiment; ten normal subjects with no hearing or vestibular deficits recruited from staff at the Prince of Wales Hospital and a small sample of patients with inner ear deficits. These included a patient with Tullio phenomenon, characterized by hypervestibular symptoms and a low threshold for acoustic activation of the vestibular apparatus (Colebatch *et al.*, 1998), a unilaterally deaf subject with a cochlear lesion but with intact vestibular function, and a hypovestibular patient with normal hearing.

B. Apparatus

Stimuli were generated by means of customized software, using a laboratory interface (1401 plus, Cambridge Electronic Design, Cambridge, UK) and custom amplifier, and delivered by a commercial bone-conductor (B71, Radioear Corp., New Eagle, PA). EMG recordings of VEMPs were made using active surface electrodes placed over the middle of the sternocleidomastoid (SCM) muscle belly and reference electrodes placed on the medial clavicle. Each subject's threshold for the VEMP response was determined and used as the reference (V_T) for the cortical recordings. EMG and EEG amplification and analog filtering were made by a

bank of D150 amplifiers (Digitimer Co, Welwyn Garden City, UK). EMG and EEG were sampled by means of a second CED 1401 plus using SIGAVG software.

C. Stimuli

Stimuli were bone-conducted 500-Hz and 5-kHz, 6-ms tone bursts (1-ms rise time, 4-ms hold time, 1-ms fall time). Tone bursts were generated by custom software with alternating polarity. The intensity of the stimuli was measured in terms of peak-to-peak amplitudes. The maximum output within the acceptable linear range was 20 V pp.

D. Procedure

Before the experiment the subjects were informed of the procedure and asked to sign a consent form. Screening audiograms were obtained to exclude significant hearing loss. The thresholds for auditory perception of the bone-conducted stimuli were measured for both 500 Hz and 5 kHz. The experiment was conducted in two parts, the first to obtain the VEMP threshold V_T and the second to obtain averaged EEG at multiple intensities relative to V_T. All subjects were stimulated on the left and right mastoids.

1. Bone-conducted tone burst evoked VEMPs

Subjects were tested lying supine on a couch, with the backrest tilted to approximately 30°–45° from the horizontal, and lifted their heads against gravity to activate the SCM muscles. Bone-conducted tone bursts were delivered to the mastoid process (approximately 3 cm posterior and 1 cm superior to the external auditory meatus) via the bone-conductor for a total of 256 stimuli at a rate of 5 Hz. VEMP recordings were made in response to both 500-Hz and 5-kHz stimuli. Surface EMG was measured from the ipsilateral SCM using self-adhesive Ag/AgCl electrodes. An earth electrode was placed over the sternum. EMG was amplified and bandpass filtered (8 Hz to 1.6 kHz). The EMG was sampled at 5 kHz from 20 ms before to 100 ms following stimulus onset and averaged. Tone burst evoked VEMP amplitude was determined by calculating the peak-to-peak amplitude (by addition of the absolute amplitudes at p13 and n23) and dividing by the mean level of baseline EMG activity (calculated from the mean rectified EMG in the 20 ms preceding stimulus onset). Threshold values for V_T were obtained by successively reducing the voltage of stimulation in 3-dB steps. The procedure was repeated for left and right ears independently.

2. Cortical potentials

Cortical evoked potentials were recorded with subjects reclining as described above, but relaxed and with their eyes shut. Bone-conducted tone bursts were presented unilaterally as described above to a total of 1000 stimuli per trial, at a rate of 5–10 Hz. EEG was recorded via 9 mm Ag/AgCl electrodes placed at Fpz, F3, F4, F7, F8, Cz, T3, and T4 and referred to linked earlobes. [Although this is a commonly used reference, which avoids ECG artifact and limits post-auricular myogenic response (PAMR) artifact, it does pick up some activity resembling that in Fpz.] An earth electrode was placed on the sternum or the inner forearm. EEG was sampled at 10 kHz for a period of 60 ms, from 10 ms before to 50 ms following stimulus onset. The signal was amplified and bandpass filtered (5–300 Hz, typical for MLR recordings). Electrode impedance was maintained below 5 kΩ in all trials. Data were collected using the same analog to digital interface and software as described above. Amplitudes were measured after removal of any amplifier DC offset.

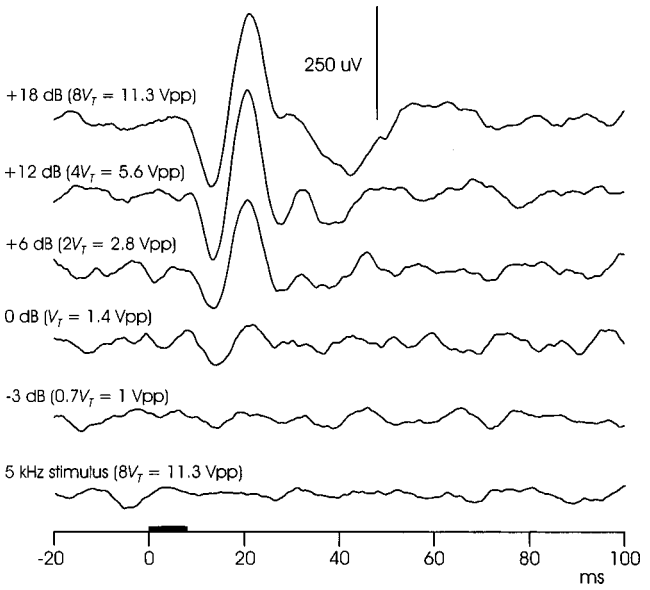


FIG. 1. Vestibular evoked myogenic responses from one subject at levels below, at and above threshold (−3, 0, +6, +12, and +18 dB *re* V_T). The bone conductor was placed on the right mastoid and the traces were recorded from the right sternocleidomastoid muscle. The lowermost trace shows the response in the same subject to a 5-kHz stimulus matched by voltage (peak to peak tone burst amplitude) to the 500-Hz stimulus level at +18 dB [in this case 11.3 V peak to peak (Vpp)].

Stimuli were presented at several intensities from 12 dB below V_T to 18 dB above V_T in 6-dB steps (−12, −6, 0, +6, +12, and +18 dB *re* V_T, equivalent to the following linear measures: $V_T/4$, $V_T/2$, V_T, $2V_T$, $4V_T$, $8V_T$) independently for left and right ears. In the case of subjects with a high threshold, intensity was increased in 6-dB steps until the maximum of 20 V pp was reached. In the case of the avestibular patient, where no VEMP threshold was obtainable, the stimuli were delivered in the same manner relative to an assigned threshold of 2 V pp (the average VEMP threshold). In the Tullio case, which exhibited a very low threshold, the intensity was increased beyond +18 dB until a maximum of 20 V pp. EEG responses were averaged over 1000 repetitions of both 500-Hz and 5-kHz stimuli at matching amplitudes.

III. RESULTS

A. Normals

The ten normal subjects all showed VEMPs in response to the 500-Hz stimuli, with thresholds which varied from 1 to 5.6 V pp (mean=2 V pp) (Fig. 1). These thresholds were equivalent to a mean of 31-dB sensation level (SL). Two subjects were not included in further analysis because their thresholds were too high to allow a full set of observations for the cortical potentials. Of the eight subjects included in

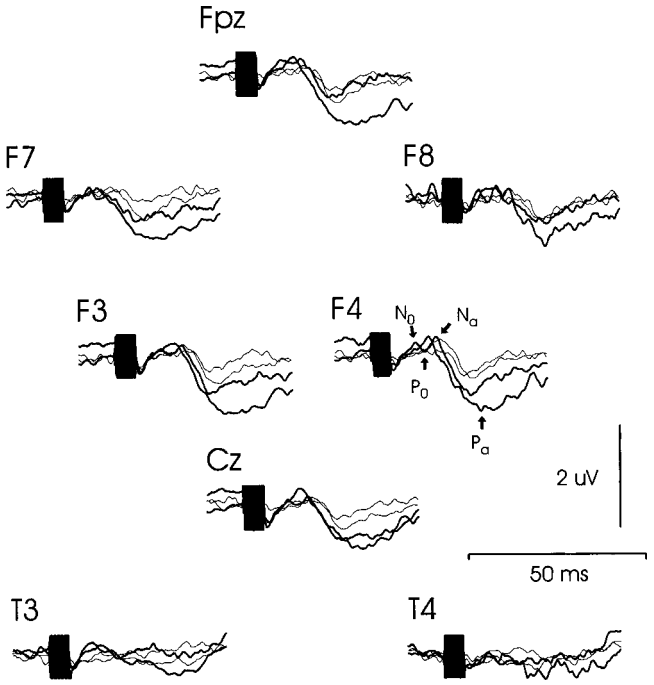


FIG. 2. Grand mean of averaged EEG responses in normal subjects following 5-kHz, 6-ms tone burst stimulation at +18 dB (bold) and −6 dB (thin) *re* V_T. Responses to both left and right ear stimulation are shown. At +18 dB $n=7$ for left and right stimulation; at −6 dB $n=6$ for right and left stimulation. A normal MLR N_0, P_0, N_a, P_a waveform can be seen, which is of a similar shape and magnitude at each electrode. The artifact indicates the time of stimulation, which began 10 ms after the beginning of the sweep. In this and subsequent figures, the size of the stimulus artifact has been limited for clarity.

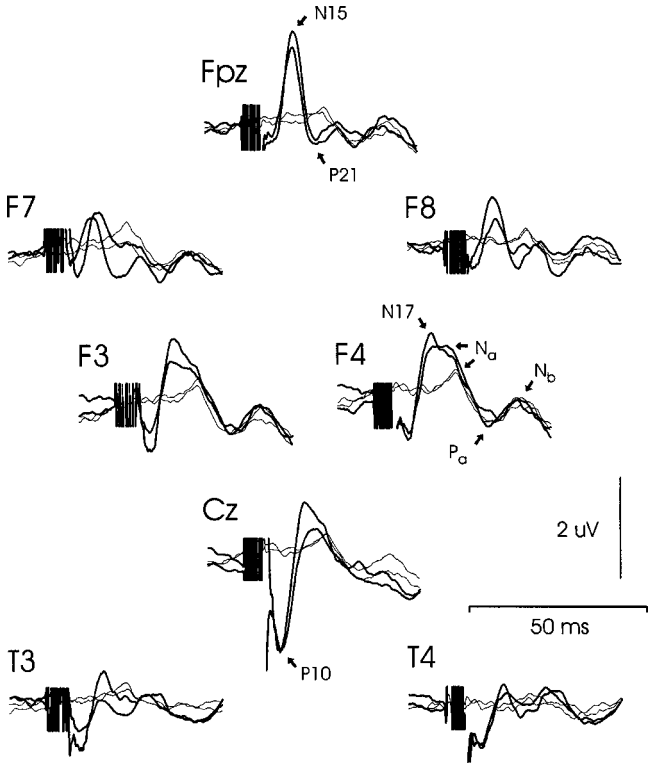


FIG. 3. Grand mean of averaged EEG responses in normal subjects following 500-Hz, 6-ms tone burst stimulation at +18 dB (bold) and −6 dB (thin) *re* V_T. Responses to both left and right ear stimulation are shown. At +18 dB $n=7$ and 8 for right and left stimulation, respectively; at −6 dB $n=8$ for both right and left stimulation. Above V_T (at +18 dB) a series of non-MLR waves appear: a P10, maximal at Cz, an N15 maximal at Fpz, an N17 maximal at F3, F4, and Cz and a P21 at Fpz.

the experiment, we confirmed that seven had no VEMP response on either side to the most intense 5-kHz stimulus. The intensity of the 5-kHz stimulus was matched to the amplitude of the 500-Hz stimulus at 18 dB above threshold. One subject was not available for this part of the experiment.

Properties of the grand average of EEG responses

The averaged EEG response is illustrated in Fig. 2 for the 5 kHz stimulus. Grand averages are for eight subjects, unless otherwise stated. While there was evidence of differences in individual responses, the 5-kHz stimulus at both −6 and +18 dB *re* V_T evoked clear N_a and P_a potentials, which were widely distributed in the centrofrontal electrodes. In particular, the responses in the two midline electrodes, Fpz and Cz, were similar at both intensities. With increasing intensity, the peaks became larger and earlier.

In the case of the 500-Hz stimulus (Fig. 3), a dramatic change in morphology occurs between the two intensities. Below V_T (−6 dB) it was possible to discern an N_a-P_a-N_b pattern, with latencies at Cz of 23.1, 31.7, and 41.0 ms on the left, and 23.4, 31.3, and 41.0 ms on the right. This MLR appears with similar morphology in all electrodes but is maximal over the centrofrontal electrodes, Cz, F3, and F4, and smaller over T3 and T4. Above V_T a quite distinct pattern appears, which contrasts with the MLR. This pattern is characterized by clear differences between the two midline electrodes, such that there is a positivity maximal at Cz and a negativity maximal at Fpz.

At +18 dB *re* V_T the positivity starts at about 7 ms and reaches a peak in Cz at 10.5 and 10.3 ms on left and right, respectively (hereafter referred to as P10, Table I). The positivity is also evident in T3, T4, F3, and F4, but is small or absent more anteriorly in F7, F8, and Fpz. An ANOVA of the amplitude of individual P10 responses at +18 dB indicates a main effect of electrode, $F(4,24)=16.7$, $p<0.001$, and a significant electrode/ear interaction, $F(4,24)=5.9$, $p<0.01$. The latter we interpret as right ear stimulation producing a more sharply centrally focused response than left stimulation. Excluding the midline electrode there is a small ipsilateral advantage of 0.68 vs 0.53 μV, albeit nonsignificant for this number of subjects, $F(1,6)=2.8$, $p=0.16$. There are no significant latency effects across electrodes, $F(4,24)=1.7$ ns.

At +18 dB *re* V_T the negativity at Fpz starts at about 8 ms and reaches a peak at 14.6 and 14.5 ms on the left and right, respectively (also present at F7 and F8 and hereafter referred to as N15, Table II). An ANOVA of the individual N15 amplitudes at +18 dB indicates a significant main effect of electrode, $F(2,12)=41.3$, $p<0.001$, and a significant ear/electrode interaction, $F(2,12)=4.8$, $p<0.05$, which we interpret as indicating a more asymmetrical response to left ear stimulation. Excluding the midline electrode gives a close to significant effect of lateralization, $F(1,6)=4.8$, $p=0.07$. ANOVA of latencies shows no main effect of electrode, $F(2,21)=0.1$, $p=0.8$, but a close to significant effect of lateralization when the midline electrode is excluded, $F(1,6)$

TABLE I. Latencies (above, ms) and amplitudes (below, μV) of the P10 response in the centrofrontal and temporal electrodes. Measurements were made from the grand average EEG trace at three intensities above threshold (V_T) and from individual traces at 18 dB above threshold.

	T3		F3		Cz		F4		T4	
	L	R	L	R	L	R	L	R	L	R
Latency										
Grand average										
6 dB	11.0	13.0	11.2	10.1	11.3	11.8	10.9	11.3	12.1	12.1
12 dB	11.5	11.4	11.2	9.8	10.9	10.8	10.7	10.0	10.9	10.8
18 dB	10.0	9.6	9.7	9.6	10.5	10.3	9.9	9.9	10.3	10.8
Individual (18 dB)										
Mean	10.5	10.9	9.9	10.5	10.6	10.7	10.1	10.5	10.3	10.7
SD	0.86	1.46	0.55	0.78	0.45	0.98	0.41	0.64	1.17	1.18
N	7	6	7	6	8	7	8	7	8	7
Amplitude										
Grand average										
6 dB	0.34	0.14	0.25	0.17	0.64	0.46	0.34	0.33	0.30	0.18
12 dB	0.59	0.45	0.60	0.34	1.22	1.03	0.60	0.56	0.39	0.82
18 dB	1.00	0.64	0.72	0.69	1.65	1.93	0.64	0.90	0.79	0.68
Individual (18 dB)										
Mean	0.75	0.56	0.92	0.43	0.95	1.47	0.55	0.44	0.53	0.57
SD	0.38	0.41	0.51	0.28	0.32	0.47	0.27	0.30	0.28	0.35
N	8	7	8	7	8	7	8	7	8	7

$=5.7$, $p=0.054$, indicating a trend for the N15 to begin earlier on the side contralateral to the stimulus (14.4 vs 13.9 ms).

A negative peak also occurs in the more posterior leads with the more intense 500-Hz stimulus, with latencies which cluster around a longer value of about 17 ms. At F3 and F4 a negative peak occurs with latencies 16.4 and 16.5 ms after left ear stimulation and 15.5 and 16.9 ms after right ear stimulation. At Cz a negative peak occurs at 17.2 and 19.6 ms after left and right ear stimulation, respectively, with amplitudes of 1.3 and 0.5 μV. At T3 and T4 the negative peak occurs at 16.8 and 16.5 ms after left and at 15.4 and 17.1 ms after right ear stimulation. Thus in terms of latency there are possibly two sources which produce a negativity, the first with a mean latency of 14.4 ms, the N15, distributed over prefrontal and lateral frontal electrodes, and a second with a mean latency of 16.8 ms (hereafter N17). N17 is distributed over the centrofrontal and temporal electrodes, but is maximal over the centrofrontal electrodes. At $+18$ dB $re\ V_T$ a second non-MLR positivity also occurs in F7, Fpz and F8 at 21.7, 21.1, and 21.3 ms, respectively, to left ear stimulation and 21.0, 21.4 and 19.7 ms, respectively, to right ear stimulation (hereafter P21).

TABLE II. Latencies (ms) and amplitudes (μV) of the N15 response in the frontal electrodes. Measurements were made from the grand average EEG trace at three intensities above threshold (V_T) and from individual traces at 18 dB above threshold.

	F7		Fpz		F8	
	L	R	L	R	L	R
Latency						
Grand average						
6 dB	16.7	13.9	15.9	15.5	16.4	18.6
12 dB	16.0	14.5	15.6	14.8	14.7	15.6
18 dB	15.4	13.2	14.6	14.5	13.8	14.6
Individual (18 dB)						
Mean	14.4	13.8	14.6	14.5	14.2	14.6
SD	1.21	1.13	1.01	0.71	1.28	0.85
N	8	7	8	7	8	7
Amplitude						
Grand average						
6 dB	-0.31	-0.20	-0.44	-0.50	-0.33	-0.38
12 dB	-0.54	-0.52	-0.93	-1.19	-0.45	-0.50
18 dB	-0.83	-0.71	-1.96	-1.52	-1.07	-0.54
Individual (18 dB)						
Mean	-0.94	-0.86	-2.22	-1.64	-1.34	-0.97
SD	0.78	0.54	1.04	0.73	0.61	0.58
N	8	7	8	7	8	7

Although the non-MLR waves are predominant at the highest intensity, the auditory N_a is still apparent as a shoulder at 22–23 ms following the negativity at 17 ms, especially in F3, F4, and Cz. At the lowest intensity the N_a is clear in all frontal electrodes, but is gradually suppressed in Fpz, F7, and F8 by the P21. A regression of amplitude versus stimulus intensity was performed using individual measurements from the eight subjects (for P10 the amplitude at Cz was averaged over left and right stimulation; for N15 the amplitude at Fpz was averaged over left and right stimulation). The regression gave thresholds for P10 of -3.7 dB ($R^2=0.46$) and for N15 of $+2.7$ dB ($R^2=0.45$), both close to our estimate of the vestibular threshold (i.e., 0 dB$=V_T$).

Figure 4 shows a direct comparison between the effects of the same amplitude stimulation, equivalent to 18 dB above V_T, for 500-Hz and 5-kHz stimulation. The responses to the 500-Hz stimulus are larger and clearly differ from those to 5 kHz at most electrodes, particularly at Fpz and Cz.

B. Hearing-impaired case

This subject was profoundly deaf in the right ear with a flat loss of about 90 dB. However, he exhibited a normal

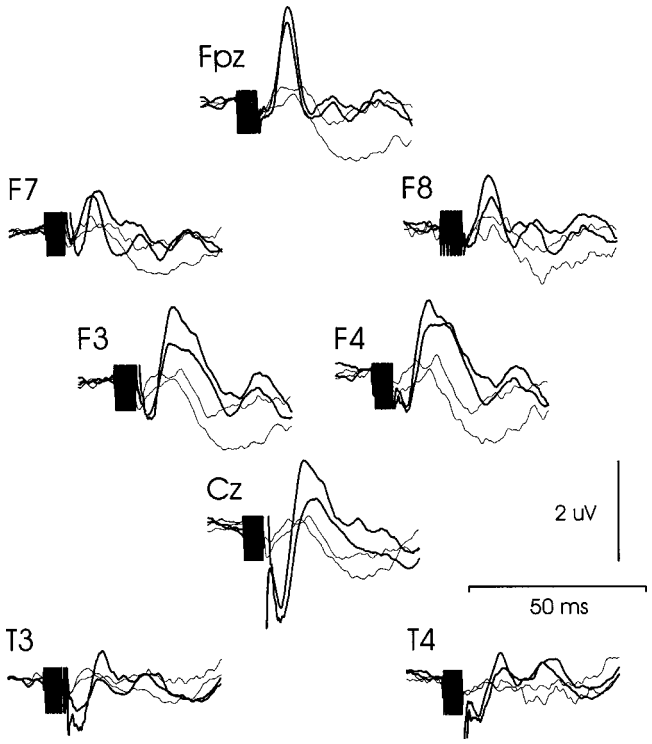


FIG. 4. Grand mean of averaged EEG responses in normal subjects following 500-Hz (bold) and 5-kHz (thin), 6-ms tone burst stimulation at +18 dB *re* V_T. Responses to both left and right ear stimulation are shown. At 500 Hz $n=7$ and 8 for right and left stimulation, respectively; at 5 kHz $n=7$ for both right and left stimulation. The figure demonstrates the marked difference in the morphology and distribution of the responses following stimulation matched for amplitude but differing in frequency (i.e., 500 Hz versus 5 kHz).

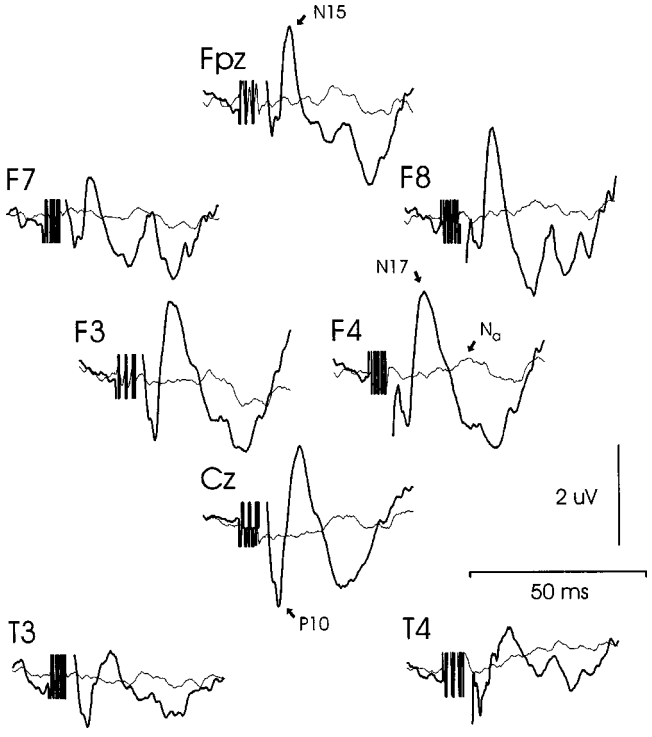


FIG. 5. Averaged EEG response in the right ear of a hearing-impaired subject following 500-Hz, 6-ms tone burst stimulation at +18 dB (bold) and −6 dB (thin) *re* V_T. Below V_T there is a probable MLR late N_a at 27 ms, but there is no evidence of non-MLR waves. Above V_T the non-MLR waves, P10, N15, and N17, are present.

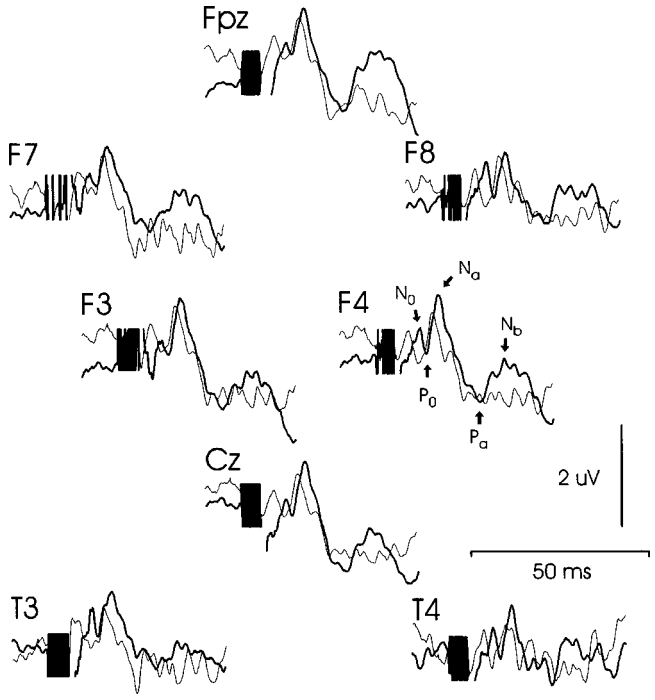


FIG. 6. Averaged EEG response in the left ear of a hypovestibular subject following 500-Hz (bold) and 5-kHz (thin), 6-ms tone burst stimulation at 16 V pp (near maximum stimulus). Responses to the amplitude matched stimuli are similar for both frequencies (500 Hz and 5 kHz) and at each electrode site. A MLR N_0, P_0, N_a, P_a pattern can be discerned in both cases, but at a greater latency following 500-Hz than 5-kHz stimulation.

VEMP indicative of an intact vestibular system. Averaged EEG responses to the 500-Hz stimulus at +18 dB and at −6 dB *re* V_T are shown in Fig. 5.

At −6 dB *re* V_T there is a probable N_a, at about 27 ms, and at higher intensities other MLR are present, presumably mediated by transmission across the skull to the intact cochlea. At +18 dB the P10 at Cz has a peak latency of 11.3 ms with an amplitude of 1.8 μV, values similar to the normal subjects. The N15 at Fpz has a peak latency of 14.5 ms and an amplitude of 1.5 μV, again similar to normals. In F3, F4, and Cz the negativity arrives at 15.8, 16.2, and 17.5 ms, with amplitudes of 1.9, 2.0, and 1.5 μV respectively. An auditory Na is evident at +18 dB as a shoulder after the N17 in F3, F4, and Cz.

C. Hypovestibular case

This subject did not have a VEMP response to the 500-Hz or 5-kHz stimulus in the left ear even at the maximum stimulus intensity of 20 V pp, consistent with a lack of vestibular function in the left ear. Further, this subject was reported in a previous study to have a marked reduction of caloric responses bilaterally and decreased vestibuloocular reflex (VOR) in the planes of the semicircular canals bilaterally (Welgampola *et al.*, 2003). In contrast the subject had normal hearing. Figure 6 shows the averaged EEG response across the eight electrodes for both 500-Hz and 5-kHz stimuli at 16 V pp (18 dB above the mean V_T). The responses at both frequencies are similar and, in particular, the midline electrode responses at each frequency are almost identical, other than a phase delay between the two stimuli.

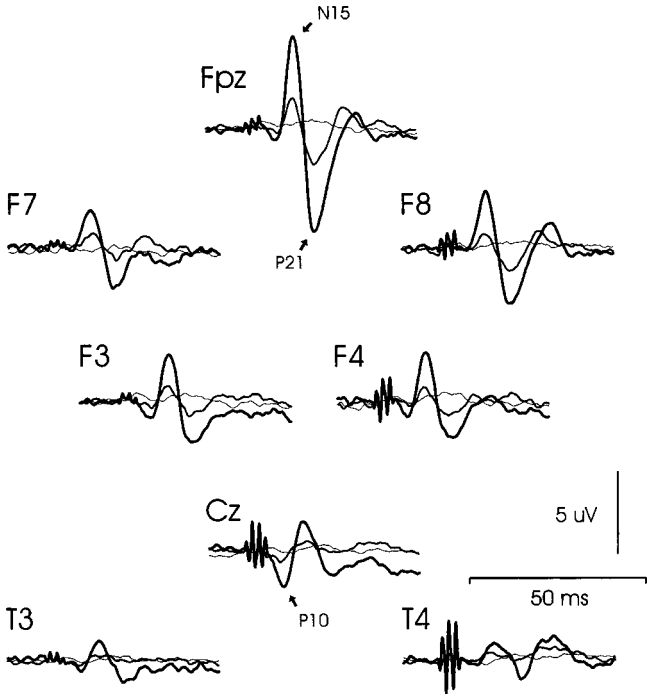


FIG. 7. Averaged EEG response in the right ear of a hypervestibular subject following 500-Hz, 6-ms tone burst stimulation at +30 dB (dark bold), +18 dB (bold), and −6 dB (thin) $re\ V_T$. Below V_T there are no non-MLR components; at +18 dB a P10, N15, P21 non-MLR waveform is present, which is much larger at +30 dB and shows increasing dominance by the N15/P21 component in the frontal electrodes. Note the different amplitude scale from other figures.

The 5-kHz, 16-V pp stimulus evokes a MLR N_0, P_0, N_a, P_a pattern. In F3, F4, and Cz the N_a has a latency of 16.1, 16.3, and 16.1 ms, respectively. The 500-Hz stimulus also evokes an auditory MLR N_0, P_0, N_a, P_a pattern. At F3, F4, and Cz the latencies of N_a are 17.4, 17.6, and 17.8 ms. As with the 5-kHz stimulus, this is somewhat earlier than the classical N_a. With both 5-kHz and 500-Hz stimuli it is clear, however, that there is no sign of the P10, N15, or P21 responses.

D. Hypervestibular case

This subject with superior canal dehiscence and the Tullio phenomenon had a low VEMP threshold of 0.125 V pp in the right ear, approximately 20 dB below the average V_T for normal subjects. This low threshold meant that correspondingly more intense levels of stimulation relative to V_T could be achieved. Hearing levels were within the normal range.

At higher intensities large N15 and P10 are present in the two midline electrodes (e.g., +18 and +30 dB $re\ V_T$, illustrated in Fig. 7). At +18 dB the dominance of the prefrontal/lateral frontal N15/P21 is complete. The N15, which has a latency of 14.7 ms at Fpz, appears at F3 and F4, with latencies 14.8 and 14.7 ms. At Fpz the N15 has an amplitude of 1.8 μV while the P21 has an amplitude of 2.2 μV. N17 and N_a are present at Cz with latencies of 18.1 and 22.2 ms, and amplitudes 0.5 and 0.1 μV.

Given the abnormally low VEMP threshold of this subject it was possible to examine EPs at intensities up to +42 dB $re\ V_T$. At +30, +36, and +42 dB the amplitudes of the N15 at Fpz were respectively 6.0, 9.7, and 11.8 μV, with a constant latency of 14.8 ms. At +30, +36, and +42 dB the N_a at Cz is completely swamped by a single negativity with latencies of 16.5, 16.4, and 16.0 ms and amplitudes of 1.6, 4.2, and 6.0 μV. The latency of the negativity at Cz, between 15 and 17 ms, suggests that it may be a composite of the prefrontal/lateral frontal N15 and centrofrontal N17.

The P10 appeared in Cz in this subject at +6 dB $re\ V_T$ with a latency of 11.6 ms. At +12 dB the P10 was present in all electrodes, other than F7, Fpz, and F8, and maximal at Cz where the latency was 10.7 ms and amplitude 0.4 μV. At +18 dB the latency at Cz is 10.3 ms with an amplitude of 0.7 μV (Fig. 7). At the higher intensities of +30, +36, and +42 dB the P10 grew rapidly but reaches a plateau with amplitudes of 2.4, 4.3, and 4.7 μV and a constant latency of 11.0 ms. In the other electrodes in which the P10 appeared there is a slight ipsilateral advantage. Thus in T3 and T4 the amplitudes in response to right stimulation are 0.8 and 0.5 μV at +30 dB, 0.8 and 1.7 μV at +36 dB, and 1.3 and 1.8 μV at +42 dB.

IV. DISCUSSION

The results may be summarized as follows. In the normals below 0 dB (V_T) in both the 500-Hz case and the 5-kHz case normal auditory MLR waves were observed, consisting of the usual N_0, P_0, N_a, P_a, N_b pattern, though the latencies in the 500-Hz case were some 2–4 ms longer than the 5-kHz case. The N_a arrived at about 18–19 ms for 5 kHz and 22–23 ms for 500 Hz. The MLR waves were maximally distributed over the centrofrontal electrodes and the morphology of waves recorded at midline electrodes was similar.

At and above 0 dB $re\ V_T$ with the 500-Hz stimulus, a number of other non-MLR waves appeared. These were a P10 which was maximal at Cz and centrally distributed, an N15 which was maximal at Fpz and was distributed over prefrontal/lateral frontal electrodes, an N17 which was centrofrontally distributed, and a P21 which was also distributed over the prefrontal/lateral frontal electrodes. The P10 and N15 had linear growth functions indicative of a threshold near 0 dB. The non-MLR waves were present in the hearing impaired subject but absent in the patient with severely impaired vestibular function. In the hypervestibular patient in the 500-Hz condition, all non-MLR waves were present but the N15/P21 in particular became increasingly dominant with increasing intensity.

A. Likely origin of non-MLR waves

The fact that the non-MLR waves only appeared in circumstances when the vestibular system was known to be acoustically activated and did not appear in the avestibular case indicates that they are vestibular, possibly saccular, in origin. We therefore refer to the non-MLR waves hereafter as short latency vestibular evoked potentials (VsEP). The question is raised therefore of what neurogenic, myogenic or other sources upstream of the vestibular end-organs may be responsible for the measured potentials.

The fact that the P10 is maximal at the vertex and has an uneven spatio-temporal distribution suggests that it is most likely of cortical origin. The vertex maximum of the P10 could be the result of the summation of bilateral dipole sources, possibly from previously demonstrated vestibular projections to temporal, temporoparietal, or insular cortices (Paulesu *et al.*, 1997). The small ipsilateral advantage may be due to a slightly larger ipsilateral source, as suggested by de Waele *et al.* (2001) and the asymmetry due to the right dipole being more vertically oriented. Alternatively, a single dipole source located near the midline could also explain the vertex maximum. de Waele *et al.* (2001) also reported that their third dipole was located at the anterior end of the supplementary motor area (SMA) and this, or a cingulate source, could be responsible.

In the case of the N15, which is symmetrically distributed over the prefrontal/lateral frontal electrodes, an ocular source cannot be ruled out. Halmagyi *et al.* (2003), who also studied patients with canal dehiscence and low click thresholds, demonstrated short latency vertical ocular movements induced by clicks. The view that the N15 is ocular in origin is supported by the fact that the hypervestibular patient exhibited such large potentials, as much as 10 μV, and complained of oscillopsia at the higher intensities. However, de Waele *et al.* (2001), who also found a prefrontal focus, felt an ocular origin was excluded because the potentials were preserved despite paralysing the patients. The P21 which follows the N15 is most likely from a similar origin since it had a very similar electrode distribution.

The origin of the N17 is not obvious. It has a very similar centrofrontal focus to the auditory N_a, consistent with earlier work by Picton *et al.* (1974), but is not apparent below 0 dB *re* V_T although it had an extrapolated threshold of about -10 dB. One possibility might be that the N17 represents a saccular input to the auditory pathway, which is plausible given that there are anatomical studies which indicate a saccular projection to the cochlear nucleus (e.g., Burian and Gstoettner, 1988) and it is known that there is vestibular input to the medial geniculate nucleus (e.g., Blum *et al.*, 1979). This might explain the similar scalp distribution to and apparent interaction with the N_a.

Both the P10 and N15 show some similarities with the short latency VsEPs reported by de Waele *et al.* (2001). Our latencies are also consistent with a trisynaptic vestibular projection. The relatively large frontal distribution of the N15 is similar to the prefrontal peak shown in Fig. 2 of de Waele *et al.* However, there are also important differences. In particular, the P10 and N15 waves shown here do not have the phase reversals which are apparent, for example, between F7 and F3 or between Cz and T4. Further, the N15 does not show the lateralization exhibited in the prefrontal potential of de Waele *et al.* There are several reasons why the VsEPs reported here might differ from those of de Waele *et al.* (2001). The first is that acoustic stimulation used here most likely only activates otolithic receptors whereas direct galvanic stimulation is likely to activate most vestibular afferent species including canal receptors. Canal projections are likely to differ to a certain extent from otolithic projections and any EOG manifestation of the VOR may be different with greater horizontal components. It is also possible that the linked earlobe reference may reduce any apparent lateralization compared to that of a nose reference.

B. Implications for acoustic perception of loud sounds and evolutionary significance

One striking result of these experiments is that the threshold difference between the auditory MLR and the short latency VsEP at 500 Hz is of the order of 30 dB. This is almost identical to the theoretical estimate of Welgampola *et al.* (2003) who, using RETFL levels for the perception of continuous sound, calculated that their VEMP threshold for bone conduction would be equivalent to 30.5 dB HL at 250 Hz. There can be no doubt that vestibular acoustic sensitivity for bone-conducted sound is remarkable. It is plausible therefore that perception of low frequency sound via the sacculus may explain the 6–10-dB advantage of bone-conducted sound over air conducted sound in loudness matching experiments (Stenfelt and Håkansson, 2002). Given that the air–bone gap (Berger, 1983) is about 40 dB, then it is also plausible that the vestibular system may be activated in a free-field environment by intensities as low as 70 dB HL, especially for very low frequencies ($<$500 Hz), below the skull antiresonance.

In Todd (2001) evidence was provided that saccular acoustic sensitivity may contribute to the quality of loud sound and especially loud low-frequency sounds. In many loud music environments, particularly dance club environments, it is the very low frequencies which seem to be most rewarding and which are actively sought. These observations are consistent with the excellent sensitivity of the sacculus to low-frequency vibrations. In loud music environments substrate-borne sound through the floor is also likely to make a contribution to vestibular acoustic activation.

That the effective stimulus for the sacculus appears to be substrate-borne rather than air-borne sound is consistent with the evolutionary history of the ear (Lewis *et al.*, 1985). In fish sound is transmitted from water to the ear via the body which has approximately the same density as water (Popper *et al.*, 1982). In some species of anuran amphibians substrate-borne sound is transmitted to the ear by special adaptation of the front legs (Lewis and Narins, 1985). In the case of water living reptiles, particularly the crocodilians, which have loud low-frequency vocalizations, the sacculus is likely to play a role in the detection of such vocalizations. Some mammals, such as elephants (Reuter *et al.*, 1998), also show adaptations for the detection of substrate-borne sound. In all cases a primary function of substrate-borne acoustic communication is reproductive behavior. Thus the good vestibular sensitivity to and reward function of vibrational sound is also consistent with the conservation of a primitive central acoustic pathway in vertebrates (McCormick, 1992). Such a pathway could plausibly be a trisynaptic projection via the parabrachial nucleus (Balaban *et al.*, 2002), the mammalian homologue of the fish and amphibian second isthmal nucleus (ten Donkelaar, 1998), to the prosencephalic centers underlying reward and reproductive behavior, including the orbitofrontal cortex.

V. SUMMARY AND CONCLUSION

In this paper we have shown that short latency vestibular evoked potentials (VsEPs) can be measured using conventional surface EEG recordings with a noninvasive stimulus. These VsEPs consist of a number of components: a P10 which is maximal at the vertex, and which might be of vestibular cortex origin; an N15/P21 which is maximal at prefrontal electrodes and may be a manifestation of VOR-induced ocular displacement; and an N17 which has a centrofrontal distribution and may represent a saccular-auditory cross-over. These VsEPs show some similarities to VsEPs produced by direct vestibular nerve stimulation, most notably in latency, but differ in other respects, including spatiotemporal distribution. Although further investigations are required, the short latency VsEP may prove to be a useful clinical test of vestibular function. Currently the VEMP procedure has become popular as a simple test of otolith function. However, since the VEMP requires some background contraction it is difficult to use with certain kinds of patients, including neonates, young children, and the infirm, because it requires sustained neck muscle contraction. The VsEPs, by contrast, require relaxation. In addition, the stimulus is innocuous and the procedure tests the integrity of upstream vestibular projections, including to the cortex.

ACKNOWLEDGMENTS

This research is supported by a grant from the National Health and Medical Research Council of Australia and a travel award from the UK Wellcome Trust.

Balaban, C. D., McGee, D. M., Zhou, J., and Scudder, C. A. (**2002**). ''Reponses of primate caudal parabrachial nucleus and Kölliker-Fuse nucleus neurons to whole body rotation,'' J. Neurophysiol. **88**, 3175–3193.

Benson, A. J. (**1982**). ''The vestibular sensory system,'' in *The Senses*, edited by H. B. Barlow and J. D. Mollon (Cambridge U.P., Cambridge), pp. 333–368.

Berger, E. H. (**1983**). ''Laboratory attenuation of earmuffs and earplugs, both singly and in combination,'' J. Am. Ind. Hyg. Assoc. **44**, 321–329.

Blum, P. S., Day, M. J., Carpenter, M. B., and Gilman, S. (**1979**). ''Thalamic components of the ascending vestibular system,'' Exp. Neurol. **64**, 587–603.

Burian, M., and Gstoettner, W. (**1988**). ''Projection of primary vestibular afferent fibres to the cochlear nucleus in the guinea pig,'' Neurosci. Lett. **84**, 13–17.

Colebatch, J. G., Halmagyi, G. M., and Skuse, N. F. (**1994**). ''Myogenic potentials generated by a click-evoked vestibulocollic reflex,'' J. Neurol., Neurosurg. Psychiatry **57**, 190–197.

Colebatch, J. G., Day, B. L., Bronstein, A. M., Davies, R. A., Gresty, M. A., Luxon, L. M., and Rothwell, J. C. (**1998**). ''Vestibular hypersensitivity to clicks is characteristic of the Tullio phenomenon,'' J. Neurol., Neurosurg. Psychiatry **65**, 670–678.

de Waele, C., Baudonnière, P. M., Lepecq, J. C., Tran Ba Huy, P., and Vidal, P. P. (**2001**). ''Vestibular projections in the human cortex,'' Exp. Brain Res. **141**, 541–551.

Halmagyi, G. M., McGarvie, L. A., Aw, S. T., Yavor, R. A., and Todd, M. J. (**2003**). ''The click-evoked vestibulo-ocular reflex in superior semicircular canal dehiscence,'' Neurology **60**, 1172–1175.

Hood, J. D., and Kayan, A. (**1985**). ''Observations upon the evoked responses to natural vestibular stimulation,'' Electroencephalogr. Clin. Neurophysiol. **62**, 266–276.

Lewis, E. R., and Narins, P. M. (**1985**). ''Do frogs communicate with seismic signals?'' Science **227**, 187–189.

Lewis, E. R., Leverenz, E., and Bialek, W. (**1985**). *The Vertebrate Inner Ear* (CRC, Boca Raton).

McCormick, C. A. (**1992**). ''Evolution of central auditory pathways in anamniotes,'' in *The Evolutionary Biology of Hearing*, edited by D. B. Webster, R. R. Fay, and A. N. Popper (Springer-Verlag, New York), pp. 323–350.

Paulesu, E., Frackowiak, R. S. J., and Bottini, G. (**1997**). ''Brain mapping studies of the vestibular system and its interactions with mental representations of the body,'' in *Human Brain Function*, edited by R. Frackowiak, K. Friston, C. Frith, R. Dolan, and J. Mazziotta (Academic, San Diego), pp. 216–230.

Picton, T. W., Hillyard, S. A., Krausz, H. I., and Galambos, R. (**1974**). ''Human auditory evoked potentials. I: Evaluation of components,'' Electroencephalogr. Clin. Neurophysiol. **36**, 179–190.

Popper, A. N., Platt, C., and Saidel, W. M. (**1982**). ''Acoustic functions in the fish ear,'' Trends Neurosci. **5**, 276–280.

Reuter, T., Nummela, S., and Hemilä, S. (**1998**). ''Elephant hearing,'' J. Acoust. Soc. Am. **104**, 1122–1123.

Sheykholeslami, K., Kermany, M. H., and Kaga, K. (**2001**). ''Frequency sensitivity range of the saccule to bone-conducted stimuli measured by vestibular evoked myogenic potentials,'' Hear. Res. **160**, 58–62.

Sohmer, H., Elidan, J., Rodionov, V., and Plotnik, M. (**1999**). ''Short and middle latency vestibular evoked potentials to angular and linear acceleration,'' Electroencephalogr Clin. Neurophysiol. Suppl. **50**, 226–234.

Stenfelt, S., and Håkansson, B. (**2002**). ''Air versus bone conduction: an equal loudness investigation,'' Hear. Res. **167**, 1–12.

Stenfelt, S., Håkansson, B., and Tjellström, A. (**2000**). ''Vibration characteristics of bone conducted sound *in vitro*,'' J. Acoust. Soc. Am. **107**, 422–443.

Stenfelt, S., Hato, N., and Goode, R. L. (**2002**). ''Factors contributing to bone conduction: The middle ear,'' J. Acoust. Soc. Am. **111**, 947–959.

Stenfelt, S., Wild, T., Hato, N., and Goode, R. L. (**2003**). ''Factors contributing to bone conduction: The outer ear,'' J. Acoust. Soc. Am. **113**, 902–913.

ten Donkelaar, H. J. (**1998**). ''Anurans,'' in *The Central Nervous System of Vertebrates, Vol. 2*, edited by R. Nieuwenhuys, H. J. ten Donkelaar, and C. Nicholson (Springer, New York), pp. 1151–1314.

Todd, N. P. M. (**2001**). ''Evidence for a behavioral significance of saccular acoustic sensitivity in humans,'' J. Acoust. Soc. Am. **110**, 380–390.

Todd, N. P. M., and Cody, F. W. (**2002**). ''Adaptation in saccular acoustical responses,'' Int. J. Audiol. **41**, 236.

Todd, N. P. M., Cody, F. W. J., and Banks, J. R. (**2000**). ''A saccular origin of frequency tuning in myogenic vestibular evoked potentials?: implications for human responses to loud sounds,'' Hear. Res. **141**, 180–188.

Tonndorf, J. (**1966**). ''Bone conduction. Studies in experimental animals,'' Acta Oto-Laryngol., Suppl. **213**, 1–132.

Welgampola, M. S., Rosengren, S. M., Halmagyi, G. M., and Colebatch, J. G. (**2003**). ''Vestibular activation by bone-conducted sound,'' J. Neurol., Neurosurg. Psychiatry **74**, 771–778.

Psychometric functions for informational masking

Robert A. Lutfi,[a] Doris J. Kistler,[b] Michael R. Callahan,[c] and Frederic L. Wightman[b]
Waisman Center, University of Wisconsin, Madison, Wisconsin 53705

(Received 3 January 2003; revised 24 September 2003; accepted 3 October 2003)

The term *informational masking* has traditionally been used to refer to elevations in signal threshold resulting from masker uncertainty. In the present study, the method of constant stimuli was used to obtain complete psychometric functions (PFs) from 44 normal-hearing listeners in conditions known to produce varying amounts of informational masking. The listener's task was to detect a pure-tone signal in the presence of a broadband noise masker (low masker uncertainty) and in the presence of multitone maskers with frequencies and amplitudes that varied at random from one presentation to the next (high masker uncertainty). Relative to the broadband noise condition, significant reductions were observed in both the slope and the upper asymptote of the PF for multitone maskers producing large amounts of informational masking. Slope was affected more for some listeners and conditions while asymptote was affected more for others; consequently, neither parameter alone was highly predictive of individual thresholds or the amount of informational masking. Mean slopes and asymptotes varied nonmonotonically with the number of masker components in a manner similar to mean thresholds, particularly when the estimated effect of energetic masking on thresholds was subtracted out. As in past studies, the threshold data were well described by a model in which trial-by-trial judgments are based on a weighted sum of levels in dB at the output of independent auditory filters. The psychometric data, however, complicated the model's interpretation in two ways: First, they suggested that, depending on the listener and condition, the weights can either reflect a fixed influence of masker components on each trial or the effect of occasionally mistaking a masker component for the signal from trial to trial. Second, they indicated that in either case the variance of the underlying decision variable as estimated from PF slope is not by itself great enough to account for the observed changes in informational masking.
 [DOI: 10.1121/1.1629303]

PACS numbers: 43.66.Ba, 43.66.Fe [MRL] Pages: 3273–3282

I. INTRODUCTION

In recent years there has been a growing interest in the detection of signals in uncertain noise backgrounds. The interest is largely motivated by a desire to understand detection in real-world settings where background noise varies unexpectedly from one moment to the next. Typically in studies, the effect of noise uncertainty has been expressed as an elevation in threshold for the signal relative to a ''minimal uncertainty'' condition in which the noise is either fixed or varies little from one presentation to the next. The term *informational masking* was first used in this context to describe the elevations in threshold resulting from masker uncertainty (Pollack, 1975; Watson and Kelly, 1981; Lutfi, 1990). Because informational masking has been commonly associated with the effect of attentional processes on threshold, it is often distinguished from *energetic masking*, which is attributed to the competition of signal and noise at the auditory periphery (i.e., cochlea or auditory nerve). Some authors, in fact, have argued that the term informational masking might be better used to refer to any masking that is not energetic (Durlach *et al.*, 2003). In this paper we use the term according to its original definition inasmuch as the focus is on conditions of masker uncertainty. However, the two definitions are not mutually exclusive, and the actual amount of informational masking we report in different conditions of this study is an estimate of the amount of masking that is not energetic (cf. Oh and Lutfi, 1998).

What is currently known about informational masking comes almost exclusively from results of adaptive psychophysical procedures that estimate signal strength corresponding to a single level of performance. Much has been learned from these data. However, for understanding the nature of the decision process it is more useful to know how performance varies over a range of signal levels, as given by the psychometric function (PF). Two features of the PF are of particular interest: its slope and its upper asymptote. The slope or spread of the PF reflects variability of the underlying decision process. Values of slope can, therefore, be used to provide strong tests of models that assume a particular form for the decision variable. Asymptotic performance is also of interest in cases wherein the listener's attention might stray from the psychophysical task, or wherein the noise may simply be ''confused'' for the signal on some number of trials. In such cases, asymptotic performance provides a measure of lapse rate or frequency of confusions. Lapses and

[a]Also at Department of Communicative Disorders, University of Wisconsin—Madison.
[b]Current address: Dept. of Psychological and Brain Sciences, University of Louisville, Life Sciences Bldg., Louisville, KY 40292.
[c]Current address: Epic Systems Corp. 5301 Tokay Blvd., Madison, WI 53711.

confusions also affect PF slope to the extent that they add variance to the underlying decision variable (cf. Green, 1995).

Both parameters of the PF are relevant to the understanding of informational masking. To date, however, no clear picture has emerged regarding the effects of masker uncertainty on these parameters. Perhaps the most extensive data come from a study by Allen and Wightman (1995). These investigators obtained PFs from 17 children (3–4 years of age) and 13 adults in a task involving the detection of a fixed-frequency, pure-tone signal in the presence of noise and a random-frequency, pure-tone masker. They used a logistic function to fit PFs to the data of individual listeners and reported only the results of the fits. On average, masker uncertainty appeared to have the effect of decreasing the slope of the PF, though the results were variable; many listeners actually showed shallower PFs in the absence of the masker. One problem with this study was that the fitted functions were forced to have an upper asymptote of 100 percent. Hence, the estimates of PF slope were to some extent confounded with asymptotic performance. Other tone-detection studies reporting PFs have used a random-frequency multitone masker instead of a pure tone; the results of these studies are mixed. Neff and Callaghan (1987), using simultaneous masker–tone complexes, and Watson and Kelly (1981), using sequential masker–tone complexes (see their Fig. 3.6), report a systematic decrease in the slope of the PF with increasing masker uncertainty. Tang and Richards (2002) and Wright and Saberi (1999), however, report significant individual differences in the effect of masker uncertainty on slope for simultaneous-masker complexes. These studies used many fewer listeners and also did not estimate upper asymptotic performance in conjunction with slope. Moreover, the latter two studies used adaptive procedures that can potentially lead to bias in estimates of slope (Leek, 2001). Finally, results from discrimination and identification studies using nonadaptive procedures have generally reported a reduction in PF slope with increasing masker uncertainty (Arbogast, Mason, and Kidd, 2002; Kidd *et al.*, 1994; Kidd, Mason, and Arbogast, 2002), although here too there are some discrepancies (Watson and Kelly, 1981; see their Fig. 3.3). To our knowledge, no study to date has reported systematic estimates of upper asymptotic performance in conditions of informational masking.

The goal of the present study was twofold: to provide more extensive documentation than past studies of the effects of masker uncertainty on the PF, and to use measures of slope and asymptote to test two interpretations of informational masking. To this end, PFs from 44 normal-hearing listeners were measured in a minimal-uncertainty condition and in comparable conditions known to produce high levels of informational masking. The task of the listener was to detect a pure-tone signal in the presence of a broadband noise masker (minimal uncertainty condition) and in the presence of a multitone masker with frequencies that varied at random from one presentation to the next. Broadband noise is commonly identified as a ''minimal uncertainty'' masker in such cases, as the spectra of individual samples appear qualitatively more similar to one another than do the spectra of the individual multitone maskers (Neff and Green, 1987; and for a quantitative account see Lutfi, 1990; and Oh and Lutfi, 1998).

With regard to the test of interpretations, the data are analyzed within the context of a model that successfully describes thresholds in these conditions (Lutfi, 1993; Oh and Lutfi, 1998). According to the model, listeners are assumed to make trial-by-trial judgments based on the weighted sum of levels (in dB) at the outputs of independent auditory filters. Here, the weights are intended to reflect the role of attentional factors that are believed to play a major role in informational masking. They have, however, at least two different interpretations. The first is that they represent a constant proportional contribution of each auditory filter to the decision variable on each trial. We refer to this as the fixed-weight hypothesis. Fixed-weight models have been used successfully to predict threshold in a variety of conditions of informational masking (Lutfi, 1992, 1993; Oh and Lutfi, 1998; Wright and Saberi, 1999), as well as to draw inferences regarding internal noise in these conditions (Berg, 1990; Richards, Tang, and Kidd, 2002; Tang and Richards, 2002). These models are the most common in the literature on informational masking. An equally viable interpretation, however, is that the weights represent the exclusive contribution of a single but different auditory filter on each trial. We refer to this as the variable-weight hypothesis. In effect, the variable-weight hypothesis proposes that a nonsignal filter is mistaken for the signal filter on some proportion of trials. Lutfi (1992, 1993) has analyzed the predictions of this hypothesis, treating the weights as a vector of random binomial variables rather than as a vector of constants.[1] Evidence in support of the hypothesis comes from the observation that high levels of informational masking often occur in conditions where the signal and masker are perceptually similar to one another and so are more likely to be confused (Arbogast *et al.*, 2002; Brungart, 2001; Kidd *et al.*, 2001; Kidd *et al.*, 1994; Kidd *et al.*, 2002; Neff, 1995). Particularly relevant in this regard is the report of plateaus in the PF (asymptotic-like behavior) where signal and masker components are of near-equal level (Brungart, 2001).

The fixed- and variable-weight hypotheses make similar predictions for signal thresholds; however, they make different predictions for the psychometric function. For fixed weights the random-frequency multitone masker must produce greater variance in the decision variable than does the broadband noise. For the broadband noise the output of any auditory filter will change little from one presentation to the next, whereas for the multitone masker it will vary greatly depending on whether or not a tone falls in the bandpass of the filter. The fixed-weight hypothesis, therefore, predicts that the slopes of the PFs for multitone maskers will be much shallower than those for the broadband noise. The fixed-weight hypothesis also makes a strong prediction for how the slopes of the PFs will change with the number of tones in the multitone masker. If the range of potential frequencies is fixed, there should be some critical number of masker components for which the probability is 0.5 that one or more components will fall within a given filter. At this point the variance of filter outputs with fixed weights will approach a

maximum and the slope of the PF will have its lowest value. For greater or fewer masker components the variance of filter outputs will be less and the slopes of the PF will increase. The fixed-weight hypothesis, therefore, predicts a nonmonotonic relation between the slope of the PF and the number of masker components similar to what is seen in the threshold data (Neff and Green, 1987; Oh and Lutfi, 1998).[2] The variable-weight hypothesis also predicts some effect on slope, but differs fundamentally from the fixed-weight hypothesis in that the predominant effect is predicted to be a change in upper asymptote. Because the outputs of nonsignal auditory filters are assumed to be independent of the output of the signal filter, mistaking a nonsignal filter for a signal filter has much the same effect as a momentary lapse in attention. Hence, while the variable-weight hypothesis predicts some differences in PF slopes for multitone and broadband noise maskers, it predicts clearly lower asymptotic performance for multitone maskers. These specific predictions are tested in the following experiment.

II. METHODS

A. Stimuli

The signal was a 1000-Hz pure tone presented simultaneously with the maskers. Both the multitone maskers and the broadband noise maskers consisted of frequency components for which the distribution of component amplitudes was Rayleigh and the distribution of component phases was rectangular. Hence, as the number of components of the multitone masker increased the multitone maskers began to approximate samples of the broadband noise masker. The number of components comprising the multitone maskers ($m = 2,10,20,40,200,400,906$) was fixed within blocks of trials but varied across blocks. For each presentation, the multitone maskers were created by choosing m components (represented as complex numbers) at random from the fast-Fourier transforms (FFTs) of a sample of Gaussian noise. The frequencies were sampled uniformly over the range of 0.1 to 10 kHz, excluding the frequencies between 920–1080 Hz. Sampling was restricted such that no two neighboring frequencies were separated by less than 11 Hz. The multitone maskers were then synthesized by computing the inverse FFT of the m randomly selected components. The duration of all maskers was 370 ms. Hence, for the broadband noise the effective number of masker components was $m = 3608$, corresponding to a resolution of 2.7 Hz (1/370 ms). To be consistent with past studies using similar conditions, the average total power of all maskers was kept constant at 60 dB SPL regardless of the number of masker components. The signal and maskers were gated on and off together with 10-ms, cosine-squared onset and offset ramps. All stimuli were played at a sampling rate of 44.1 kHz using 16-bit, digital-to-analog conversion (Tucker Davis Technologies DD1). Stimuli were presented monaurally through Sennheiser model HD-414 headphones to individual listeners seated in a double-walled, IAC sound-attenuation chamber.

B. Procedure

Psychometric functions for detection of the signal in the multitone maskers, in the noise masker, and in quiet were obtained using a cued, two-interval, forced-choice (2IFC), staircase procedure. The listener heard three sounds on each trial separated by 1 half-second. The first sound was the signal presented at 60 dB SPL in isolation as a cue to aid detection. The subsequent two sounds both contained maskers but only one contained a signal. The listener's task was to identify the sound containing the signal, which was equally likely to be either sound. Feedback indicating the correct response was given after each trial. In this regard the procedure was typical of adaptive procedures commonly used in past studies of informational masking. The primary difference was that the level of the signal on each trial did not depend on the listener's response, but rather was predetermined for each block. The sequence began with the signal at a clearly audible level; for the next three trials it was reduced in 8-dB steps, then for the next three trials it was increased in 8-dB steps so as to return to the starting level. This up–down pattern was repeated five times for each block, yielding 40 trials per block. Practice trials were given until the listener appeared familiar with the task (almost always one block of trials). Listeners then completed three or more trial blocks for each condition, yielding at least 15 trials each for the highest and lowest signal levels, and 30 trials for each intermediate signal level. The starting level of the signal was selected such that performance for the intermediate levels fell between chance and best performance. If this criterion was not met an additional block of trials was obtained with the starting level adjusted to achieve the desired range of performance. Also, in rare cases where the slope of the PF appeared to be unusually shallow, an additional block of trials was collected to provide a more reliable estimate of upper asymptotic performance. An additional block of trials was necessary for approximately 5 to 10 percent of all runs, and then typically only for the first run of a condition. The order of conditions was quiet, followed by broadband noise, followed by the seven multitone masker conditions. The multitone masker conditions were presented in random order. Listeners participated in 1.5-h sessions on each day with breaks and required approximately 5 min to complete each block of 40 trials. Completion of all conditions required 3–4 sessions, depending on the listener.

C. Listeners

The data analyzed in this study are a subset of data previously collected as part of a larger study of individual differences in informational masking (see Lutfi *et al.*, 2003). The data are from 44 listeners (14 males and 30 females) of a total of 84 listeners who had participated in that study.[3] Of these 84 listeners, 30 were not included because their data were collected using an adaptive procedure rather than the method of constant stimuli. Another ten were not included because they were of an age (less than 6 years) for which thresholds in these conditions tend to be significantly elevated relative to other age groups (cf. Lutfi *et al.*, 2003). All of the listeners had pure-tone thresholds less than 15 dB HL

at the octave frequencies from 250 to 4000 Hz, as determined by standard audiometric tests (ANSI, 1989). Sixteen of the listeners were adults (19–30 years), 16 were late school-aged (11–16 years), and 12 were early school-aged (6–10 years). Though age is not a factor under consideration in this study, it was anticipated, based on previous work (e.g., Oh *et al.*, 2001), that within- and between-listener variability in slope and asymptote estimates would be large, and that a large pool of listeners would therefore be necessary to observe significant differences across conditions. The pooling of data across age seemed justified inasmuch as Lutfi *et al.* (2003) showed only small, statistically significant differences in mean thresholds across the age groups selected here. Moreover, that study concluded that the same factor responsible for differences across age groups is responsible for individual differences within age groups.

D. The psychometric function

The data relating proportion of correct responses p to signal level L was used to estimate psychometric functions for each listener and each condition using a maximum-likelihood criterion (cf. Wichmann and Hill, 2001). The assumed form of the PF between chance and asymptotic performance was a logistic. This form is commonly used in studies of auditory detection and typically yields excellent fits. The complete expression for the PF is

$$\hat{p} = 0.5 + (0.5 - \lambda)(1 + e^{-(L-\alpha)/\beta})^{-1}, \tag{1}$$

where $\hat{p}$ denotes the estimate of p, $0 \leqslant \lambda \leqslant 0.5$ is the lapse parameter determining upper asymptotic performance, α is a threshold parameter corresponding to the signal level yielding performance halfway between chance and asymptotic performance, and $\beta > 0$ is the slope parameter (larger values corresponding to smaller slopes). With the exception of the specified bounds for λ and β, no constraints were placed on any of the free parameters.

Figure 1 shows representative fits from two listeners for the $m = 20$ multitone masker condition (unfilled symbols) and the broadband noise minimal uncertainty condition (filled symbols). These fits demonstrate how λ and β capture the effect of masker uncertainty in terms of the change in upper asymptote and slope of the psychometric function, respectively. The PFs of the first listener (top panel) are consistent with the fixed-weight hypothesis; the PFs of the second listener (bottom panel) are consistent with the variable-weight hypothesis. For the purposes of this study the analysis focuses on how λ and β change with the number of components in the multitone masker, and whether one or the other of these parameters is significantly greater for the mutlitone maskers than for the broadband noise. The variable-weight hypothesis predicts changes in both β and λ across conditions, but predicts that the greatest changes will occur in λ. The fixed-weight hypothesis predicts significant changes in β alone. Moreover, the fixed-weight hypothesis predicts a nonmonotonic relation between β and the number of masker components mirroring that obtained for the masked thresholds.

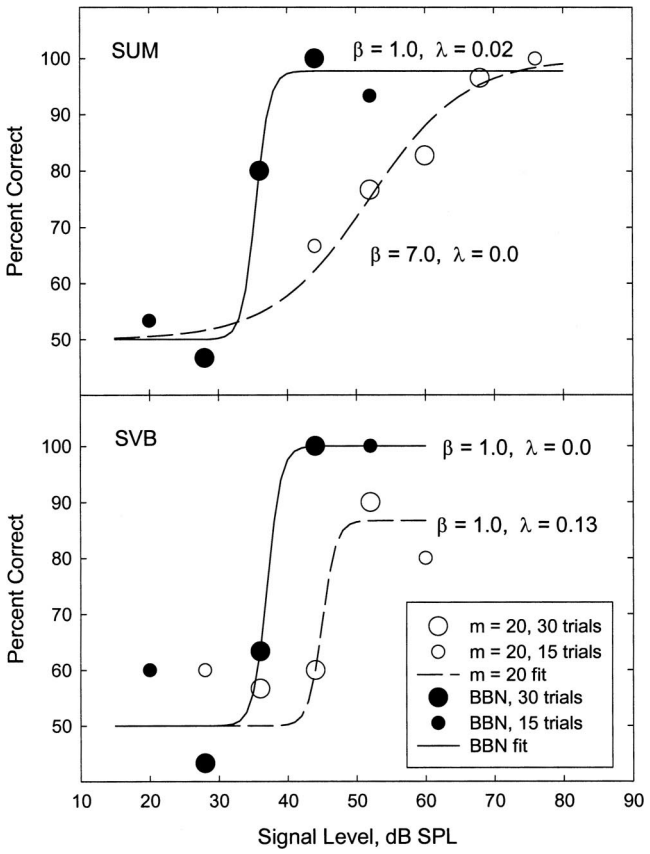


FIG. 1. Representative PFs from two listeners fitted according to Eq. (1). Data are from the $m = 20$-multitone masker condition (unfilled symbols) and the broadband noise minimal uncertainty condition (filled symbols). The top panel (listener SUM) shows data consistent with the fixed-weight hypothesis. The effect of masker uncertainty in this case is a reduction in the slope of the PF for the multitone masker. The second panel (listener SVB) shows data consistent with the variable-weight hypothesis. The effect of masker uncertainty in this case is a reduction in upper asymptote of the PF for the multitone masker.

III. RESULTS

A. Goodness of fit

Goodness of fit was evaluated using deviance, a measure that is a monotonic transformation of the likelihood of obtaining a particular set of data given the presumed form of the PF is correct (cf. Wichman and Hill, 2001). For the present application the deviance measure is

$$D = 2\sum_{i=1}^{K} \left\{ n_i p_i \log\left(\frac{p_i}{\hat{p}_i}\right) + n_i(1-p_i)\log\left(\frac{1-p_i}{1-\hat{p}_i}\right) \right\}, \tag{2}$$

where $K = 5$ is the number of data points for each PF, and n_i is the number of trials for the ith data point (15 or 30). Slightly greater than 5 percent of the fitted functions were estimated to have a likelihood of less than 5 percent assuming the present model is correct. Only one of these functions had estimated parameters that identified it as a clear outlier. Eliminating these data did not essentially change the pattern of results, so they were included in the analysis to offset any bias toward underdispersion in the data.

In addition to estimating the deviance for each fitted function, maximum-likelihood estimates of confidence intervals were obtained for each estimate of α ('threshold') and

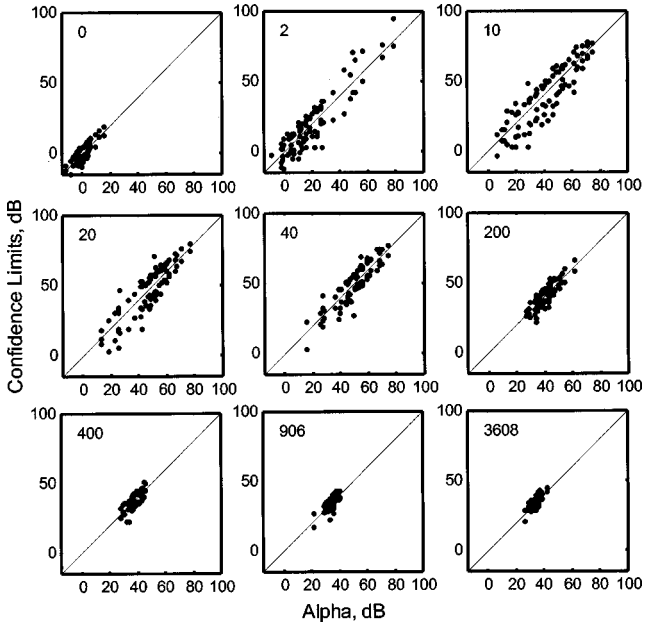


FIG. 2. Scatterplot of the estimated upper and lower confidence limits of α obtained by ''bootstrapping'' technique (see the text for details). Each pair of points above and below the diagonal represents the upper and lower limit of the estimated confidence interval for a single listener and a single condition, respectively. The number of masker components for each condition is given in the upper-left-hand corner of each plot.

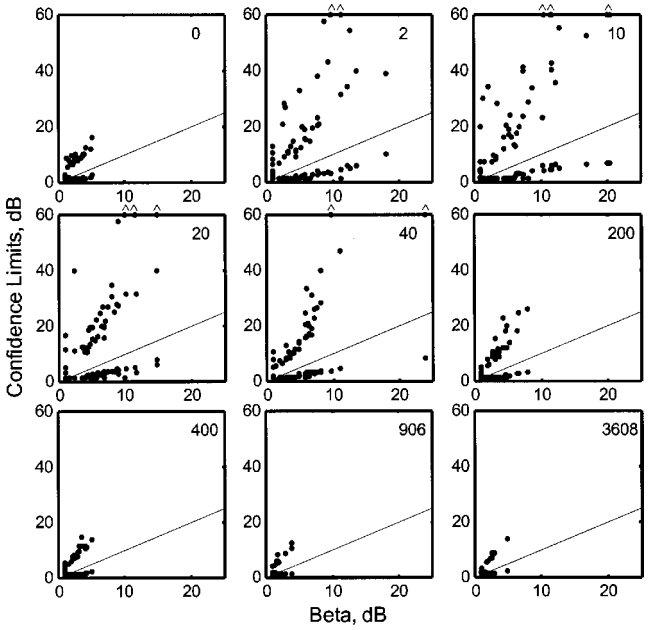


FIG. 3. Scatterplot of the estimated upper and lower confidence limits of β plotted as in Fig. 2. Arrows indicate values falling above the upper limit of the plot at 60 dB.

β. Confidence intervals were not obtained for λ as these would be rather meaningless given the strict upper and lower bounds on this parameter. The procedure used to estimate confidence limits is described by Wichmann and Hill (2001). The approach is to use a ''bootstrapping'' technique to estimate the sampling distributions of α and β, and then to take confidence limits of the estimated distributions. Using the values of percent correct at each signal level from each fitted function, and assuming binomially distributed percent correct values, a simulated percent correct at each signal level was randomly drawn and a logistic was fit to these values. This was repeated 10 000 times to provide 10 000 estimates of each of the three logistic parameters. The 2.5% and 97.5% points from the sampling distributions of α and β were then taken as the 95% confidence limits. Figures 2 and 3 give scatterplots of these values for each listener (different points) and each condition (different panels). The intent of these figures is to show the distribution of estimates for the upper and lower bounds, rather than to identify individual confidence intervals. Nonetheless, individual intervals can often be identified by a pair of points located horizontally from one another. Figure 2 shows the confidence intervals to be roughly symmetric about α, changing in size with number of masker components. For 2–40 masker components the size of the intervals is in the vicinity of 20 dB; for the quiet threshold condition (0 components) and for maskers comprised of 200 or more components they are in the neighborhood of 10 dB. Figure 3 shows the confidence intervals for β to be highly asymmetric, with a dramatic increase in upper bound as β increases. This is not surprising given the 0 lower bound for this parameter and the inverse exponential relation between β and PF slope; because of this relation, large increases in β have little impact on slope when β is already large. Again, it is because of the difficulty associated with getting reliable estimates of β when PF slope is shallow that we have chosen in this study to evaluate effects by obtaining estimates of β averaged over a large number of listeners.

B. Analysis of mean parameter values

Figure 4 shows the mean thresholds across listeners (filled symbols) plotted against the number of masker components. Thresholds correspond to the 71-percent-correct point of the individually fitted PFs; 95-percent confidence

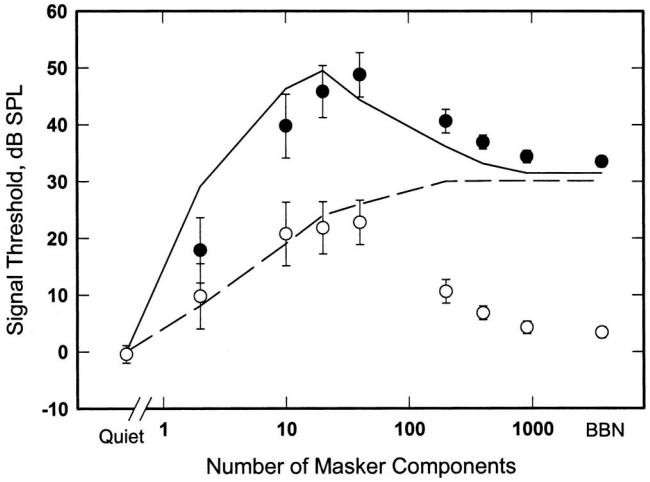


FIG. 4. Mean signal thresholds corresponding to the 71%-correct point on the PF are plotted as a function of the number of masker components m, where $m=0$ corresponds to the quiet-threshold condition and $m=3608$ corresponds to the broadband-noise condition. The data are means of 44 listeners with vertical bars representing the 95%-confidence limits. The dashed curve is an estimate of energetic masking for these conditions, and the continuous curve is a prediction for the total amount of masking based on the standard deviation σ of level at the output of a typical nonsignal auditory filter, as computed from the trial-by-trial stimuli (see the text for details).

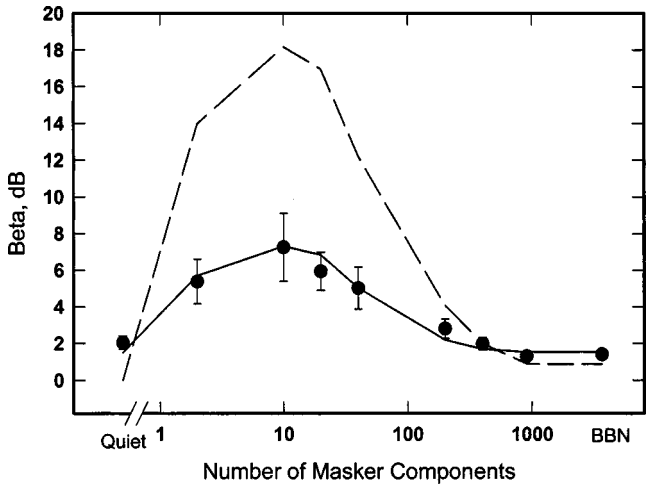


FIG. 5. Same as Fig. 4 except data are the mean slope parameter (β) estimates of the 44 listeners. The dashed curve is the standard deviation σ of the level at the output of a typical nonsignal auditory filter as computed from the trial-by-trial stimuli. The continuous curve is a prediction for $\beta = \sqrt{(n\sigma^2+\sigma_e^2)/2}$, where σ_e is an internal noise estimate corresponding to β for the quiet-threshold condition.

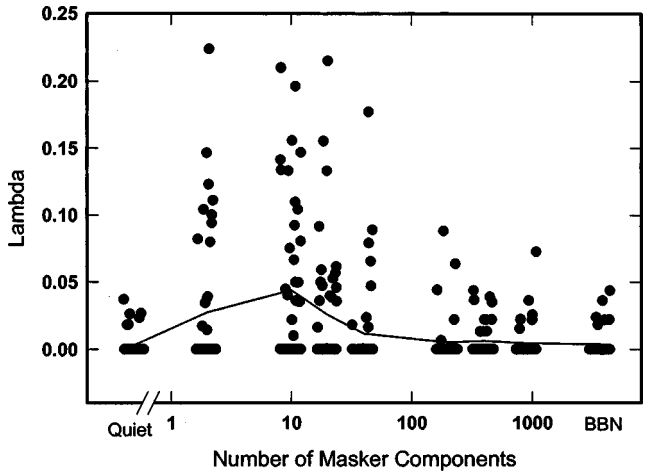


FIG. 6. Same as Fig. 4 except individual mean values of the lapse parameter (λ) represented by the continuous curve are accompanied by the individual values of this parameter. Values have been jittered along the abscissa for visibility.

limits are also shown. We report threshold values corresponding to 71 percent correct to permit comparison to past studies using adaptive procedures; however, the values obtained were essentially identical to those obtained for α.[4] A two-way ANOVA of the individual thresholds revealed a significant main effect of age group [$F(2,41)=8.48$, $p<0.001$], with younger listeners having on average higher thresholds than adults, and a significant interaction between age group and number of masker components [$F(16,328)=293$, $p<0.001$], with younger listeners having higher mean thresholds when the number of masker components was small ($m\leqslant 20$). These effects were quite small, however, so only the mean thresholds across age groups are shown (for further information on age effects on threshold the reader is referred to the paper by Lutfi *et al.*, 2003). The main effect of number of masker components was significant [$F(8,328)=151$, $p<0.001$] and replicates that obtained for very similar conditions using adaptive psychophysical procedures (cf. Neff and Green, 1987; Oh and Lutfi, 1998). In particular, the data show the characteristic nonmonotonic relation between threshold and number of masker components with highest thresholds for maskers comprised of 20–40 components. The highest thresholds are 12–15 dB greater than thresholds for the broadband noise condition, which also compares favorably with the mean magnitude of effects observed in past studies. The dashed line in Fig. 4 gives the estimate of energetic masking for these conditions using a roex auditory filter model (see Oh and Lutfi, 1998 for details regarding this derivation). The unfilled symbols give the difference between the obtained thresholds and the estimated amount of energetic masking. Hence, the unfilled symbols are estimates of informational masking alternatively defined as the amount of masking that is not energetic.

Figure 5 plots the mean β values in similar fashion to Fig. 4. A two-way ANOVA of the individual β's revealed no significant main effect for age group [$F(2,41)=0.381$, $p=0.686$] and no significant interaction of age group and number of masker components [$F(16,328)=0.732$, $p=0.761$]; hence, only the mean β's are shown. The absence of a significant age effect might appear inconsistent with the results of Allen and Wightman (1995) except that the children in Allen and Wightman's study were much younger (3–4 years versus greater than 6 years in the present study). Again, we intentionally chose our listeners so as to minimize the age effect on thresholds and on β based on the results of Lutfi *et al.* (2003). The main effect of number of masker components was significant [$F(8,328)=20.27$, $p<0.001$], with the broadband noise yielding significantly smaller values of β than the multitone maskers comprised of less than 400 components. The largest values of β occur for the 10–20-component maskers, unlike the highest thresholds, which occur for 20–40 component maskers. Also, the curve for β is nearly symmetric about the maximum, whereas the curve for threshold is elevated when maskers are comprised of 40 or more components. These differences are to be expected as the added effect of energetic masking (dashed curve in Fig. 4) is predicted to elevate thresholds with little effect on β when maskers are comprised of 40 or more components (cf. Oh and Lutfi, 1998). The mean values of β agree much more favorably with the estimates of informational masking given by the unfilled symbols in Fig. 4—this in agreement with the fixed-weight hypothesis.

Figure 6 shows the mean values of λ (continuous curve) again plotted in the fashion of Figs. 4 and 5. The conversion to upper asymptotic performance as percent correct is $100(1-\lambda)$ and the conversion to lapse rate or confusion rate as a percentage of trials is 200λ. Again, confidence intervals are not shown here because of the upper and lower bounds on λ. Instead, we give the mean (continuous curve) and individual values for each condition (note values have been jittered along the abscissa for visibility). The range of mean values is quite small relative to the range of individual values, 0–0.05 versus 0–0.25 for the individual values. The range for the individual values corresponds to upper asymptotic performance between 75 and 100 percent correct or confusion rates between 0 and 50 percent. Once again there was no signifi-

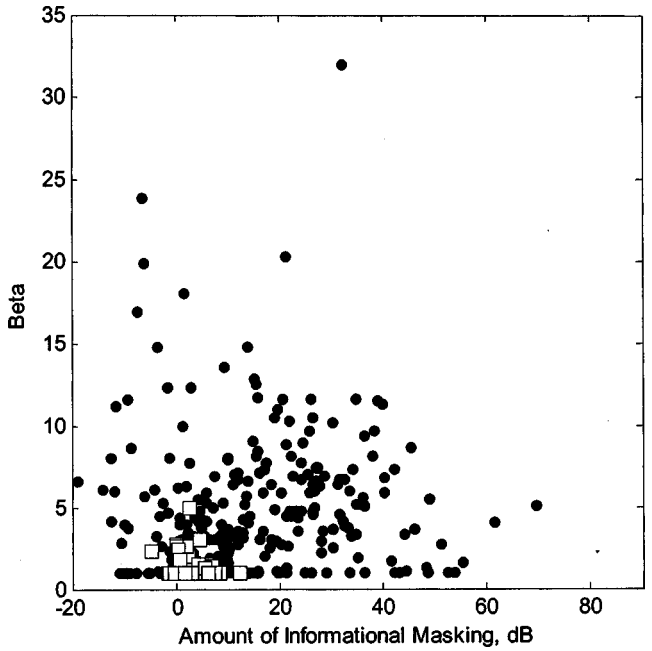


FIG. 7. Scatterplot of individual slope parameter (β) estimates against individual amounts of informational masking. Filled circles designate multitone-masker conditions; unfilled squares, the broadband noise (minimal uncertainty) condition.

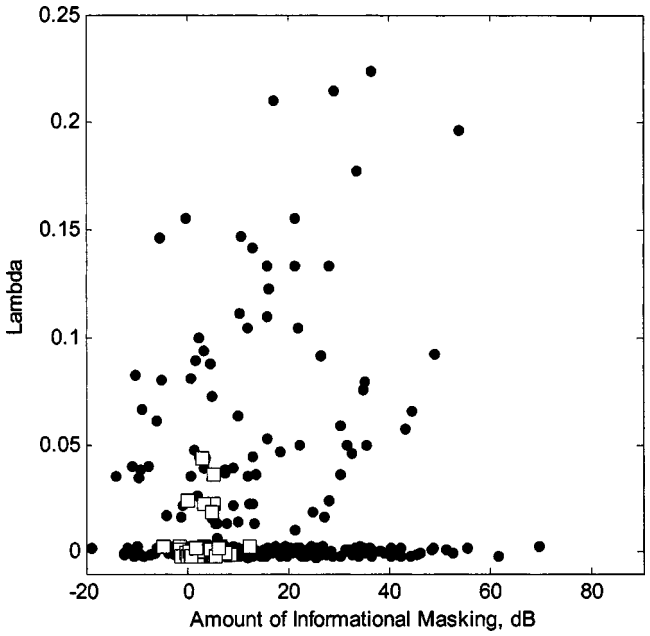


FIG. 8. Scatterplot of individual lapse parameter (λ) estimates against individual amounts of informational masking. As in Fig. 7, filled circles designate multitone-masker conditions; unfilled squares, the broadband noise (minimal uncertainty) condition.

cant main effect for age group [$F(2,41)=1.654$, $p=0.204$] and no significant interaction of age group and number of masker components [$F(16,328)=0.677$, $p=0.817$]. The mean λ's also changed significantly with the number of masker components in a manner similar to that of the mean thresholds [$F(8,328)=7.454$, $p<0.0001$] and the estimated amount of informational masking. In this respect, the data are also consistent with the variable-weight hypothesis.

C. Analysis of individual parameter values

Since the mean values of β and λ, and the estimated amount of informational masking all appeared to vary in a similar fashion with the number of masker components, we wished to determine whether there might be significant correlations between the individual values of these quantities. One question relevant to the fixed- versus variable-weight distinction is which parameter of the PF, β or λ, correlates more highly with the amount of informational masking. Figures 7 and 8 give the individual values of β and λ plotted against the estimated amount of informational masking for all listeners and all conditions. The amount of energetic masking in this calculation was assumed to be the same for all listeners as given by the dashed curve in Fig. 4. Solid circles and unfilled squares represent the multitone masker, and broadband noise masker conditions, respectively. The figures show the values β and λ clustering at or near zero, as might be expected since zero is the lower bound of these parameters. The largest estimates of β and λ, and the greatest variability in these estimates, occur for the multitone masker conditions. Interestingly, the figures reveal little or no discernible relation between the estimated amount of informational masking and either of the two parameters of the PF. This result would be difficult to explain except for the possibility that changes in informational masking are mediated differently for different listeners and/or conditions. In particular, no strong relation with either parameter would be expected if for some listeners and conditions the weights are variable, predominantly affecting asymptote, while for other listeners and conditions the weights are fixed, affecting only slope.

Figure 9 shows the individual estimates of β plotted against λ for all listeners and conditions. The multitone

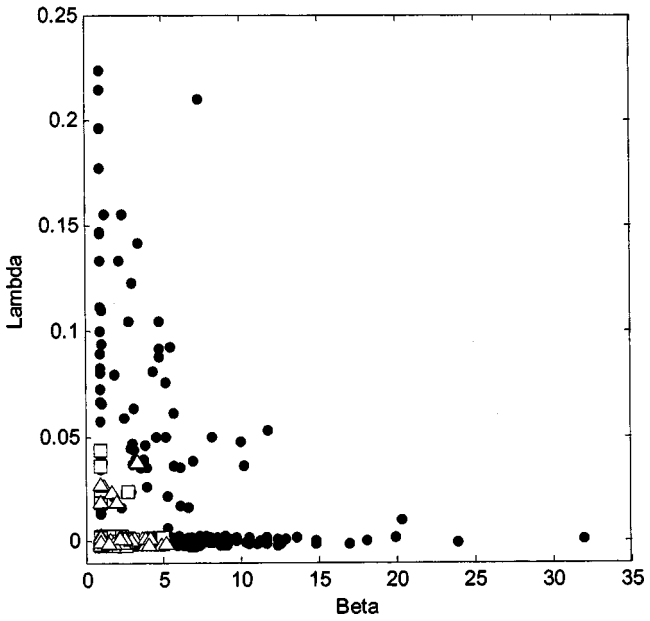


FIG. 9. Scatterplot of individual slope parameter (β) and lapse parameter (λ) estimates for all conditions and all listeners. As in Fig. 7, filled circles designate multitone-masker conditions; unfilled squares, the broadband noise (minimal uncertainty) condition; and unfilled triangles, the threshold in quiet condition.

masker and broadband noise conditions are represented by the filled circles and unfilled squares as before; the quiet condition, which was not shown in Figs. 7 and 8, is represented by the unfilled triangles. Taken across all conditions, the two parameters of the PF appear nearly orthogonal to one another (Pearson correlation of-0.11) consistent with the interpretation of mutually exclusive effects. Moreover, the orthogonal relation is most evident for the multitone maskers for which thresholds vary most widely. Within-listener variability in the estimates of β and λ was too great and the effect of number of masker components on these parameters was too small to discern any systematic dependence of this result on the number of masker components for individual listeners. Moreover, individual listener data were not always characterized exclusively by changes in one or the other parameter.

IV. DISCUSSION

The results of the present study reveal significant reductions in both the slope and upper asymptote of the PF for conditions of informational masking. Variability in estimates was too large to discern any systematic dependence of slope and asymptote on the number of masker components for individual listeners. However, the mean values of these parameters taken across listeners varied nonmonotonically with the number of masker components in a manner similar to that of the mean thresholds, particularly when the estimated effect of energetic masking was subtracted from thresholds. Despite the common relation of mean slope and asymptote to mean threshold, there was little correlation between the individual values of slope and asymptote across listeners and conditions. Indeed, when the individual values of these parameters were plotted against one another, the two parameters appeared nearly orthogonal to one another. The results indicate that, depending on the particular listener and condition, masker uncertainty has primarily one of two effects on the PF: a reduction in slope with little change in asymptote, or a reduction in asymptote with comparatively little change in slope. An important consequence of this result is that neither the values of slope alone nor the values of asymptote alone are generally predictive of the amount of informational masking.

The results are consistent with a model that assumes trial-by-trial judgments are based on a weighted sum of levels at the output of independent auditory filters. They suggest, however, that the weights are determined differently for different listeners in different multitone masker conditions. For some listeners and conditions they appear to reflect a fixed contribution of signal and nonsignal auditory filters on each trial (fixed-weight hypothesis). In such cases, the effect of masker uncertainty is a reduction in slope. For other listeners and conditions the weights appear to reflect some likelihood that a nonsignal auditory filter will be confused for the signal filter on any given trial (variable-weight hypothesis). In these cases, the effect of masker uncertainty is predominantly a reduction of upper asymptote. Both effects occurred with near-equal frequency for the different listeners and multitone masker conditions of this experiment. The results suggest that confusions are a significant component of informational masking. Consistent with this view are the results of numerous studies identifying perceptual similarity between the signal and masker as an important factor in informational masking (Arbogast *et al.*, 2002; Kidd *et al.*, 1994, 2001, 2002; Neff, 1995). Perhaps the most compelling evidence of this type comes from a study by Lutfi and Alexander (2002), in which some listeners were shown to make significant numbers of errors on randomly intermixed trials in which the signal was presented alone.

The effect of confusions and, more generally, the distinction between fixed- and variable weights has important implications for predicting informational masking. In fixed-weight models the listener's response is largely stimulus driven such that, if the weights are known, one should be able to predict the listener's response on any given trial from the parameters of the signal and masker on that trial (cf. Alexander and Lutfi, 2002). In variable-weight models the trial-by-trial responses are not so easily predicted since it is not possible to know on any given trial whether the listener is likely to make a confusion. This distinction allows us to get a sense of the relative importance of these two processes by determining the degree to which the data can be strictly predicted from the values of stimulus parameters. Following Oh and Lutfi (1998), we summarize the effect of fixed weights with a single free parameter n representing the number of nonsignal auditory filters contributing equally to the decision variable on each trial. This greatly simplifies the approach while still providing accurate predictions for threshold. Accordingly, the variance in the decision variable for the 2IFC task is

$$\frac{n\sigma^2+\sigma_e^2}{2}=\beta^2, \tag{3}$$

where $\frac{1}{2}\sigma_e^2$ is an internal noise variance given by β^2 for the quiet-threshold condition, and σ^2 is the variance in level at the output of a typical nonsignal filter. The value of σ^2 is the stimulus-driven component of the fixed-weight model; it is computed directly from the spectra of the maskers on each trial and varies with the number of masker components (cf. Oh and Lutfi, 1998). The dashed curve of Fig. 5 gives σ, while the continuous curve is a prediction for β obtained according to Eq. (3) with n as the only free parameter; n in this case has a value of 0.3. (Note: a value less than 1 implies that the combined weight for nonsignal filters is less than that for the signal filter.) The continuous curve of Fig. 4 gives the fixed-weight model prediction for threshold. The prediction is $E+d'\sqrt{n}\sigma$, where E is an estimate of energetic masking (dashed curve) and $d'=0.78$ is the index of sensitivity corresponding to 71-percent correct in the 2IFC task (for complete details regarding this derivation see Oh and Lutfi, 1998). In this case the best-fitting value of n is 3.5, in agreement with past data (cf. Oh and Lutfi, 1998). The two estimates of n should agree if the fixed-weight model were correct. That the two estimates are largely discrepant from one another suggests that there is another factor affecting threshold that is not tied to variance in the stimulus. To evaluate the extent to which this second factor is the effect of confusions we repeated the analysis, this time using only

data for which lambda was less than 0.001 (for which there were few if any confusions). If confusions alone are responsible for the discrepancy the two estimates should agree this time. The resultant estimate of n for thresholds remained essentially the same, 3.6, while the estimate for beta increased to 0.6. This result suggests that confusions are, at best, only partly responsible for the discrepancy. More importantly, it suggests that while changes in PF slope appear on average to covary with the amount of informational masking, the changes in slope are not large enough to account for informational masking in terms of the fixed-weight model.

The clear failure of the fixed-weight model in this case is noteworthy given that many past studies have nonetheless made accurate predictions for individual and mean thresholds assuming fixed weights. The outcome can be understood, however, in light of the fact that past predictions have been made for thresholds without constraints imposed by the underlying psychometric functions. When making predictions for threshold alone, the variance of the decision variable, $n\sigma^2$, need only be scaled by the free parameter n to yield the best fit. However, when additionally evaluating the PFs, the variance of the decision variable must agree with PF slope. Remarkably, the data from the present study do show good agreement between the two, except for the scalar n. This result complicates the interpretation of fixed-weight models despite their demonstrated ability to accurately describe individual thresholds in many studies. This outcome, together with the finding of less than perfect asymptotic performance for many PFs, suggests that fixed-weight assumptions are unlikely to yield meaningful estimates of other parameters of the decision process in conditions of informational masking (cf. Berg, 1990; Richards, Tang, and Kidd, 2002; Tang and Richards, 2002).

V. CONCLUSIONS

(1) Conditions of informational masking produce significant reductions in both the slope and the upper asymptote of the PF. Slope is affected more for some listeners and conditions, while asymptote is affected more for other listeners and conditions.
(2) Individual values of slope and upper asymptote are nearly orthogonal; hence, neither parameter alone is highly predictive of individual thresholds or the amount of informational masking.
(3) Mean slope and asymptote parameters of the PF vary nonmonotonically with the number of masker components in a manner similar to the amount of informational masking, where the amount of informational masking is given as the difference between masked threshold and the estimated amount of energetic masking.
(4) Masked thresholds are well described by a fixed-weight model. However, the effect of confusions and the failure to obtain similar values of model parameters from threshold and psychometric data complicate the interpretation of this model.

ACKNOWLEDGMENTS

The research was supported by grants from the NIDCD (R01 DC1262-10 and R01 HD23333-08). The authors wish to acknowledge the helpful comments of Joshua Alexander, Dr. Marjorie Leek, and two anonymous reviewers.

[1]Lutfi (1992) provides an analytic development of two possible versions of a variable-weight model, one treating the weights as a binomial random variable, the other as a normal random variable. Only the former is tested here.
[2]The relation is predicted to be similar to but not identical to that of the threshold data because of the effect of energetic masking (cf. Oh and Lutfi, 1998), a point that we revisit in the Results and Discussion sections.
[3]Neff, Kessler, and Dethlefs (1996) report larger amounts of informational masking in female listeners.
[4]Note, α marks the midpoint of the PF, the point about which PF slope is symmetric. It is for this reason that confidence intervals were estimated for individual values of α rather than for individual values of threshold defined by the 71%-correct point.

Alexander, J. M., and Lutfi, R. A. (**2002**). ''Sensation level and decision weights as factors underlying informational masking in hearing-impaired listeners,'' Abstracts of the Assoc. for Research in Otolaryngology, Vol. 25, p. 180.
Allen, P., and Wightman, F. (**1995**). ''Effects of signal and masker uncertainty on children's detection,'' J. Speech Hear. Res. **38**, 503–511.
Arbogast, T. L., Mason, C. R., and Kidd, Jr., G. K. (**2002**). ''The effect of spatial separation on informational and energetic masking of speech,'' J. Acoust. Soc. Am. **112**, 2086–2098.
Berg, B. G. (**1990**). ''Observer efficiency and weights in a multiple observation task,'' J. Acoust. Soc. Am. **88**, 149–158.
Brungart, D. S. (**2001**). ''Informational and energetic masking effects in the perception of two simultaneous talkers,'' J. Acoust. Soc. Am. **109**, 1101–1109.
Durlach, N. I., Mason, C. R., Kidd, Jr., G., Arbogast, T. L., Colburn, H. S., and Shin-Cunningham, B. (**2003**). ''Note on informational masking,'' J. Acoust. Soc. Am. (in press).
Green, D. M. (**1995**). ''Maximum-likelihood procedures and the inattentive observer,'' J. Acoust. Soc. Am. **97**, 3749–3760.
Kidd, Jr., G., Arbogast, T. L., Mason, C. R., and Walsh, M. (**2001**). ''Informational masking in listeners with sensorineural hearing loss,'' J. Assoc. Res. Otolaryngol. **3**, 107–119.
Kidd, Jr., G., Mason, C. R., and Arbogast, T. L. (**2002**). ''Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns,'' J. Acoust. Soc. Am. **111**, 1367–1376.
Kidd, Jr., G., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, S. H. (**1994**). ''Reducing informational masking by sound segregation,'' J. Acoust. Soc. Am. **95**(6), 3475–3480.
Kidd, Jr., G., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1994**). ''Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns,'' J. Acoust. Soc. Am. **104**(1), 422–431.
Leek, M. R. (**2001**). ''Adaptive procedures in psychophysical research,'' Percept. Psychophys. **63**(8), 1279–1292.
Lutfi, R. A. (**1990**). ''How much masking is informational masking?'' J. Acoust. Soc. Am. **88**, 2607–2610.
Lutfi, R. A. (**1992**). ''Informational processing of complex sound. III. Interference,'' J. Acoust. Soc. Am. **91**, 3391–3401.
Lutfi, R. A. (**1993**). ''A model of auditory pattern analysis based on component-relative entropy,'' J. Acoust. Soc. Am. **94**, 748–758.
Lutfi, R. A., and Alexander, J. M. (**2002**). ''Informational masking without maskers,'' J. Acoust. Soc. Am. **111**, 2470.
Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., and Callahan, M. R. (**2003**). ''One factor underlies individual differences in auditory informational masking within and across age groups,'' Percept. Psychophys. **65**(3), 396–406.
Neff, D. L., and Callahan, B. P. (**1987**). ''Psychometric functions for multicomponent maskers with spectral uncertainty,'' J. Acoust. Soc. Am. Suppl. 1 **81**, S53.
Neff, D. L., and Green, D. M. (**1987**). ''Masking produced by spectral uncertainty with multicomponent maskers,'' Percept. Psychophys. **41**, 409–415.

Neff, D. L., Kessler, C. J., and Dethlefs, T. M. (**1996**). ''Sex differences in simultaneous masking with random-frequency maskers,'' J. Acoust. Soc. Am. **100**, 2547–2550.
Oh, E. L., and Lutfi, R. A. (**1998**). ''Nonmonotonicity of informational masking,'' J. Acoust. Soc. Am. **104**, 3489–3499.
Oh, E. L., Wightman, F., and Lutfi, R. A. (**2001**). ''Children's detection of pure-tone signals with random multitone maskers,'' J. Acoust. Soc. Am. **109**, 2888–2895.
Pollack, I. (**1975**). ''Auditory informational masking,'' J. Acoust. Soc. Am. **57**, S5.
Richards, V. M., Tang, Z., and Kidd, Jr., G. (**2002**). ''Informational masking with small set sizes,'' J. Acoust. Soc. Am. **111**, 1359–1366.
Tang, Z., and Richards, V. A. (**2002**). ''Internal noise invariance across two informational masking tasks,'' J. Acoust. Soc. Am. **111**, 2470.
Watson, C. S., and Kelly, W. J. (**1981**). ''The role of stimulus uncertainty in the discrimination of auditory patterns,'' in *Auditory and Visual Pattern Recognition*, edited by D. J. Getty and J. H. Howard, Jr. (Erlbaum, Hillsdale, NJ).
Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). ''Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty,'' J. Acoust. Soc. Am. **60**, 1176–1188.
Wichmann, F. A., and Hill, N. J. (**2001**). ''The psychometric function. I. Fitting, sampling, and goodness of fit,'' Percept. Psychophys. **63**(8), 1293–1313.
Wright, B. A., and Saberi, K. (**1999**). ''Strategies to detect auditory signals in small sets of random maskers,'' J. Acoust. Soc. Am. **105**, 1765–1775.

Tone decay for hearing-impaired listeners with and without dead regions in the cochlea

Martina Huss and Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB, England

(Received 10 December 2002; revised 22 August 2003; accepted 15 September 2003)

For people with normal hearing, a sustained tone with a frequency within the standard audiometric range remains audible when presented at a level well above threshold. However, for a pure tone with frequency close to the upper limit of hearing (well above 8 kHz), the loudness may decrease within seconds and the tone may decay to inaudibility, even when presented at a level between 20 and 40 dB SL. Scharf [in *Hearing Research and Theory*, edited by J. V. Tobias and E. D. Schubert (Academic, New York, 1983), Vol. 2, pp. 1–53] suggested that marked loudness adaptation only occurs when the excitation pattern evoked by a tone is spatially limited. The upper limit of hearing may be comparable to the boundary of a ''dead region,'' which is a region with a complete loss of inner hair cell (IHC) and/or neural function. The present study investigated the perceived decay of pure tones for 9 normal-hearing subjects and 12 subjects with moderate to severe sensorineural hearing loss, using a wide range of frequencies (0.125–12 kHz). A dead region was diagnosed for 8 of the 12 subjects. No consistent association was found between the degree of tone decay and the presence of a dead region. Subjects with dead regions did not experience significantly more tone decay than subjects with comparable absolute thresholds but without a dead region, even when the frequency of the tone fell within or close to the edge of a dead region. For severely hearing-impaired subjects, spatial restriction of the excitation pattern was neither necessary nor sufficient to lead to tone decay. The prevalence of tone decay was not well predicted by the audiometric threshold at the test frequency. It is proposed that tone decay depends on the physiological condition of the place in the cochlea where the tone is detected, which, in a case involving a dead region, is the place adjacent to the dead region. The prevalence of tone decay increased when the audiometric threshold was above 50 dB HL in the frequency region where the tone was detected. [DOI: 10.1121/1.1624063]

PACS numbers: 43.66.Cb, 43.66.Sr, 43.66.Mk [MRL] Pages: 3283–3294

I. INTRODUCTION

For normal-hearing listeners, tones with frequencies within the usual audiometric frequency range (below 8 kHz) remain audible when presented at a level well above threshold (Bray *et al.*, 1973), and the loudness does not decrease greatly over time (Scharf, 1983; Hellman *et al.*, 1997). However, in some cases of hearing impairment, tone decay occurs: the loudness of a continuously presented tone progressively reduces, until the tone becomes inaudible. Marked tone decay is commonly interpreted as being symptomatic of retrocochlear dysfunction, and measures of it have been used as a diagnostic tool (Carhart, 1957); for a review, see Green (1985). Tone decay is considered as abnormal when the perception of a sustained pure tone with a level above 5 dB SL is not maintained throughout the whole duration of presentation (typically 60 seconds). Abnormal tone decay can be caused by neural degeneration, inflammation, trauma, and by space-occupying lesions, such as tumors, involving injury of the eighth nerve (Green, 1985). The role of more peripheral factors in causing or contributing to tone decay and loudness adaptation remains unclear, although some researchers have reported moderate amounts of tone decay in subjects with cochlear hearing loss, which suggests that its origin is not always retrocochlear (Green, 1985).

While extreme forms of tone decay were thought to occur only in patients with retrocochlear dysfunction, early experiments using normal-hearing subjects appeared to show considerable adaptation using a procedure of loudness balancing between the two ears. An interrupted pure tone of the same frequency as the continuous test tone was commonly used as a comparison stimulus. The comparison tone was presented either in the opposite ear to the test tone (Hood, 1950; Egan, 1955; Jerger, 1957; Palva and Kärjä, 1969) or in the same ear (Davis and Weiler, 1976; Feaster and Weiler, 1975; Weiler and Gross, 1976). However, it was later suggested that adaptation measured in this way was somehow induced by the presence of the interrupted comparison tone (Bray *et al.*, 1973; Scharf, 1983). Later experiments avoiding the use of a comparison tone have indicated that tone decay and loudness adaptation do occur for normal-hearing listeners, but are restricted to soft stimuli (Green, 1985; Hellman *et al.*, 1997) or stimuli with very high frequencies (Ward, 1973; Scharf, 1983; Viemeister and Bacon, 1983; Hellman *et al.*, 1997).

Scharf (1983) reported that the extent of tone decay does not seem to depend strongly on the subject's age or absolute threshold, although children under 16 years of age show less adaptation than adults (Scharf, 1983). Miśkiewicz *et al.* (1993) showed that a 16-kHz tone presented at 40 dB SL became inaudible within the first minute of exposure for

about one-third of their normal-hearing subjects. Another group of subjects with elevated thresholds at 16 kHz (thresholds >50 dB SPL) showed significantly more tone decay at 14 kHz than the group of normal-hearing subjects (thresholds <50 dB SPL at 16 kHz).

Miśkiewicz *et al.* (1993) interpreted these observations as support for the ''restricted excitation pattern'' hypothesis proposed by Scharf (1983). The hypothesis is that marked tone decay occurs when the excitation pattern evoked by a steady tone is highly spatially restricted. This is the case for tones of either low sensation level or with a signal frequency exciting the very basal end of the cochlea at the upper frequency limit of hearing. According to Miśkiewicz *et al.* (1993), this restriction of the excitation pattern to the base of the cochlea is required in normally hearing listeners to induce adaptation. Listeners with elevated thresholds at 16 kHz are assumed to have a more restricted high-frequency range of hearing than ''normal-hearing'' subjects. According to Miśkiewicz *et al.* (1993), the restricted excitation pattern hypothesis is supported by the finding that adaptation occurs for an intense 4-kHz tone when the tone's excitation pattern is partially masked by a high-pass or bandpass noise (Wright, 1959; Scharf, 1983; Moore and Peters, 1997). However, this effect occurs only for high frequencies.

Viemeister and Bacon (1983) and Bacon and Viemeister (1994) suggested that high-frequency tone decay reflects neural long-term adaptation, perhaps mediated by effects at the level of hair cells or their synapses. Evidence consistent with this idea has been reported by Javel (1996). He measured substantial adaptation in cat auditory-nerve fiber responses to sustained tones, the discharge rates reaching values below spontaneous rates. As similarly tuned auditory-nerve fibers showed varying amounts of adaptation, Javel concluded that the long-term adaptation does not arise from cochlear mechanics. A time-dependent reduction in peak basilar-membrane displacement would affect all fibers with similar characteristic frequencies (CFs). Also, auditory-nerve fibers maintained spike rates over a long duration when stimulated electrically. Therefore, the differences between fibers were argued to arise from mechanisms more peripheral than spike generation in the spiral ganglion cells. Consistent with this argument, fibers with low spontaneous firing rates exhibited pronounced long-term adaptation, whereas fibers with high spontaneous rates adapted less.

Javel (1996) claimed that his findings were in general agreement with perceptual findings in human subjects by Miśkiewicz *et al.* (1993) with respect to the time course and magnitude of adaptation. However, the perception of a steady low- to midfrequency pure tone is maintained indefinitely, even though the majority of fibers of all CFs adapt substantially. Also, it is difficult to explain why the effect of loudness adaptation, when measured perceptually, decreases or disappears with increasing sensation level in humans, whereas physiologically, adaptation measured in the auditory nerve in cats increases significantly with increasing sensation level. Miśkiewicz *et al.* (1993) showed that adaptation at very high frequencies depends on the absolute threshold, whereas Javel found no significant correlation between neural adaptation and absolute threshold of the fibers.

Javel (1996) proposed an explanation for the more pronounced perceptual adaptation at high frequencies found in humans: as the inner hair cells (IHCs) at the extreme cochlear base are at an increased distance from the *stria vascularis*, they might be more vulnerable to metabolically mediated effects. Depletion of the metabolic energy source with age could affect IHC or synaptic function at the cochlear base, leading to loss of sensitivity to high-frequency tones (presbyacusis) (Walsh *et al.*, 1984; Schmiedt, 1996). Associated with the decrease in IHC function would be an increase in adaptation. However, this explanation does not seem consistent with Schuknecht's (1994) finding that, in humans, lesions in the *stria vascularis* lead to flat hearing losses rather than high-frequency losses.

It is possible that adaptation and tone decay depend partly on the status of the hair cells within the cochlea. Later in this paper we propose that the status of the IHCs is especially important. It has been suggested that the overall extent of a cochlear hearing loss at a given frequency can be modeled as a component due to IHC (or neural) damage and a component due to outer hair cell (OHC) damage: $HL_{OHC} + HL_{IHC} = HL_{TOTAL}$ (Moore and Glasberg, 1997). Damage to the OHCs results in reduced effectiveness of the ''active mechanism'' and a reduced amplitude of vibration on the basilar membrane for low level inputs (Ruggero and Rich, 1991). Damage to the IHCs results in reduced efficiency of transduction. Neural damage or degeneration may result in effects similar to damage to the IHCs. The maximum gain of the active mechanism is assumed to be about 55 dB for frequencies below 2 kHz and 65 dB for frequencies above 2 kHz (Yates *et al.*, 1992; Moore and Glasberg, 1997). Hearing loss exceeding these values implies loss of IHC and/or neural function. Hearing loss less than this may or may not reflect IHC/neural damage. Moore and Glasberg (1997) proposed that, on average, $HL_{OHC} = 0.8HL_{TOTAL}$ for losses up to 55–65 dB.

Presbycusis is often associated with damage to the OHCs, IHCs, and neurons (Schuknecht, 1994; Moore, 1998). Sometimes there may be complete loss of function of the IHCs and/or neurons, a ''dead region,'' especially over the basal region of the cochlea (Moore, 2001). A dead region can be defined in terms of the CFs of the IHCs and/or neurons immediately adjacent to the dead region (Moore *et al.*, 2000; Moore, 2001). A tone with a frequency falling into a dead region may be detected via spread of basilar-membrane vibration to a place adjacent to the dead region. In such a case, a major part of the vibration on the basilar membrane evoked by the tone falls within the region lacking functional IHCs/neurons. Dead regions are therefore associated with a restricted region of transduced excitation on the basilar membrane.

If the decrease in the upper-frequency limit of hearing with increasing age is associated with progressive loss of function of IHCs/neurons at the base of the cochlea, then presentation of a tone with frequency just below the upper limit of hearing may be comparable to presenting a tone with frequency close to the boundary of an extensive dead region. The restricted excitation pattern hypothesis leads to the prediction that marked adaptation will occur for tones falling

into or close to the boundaries of a dead region at *any* place on the basilar membrane. This study tests this prediction by investigating the incidence of tone decay over a wide frequency range for subjects with and without dead regions.

II. METHOD

A. Procedure

The subject was seated in a large, sound-isolated booth. Absolute thresholds were measured using manual audiometry, applying the procedure proposed by Carhart and Jerger (1959). The signals were generated and controlled by a Grason-Stadler GSI-16 audiometer and presented via Telephonics TDH50 earphones (except when there was a large interaural asymmetry in absolute thresholds, when ER3A insert earphones were used; see below for details). Dead regions were previously determined and localized using psychophysical tuning curves (PTCs) and the threshold-equalizing noise (TEN) test (Moore *et al.*, 2000; Moore and Alcántara, 2001). The underlying assumption of both tests is that a tone with a frequency falling in a dead region is detected via IHCs/neurons with CFs remote from the signal frequency. The basilar membrane generally vibrates maximally at the region where the CF corresponds to the signal frequency, and it vibrates less at the remote region where the tone is detected. Therefore, a masker with frequency components close to the CF of the functioning part of the basilar membrane where the tone is detected will produce more masking than would normally be the case. Both tests are described in detail in Moore *et al.* (2000). For the TEN test, the criteria for diagnosis of a dead region at a specific frequency are: (1) The masked threshold should be 10 dB or more above the noise level per ERB; (2) The masked threshold should be 10 dB or more above the absolute threshold. For PTCs, a shift in the tip frequency away from the signal frequency was taken as evidence for a dead region.

Tone decay was measured using a modification of the procedure described by Carhart (1957). The test entailed the following steps.

(1) The subject's absolute threshold was measured for frequencies of 125, 250, 500, 750, 1000, 1500, 2000, 3000, 4000, 6000, 8000, and 12 000 Hz, using a pure tone lasting about 1 s and starting with the highest frequency.
(2) A sustained tone of fixed frequency was presented at 10 dB SL.
(3) The subject was instructed to press a button as long as the tone was audible and to release it if the tone faded to inaudibility.
(4) A stop watch was started when the tone was turned on. If the tone was heard for 60 s, the test was concluded.
(5) If the subject indicated that the tone decayed to inaudibility before the 60 s were over, the level of the tone was raised by 5 dB without interrupting the tone and the time was recorded. The stop watch was set back to zero and the timing restarted. If the subject stopped pressing the button again within 60 s, another 5-dB increment in level was introduced, and so on. For each level, a record was kept of the number of seconds the tone was audible.
(6) The level of the tone was raised in 5-dB steps, until a level was reached that allowed the subject to perceive the tone for 60 s.
(7) This procedure was repeated for each test frequency in turn, starting with the highest frequency and working towards the lower frequencies.

The test was sometimes limited by the highest permissible tone level, which was determined either by the intensity-duration limits prescribed by industrial noise exposure regulations or by the uncomfortable loudness level (ULL) of the subject, whichever was the lower. The maximum exposure time never exceeded 1 min at 105-dB SPL and was less than 5 s at 120-dB SPL.

The whole procedure was repeated three times on different days. Subjects were instructed not to move their head or to replace the headphones until testing was completed for all frequencies. They were asked not to swallow while measurement of tone decay was taking place for one frequency, but only between measurements. Absolute thresholds were measured at the start of each session and were consistent within the 5-dB tolerance for manual audiometry. In preliminary trials, the (mostly elderly) subjects reported difficulties in determining the exact time of cessation of the tone due to problems in sustaining concentration. To counter this problem, the subjects were instructed to continuously assess the quality of the tone. In an initial trial, subjects were asked to vocalize these descriptions. Subsequent trials were conducted without vocalization, although subjects were still asked to think about the quality of the tone.

TABLE I. Etiologies and ages of the hearing-impaired subjects. Information about dead regions is given in the two columns on the right. When a dead region was present, the number indicates the estimated low-frequency edge, f_e, of the region.

Subject	Etiology	Age	Dead region edge (kHz)	
			LE	RE
KC	Noise-induced (machinery)	64	None	None
CH	Since childhood, hereditary	31	None	None
MR	Noise-induced (shooting) or hereditary	47	None	None
TB	Presbyacusis	57	None	None
DO	Hereditary and noise-induced (machinery)	72	2	None
NC	Noise-induced (shooting)	81	1.8	0.85
PR	Since childhood, possibly congenital	77	1.7	1.5–2
JM	Unknown	79	3.5	3.5
PJ	Hereditary and noise-induced (machinery)	69	3.5	3
LR	Since childhood, possibly congenital	48	1.7	1
CA	Noise-induced (shooting) or hereditary	54	0.85	0.5
RC	Viral infection/presbyacusis	70	1.1	Deaf

B. Subjects

Twelve subjects with sensorineural hearing impairment of varying degrees were tested. Ten subjects were tested using each ear separately and two were tested using one ear only. Subject details are given in Table I. Bone-conduction audiometry and impedance audiometry confirmed that there was no conductive component to the losses. Subjects with highly asymmetric losses were excluded. In the great majority of cases, the hearing losses were reasonably symmetrical

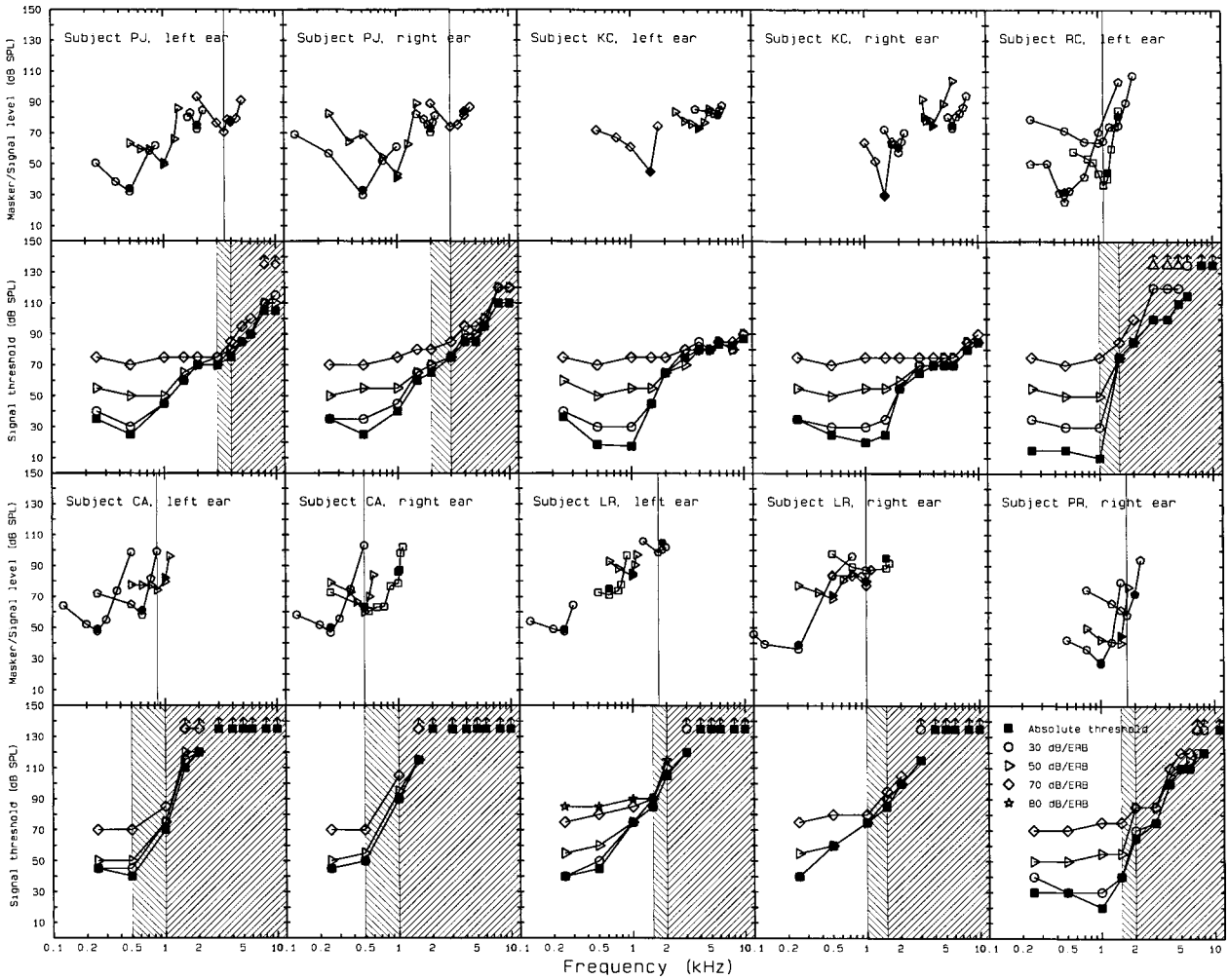


FIG. 1. Results of the TEN test and PTCs for ten ears of six subjects. Panels are grouped in pairs, with the PTCs for a given ear at the top and the results of the TEN test at the bottom. For the TEN test, filled squares indicate absolute thresholds and open symbols indicated masked thresholds in the TEN. The left-most shaded area (when present) indicates the range of frequencies over which the edge frequency, f_e, of the dead region might be located (this covers the range between the highest frequency at which the TEN test criteria were not met and the lowest frequency at which the criteria were met). The right-most shaded area indicates the frequency region over which the criteria were met. Up-pointing arrows indicate that the threshold was too high to be determined (above 120 dB HL). PTCs were determined using signals at 8 dB SL; signal levels and frequencies used to determine the PTCs are indicated by solid symbols. When a thin vertical line is present, this indicates the frequency at the tip of a PTC with a tip shifted away from the signal frequency; the frequency at the tip is assumed to correspond to f_e.

across the two ears, with differences in audiometric threshold typically around 10 dB and maximally 35 dB (subject NC at 1000 Hz). In these cases, it is unlikely that the results were affected by ''cross hearing,'' given that the interaural attenuation of the TDH50 earphones is at least 40 dB. There was one case where the hearing loss differed by 50 dB across ears (subject DO at 500 Hz). For that case, and for NC at 1000 Hz, testing was conducted using ER3A insert earphones, which have higher interaural attenuation (60 dB or more).

Results of the TEN test and PTCs are presented in Figs. 1 and 2. Except for the bottom-right section of Fig. 2, panels are grouped in pairs, with the PTCs for a given ear at the top and the results of the TEN test at the bottom. For the TEN test, the left-most shaded area (when present) indicates the range of frequencies over which the edge frequency, f_e, of the dead region might be located (this covers the range between the highest frequency at which the TEN test criteria were not met and the lowest frequency at which the criteria were met). The right-most shaded area indicates the frequency region over which the criteria were met. For the PTCs the thin vertical line indicates the frequency at the tip of a PTC with a tip shifted away from the signal frequency; the frequency at the tip is assumed to correspond to f_e. The final estimated value of f_e was determined by the position of the thin vertical line.

Fourteen ears of eight subjects were diagnosed as having dead regions. For 12 of these ears, the diagnosis was based both on PTCs and TEN test. In one case (PR left ear) the definition of the dead region relied on the results of the TEN test only. PTCs were not obtained for four ears showing no evidence of a dead region based on the TEN test (DO right ear, TB left ear, CH left and right ears). The results for the TEN test in these cases are shown towards the bottom right of Fig. 2.

Supplementary measurements were conducted with nine normally hearing subjects (absolute thresholds $\leqslant$20 dB HL at all audiometric frequencies) and one subject with moderate high-frequency hearing loss (absolute thresholds $\leqslant$50 dB

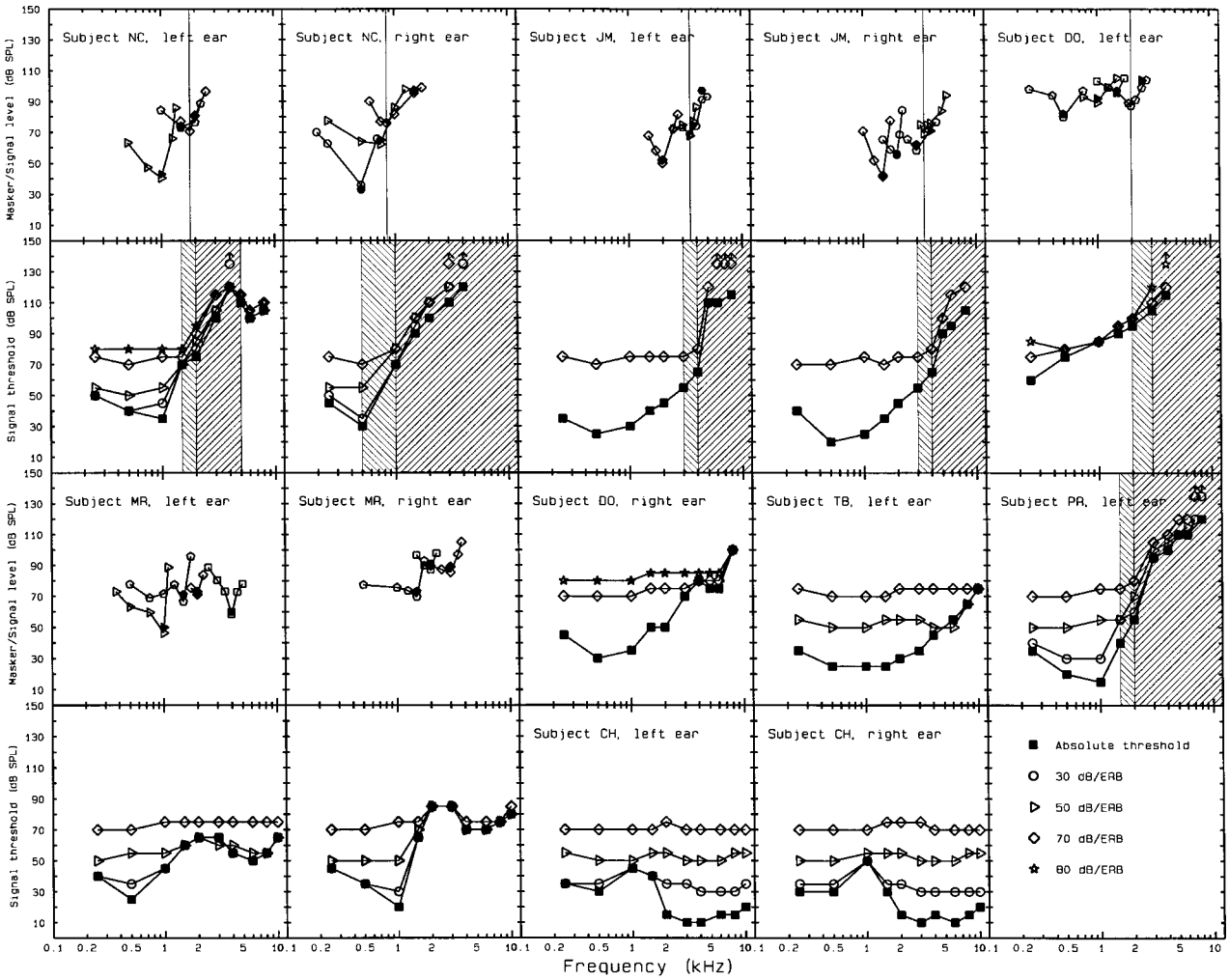


FIG. 2. As Fig. 1, but for 12 ears of eight subjects. For five ears, PTCs were not determined. TEN-test results for these ears are shown in the bottom right of the figure.

HL at all audiometric frequencies). These ten subjects were divided into four groups according to their absolute thresholds; see Table II.

III. RESULTS

For all subjects, results were reasonably consistent across the three repetitions of the tone-decay test for each frequency. When the tone was reported to become inaudible within 60 s, the time at which this occurred nearly always fell within a 10-s range across repetitions. The average of the standard deviations for all the subjects was 3.2 s. In what follows, we present only mean results across repetitions.

TABLE II. Absolute thresholds and ages of the subjects in the four groups described in the text. Subjects were grouped according to their absolute thresholds at high frequencies.

Group	Subject	Absolute threshold	Age
1	JN	≤5 dB HL up to 12 kHz	29
	RJ	≤10 dB HL up to 12 kHz	27
	VN	≤10 dB HL up to 12 kHz	25
2	RV	≤10 dB HL up to 8 kHz 15 dB HL at 12 kHz	28
	GM	≤10 dB HL up to 8 kHz 15 dB HL at 12 kHz	28
	BD	≤15 dB HL up to 12 kHz	26
3	TH	≤20 dB HL up to 12 kHz	23
	DS	≤15 dB HL up to 8 kHz 35 dB HL at 12 kHz	27
	JG	≤25 dB HL up to 12 kHz	29
4	BM	≤50 dB HL up to 8 kHz 75 dB HL at 12 kHz	57

A. Tone decay for subjects with moderate to severe hearing loss

Results for the 12 subjects with high-frequency hearing loss are shown in Figs. 3–10. The sequence in which the individual results are displayed was chosen to facilitate discussion of different aspects of the relation between audiogram shape, the presence of a dead region, and tone decay. In all of these figures, short lines without symbols show tone-decay results. The ordinate indicates the level of the test tone in dB HL. Each line is centered at the abscissa value corresponding to the test frequency. The abscissa indicates both the frequency of the test tone in kHz and its time of presentation in seconds. The left-hand end of each short horizontal line corresponds to the time of the start of the presentation. The audible duration of the tone at a given level is indicated by the length of the horizontal line, whereby 0.8 of a division on the abscissa corresponds to 60 s, as indicated by the key. Vertical segments indicate that a tone decayed to inaudibility

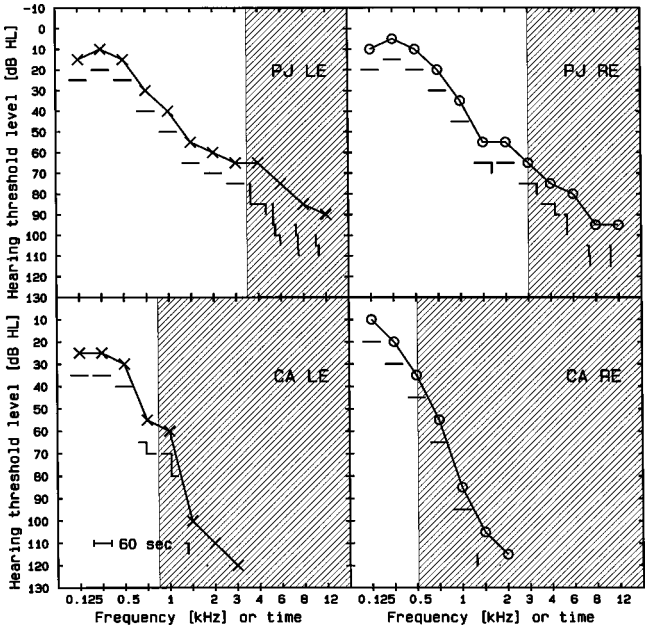


FIG. 3. Results for two subjects (four ears) with high-frequency hearing loss and a high-frequency dead region. Circles/crosses connected by lines show the mean of three measurements of audiometric thresholds. A shaded area indicates the presence and extent of a dead region. Short lines without symbols show tone-decay results (the means taken from three measurements). The center of the uppermost horizontal line indicates the frequency of the test tone. The left-hand end of each short horizontal line corresponds to the time of the start of the presentation. The mean audible duration of the tone (across the three runs) is indicated by the length of the horizontal part of the line; 0.8 of a division on the abscissa corresponds to 60 s, as indicated by the key. Vertical segments indicate that the tone became inaudible within 60 s and that the tone's level was subsequently increased (see the text for details).

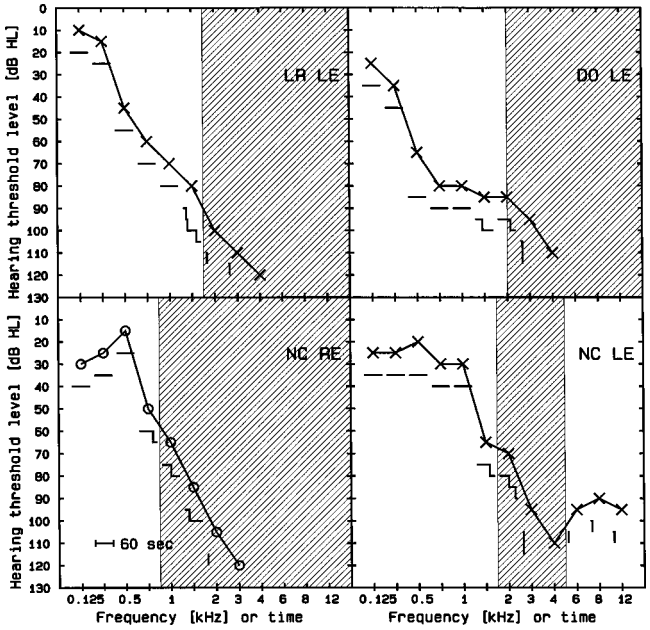


FIG. 4. As Fig. 3, but showing results for four more ears with dead regions.

within 60 s and that the tone's level was consequently increased. If a tone had decayed at a given level, the time it was audible at the next higher level is represented by adding the lengths of the horizontal lines for those two levels. A tone with a higher level was never perceived to decay faster than at a lower level. If a tone decayed within the same time for consecutive levels, this is indicated by a line whose vertical extent exceeds 5 dB. For example, if tone decay occurred after 10 s for three consecutive levels separated by 5 dB, this would be indicated by a vertical line with a length corresponding to 15 dB, positioned at 10 s (0.133 of a division) relative to start point (the left-hand end) of the line. In cases where the entire measurement could be carried out using acceptable tone levels, the addition of all the horizontal lines for a given frequency results in 0.8 of a division, corresponding to 60 s.

Figure 3 displays results for two subjects (four ears) selected to show that tone decay can occur within and close to f_e, even when f_e is quite low. Both subjects had high-frequency hearing loss and a high-frequency dead region in both ears. Circles (right ear)/crosses (left ear) connected by lines show the mean of three measurements of audiometric thresholds. A shaded area indicates the presence and extent of a dead region. Both subjects experienced tone decay in both ears for frequencies close to their individual upper limit of hearing. Subject PJ has a high-frequency dead region in the left ear with $f_e = 3.5$ kHz. For frequencies up to 3 kHz, no tone decay was measured. However, marked tone decay was found for all frequencies within the dead region; the audible duration of tones initially presented at a level of 10 dB SL was less than 10 s. The perception of tones with frequencies of 6, 8, and 12 kHz was not maintained for 60 s at levels 20 to 30 dB above absolute threshold. The results for the right ear showed a similar pattern. Tone decay occurred for all tones with frequencies above f_e (3 kHz). However, even at 1.5 kHz, one octave below f_e, tone decay occurred. Subject CA has a high-frequency dead region in the left ear with $f_e = 0.85$ kHz. For this ear, tone decay occurred for all frequencies above $0.5 f_e$. In the right ear, f_e is about 0.5 kHz. Tone decay for this ear occurred only for frequencies well inside the dead region (1.5 and 2 kHz). Overall, these results are broadly consistent with the restricted excitation pattern hypothesis, as tone decay mostly occurred for frequencies within or close to f_e.

Figure 4 shows the results for three hearing-impaired subjects (four ears), whose results were also broadly consistent with the restricted excitation pattern hypothesis. All these subjects had dead regions, and all experienced tone decay consistently for tones with frequencies within the dead region and up to one octave below f_e (DO, left ear). However, for NC (left ear), who had a dead region of limited extent, tone decay was similar for frequencies falling inside and above the dead region.

Figure 5 shows results for three subjects (four ears) without dead regions for whom tone decay was observed. In these cases, there should be at least *some* remaining IHC/neural function over the whole frequency range tested. Therefore, the tone's excitation pattern should not be severely restricted. The finding of tone decay is inconsistent with the hypothesis that a restricted excitation pattern is required for tone decay to occur. For all four ears, tone decay was found only for frequencies at which the absolute threshold exceeded 40 dB HL.

The results in Fig. 6 show that there is not a consistent association between the presence of dead regions and tone

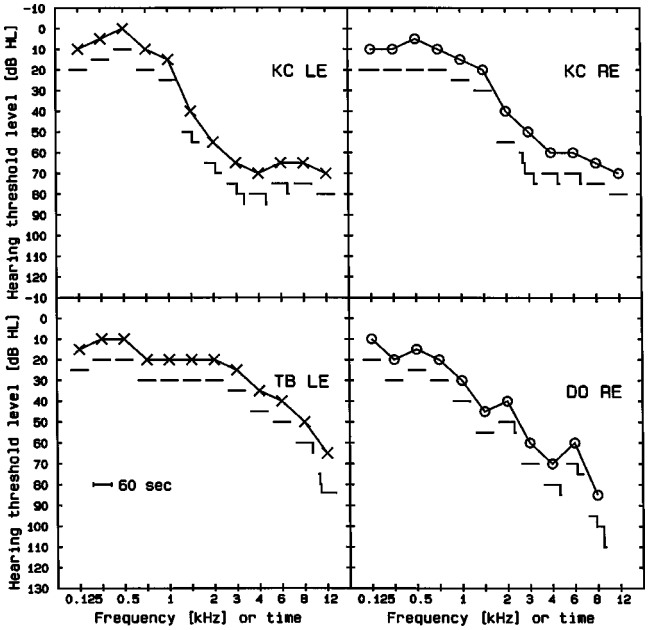


FIG. 5. As Fig. 3, but showing results for four ears without a dead region.

decay. For all of the ears shown, dead regions were present, but only one of the ears (LR, right ear) showed any tone decay. This is inconsistent with the restricted excitation pattern hypothesis. Note that the audiograms of subjects JM (both ears) and RC (left ear) are steeply sloping. The absolute thresholds for frequencies just below f_e were always equal to or below 40 dB for subject JM and 10 dB for subject RC. If pure tones with frequencies falling in the dead region are detected via IHCs/neurons just outside the boundary of the dead region, then these tones would be transduced by IHCs originating from an area where the cochlea is relatively healthy. For subject LR, the absolute threshold for frequencies just below f_e was about 70 dB HL. The IHCs transducing the excitation or neurons signaling the excitation for tones falling into the dead region were probably functioning less well than for JM and RC. This may explain why LR shows tone decay, but JM and RC do not. We will return to this point later.

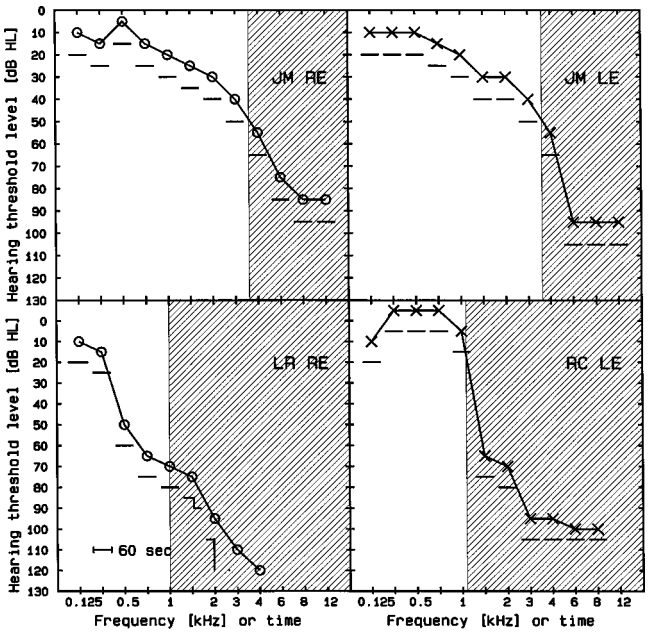


FIG. 6. As Fig. 3, but showing results for four ears with dead regions, but mostly not showing tone decay.

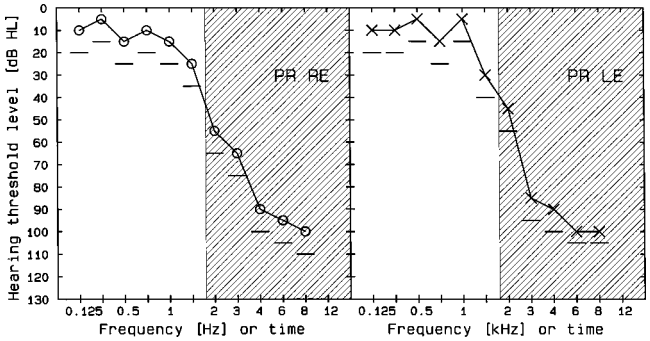


FIG. 7. As Fig. 3, but showing results for two ears with dead regions, but no tone decay.

The results in Fig. 7 again show no consistent association between the presence of dead regions and tone decay. For the two ears displayed, dead regions were present, but neither showed any tone decay. This is inconsistent with the restricted excitation pattern hypothesis. Note that the audiograms of subject PR (both ears) are particularly steeply sloping. The absolute thresholds for frequencies just below f_e were equal to or better than 30 dB.

Figure 8 shows results for two subjects (four ears) without dead regions, and with audiograms showing notches or dips. One subject (MR) did not show tone decay for any frequency, even for frequencies where the absolute threshold was as high as 80 dB HL. The other subject showed tone decay for frequencies just above the notch in the audiogram, at which the audiometric threshold was normal (0–5 dB HL).

In summary nine of the 14 ears with dead regions showed tone decay and five did not, for any frequency. Six

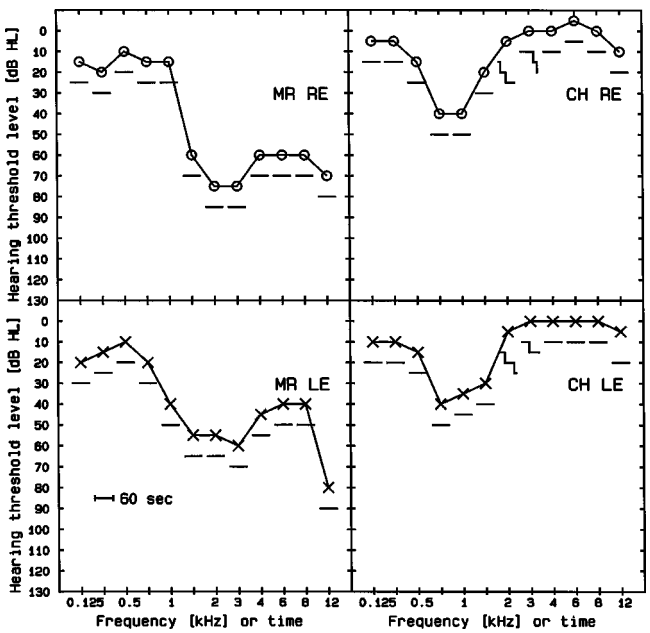


FIG. 8. As Fig. 3, but showing results for four ears (two subjects) without a dead region but with ''notches'' in their audiograms.

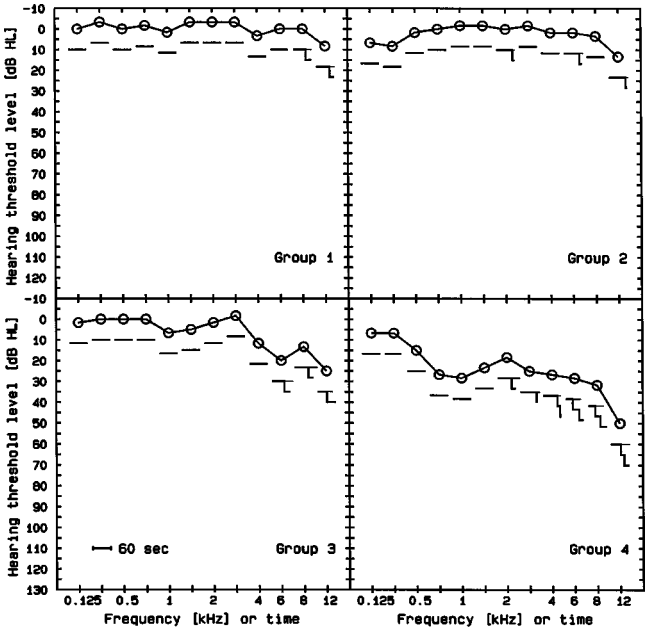


FIG. 9. Results for four groups of subjects, all with normal or near-normal hearing at low frequencies, but with varying degrees of hearing loss at high frequencies. Mean audiometric thresholds are shown by circles. The high-frequency loss increased from group 1 to group 4. The thin extensions to the thick horizontal lines indicate the standard deviation of the audible duration across subjects.

out of eight ears with moderate to severe hearing loss without a dead region showed tone decay and two did not, for any frequency. Thus, tone decay is neither always associated with a dead region in the cochlea nor always absent when there is remaining IHC/neural function over the whole range of audiometric frequencies. Tone decay seems not to be predictable from the audiogram alone. However, the results of the hearing-impaired subjects showed an increase in probability of tone decay with increasing absolute threshold when the latter exceeded about 50 dB HL. We return to this point later.

B. Tone decay for subjects with normal hearing or mild hearing loss

Figure 9 shows group results for subjects with normal hearing or mild hearing loss. Groups 1–3 are as defined in Table II. These groups are distinguished only by their absolute threshold at high frequencies, which increases across the groups. Group 4 was composed of subject BM (Table II) plus subjects CH and TB from the hearing-impaired subjects described earlier (Table I). None of the subjects with normal or near-normal hearing (groups 1–3) experienced tone decay for frequencies below 1.5 kHz. Also, the adaptation effects displayed in the upper right panel at 2 and 6 kHz were due to one subject's responses only. Across the three groups, the majority of the subjects experienced tone decay only for the very high frequencies and only for the lowest sensation level (10 dB SL); following the increase of the test tone level to 15 dB SL, the sensation persisted for 60 s for all subjects. The fourth group, with mild to moderate midfrequency and/or high-frequency hearing loss, showed more tone decay for mid- to high-frequency tones.

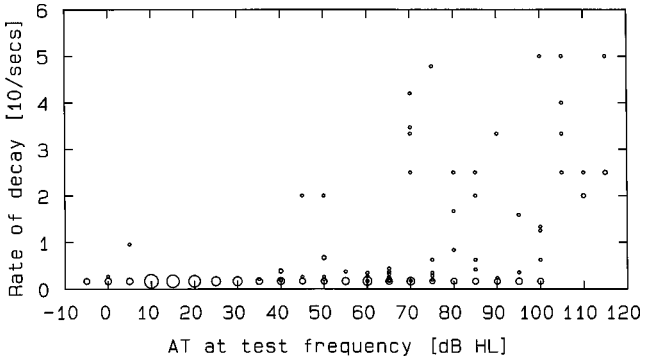


FIG. 10. The rate of tone decay (expressed as 10 divided by the audible duration of the 10-dB SL pure tone) plotted as a function of the absolute threshold at the test frequency for the 11 hearing-impaired subjects. The data are collapsed across all frequencies tested. The area of each point is proportional to the number of cases falling at that position.

The present data suggest that there is a relation between absolute threshold and tone decay for hearing-impaired subjects. This conflicts with the statement of Scharf (1983) that no relation had been found between the degree to which a person adapts, and individual characteristics such as absolute threshold, age, and sex. The present results support the finding of Miśkiewicz *et al.* (1993) that adaptation magnitude is related to absolute threshold for high frequencies.

To obtain a convenient measure of the rate of tone decay, we took the reciprocal of the audible duration of a 10-dB SL tone, and multiplied it by 10. If the tone did not decay to inaudibility within 60 s, the measure was taken as 10/60. Figure 10 shows the rate of tone decay plotted as a function of the absolute threshold at the signal frequency, for the 11 hearing-impaired subjects. Results are collapsed across signal frequency. The area of each circle is proportional to the number of measurements falling at that point. Tone decay was very rarely experienced for tones giving absolute thresholds below about 50 dB HL. Once the absolute threshold reached about 50 dB HL, the proportion of cases of tone decay increased with increasing absolute threshold. Note the huge range of different rates of tone decay for a specific absolute threshold. The scatter may occur because a given degree of hearing loss can have different physiological causes. Note also that, even when the absolute threshold is as high as 90–100 dB HL, there are cases where no tone decay occurs.

As mentioned earlier, when the frequency of a test tone falls within a dead region, the presence or absence of tone decay may be determined by the status of IHCs immediately adjacent to the dead region, assuming that the tone is detected via those IHCs. To assess this idea, for each tone frequency, f_t, that fell within a dead region, we determined the audiometric threshold for the test frequency, f_o, just outside the dead region. The rate of tone decay for f_t was then plotted against the absolute threshold for f_o. Figure 11 shows the results of this analysis, but includes cases where the test frequency did not fall inside a dead region (for those cases, as in Fig. 10, the rate of decay at the test frequency is plotted against the audiometric threshold at that frequency). When the rate of decay is plotted as a function of absolute threshold at the "detection frequency," there are relatively

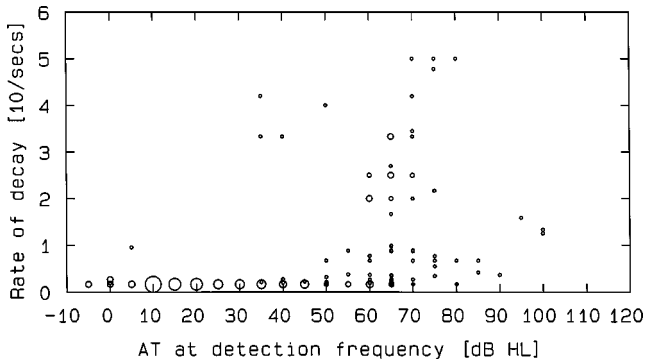


FIG. 11. As Fig. 10, except that, when the frequency of the test tone fell within a dead region, the absolute threshold was taken for the frequency just below the edge frequency of the dead region.

few cases where the threshold is higher than 50 dB HL but there is no tone decay. In other words, when the absolute threshold is high (≥50 dB HL) for the frequency region *where the test tone is likely to be detected*, then the proportion of cases where tone decay occurs is relatively high (0.63). Since Fig. 11 appears to show a kind of ''breakpoint'' at around 50 dB HL, we use that value to classify the data for subsequent analyses.

Table III shows the number of cases of tone decay, collapsed across ears and test frequencies, for various different methods of classification. This was done to test three different hypotheses. The first hypothesis is the restricted excitation pattern hypothesis, which leads to the prediction that tone decay should occur consistently for tones whose frequencies fall in a dead region or somewhat below f_e, and not otherwise. We chose to define ''somewhat below'' as an octave below, since, according to the model presented by Moore and Glasberg (1997), the audible excitation evoked by a tone at 10 dB SL in an ear with moderate hearing loss extends to center frequencies about one octave above the frequency of the tone. Part A1 of the table shows the extent to which the prediction is true. For frequencies within a dead region or between $0.5f_e$ and f_e, tone decay occurred in 36 cases out of 81. When there was not a dead region or the test frequency was below $0.5f_e$, tone decay occurred in 22 cases out of 152. The difference in incidence was statistically significant ($\chi^2=23.82$,df$=1$,$p<0.001$).

These results suggest that the restricted excitation pattern hypothesis might be partly correct. However, in this classification method, the presence or absence of dead regions is confounded with the absolute threshold at the test frequency; dead regions are associated with high thresholds. If the restricted excitation pattern hypothesis is correct, it should apply even when the analysis is restricted to cases where the absolute threshold at the test frequency was above 50 dB HL. Part A2 of Table III is similar to part A1, but with results restricted to cases where the absolute threshold at the test frequency was above 50 dB HL. When the test frequency fell within a dead region or between $0.5f_e$ and f_e, tone decay occurred in 34 cases out of 67. When there was not a dead region or the test frequency was below $0.5f_e$, tone decay occurred in 14 cases out of 35. The difference in incidence was not statistically significant ($\chi^2=0.67$,df$=1$,$p>0.05$). This is not consistent with the restricted excitation pattern hypothesis.

The second hypothesis is that tone decay is related to the audiometric threshold at the test frequency. To test this, the results were subdivided according to whether the absolute threshold for frequencies showing adaptation was below (better than) 50 dB HL or equal to or above 50 dB HL. Part B of the table shows that tone decay occurred in 6 out of 132 cases when the absolute threshold was less than 50 dB HL, and in 48 out of 101 cases when the absolute threshold was

TABLE III. Incidence of tone decay, collapsed across ears and frequencies: (A1) According to whether the test frequency fell within or just below a dead region (DR); (A2) As A1, but restricted to frequencies for which the absolute threshold (AT) was ≥50 dB HL; (B) According to whether ATs at the test frequency were below or above 50 dB HL; and (C) According to whether ATs in the frequency region where the tone was detected were below or above 50 dB HL. For tone frequencies falling in a dead region, the relevant AT was taken as the value for a frequency just below f_e. See the text for details.

Hypothesis		Within DR	Not within DR	χ^2	Significance
(A1)	Tone decay	36	22		
	No tone decay	45	130		
	Totals	81	152	23.82	$p<0.001$
Hypothesis		Within DR	Not within DR	χ^2	Significance
(A2)	Tone decay	34	14		
	No tone decay	33	21		
	Totals	67	35	0.67	$p>0.05$ (n.s.)
Hypothesis		<50 dB HL	≥50 dB HL	χ^2	Significance
(B)	Tone decay	6	48		
	No tone decay	126	53		
	Totals	132	101	56.96	$p<0.001$
Hypothesis		<50 dB HL	≥50 dB HL	χ^2	Significance
(C)	Tone decay	8	47		
	No tone decay	151	27		
	Totals	159	74	92.54	$p<0.001$

equal to or greater than 50 dB HL. The incidence of tone decay was significantly greater when the hearing loss exceeded 50 dB HL at the test frequency ($\chi^2=56.96$,df$=1$,$p<0.001$), which is broadly consistent with the hypothesis.

The third hypothesis is that tone decay is related to the absolute threshold in the frequency region where the tone is detected. When the frequency of the tone falls in a dead region, this corresponds to the absolute threshold for a frequency adjacent to the dead region, f_o. Part C of the table shows that tone decay occurred in 8 out of 159 cases when the absolute threshold at the ''detection frequency'' was less than 50 dB HL and in 47 out of 74 cases when the absolute threshold was equal to or greater than 50 dB HL. The occurrence of tone decay was significantly greater when the hearing loss exceeded 50 dB HL at the ''detection frequency'' than when it did not ($\chi^2=92.54$, df$=1$, $p<0.001$).

To assess the relative merit of the three hypotheses, we compared the number of cases predicted by each one. The first, second, and third hypotheses (parts A1, B, and C of Table III) predicted 166, 174, and 198 cases out of 233, respectively. The proportion of cases predicted was significantly greater for the third than for the first hypothesis ($\chi^2=12.06$,df$=1$,$p<0.001$) and for the third than for the second hypothesis ($\chi^2=7.05$,df$=1$,$p<0.01$).

IV. DISCUSSION

We have argued that our results are not generally consistent with the restricted excitation pattern hypothesis proposed by Scharf (1983). However, the hypothesis as originally proposed is somewhat more complex than described in later papers (Miskiewicz *et al.*, 1993; Bacon and Viemeister, 1994; Viemeister *et al.*, 1985) and in the Introduction. Scharf (1983) suggested that for the percept of a steady tone to be maintained there should be ''variations in the pattern of excitation across fibers,'' and that this occurs ''at higher levels where the pattern evoked by a tone is widespread.'' In other words, differences in responses across neurons are important for the percept to be maintained. We propose here an extension of that idea: the critical factors are the number, distribution, and timing of nerve spikes (across CF) evoked by the tone.

Tone decay may be greater for high than for low frequencies because, for the former, phase-locking cues are not available; decay of neural firing rate will lead to decay of sensation. For low frequencies, phase locking to the signal frequency will occur, and this, plus the contrast in phase locking across CFs, may help to maintain a stable percept (Moore and Peters, 1997).

For subjects with hearing impairment, other factors may play a role. Consider first subjects without any dead region. For a frequency where the hearing loss is less than about 50 dB, the hearing loss may arise mainly from OHC dysfunction (Moore, 1998). The number of spikes evoked by a tone at 10 dB SL would probably be close to normal, and neural responses would not be expected to show abnormal adaptation. Under these conditions, the hearing-impaired listener may behave almost like a normal-hearing listener, and show little tone decay.

For a frequency where the hearing loss exceeds 50 dB, IHC dysfunction is probably involved. As noted earlier, Moore and Glasberg (1997) suggested that, for moderate hearing losses, the proportion of hearing loss due to IHC damage was typically around 20%. Thus, for an overall loss of 50 dB, the loss due to IHC damage would typically be 10 dB, implying that the input to the IHCs needs to be ten times as intense as normal to evoke the same neural response at threshold. Such a degree of IHC damage may be sufficient to lead to increased tone decay. In addition, for low-frequency tones, the pattern of phase locking may be disrupted, or the temporal distribution of spikes may be more sparse (Miller *et al.*, 1997), and this may lead to tone decay even for low-frequency tones, as we sometimes observed.

Consider now subjects with dead regions. When the frequency of a tone falls within a dead region, the absolute threshold at the test frequency does not indicate the functional status of the IHCs/neurons responsible for detection of the test tone. Rather, as argued above, the test tone is detected via IHCs/neurons close to the boundary of the dead region, and it is the absolute threshold close to the boundary that indicates the status of the relevant IHCs/neurons. When the absolute threshold close to the boundary of the dead region was greater than 50 dB HL, tone decay was commonly observed for tones whose frequencies fell within or just below the boundary of the dead region. This is consistent with the pattern of results for subjects without dead regions, and suggests that tone decay will occur when the IHCs/neurons just below the boundary of the dead region are poorly functioning. In contrast, when the absolute threshold close to the boundary of the dead region was less than 50 dB HL, tone decay was not usually observed.

If it is the case that tone decay depends upon the functional status of the IHCs/neurons responsible for detection of a tone, then, at first sight, it might be expected that tone decay would be the same for all tones falling very close to f_e or within a dead region, since all such tones would be detected at the same place. However, the data show several cases where this does not occur. Often, tone decay is greater for tones falling well inside a dead region than for tones close to f_e; see for example, the results for PJ, left and right ear (Fig. 3) and LR, right ear (Fig. 6). There are a number of possible explanations for this. One is connected with the possibility, discussed above, that phase locking helps to prevent tone decay; tone decay may be more marked for frequencies where phase locking is less precise, i.e., higher frequencies (Palmer and Russell, 1986). Frequencies well inside a dead region are also higher frequencies, and this might account for the increased tone decay in such cases. In addition, it is possible that phase-locking information can only be ''decoded'' appropriately when it occurs roughly in the right place within the cochlea; in other words, there needs to be a correspondence between place and temporal information for the temporal information to be extracted (Evans, 1978; Loeb *et al.*, 1983; Moore and Carlyon, 2003; Shamma and Klein, 2000). For a tone falling well inside a dead region, the discrepancy between the place and the temporal information may be too large for the temporal information to be decoded (Moore, 2001).

A second possible explanation for the increase in tone decay for tones with frequencies falling well inside a dead region is connected with the shape of the excitation pattern in the region where the tones are detected; this region is approximately triangular in shape. As the frequency of a tone is moved further inside a dead region, keeping its sensation level constant, the triangle increases in height, but decreases in horizontal extent (see Fig. 2 of Moore, 2001, and Fig. 8 of Moore and Alcántara, 2001). Thus, for tones with frequencies well inside a dead region, the number of IHCs/neurons involved in detecting the tone decreases, and this may lead to greater tone decay.

In summary, we have proposed an extension of Scharf's (1983) restricted excitation pattern hypothesis: we propose that the critical factor determining tone decay is the number, distribution, and timing of nerve spikes (across CF) evoked by the tone. We argue that the status of the IHCs/neurons responsible for the detection of the tone is the critical factor. Tone decay may occur when the IHCs/neurons are functioning poorly, and IHC/neural dysfunction is typically associated with audiometric thresholds exceeding 50 dB HL. For a person with a dead region, the relevant audiometric threshold is the one for a signal frequency just below f_e. This extended hypothesis is consistent with most of our results; tone decay is usually observed when the audiometric threshold in the frequency region where the tone is detected is greater than 50 dB HL, and is usually not observed when the audiometric threshold is below 50 dB HL. Thus, the degree of tone decay may provide a measure of the functional status of the IHCs/neurons responsible for detecting the test tone.

The results for two subjects appear inconsistent with our hypothesis. One subject (CH) showed tone decay for a frequency where the audiometric threshold was normal. This person has a genetic hearing loss and it is possible that this is associated with IHC damage (but not a dead region). The other subject (MR) showed no tone decay for frequencies where the audiometric threshold exceeded 50 dB HL. We have no explanation for the lack of tone decay for this subject.

V. CONCLUSIONS

(1) For subjects with cochlear hearing loss, tone decay is not restricted to high frequencies. Some of the moderately to severely hearing-impaired subjects experienced tone decay for pure tones with frequencies below 1 kHz, and when those tones were presented at levels up to 30 dB SL.
(2) No consistent association was found between the degree of tone decay and the presence of a dead region. Tone decay was not always observed for tone frequencies that fell within or close to a dead region, and tone decay did sometimes occur for subjects without dead regions. The results do not support the idea that, for hearing-impaired subjects, the key factor producing tone decay is a restricted spread of audible excitation. A restricted spread of the tone's excitation pattern was neither necessary nor sufficient for tone decay to occur.
(3) For subjects with normal or near-normal hearing at low and medium frequencies, there was a trend for greater tone decay to be associated with greater hearing loss at high frequencies.
(4) For moderately to severely hearing-impaired subjects, the incidence and extent of tone decay increased with increasing hearing loss above 50 dB HL.
(5) The predictability of tone decay from the audiogram alone is poor. However, there does seem to be a reasonably orderly relationship between tone decay and the audiometric threshold in the frequency region where the tone is detected; for subjects with dead regions, this corresponds to the frequency just below the edge of the dead region. The results were consistent with the hypothesis that absolute thresholds above about 50 dB HL in the frequency region where the tone is detected are associated with an increased probability of tone decay.
(6) ''Abnormal'' tone decay occurred quite often for our hearing-impaired subjects, even though there was no evidence that they had a retrocochlear disorder. Thus, the occurrence of tone decay does not necessarily indicate a tumor pressing on the auditory nerve.

ACKNOWLEDGMENTS

We thank Michael Stone for helpful comments over the whole course of this study. We also thank Brian Glasberg, Tom Bear, Sheila Flanagan, Geoffrey Moore, and Thomas Stainsby for helpful discussions, Brian Glasberg for help with statistics, and Marjorie Leek and two anonymous reviewers for helpful comments on an earlier version of this paper. This work was supported by the BBSRC and the MRC (UK).

Bacon, S. P., and Viemeister, N. F. (**1994**). ''Intensity discrimination and increment detection at 16 kHz,'' J. Acoust. Soc. Am. **95**, 2616–2621.
Bray, D. A., Dirks, D. D., and Morgan, D. E. (**1973**). ''Perstimulatory loudness adaptation,'' J. Acoust. Soc. Am. **53**, 1544–1548.
Carhart, R. (**1957**). ''Clinical determination of abnormal auditory adaptation,'' Arch. Otolaryngol. **65**, 32–39.
Carhart, R., and Jerger, J. F. (**1959**). ''Preferred method for clinical determination of pure-tone thresholds,'' J. Speech Hear Disord. **24**, 330–345.
Davis, J. M., and Weiler, E. (**1976**). ''Monaural auditory adaptation as measured by simple reaction time,'' Br. J. Audiol. **10**, 102–106.
Egan, J. P. (**1955**). ''Perstimulatory fatigue as measured by heterophonic loudness balances,'' J. Acoust. Soc. Am. **27**, 111–120.
Evans, E. F. (**1978**). ''Place and time coding of frequency in the peripheral auditory system: Some physiological pros and cons,'' Audiology **17**, 369–420.
Feaster, S. A., and Weiler, E. M. (**1975**). ''The effects of monaural auditory adaptation in the speech range,'' Br. J. Audiol. **9**, 81–83.
Green, D. S. (**1985**). ''Tone decay,'' in *Handbook of Clinical Audiology*, edited by J. Katz (Williams & Wilkins, New York).
Hellman, R. P., Miskiewicz, A., and Scharf, B. (**1997**). ''Loudness adaptation and excitation patterns: Effects of frequency and level,'' J. Acoust. Soc. Am. **101**, 2176–2185.
Hood, J. D. (**1950**). ''Studies in auditory fatigue and adaptation,'' Acta Oto-Laryngol., Suppl. **92**, 1–57.
Javel, E. (**1996**). ''Long-term adaptation in cat auditory-nerve fiber responses,'' J. Acoust. Soc. Am. **99**, 1040–1052.
Jerger, J. F. (**1957**). ''Auditory adaptation,'' J. Acoust. Soc. Am. **29**, 357–363.
Loeb, G. E., White, M. W., and Merzenich, M. M. (**1983**). ''Spatial cross correlation: A proposed mechanism for acoustic pitch perception,'' Biol. Cybern. **47**, 149–163.

Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D. (**1997**). ''Effects of acoustic trauma on the representation of the vowel /ε/ in cat auditory-nerve fibers,'' J. Acoust. Soc. Am. **101**, 3602–3616.
Mińkiewicz, A., Scharf, B., Hellman, R., and Meiselman, C. (**1993**). ''Loudness adaptation at high frequencies,'' J. Acoust. Soc. Am. **94**, 1281–1286.
Moore, B. C. J. (**1998**). *Cochlear Hearing Loss* (Whurr, London).
Moore, B. C. J. (**2001**). ''Dead regions in the cochlea: Diagnosis, perceptual consequences, and implications for the fitting of hearing aids,'' Trends Amplif. **5**, 1–34.
Moore, B. C. J., and Alcántara, J. I. (**2001**). ''The use of psychophysical tuning curves to explore dead regions in the cochlea,'' Ear Hear. **22**, 268–278.
Moore, B. C. J., and Carlyon, R. P. (**2003**). ''Perception of pitch by people with cochlear hearing loss and by cochlear implant users,'' in *Pitch Perception*, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Springer, New York) (in press).
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., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcántara, J. I. (**2000**). ''A test for the diagnosis of dead regions in the cochlea,'' Br. J. Audiol. **34**, 205–224.
Moore, B. C. J., and Peters, R. W. (**1997**). ''Detection of increments and decrements in sinusoids as a function of frequency, increment and decrement duration and pedestal duration.,'' J. Acoust. Soc. Am. **102**, 2954–2965.
Palmer, A. R., and Russell, I. J. (**1986**). ''Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells,'' Hear. Res. **24**, 1–15.
Palva, T., and Kärjä, J. (**1969**). ''Suprathreshold auditory adaptation,'' J. Acoust. Soc. Am. **45**, 1018–1021.
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.
Scharf, B. (**1983**). ''Loudness adaptation,'' in *Hearing Research and Theory*, edited by J. V. Tobias and E. D. Schubert (Academic, New York), Vol. 2.
Schmiedt, R. A. (**1996**). ''Effects of aging on potassium homeostasis and the endocochlear potential in the gerbil cochlea,'' Hear. Res. **102**, 125–132.
Schuknecht, H. F. (**1994**). ''Auditory and cytocochlear correlates of inner ear disorders,'' Otolaryngol.-Head Neck Surg. **110**, 530–538.
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.
Viemeister, N. F., and Bacon, S. P. (**1983**). ''Increment detection, intensity discrimination, and tone decay at high frequencies,'' J. Acoust. Soc. Am. Suppl. 1 **73**, S93.
Viemeister, N. F., Javel, E., and Bacon, S. P. (**1985**). ''Psychophysical and psychological correlates of high-frequency tone decay,'' Abstract from the 8th Midwinter Meeting Assoc. Res. Otolaryngol.
Walsh, E. J., McGee, J., and Javel, E. (**1984**). ''Auditory nerve responses following middle-ear disarticulation in kittens,'' J. Acoust. Soc. Am. **75**, 13.
Ward, W. D. (**1973**). ''Adaptation and fatigue,'' in *Modern Developments in Audiology*, edited by J. Jerger (Academic, New York).
Weiler, E. M., and Gross, L. F. (**1976**). ''A positive measure of monaural heterophonic auditory adaptation,'' J. Audit. Res. **16**, 20–23.
Wright, H. N. (**1959**). ''Auditory adaptation in noise,'' J. Acoust. Soc. Am. **31**, 1004–1012.
Yates, G. K., Johnstone, B. M., Patuzzi, R. B., and Robertson, D. (**1992**). ''Mechanical preprocessing in the mammalian cochlea,'' Trends Neurosci. **15**, 57–61.

Temporal masking of multidimensional tactual stimuli

Hong Z. Tan[a)]
Haptic Interface Research Laboratory, Electrical Engineering Building, 465 Northwestern Avenue, Purdue University, West Lafayette, Indiana 47907

Charlotte M. Reed, Lorraine A. Delhorne, Nathaniel I. Durlach, and Natasha Wan
Research Laboratory of Electronics, Room 36-751, 77 Massachusetts Avenue, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 10 January 2003; revised 22 August 2003; accepted 15 September 2003)

Experiments were performed to examine the temporal masking properties of multidimensional tactual stimulation patterns delivered to the left index finger. The stimuli consisted of fixed-frequency sinusoidal motions in the kinesthetic (2 or 4 Hz), midfrequency (30 Hz), and cutaneous (300 Hz) frequency ranges. Seven stimuli composed of one, two, or three spectral components were constructed at each of two signal durations (125 or 250 ms). Subjects identified target signals under three different masking paradigms: forward masking, backward masking, and sandwiched masking (in which the target is presented between two maskers). Target identification was studied as a function of interstimulus interval (ISI) in the range 0 to 640 ms. For both signal durations, percent-correct scores increased with ISI for each of the three masking paradigms. Scores with forward and backward masking were similar and significantly higher than scores obtained with sandwiched masking. Analyses of error trials revealed that subjects showed a tendency to respond, more often than chance, with the *masker*, the *composite* of the masker and target, or the combination of the target and a *component* of the masker. The current results are compared to those obtained in previous studies of tactual recognition masking with brief cutaneous spatial patterns. The results are also discussed in terms of estimates of information transfer (IT) and IT rate, are compared to previous studies with multidimensional tactual signals, and are related to research on the development of tactual aids for the deaf.
[DOI: 10.1121/1.1623788]

PACS numbers: 43.66.Dc, 43.66.Wv, 43.66.Ts [MRL] Pages: 3295–3308

I. INTRODUCTION

A goal of recent research on the development of tactual aids for the deaf and deaf–blind is to demonstrate communication rates comparable to those observed for natural methods of tactual communication (e.g., see Reed, Durlach, and Delhorne, 1992). One approach to improving the information-transfer rate capacity of artificial tactual devices is through modifications designed to increase the perceptual richness or dimensionality of the output display (e.g., see Reed *et al.*, 1985; Leotta *et al.*, 1988; Eberhardt *et al.*, 1994; Tan and Rabinowitz, 1996; Rinker, Craig, and Bernstein, 1998; Tan *et al.*, 1999). In addition to providing stimulation of the cutaneous component of the tactual sensory system, similar to most older tactile devices (e.g., see reviews by Kirman, 1973; Reed, Durlach, and Braida, 1982; Reed *et al.*, 1989; Bernstein, 1992), the more recent displays have also incorporated stimulation of the kinesthetic component of the tactual system.[1] The design of these displays is motivated in part by the display characteristics of natural methods of tactual communication employed by deaf–blind individuals (Reed *et al.*, 1992), by basic principles governing information transfer (Garner, 1962), and by physiological and psychophysical studies of the separate channels of the tactual system and their interactions (Bolanowski *et al.*, 1988; Johnson, Yoshioka, and Vega-Bermudez, 2000). The current study is concerned with further investigation of the perception of multidimensional tactual signals presented through a multifinger tactual display designed to provide stimulation of the tactual sensory system along a continuum from kinesthetic movements to cutaneous vibrations (Tan, 1996; Tan and Rabinowitz, 1996).

Previous research (Tan, 1996; Tan *et al.*, 1997; Tan *et al.*, 1999) investigated information transfer (IT) for sets of multidimensional stimuli created using signal components selected from each of three perceptually distinct regions: slow motion (from dc to roughly 6 Hz), fluttering motion (in the region of roughly 10–70 Hz), and smooth vibration (above roughly 150 Hz). One or two (highly discriminable) frequencies were selected from each of these three spectral regions. Stimuli were constructed using single-frequency waveforms, double-frequency waveforms (composed of one frequency component from each of two different spectral regions), and triple-frequency waveforms (composed of one frequency component from each of three different spectral regions). Three stimulus sets were constructed at three signal durations of 125, 250, and 500 ms. (At some values of frequency and duration, amplitude and onset direction of movement were also used as distinguishing characteristics of the stimulus waveforms.) The total number of distinct waveforms was 30 for the 250-ms and 500-ms stimulus sets and 19 for the 125-ms set. In addition, each waveform could be

a)Electronic mail: hongtan@purdue.edu

presented at one of four possible locations (thumb, index finger, middle finger, or all three digits simultaneously), leading to stimulus uncertainty of 6.25 bits (at 125 ms) or 6.91 bits (at the two longer durations).

Estimates of static IT were derived from percent-correct scores (using an empirically derived formulation described by Tan *et al.*, 1999) obtained under a one-interval, forced-choice identification procedure with 120 alternatives (for the two 250- and 500-ms stimulus sets) or 76 alternatives (for the 125-ms stimulus set). Static IT, averaged across three subjects, was estimated to be 5.6 to 6.5 bits for the three stimulus sets.

Estimates of IT *rate* were obtained using a highly simplified procedure (referred to as an AXB paradigm) that required subjects to identify a signal X that was preceded by signal A and followed by signal B. These experiments employed the stimulus sets described above with the exception that waveforms were presented at one of three (rather than four) possible locations. Signals and maskers were always of equal duration, did not overlap in time, and were separated by the same value of inter-stimulus interval (ISI), that is, the time between the offset of one stimulus and the onset of the next stimulus within a given trial. Different presentation rates were achieved at each signal duration by varying the value of ISI, also expressed in terms of stimulus-onset asynchrony (SOA) which is defined as the time between the onsets of two adjacent stimuli in a given trial. IT rate in bits/s was calculated as the product of IT in bits/item (estimated from percent-correct scores) and the presentation rate in items/s (the reciprocal of the SOA). Maximum values of IT rate achieved in these previous experiments were roughly 12 bits/s and occurred at SOA values ranging from 350–500 ms (i.e., 2–3 items/s and consistent with previous observations of Garner, 1962, pp. 90–93). These estimated IT rates fall within the range of estimates of maximal IT rates for natural methods of tactual communication (see Reed and Durlach, 1998) and are among the highest observed to date with an artificial tactual display. It should be noted, however, that these estimates are based on the strong assumption that the identification performance observed in the AXB paradigm could (with sufficient training) be maintained for identifying streams of signals (such as those that would be encountered in using signals such as these to encode the acoustic speech signal). These estimates can reasonably be assumed to represent an upper bound on IT rate for continuous signals. The relation between these upper-bound estimates of IT rate and the actual IT rate that can be achieved for continuous streams of signals, however, requires further experimental study.

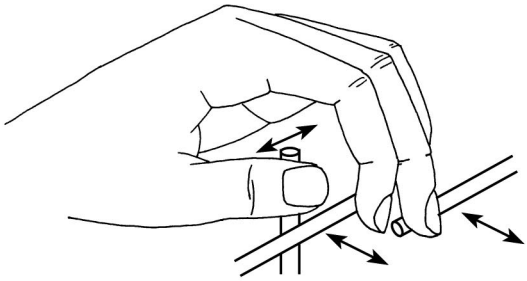
The current research is concerned with further investigation of the interactions between adjacent signals in the identification of multidimensional tactual stimuli presented in sequences. In the AXB paradigm used by Tan *et al.* (1999) to derive estimates of IT rate, identification of the target signal X was based on the combined effects of a forward masker (A) and a backward masker (B) at stimulation sites that were selected randomly (from a set of three possible sites) for each target and masker. The purpose of the current study was to investigate the separate effects of forward and backward masking (as well as their combined effects using the AXB paradigm employed previously) on the identification of tactual stimuli at a single site (which would presumably maximize temporal masking effects). The stimuli employed in these studies were a small subset of the multidimensional stimuli developed by Tan (1996) and were presented at only one site (the left index finger). Two stimulus sets were constructed at each of two signal durations (125 and 250 ms). Each stimulus set was made up of seven waveforms: three single-frequency waveforms (one from each of the three distinct spectral regions described above), three double-frequency waveforms (resulting from all possible combinations of the three single waveforms), and one triple-frequency waveform (composed of the three single-frequency waveforms). Under each of the three masking paradigms, target identification was studied as a function of interstimulus interval for seven values in the range 0–640 ms. Targets and maskers were always selected at random from the same stimulus set. The results are summarized by functions describing overall performance in percent-correct performance and IT as a function of SOA. In addition, error trials are analyzed to gain insight into the underlying structure of the errors. The results of the current study are compared both to our own previous studies of IT and IT rate for multidimensional tactual stimuli (Tan *et al.*, 1999) and to previous studies of tactual recognition masking using cutaneous spatial-pattern displays (e.g., Craig, 1983, 1995; Evans and Craig, 1986; Evans, 1987).

II. METHODS

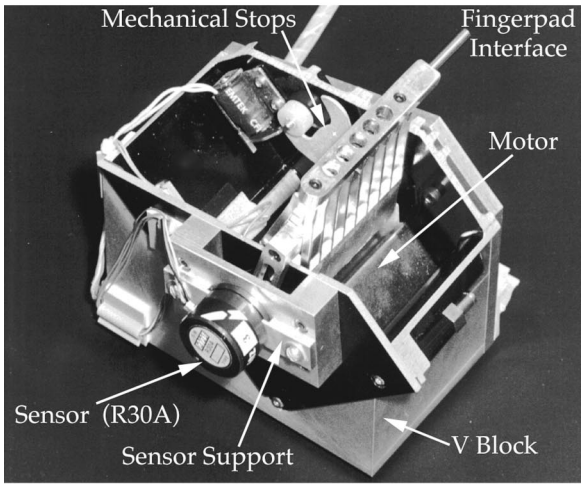
A. Apparatus

The experimental apparatus (the Tactuator) consists of three independent, point-contact, one-degree-of-freedom actuators interfaced individually with the fingerpads of the thumb, the index finger, and the middle finger [Fig. 1(a)]. The motion trajectory for the thumb is perpendicular to that of the index and middle fingers, thereby maintaining an approximately natural hand configuration with the wrist resting in its neutral position. The range of motion provided by the display for each digit is about 26 mm. All motions begin and end with each of the three digits at the middle of its respective range of motion. Each digit can thus be moved either outward (extension) or inward (flexion).

Each actuator utilizes a disk-drive head-positioning motor augmented with angular position feedback from a precision rotary variable differential transformer which has a response bandwidth (-3 dB) of 1 kHz and effectively infinite resolution due to its electromagnetic coupling [Fig. 1(b)]. Real-time positional control is provided by a digital PID controller implemented on a floating-point DSP system with 16-bit analog-to-digital and digital-to-analog converters. The overall system performance is well suited for this study for several reasons. First, each movement channel has a *continuous* frequency response from dc to 300 Hz (the disk-drive motor ceases to move beyond about 400 Hz). Therefore, the Tactuator can deliver stimulation in the kinesthetic (i.e., low-frequency gross motion) and cutaneous (i.e., high-frequency vibration) ranges as well as in the midfrequency range. Stimuli with any desired spectral components can be realized



(a)



(b)

FIG. 1. The Tactuator. (a) Schematic drawing illustrating finger placement on the Tactuator and the motion trajectories of the thumb, index, and middle fingers. (b) Photograph of one of the three associated motor assemblies with its components labeled. The subject rests the fingerpad on the rod labeled ''finger interface'' in (b).

within the frequency range of dc to 300 Hz. Second, across the frequency range of dc to 300 Hz, an amplitude of at least 47 dB SL *per frequency* can be achieved. This stimulus range is well matched to the human's sensory range in that stimulation levels exceeding 50–55-dB SL tend to induce discomfort and fatigue (Verrillo and Gescheider, 1992). Third, measurements with single- and multiple-frequency inputs at various amplitude levels indicate that each channel is highly linear, harmonic distortion is low, and interchannel crosstalk is small. This allows high-fidelity delivery of waveforms of arbitrary frequency content and stimulation level (e.g., a 30-μm high-frequency vibration superimposed on a 26-mm low-frequency motion) simultaneously to all fingers. Fourth, ''loading'' a movement channel (i.e., resting a finger lightly on the actuator's moving bar) does not significantly alter the intended stimuli. Selected measurements indicate that loading reduces the magnitude of stimulation by an average of 1.5 dB at 2 Hz, 2.7 dB at 20 Hz, and 0.1 dB at 200 Hz. This does not pose a significant problem for the experiments reported here because the stimuli are generally strong (i.e., at least 40 dB SL) and only one amplitude value is used per spectral component of a fixed frequency.

A more detailed description of the Tactuator and its performance characteristics can be found in Tan and Rabinowitz (1996).

B. Stimuli

In our earlier study (Tan *et al.*, 1999), the following stimulus attributes associated with fixed-frequency sinusoidal movement profiles were found to be effective in creating a multiattribute stimulus set with the Tactuator: frequency, amplitude, duration, site of stimulation, and signal onset direction (for gross motions only). For example, a typical stimulus employed in that study was a 125-ms 4-Hz motion at 35 dB SL delivered to the middle finger with an onset direction of extension. Signals of different frequencies were also combined to create double-frequency and triple-frequency stimuli.

Our current study employed two stimulus sets (each of which was applied only to the left index finger), referred to as the 250-ms and 125-ms stimulus sets, that are subsets of the ones used in Tan *et al.* (1999).[2] The stimulus sets were designed so that (1) subjects could learn to identify all of the stimulus alternatives easily, and (2) errors in a recognition-masking task could be attributed to masking effects rather than to the difficulties of recognizing complex stimuli. The 250-ms stimulus set contained a 2-Hz low-frequency component at 44 dB SL, a 30-Hz midfrequency component at 40 dB SL, a 300-Hz high-frequency component at 47 dB SL, and their combinations (Table I, third column). The seven stimuli in the 125-ms stimulus set were identical to those in the 250-ms set except that the 2-Hz component was replaced by 4 Hz, so that the movement started and ended at the same (resting) position (Table I, fourth column). All sensation levels were calculated from the absolute detection thresholds measured previously with the Tactuator (35, 34, 23.5, and -18 dB *re*: 1-μm peak at 2, 4, 30, and 300 Hz,

TABLE I. Signals in the 250- and 125-ms stimulus sets.

Category	Symbol[a]	Waveform at 250 ms	Waveform at 125 ms
Single-frequency	F_L	(2,44)	(4,44)
	F_M	(30,40)	(30,40)
	F_H	(300,47)	(300,47)
Double-frequency	F_L+F_M	(2,44)+(30,40)	(4,44)+(30,40)
	F_L+F_H	(2,44)+(300,47)	(4,44)+(300,47)
	F_M+F_H	(30,40)+(300,47)	(30,40)+(300,47)
Triple-frequency	$F_L+F_M+F_H$	(2,44)+(30,40)+(300,47)	(4,44)+(30,40)+(300,47)

[a]F_L, F_M, and F_H refer to low, medium, and high-frequency components, respectively. The units given within each pair of parentheses are the frequency in hertz and the amplitude in decibels (SL), respectively. The ''+'' sign indicates that the waveforms are added to form a new signal.

respectively—see Tan and Rabinowitz, 1996). The actual peak displacements at the fingertip for the four single-frequency sinusoidal signals were 8.9, 7.9, 1.5, and 0.028 mm at 2, 4, 30, and 300 Hz, respectively. All signals in both stimulus sets started in the finger-extension direction, and were delivered to the index fingertip of the left hand.

These signals could be easily discriminated and identified in isolation. The 2-Hz or 4-Hz low-frequency signal was perceived as a movement that extended the index finger and then brought it back to its resting position. The 30-Hz mid-frequency signal gave rise to a mixed flutter/rough sensation. The 300-Hz high-frequency signal felt like a smooth and penetrating vibration. When two or three of the single-frequency components were combined, the perceptual qualities associated with each component could still be discerned. For example, the percept of a 2-Hz and 30-Hz combination signal was a ''wobbly/rough'' extension and flexion of the index finger. The fact that the individual signal components retained their perceptual distinctiveness is consistent with the idea that they evoke separate mechanoreceptor channels in the skin surface (Bolanowski *et al.*, 1988; Johnson *et al.*, 2000).

C. Responses

In consideration of stimulus–response compatibility, response codes based on graphic icons, similar to those used by Tan *et al.* (1999), were employed in this study. The icon layout was placed on a digitizing tablet. Subjects used a stylus to press the appropriate icon as a response, followed by an ''ENTER'' icon. A ''DEL'' icon was available for deleting a response entered by mistake. The subject then reselected the intended response, followed by ''ENTER.'' The same response setup was used for both the 250-ms and the 125-ms stimulus sets.

D. Subjects

Three subjects participated in the experiments. S1 had participated as a subject in pilot studies that involved training on the identification of seven stimuli (similar to those used in this study) at a duration of 500 ms. S2 had previous experience in a different type of tactual experiment. S3 had participated in a previous study on the information transmission capabilities of the Tactuator (Tan *et al.*, 1999) and other tactual perception studies. S2 and S3 were paid for their participation in this study. None of the subjects reported any known tactual impairments of their hands.

E. Masking paradigms

Three masking paradigms were employed in this study: forward, backward, and ''sandwiched'' masking. In the forward-masking paradigm, the forward masker (M_F) is presented before the target signal (T). In the backward-masking paradigm, the backward masker (M_B) is presented after the target signal (T). In the ''sandwiched''-masking paradigm, the target (T) is sandwiched between a forward masker (M_F) and a backward masker (M_B). Interstimulus interval (ISI) is defined as the time between the offset of one stimulus (either a target or a masker) and the onset of the next stimulus in a given trial. Stimulus-onset asynchrony (SOA) is defined as the time between the onset of the target and the onset of the masker, with positive values of SOA indicating that the target preceded the masker (backward masking) and negative values of SOA indicating that the masker preceded the target (forward masking). In this study, the duration of the target and masker(s) was always kept the same. The maskers were selected from the same stimulus set as the stimuli. On a particular trial, each target and masker signal was selected from the seven stimuli in the 250-ms (or 125-ms) stimulus set with equal *a priori* probabilities using randomization with replacement.

F. Procedures

Prior to the masking experiment, subjects were trained to identify the seven target signals employed at each of the two durations. Training was conducted using a one-interval, seven-alternative, forced-choice procedure with trial-by-trial correct-answer feedback. The goal of this training was to familiarize the subjects with the signals and to produce near-perfect identification of signals presented in isolation. Different amounts of training were required by each of the three subjects at each duration, dependent primarily on their previous experience with the signals. Both S1 and S3 had previous experience with stimuli similar to those used in the current study, and thus required less training than S2. S1 received 10 runs (140 trials per run) with the 250-ms signals and 9 runs with the 125-ms signals. S2 received 38 runs at 250 ms and 3 runs at 125 ms. S3 received 2 runs at 250 ms and 1 run at 125 ms. Training was terminated when the subjects demonstrated consistent identification scores in the range 95%–100% correct.

The independent variables of this study were signal duration (250 or 125 ms), masking paradigm (forward, backward, or sandwiched), and ISI (interstimulus intervals: 0, 20, 40, 80, 160, 320, or 640 ms). For each subject, a total of six (6) 100-trial runs was conducted per duration–paradigm–ISI combination.[3] Each subject was tested with the 250-ms stimulus set first. For each stimulus set, the order of the experimental paradigms was always forward, backward, and sandwiched masking. For each duration–paradigm combination, the order of the ISI values was randomized for each subject.

The subject was instructed as to the temporal location of the target signal within the stimulus sequence under each of the three masking paradigms, and was told to identity only the target and to ignore the masker(s). During the experiments, the Tactuator was placed to the left of the subject's torso. It was covered by a padded wooden box that served as an armrest for the subject's left forearm. The top of the box had an opening that allowed the subject to place the left index finger on the ''fingerpad interface'' rod as shown in Fig. 1(b). Earplugs and earphones with pink noise were used to eliminate possible auditory cues. (The Tactuator produces little audible noise except in the neighborhood of 300 Hz.) Correct-answer feedback was not provided during the main experiments for two reasons. First, all subjects were well trained with the signals in the two stimulus sets. Second, requiring the subjects to attend to correct-answer feedback

TABLE II. Analysis of masker (A), composite (B), and component (C) responses for the 125-ms stimulus set with backward masking. See the text for details.

(A)	SOA (ms)	Total trials	Skipped trials	Correct trials	Error trials	$M \neq T$ trials	$R = M$ trials	$R = M$ given $M \neq T$	$R = M$ given all errors
	125	1800	52	1248	500	476	189	39.71%	37.80%

(B)	SOA (ms)	Total trials	Error trials	T/M Nonoverlapping trials	$R = T + M$ trials	$R = T + M$ given nonoverlapping T/M	$R = T + M$ given all errors
	125	1800	500	100	29	29.00%	5.80%

(C)	SOA (ms)	Error trials	$F_L \in M$ trials	$R = T + F_L$ trials	$F_M \in M$ trials	$R = T + F_M$ trials	$F_H \in M$ trials	$R = T + F_H$ trials
	125	500	35	10	49	22	45	12
	145	440	29	9	46	28	45	15
	165	368	42	11	32	18	29	13
	205	321	21	6	28	18	23	10
	285	230	8	1	15	12	21	11
	445	121	6	1	4	4	7	6
	765	87	7	0	2	2	7	5
Total			148	38	176	104	177	72
Percentage				25.7%		59.1%		40.7%

tended to break the ''rhythm'' of the experiment. Subjects were instructed to use their right hand to select one of the seven stimulus alternatives on the digitizing tablet after each trial and then press ''ENTER'' to initiate the next trial. Occasionally, however, subjects hit ''ENTER'' before selecting a response, thus leading to ''skipped trials.'' The number of skipped trials accounted for roughly 2.5% of the total trials across all subjects and conditions. These skipped trials were not included in the data processing.

G. Data analyses

1. Overall percent-correct score

The overall performance level in percent correct was computed for each subject under each experimental condition (i.e., each combination of duration, masking paradigm and SOA). The percent-correct score was calculated separately for each individual run based only on the trials to which the subject had entered a legitimate response (i.e., skipped trials were ignored). The percent-correct scores were averaged across the individual runs for each subject at each experimental condition. Scores for individual subjects were examined as a function of SOA for each of the three masking paradigms at each of the two durations.

Conditional analyses of incorrect responses were carried out to gain further insight into the structure of the errors, using data pooled across the three subjects for each experimental condition (i.e., each masking paradigm and each SOA). The data from the sandwiched-masking paradigm were analyzed twice: first by examining the effect of the forward masker on the identification of the target (referred to as sandwiched masking M_{F}) and second by examining the effect of the backward masker on the identification of the target (referred to as sandwiched masking M_{B}). Each of three separate analyses of the error responses (i.e., masker-response analysis, composite-response analysis, and component-response analysis) is described below.

2. Masker-response analysis

A ''masker response'' is defined as an error consisting of the use of the label associated with the masker as the response. Only those trials on which the masker and target consisted of different signals were included in this analysis. Of the 49 possible masker–target combinations (7 maskers$\times$7 targets), the seven combinations where the masker and target consisted of the same signals were eliminated from further consideration in this analysis, thus leaving 42 combinations of nonidentical target and masker pairs for analysis. The percentage of masker responses, ''$R = M$ given $M \neq T(\%)$'' (where R denotes response, and M refers to either the forward masker M_{F} or the backward masker M_{B}), was calculated as the ratio of the number of trials where the masker label was used as the response, over the total number of error trials where the masker and target were different.

An example of the computation of the percentage of masker responses is provided in Table IIA for one experimental condition (backward masking, 125-ms duration, 125-ms SOA). Each subject contributed 600 trials to this condition, for a total of 1800 pooled trials (the entry in column 2). These 1800 trials were sorted into three categories: skipped trials (the number of trials on which no response was provided by the subject, shown in column 3), correct trials (the number of trials on which the subject identified the target correctly, column 4), and error trials (the number of trials on which the subject entered an incorrect response, column 5). The error trials were further sorted into two categories: those where the masker and target consisted of the same stimulus ($M = T$ trials, not shown) and those where the masker and target were nonidentical stimuli ($M \neq T$ trials, shown in column 6). The $M = T$ trials were discarded in this analysis, and only those error trials in which $M \neq T$ were used to examine the use of the masker as response. The number of $M \neq T$ trials on which the subjects erroneously responded with the label of the masker is shown in column 7

($R=M$ trials) of Table IIA. The percentage of $R=M$ responses relative to the number of $M\neq T$ trials is provided in column 8 ($R=M$ given $M\neq T$). The percentage of $R=M$ responses relative to the total number of error trials is provided in the final column ($R=M$ given all errors). Data were processed in an analogous manner for all values of SOA for forward and backward masking at both signal durations. In addition, data from the sandwiched-masking paradigm were processed twice in this manner: once considering only the forward masker (designated M_{F}) and once considering only the backward masker (M_{B}).

3. Composite-response analysis

The percentage of composite responses was calculated to examine the extent of temporal integration of the target and masker. A composite response was possible only in those cases where the spectral components of the masker and target were both nonoverlapping and combined to form another member of the stimulus set. Only 12 of the 49 possible target and masker combinations met these specifications (i.e., did not have overlapping spectral components and combined to form another member of the stimulus set). Therefore, the error trials corresponding to these 12 combinations were analyzed for composite responses. The percentage of composite response, "$R=T+M$ (%)", was calculated as the ratio of the number of trials where the response was the same as the composite of the target and masker, over the total number of error trials where the target and masker did not share any spectral components.

An example of the computation of the percentage of composite responses is provided in Table IIB, again for one experimental condition (backward masking, 125-ms duration, 125-ms SOA). The total number of pooled trials (column 2) and the error trials (column 3) were calculated in the same way as those in Table IIA. For this analysis, the error trials were further sorted into two categories: those where the masker and target shared one or more components (not shown) and those where the masker and target did not share any spectral components (T/M nonoverlapping trials, column 4). The number of T/M nonoverlapping trials on which the subjects erroneously responded with the composite of the masker and the target is shown in column 5 ($R=T+M$ trials) of Table IIB. The percentage of $R=T+M$ responses relative to the number of T/M nonoverlapping trials is provided in column 6 ($R=T+M$ given nonoverlapping T/M). The percentage of $R=T+M$ responses relative to the total number of error trials is provided in column 7 ($R=T+M$ given all errors). Data were processed in an analogous manner for all values of SOA for forward and backward masking at both signal durations. In addition, data from the sandwiched-masking paradigm were processed twice in this manner: once considering only the forward masker (designated M_{F}) and once considering only the backward masker (M_{B}).

4. Component-response analysis

A component response was defined as a case in which a frequency component that was present in the masker, but not in the target, was contained in an incorrect response. The analysis was thus restricted to those error trials where the target and masker did not share any common frequency components. The masking effects of the low-, mid-, and high-frequency components were analyzed separately in terms of the percentages of F_L, F_M, or F_H component responses, using the relevant subset of error trials.

An example of the computation of the percentage of component responses is provided in Table IIC for one experimental condition (backward masking, 125-ms duration). For the analysis on low-frequency component response, the error trials shown in column 2 were further sorted into two categories: those where the masker contained a low-frequency component ($F_L\in M$ trials, column 3) and those where the masker did not contain such a component (not shown). Of those error trials where the masker contained a low-frequency component, the number of trials on which the subjects erroneously responded with the composite of the target and a low-frequency component is shown in column 4 ($R=T+F_L$ trials). The number of $F_L\in M$ and $R=T+F_L$ trials were further pooled across all SOA values (shown in the "Total" row towards the end of Table IIC). The percentage of $R=T+F_L$ responses relative to the number of $F_L\in M$ trials is provided on the last row in column 4. Data were processed in an analogous manner for midfrequency and highfrequency component responses as shown in Table IIC. The same data processing procedure was applied to forward and backward masking at both signal durations. In addition, data from the sandwiched masking paradigm were processed twice in this manner: once considering only the forward masker (designated M_{F}) and once considering only the backward masker (M_{B}).

5. Information-transmission analysis

Information transfer (IT) and information-transfer rate (IT rate) were estimated from percent-correct scores for each subject. IT was calculated as $IS\times(1-2e)$, where IS is the information in the stimulus, and e is the error rate. This formula has been used to obtain a lower-bound estimate of IT when the total number of trials is small (see Tan, 1996). When all stimulus alternatives are equally likely (as was the case in the current study), IS is simply $\log_2 k$, where $k=7$ is the number of stimulus alternatives (that is, $IS=2.81$ bits for all experiments considered here). Error rate e was calculated as $(1-pc)$, where pc stands for percent-correct scores in percentages. For cases in which pc was below 50%, IT was set to 0.

The estimated values of IT at each combination of stimulus duration and SOA were used to calculate estimates of IT rate in bits/s. Specifically, IT rate was calculated as the product of estimated IT in bits/item and presentation rate in items/s. The presentation rate is simply the reciprocal of the SOA (i.e., ISI plus stimulus duration in seconds).

III. RESULTS

A. Overall percent-correct score

The overall performance levels of each subject under the three masking paradigms are shown in Fig. 2 for the 250-ms

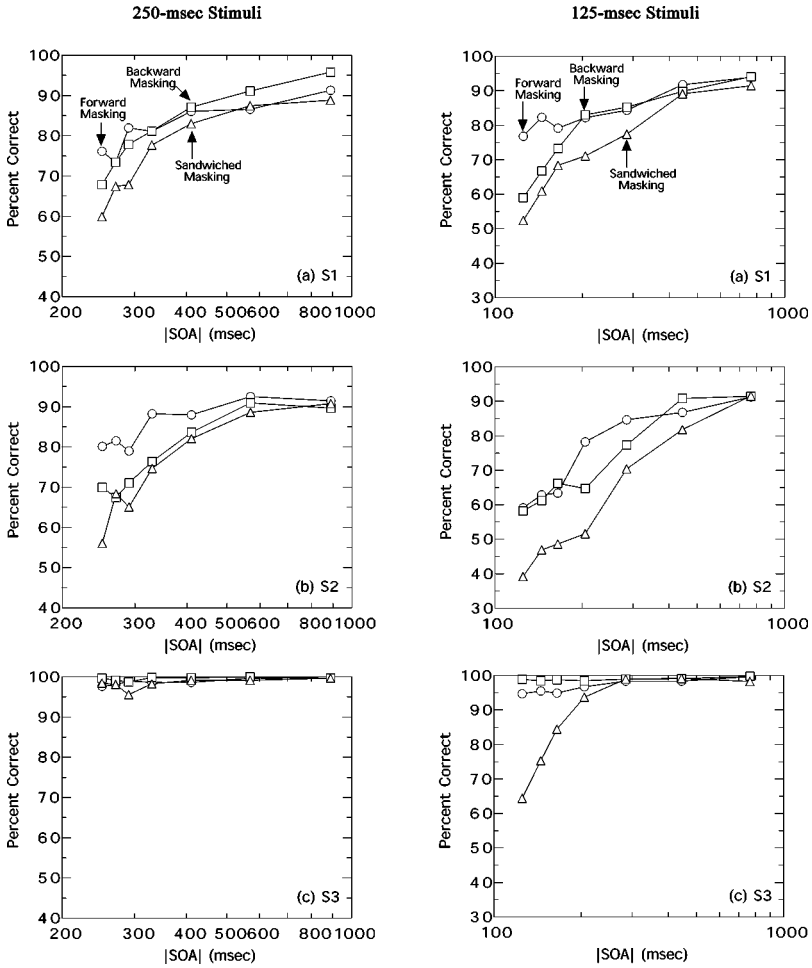


FIG. 2. Percentage-correct scores for the 250-s (left column) and 125-ms (right column) stimulus sets for subjects S1, S2, and S3 (top, middle, and bottom rows, respectively). Data from forward, backward, and sandwiched masking paradigms are shown by open circles, open squares, and open triangles, respectively. Data are plotted against the absolute value of SOA (negative for forward masking and positive for backward masking).

stimulus set (left column) and the 125-ms stimulus set (right column). At 250-ms, S3 had essentially errorless performance at all values of SOA under all three types of masking. The performance of S1 and S2, on the other hand, began to decrease steadily with a decrease in SOA under all three types of masking. For both these subjects, a tendency was observed for worst performance under sandwiched masking and for the highest performance under forward masking, particularly at the shortest values of SOA. At 125 ms, the performance of S3 was again quite high and began to deteriorate only under sandwiched masking (at the three shortest values of SOA). S1 and S2 demonstrated deteriorating performance with SOA under all three masking paradigms. For S1, performance was worst under sandwiched masking and best for forward masking at the three shortest values of SOA. For S2, performance was worst under sandwiched masking but similar for forward and backward masking conditions at the shortest values of SOA. Our observation that forward masking produced similar or better results than backward masking at the shortest SOA values is inconsistent with the general consensus in the literature that, under similar conditions, there is more interference in backward masking than in forward masking (see, for example, Craig and Evans, 1995). A trend toward lower scores for backward compared to forward masking was observed at shorter values of ISI for S1 (at both stimulus durations) and for S2 (at 250 ms only). The finding that performance level was lower with the sandwiched-masking paradigm was to be expected since signal identification in the sandwiched-masking paradigm was affected by both the forward and backward maskers.

From the individual-subject data in Fig. 2, it is clear that S3 performed near 100% correct at all SOA values (except for the sandwiched masking at 125 ms), whereas S1 and S2 made more errors at smaller SOA values. This intersubject difference may be attributed to the fact that S3 had participated in an earlier study requiring the identification of similar signals from a much larger set of stimulus alternatives.

B. Masker response

Masker responses are the percentages of error trials where the masker and target were different, and subjects responded with the masker (see the column "$R=M$ given $M \neq T$" in Table IIA). The chance performance level was 12.2%. This was calculated as the multiplication of 42/49 (probability of a target and masker combination where $T \neq M$) and 1/7 (probability of choosing the masker as the response given the target and masker combination, assuming that the subject's response was independent of the target or masker). The percentages of masker responses were well above chance level for both the 250- and 125-ms stimulus sets. For the 250-ms stimuli, the level of masker response was in the range 15%–32%. Percentage of masker responses was higher for the 125-ms stimuli, falling in the range 21%–

TABLE III. Percentages of component responses and the corresponding number of trials analyzed (in parentheses).

Components	Forward		Backward		Sandwiched (M_F)		Sandwiched (M_B)	
	250 ms	125 ms	250 ms	125 ms	250 ms	125 ms	250 ms	125 ms
F_L	11.1% (144)	26.9% (171)	11.5% (148)	25.7% (148)	13.2% (204)	19.5% (251)	13.6% (169)	25.7% (261)
F_M	60.0% (30)	75.5% (94)	43.1% (72)	59.1% (176)	39.6% (91)	45.9% (281)	43.4% (106)	44.0% (273)
F_H	64.7% (133)	52.4% (170)	51.3% (150)	40.7% (177)	62.2% (164)	28.9% (374)	52.0% (179)	31.3% (335)

45%. The result of a three-way ANOVA (on the factors of stimulus duration, masking paradigm, and ISI[4]) performed on the percentages confirmed a significant main effect of stimulus duration [$F(1,18)=58.26$, $p<0.0001$], and ISI [$F(6,18)=5.94$, $p=0.0014$]. The latter was consistent with the fact that subjects responded with the masker more often for small ISI values.

C. Composite response

Composite responses are the percentages of error trials (see the column "$R=T+M$ given nonoverlapping T/M" in Table IIB) where subjects responded with the composite of the target and masker. The chance performance level was 3.5%. This was calculated as the multiplication of 12/49 (probability of a valid target and masker combination) and 1/7 (probability of choosing the composite of the target and masker as the response given the target and masker combination, assuming that the subject's response was independent of the target or masker). The percentages of composite responses were well above chance level for both the 250- and 125-ms stimulus sets. The level of composite response was in the range 14%–48% for the 250 ms stimuli, and 25%–54% for the 125-ms stimuli. It was observed that subjects responded with the composite of the target and masker significantly more often with the 125-ms stimuli than with the 250-ms stimuli. The result of a three-way ANOVA (on the factors of stimulus duration, masking paradigm, and ISI) performed on the percentages confirmed a significant main effect only for stimulus duration [$F(1,18)=14.64$, $p=0.0012$].

D. Component response

Recall that component-response analysis examines the extent to which a particular spectral component in the masker was combined with the target response to form an incorrect response. Table III contains the percentages of component responses for the three single-frequency components and the four masking categories at 250 and 125 ms. Note that the entries in the last two rows of Table IIC are entered into column 5 of Table III (backward, 125 ms). The chance performance level for the F_L, F_M, or F_H component was 1.5%. For example, the chance level for F_L component response was the multiplication of 5/49 (probability of the joint event that the target and masker did not share any spectral components and the masker contained an F_L component) and 1/7 (probability of choosing the composite of the target and F_L as the response, assuming that the subject's response was independent of the target or masker). From Table III, it is clear that the percentage of component responses under each category was in the range 11%–76% and was well above the chance level of 1.5%, indicating that subjects indeed responded erroneously by choosing a response based on the combination of the target and a component of the masker.

A three-way ANOVA was conducted on the results presented in Table III for the three main factors of masking paradigm, stimulus duration, and frequency of the masking component. The results indicated that the rate of usage of the component response was dependent on masking paradigm [$F(3,6)=21.26$, $p<0.001$] and on the frequency of the masking component [$F(2,6)=228.76$, $p<0.0001$], but not on stimulus duration [$F(1,6)=0.36$, $p<0.57$]. A higher rate of component response was observed under forward masking (11%–76%, average 48.4%) than under the other three paradigms (12%–62%, average 36.2%). Component responses were least likely to occur with F_L (11%–27%, average 18.4%) compared to F_M (40%–76%, average 51.3%) and F_H (29%–65%, average 47.9%).

E. Information transmission

Estimated information transfer (IT) is plotted as a function of the absolute value of SOA (|SOA|) in Fig. 3 for each of the three subjects. Because the estimated IT values are derived from a multiplicative transformation of the percent-correct scores, the trends observed previously for percent-correct scores also apply to the IT measure. The IT values are limited by the stimulus uncertainty of 2.81 bits in these experiments. Across conditions for the 250-ms stimuli, estimated values of IT ranged from roughly 0.5 to 2.6 bits for S1, 0.25 to 2.25 bits for S2, and 2.6 to 2.75 bits for S3. Across conditions for the 125-ms stimuli, IT ranged from roughly 0.13 to 2.5 bits for S1, 0 to 2.25 bits for S2, and 0.75 to 2.75 bits for S3.

In Fig. 4, estimated IT rate in bits/s is plotted as a function of the actual information presentation rate. Data for each subject are plotted separately in each of the three panels of the figure. The information-presentation rate was computed as stimulus uncertainty (2.81 bits) divided by |SOA|. In general, IT rate was close to the information-presentation rate (i.e., the maximum achievable IT rate, shown by the dashed lines in Fig. 4) when information presentation rate was low (i.e., when |SOA| was large). IT rate deviates from the dashed lines as information-presentation rate increases. The

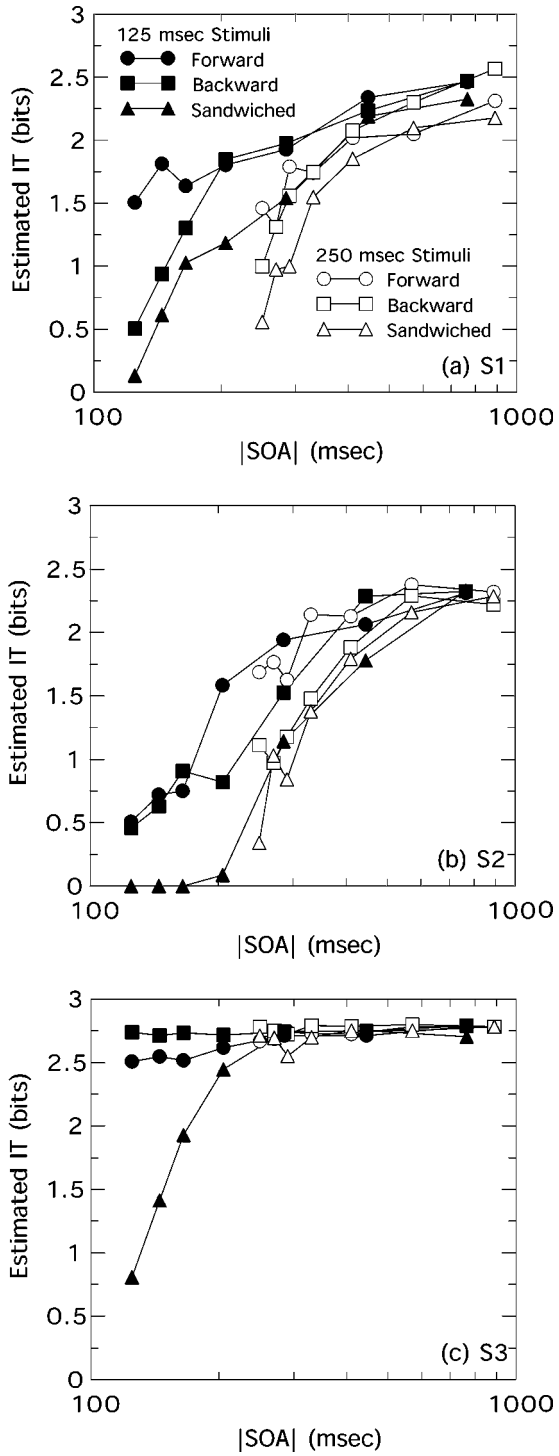


FIG. 3. Estimated IT in bits as a function of |SOA| in ms. Individual-subject data are presented in each of the three panels for 125-ms (filled symbols) and 250-ms (open symbols) stimuli for each of the three masking paradigms: forward (circles), backward (squares), and sandwiched (triangles).

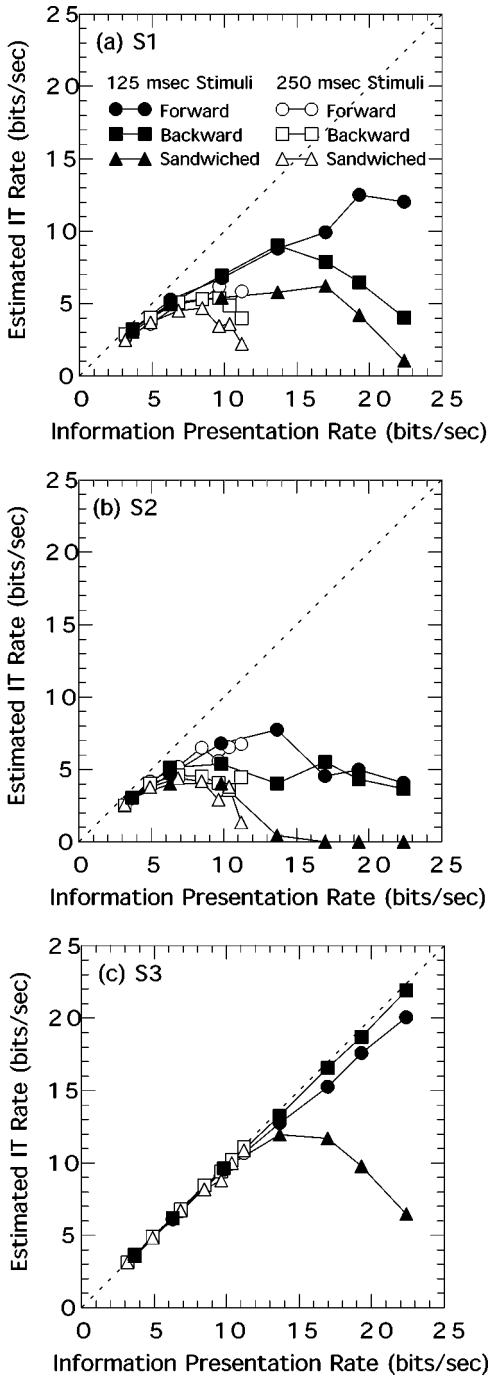


FIG. 4. Estimated IT rate in bits/s as a function of information-presentation rate in bits/s. Individual-subject data are presented in each of the three panels for 125-ms (filled symbols) and 250-ms (open symbols) stimuli for each of the three masking paradigms: forward (circles), backward (squares), and sandwiched (triangles). The diagonal dashed line represents the maximum achievable IT rate.

data of S3 in Fig. 4(c) show a striking consistency: IT rate falls on or near the diagonal dashed line except for the sandwiched-masking paradigm at 125 ms at the highest three values of information presentation rate.

Maximum values of IT rate varied across subjects and were achieved at different combinations of masking paradigm and |SOA| across subjects. For the 125-ms stimulus set, S1 reached a maximum IT rate of 12.5 bits/s at |SOA|=145 ms with the forward-masking paradigm. S2 reached a maximum IT rate of 7.72 bits/s at |SOA|=205 ms with the forward-masking paradigm. S3 reached a maximum IT rate of 21.9 bits/s at |SOA|=125 ms with the backward-masking paradigm. For the 250-ms stimulus set, S1 reached a maximum IT rate of 6.2 bits/s at |SOA|=290 ms with the forward-masking paradigm. S2 reached a maximum IT rate of 6.8 bits/s at |SOA|=250 ms with the forward-masking paradigm. S3 reached a maximum IT rate of 11.1 bits/s at |SOA|=250

ms with the backward-masking paradigm. The maximal rate achieved by S1 and S2 falls substantially below the information-presentation rate at the conditions tested. This result differs from that observed for S3, whose peak IT rate (with the exception of sandwiched masking at 125 ms) is limited primarily by the information-presentation rate achieved in the stimulus ensembles.

IV. DISCUSSION

A. Comparison with tactile temporal masking literature

The results obtained in the present study with relatively long-duration, spectrally defined haptic stimuli presented to the index finger can be compared with those obtained from masking studies employing brief spatial or spatiotemporal patterns delivered through cutaneous stimulation of the index finger (e.g., Craig, 1983, 1995, 1998, 2000; Evans and Craig, 1986; Evans, 1987). Under both forward- and backward-masking paradigms, the ability to identify the target signal has been shown to increase monotonically with SOA such that no further increase in performance is observed for SOA greater than 100 ms for backward masking and for SOA greater than roughly 1000 ms for forward masking (e.g., Craig, 1983, 1995; Evans and Craig, 1986). It is also generally accepted in the spatiotemporal masking literature that more backward than forward masking occurs at brief values of SOA (i.e., less than 100 ms) (e.g., Evans, 1987; Evans and Craig, 1986; Gescheider, Bolanowski, and Verrillo, 1989). At the same time, there is evidence to suggest that forward masking occurs over a longer time period than backward masking: although backward-masking effects are complete for SOA of 100–200 ms, forward masking appears to persist for SOAs up to roughly 1 s (Craig and Evans, 1987). As expected, the results of the current study also demonstrate an orderly increase in target-identification ability with SOA (see Fig. 2). Our results, however, do not demonstrate a consistent performance difference between forward and backward masking [except perhaps for the data of S2 at 250 ms (Fig. 2, left panel b) and the data of S1 at 125 ms (Fig. 2, right panel a)]. The amount of forward or backward masking is similar across the range of SOA values tested and appears to be complete within the first 400–500 ms following stimulus onset (Fig. 2).

The results described above for backward- and forward-masking patterns with cutaneous spatial stimuli have been explained in terms of effects associated with the storage of an internal representation of tactile stimuli that persists following their offset. In the case of short values of SOA, the internal representation associated with the target is stronger under forward masking (where the target is the trailing stimulus in the pair) than under backward masking (where it is the leading stimulus). Thus, less masking of the target occurs under forward compared to backward masking at short SOA. At longer values of SOA, however, the internal representation associated with the leading stimulus (the masker in forward masking) appears to persist for hundreds of ms and to interfere with identification of the trailing target. Differences in the relative effects of forward versus backward masking between the current study and previous studies may be due in part to differences in stimulus duration. The stimuli employed in the current study were relatively long in duration (125 or 250 ms, in order to accommodate low-frequency signals at 2 and 4 Hz) compared to the 26-ms duration used in many of the Optacon-based studies. For the relatively long stimulus durations (leading to relatively long SOA values) employed here, a strong internal representation of the target may be established whether the target is the leading or trailing stimulus in the pair, thus explaining the similarity in performance between forward and backward masking. The smallest SOA employed in the current study was 125 ms (associated with the 125-ms stimulus set and ISI=0 ms), which exceeds the 100-ms time window under which tactile masking effects appear to be the strongest.

The heterogeneous frequency composition of the stimuli employed here also likely plays a role in explaining differences between the current results and those of previous studies employing strictly cutaneous stimulation. From a neurophysiological point of view, the slowly adapting type 1 system (SA I), the rapidly adapting system (RA), and the Pacinian system (PC) are most sensitive to stimulation in the F_L, F_M, and F_H frequency ranges, respectively (Bolanowski *et al.*, 1988; Johnson *et al.*, 2000). Each of these systems is known to exhibit different temporal and spatial summation properties, thus adding complexity to the time course of the ''internal representation'' of these stimuli compared to those employing strictly cutaneous stimulation. The component-response analysis attempted to separate the masking effects of low, mid, and high-frequency components of the stimuli employed in the current study. The results indicated that the low-frequency component F_L in a masker was least likely to be combined with the target to form an erroneous response. This result was consistent with the subjective impression that only the F_M and F_H components tended to ''spread'' in time. The use of F_H in a component response was greater for the 250-ms compared to the 125-ms stimuli and for forward compared to backward masking. Of the three different afferent types of the tactual sensory system, only PC (associated with F_H stimulation) is known to exhibit temporal and spatial summation properties. Such properties may explain why an F_H component in a masker might have been perceived as belonging to a target presented before or after the masker and why this would be more likely to occur with longer-duration stimuli and forward masking. Such temporal/spatial summation properties alone, however, may not be sufficient to explain all the component-response results. Although it is known that neither the RA nor the SA I fibers exhibit temporal or spatial summation (Bolanowski *et al.*, 1988), component responses were observed with F_M as well as with F_H.

Comparisons of error patterns in the current study and previous studies reveal further similarities and differences in the processing of spatiotemporal patterns versus the spectral-based stimuli studied here. Four areas are discussed below, including the role of stimulus complexity in target identification, masker-response competition, temporal integration of

TABLE IV. Percent-correct scores for target patterns (*T*) containing one, two or three spectral components. Results are averaged across SOAs.

	Forward masking			Backward masking			Sandwiched masking		
Duration (ms)	*T* has 1 freq.	*T* has 2 freq.	*T* has 3 freq.	*T* has 1 freq.	*T* has 2 freq.	*T* has 3 freq.	*T* has 1 freq.	*T* has 2 freq.	*T* has 3 freq.
250	91.1%	85.9%	78.1%	92.2%	83.7%	75.5%	88.5%	81.0%	67.1%
125	89.7%	83.7%	76.6%	88.3%	83.4%	67.4%	79.5%	74.0%	57.9%

masker and target, and performance when target and masking signals are identical.

1. Stimulus complexity

The effects of stimulus complexity on target identification appear to be generally similar for both types of stimuli. For backward masking, Evans (1987) reported that percent-correct scores with one-line target patterns were significantly higher than those with two-line target patterns, indicating that subjects were able to identify simpler spatial patterns more accurately. Similarly, our percent-correct scores were highest with target patterns that contained one spectral component, and lowest with those that contained three spectral components (see Table IV).

2. Masker-response competition

Response competition has been widely studied as the source of errors in temporal masking (e.g., Evans, 1987; Evans and Craig, 1992; Craig and Evans, 1995; Horner, 2000). The rationale for this type of interference is based on the notion that both the target and the masker are fully processed to their respective responses, but that the subject mistakenly chooses the masker response instead of the target response at brief values of SOA. Evans (1987) reported significant use of the masker as the response under backward, but not forward, masking of spatial patterns presented to the index finger. Under backward masking, Evans (1987) observed that roughly 20%–30% of the errors made at SOAs in the range of 26–106 ms could be attributed to the use of the masker as response. Under forward masking, however, the use of the masker as response did not exceed chance levels at any value of SOA. The results of the current study demonstrate substantial use of the masker as response under both backward and forward masking. The rate of masker response on error trials for both types of masking ranged from roughly 15%–30% across SOA for the 250-ms signals and from roughly 20%–40% for the 125-ms signals. The highest rate of responding with the masker was observed for the short-duration stimuli at small values of SOA. One plausible explanation for the masker-response competition lies in the limitation imposed by temporal order discrimination threshold (e.g., Craig and Baihua, 1990).

3. Temporal integration of target and masker

The role of temporal integration of the target and masker has been widely studied as a potential source of error responses in temporal masking of tactile patterns (e.g., Evans and Craig, 1986; Evans, 1987; Craig and Evans, 1987; Craig, 1996; Mahar and Mackenzie, 1993). Originally proposed for visual masking (Felsten and Wasserman, 1980), the temporal-integration theory seems to account for spatiotemporal tactile patterns as well as the spectral-temporal haptic patterns used in the present study. Temporal integration implies that the target and masker form a composite percept that is the temporal and/or spatial sum of both signals. Evans and Craig (1986), in a study of backward masking using tactile spatial patterns, observed an increase in the perceived number of line segments in the response relative to the number in the target for brief SOAs in an identification task and also an overestimation of the number of line segments in the target in an estimation task. Evans (1987) studied the identification of 26-ms spatial patterns under backward and forward masking using a set of 13 line patterns that included stimuli that were composites of two other stimuli in the set. Evidence of use of the composite response on error trials was observed under both backward and forward masking at brief values of SOA (i.e., less than 100 ms): the percentage of composite responses ranged from 20%–70% at the shortest SOA of 26 ms and decreased monotonically as SOA increased. The results of the current study also indicate use of the composite response on error trials, although with some differences in the trends observed by Evans (1987). The rate of composite response observed here ranged from roughly 20%–50% across conditions, did not show a consistent decrease as SOA increased, and was more predominant for the 125-ms compared to the 250-ms signals. Generally, the results from experiments with both tactile and haptic signals provide evidence for the formation of a composite response arising from the properties of the masker and the target.

4. Performance on trials with identical masker and target

Trials with the same stimulus selected for masker and target are of interest based on qualitative predictions from both the response-competition and temporal-integration theories of masking. From both points of view, one would expect that identical stimuli in both intervals of a backward- or forward-masking paradigm should reduce the probability of error. Evans (1987) found very low error rates in the range 1%–4% that did not vary with SOA over trials where target =masker under both forward and backward masking. Data from the current experiments, considering only those trials on which target=masker, are plotted in Fig. 5. Our results showed a similar trend in that percent scores for trials with identical target and masker were higher than those for all trials. Specifically, the error rates associated with trials with identical target and masker for the forward- and backward-masking paradigms were in the range 3.3% to 10.4% for the

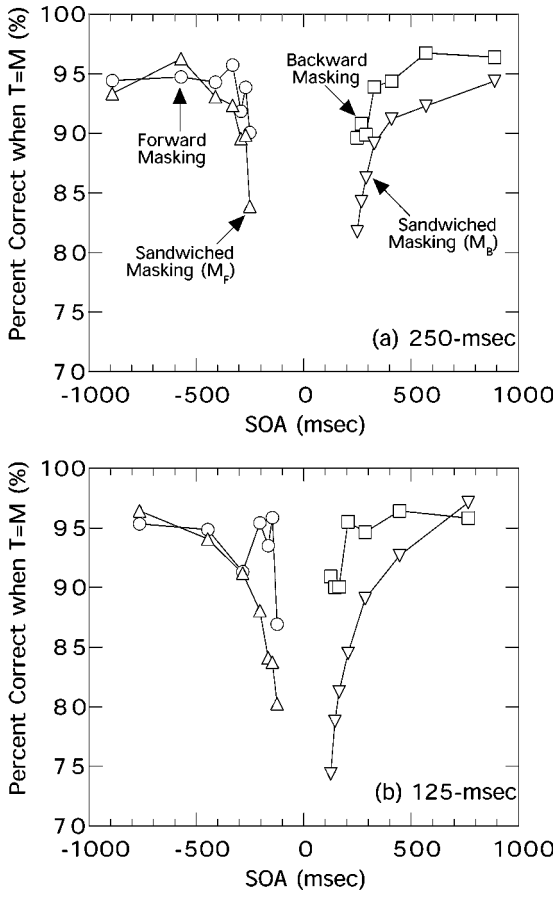


FIG. 5. Percent-correct scores for trials with identical target and masker plotted as a function of SOA (in ms) for 250-ms signals [panel (a)] and 125-ms signals [panel (b)]. Data are pooled across subjects and are plotted for forward (circles), backward (squares), and sandwiched (triangles) masking.

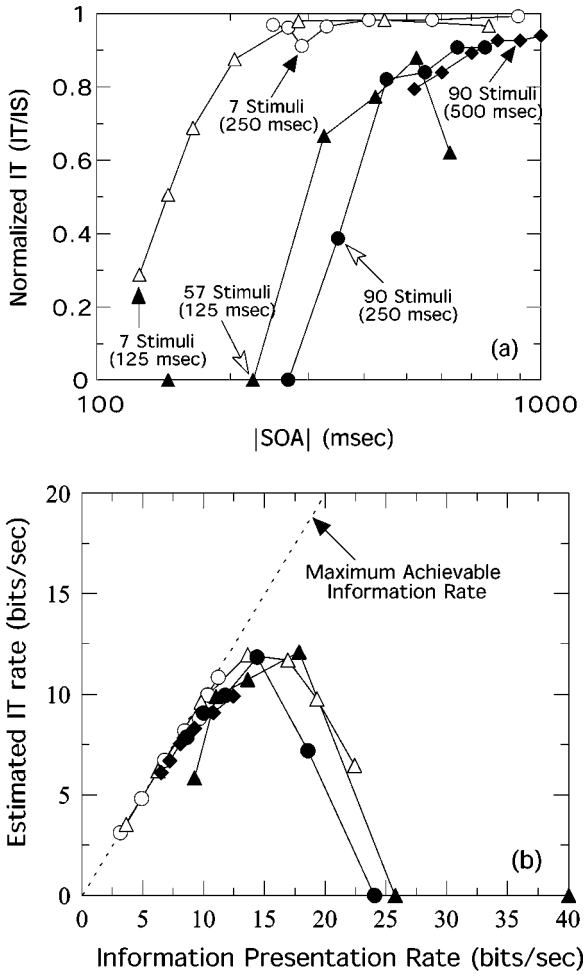


FIG. 6. Comparison of data from five different stimulus sets for subject S3. In panel (a), the normalized IT (i.e., the ratio of estimated IT to IS) is plotted as a function of |SOA| in ms for five different stimulus sets described in the figure. In panel (b), the estimated IT rate in bits/s is plotted as a function of the information-presentation rate in bits/s for the same five stimulus sets. The dashed line plots the maximum achievable IT rate.

250-ms stimulus set, and in the range 3.6% to 13.1% for the 125-ms stimulus set. These data differ from those of Evans (1987) in that the error rates shown in Fig. 5 do tend to increase with a decrease in SOA. The error rates for the sandwiched-masking paradigm in Fig. 5 were much higher, especially at short SOAs. This higher error rate can be explained by the fact that of all the trials where target=forward masker, only 14.29% (1 out of 7) of the trials contained a backward masker that was identical to the target/forward masker and therefore did not interfere with the identification of the target.

B. Information transmission

The estimates of IT rate from the current study may be compared with those reported by Tan *et al.* (1999) who measured IT and IT rate using the sandwiched-masking paradigm and three different stimulus sets composed of 57 signals at a duration of 125 ms and 90 signals at durations of 250 and 500 ms, respectively. The stimuli employed in the current study represent a subset of the signals from Tan *et al.* (1999). A comparison of the data from the two studies is provided in Fig. 6(a) and Fig. 6(b) for normalized IT (i.e., the ratio of estimated IT to IS, information in stimulus) and IT rate, respectively, for subject S3 who participated in both studies. From Fig. 6(a), it is clear that for a small number of stimulus alternatives (7; see the open symbols), S3's performance approached 100% information transmission at |SOA| values beyond 250 ms. For a large number of stimulus alternatives (57 or 90; see the filled symbols), S3's performance continued to improve even at large |SOA| values. The one datum point for 125-ms stimuli with 57 alternatives at the largest |SOA| value appears to be an anomaly, in that it does not follow the monotonically increasing trend of the rest of the data in Fig. 6(a). As can be seen in Fig. 6(b), S3's performance initially followed the diagonal line of maximum achievable IT rate, reached a peak IT rate, then decreased, as a function of information presentation rate in bits/s. The IT rate for the 90-stimulus set at 500 ms (filled diamonds) and the 7-stimulus set at 250 ms (open circles) seemed to be limited by presentation rate rather than perceptual constraints. One striking trend evidenced in Fig. 6(b) is that performance in terms of IT rate seems to be solely determined by the information presentation rate in bits/s; to a good approximation, all the curves for the different conditions coincide.

The conditions under which maximum IT rate was obtained are tabulated in Table V. The two rates marked with asterisks could potentially be improved if higher information

TABLE V. Optimal SOA, stimulus delivery rate, and the corresponding peak IT rate for data shown in Fig. 6(b). The two rates marked with asterisks seem to be limited by information presentation rate, and can potentially be improved.

Condition	Optimal SOA (ms)	Optimal delivery rate (items/s)	Peak IT rate (bits/s)
7 stimuli (125 ms)	205	4.9	11.9
7 stimuli (250 ms)	250	4.0	10.9*
57 stimuli (125 ms)	325	3.1	12.1
90 stimuli (250 ms)	450	2.2	11.8
90 stimuli (500 ms)	520	1.9	9.9*

presentation rate were used with the corresponding conditions (i.e., in these cases, the normalized IT rate saturated at unity). In general, optimal delivery rate in items/s decreased as stimulus duration increased and as size of the stimulus set (i.e., stimulus uncertainty) increased. On the other hand, the peak IT rate remained remarkably constant across the conditions (around 12 bits/s for the three unmarked values in Table V).

Our results are inconsistent with the generally accepted view that the optimal stimulus delivery rate is between 2 and 3 items/s independent of the stimulus uncertainty (Garner, 1962, p. 91, citing Klemmer and Muller, 1953). In the current study, a constant peak IT rate [see Fig. 6(b) and Table V] of roughly 12 bits/s was obtained independent of stimulus uncertainty in the range of 2.8 to 6.5 bits and independent of stimulus duration over the range 125–500 ms. These results are at variance with the conclusions reached by previous investigators (including Klemmer and Muller, 1953, and Alluisi, Muller, and Fitts, 1957) indicating that peak IT rate increases monotonically with stimulus uncertainty. A rigorous comparison of the current data with previous results is difficult due to the substantial differences in methodology that exist among the relevant studies and is beyond the scope of the current paper. These differences include stimulus modality, type of motor response required of the subject, relationship between signal duration and presentation rate, whether or not response time is taken into consideration in determining the IT rate, as well as values of stimulus uncertainty and information-presentation rate. Further investigation of the properties that contribute to optimizing IT rate is important not only in connection with the design of improved aids for individuals with sensory impairments, but also in connection with the design of improved displays for normally sensed users of data visualization systems and synthetic environments.

V. CONCLUDING REMARKS

The maximum IT rate of roughly 12 bits/s found in our current study and in that of Tan *et al.* (1999) approaches the typical rates demonstrated by natural tactual communication methods. For example, typical communication rate for the Tadoma method (Reed *et al.*, 1985) employed by experienced deaf–blind individuals is estimated to be 11.2 bits/s (Table I, Reed and Durlach, 1998). Typical IT rate for tactual reception of American Sign Language is estimated to be 11.7 bits/s (Table I, Reed and Durlach, 1998). The 12 bits/s rate is higher than the communication rates demonstrated by any other artificial tactual display, including the Optacon (4.5–9 bits/s), vibrotactile or kinesthetic reception of Morse code (0.9–2.2 bits/s), Vibratese (1.8–3.6 bits/s), and tactual reception of fingerspelling (8.1 bits/s) (see Reed and Durlach, 1998, for derivation of these IT rates and for reference to original studies). One major difference between the Tadoma method and most previous synthetic tactual displays is that a talking face provides a multidimensional information display employing a rich set of signal attributes, as compared to the purely cutaneous stimulation provided by many artificial tactual displays. The Tactuator employed in our studies spans the tactual stimulation continuum from kinesthetic movements to cutaneous vibrations. The multidimensional nature of the signals delivered through this display may be the primary factor responsible for the relatively high IT rate achieved in studies with this device. Future experiments are necessary to examine the extent to which the 12 bits/s rate can be maintained for processing continuous streams of multidimensional tactual signals.

ACKNOWLEDGMENTS

This research was supported in part by Research Grant Number 5 R01-DC00126 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. It was also supported in part by a National Science Foundation Faculty Early Career Development (CAREER) Award under Grant No. 9984991. The first author wishes to thank Seungmoon Choi for his assistance with statistical analysis. The thoughtful comments of two anonymous reviewers are greatly appreciated.

[1]Our definitions of the terms ''tactile,'' ''kinesthetic,'' and ''tactual'' follow those provided by Loomis and Lederman (1986). The term ''tactile'' refers to information acquired through surface contact factors via cutaneous sensors in the skin (e.g., information on texture obtained by relative stroking motion between skin and object, information obtained from vibratory arrays, etc.). The term ''kinesthetic'' relates to information about finger position, motion, and force obtained via sensors in the internal components of the hand, wrist, and arm such as muscles, joints, and tendons. The term kinesthetic is intended to include proprioceptive, and the term ''tactual'' includes both tactile and kinesthetic.

[2]The structure of the stimulus sets employed in the current study may be considered, in some respects, a ''spectral'' analog of the spatial-pattern stimuli studied by Evans (1987).

[3]The one exception was that subject S3 performed only three 100-trial runs for the forward 250 ms condition, because his performance level was near perfect.

[4]ISI values, not SOAs, were used in ANOVA analyses, because ISIs were identical for the two stimulus durations. Given the one-to-one correspondence between the ISI and SOA values, the statistical effects of ISI and SOA are equivalent. According to Craig (1983), SOA, rather than ISI, is the primary determinant of temporal interaction between sequentially presented vibrotactile patterns. Therefore, our results are presented as a function of SOA whenever possible.

Alluisi, E. A., Paul, F., Muller, J., and Fitts, P. M. (**1957**). ''An information analysis of verbal and motor responses in a forced-paced serial task,'' J. Exp. Psychol. **53**(3), 153–158.

Bernstein, L. E. (**1992**). ''The evaluation of tactile aids,'' in *Tactile Aids for the Hearing Impaired*, edited by I. R. Summers (Whurr, London), pp. 167–186.

Bolanowski, S. J., Gescheider, G. A., Verrillo, R. T., and Checkosky, C. M. (**1988**). ''Four channels mediate the mechanical aspects of touch,'' J. Acoust. Soc. Am. **84**(5), 1680–1694.
Craig, J. C. (**1983**). ''The role of onset in the perception of sequentially presented vibrotactile patterns,'' Percept. Psychophys. **34**, 421–432.
Craig, J. C. (**1995**). ''Vibrotactile masking: The role of response competition,'' Percept. Psychophys. **57**(8), 1190–1200.
Craig, J. C. (**1996**). ''Interference in identifying tactile patterns: Response competition and temporal integration,'' Somatosens Mot Res. **13**(3–4), 199–213.
Craig, J. C. (**1998**). ''Vibrotactile pattern isolation/integration,'' Percept. Psychophys. **60**(5), 888–899.
Craig, J. C. (**2000**). ''Processing of sequential tactile patterns: Effects of a neutral stimulus,'' Percept. Psychophys. **62**(3), 596–606.
Craig, J. C., and Baihua, X. (**1990**). ''Temporal order and tactile patterns,'' Percept. Psychophys. **47**(1), 22–34.
Craig, J. C., and Evans, P. M. (**1987**). ''Vibrotactile masking and the persistence of tactual features,'' Percept. Psychophys. **42**(4), 309–317.
Craig, J. C., and Evans, P. M. (**1995**). ''Tactile selective attention and temporal masking,'' Percept. Psychophys. **57**(4), 511–518.
Eberhardt, S. P., Bernstein, L. E., Barac-Cikoja, D., Coulter, D. C., and Jordan, J. (**1994**). ''Inducing dynamic haptic perception by the hand: System description and some results,'' in *Proceedings of the ASME Dynamic Systems and Control Division* (The American Society of Mechanical Engineers, New York), Vol. 1, pp. 345–351.
Evans, P. M. (**1987**). ''Vibrotactile masking: Temporal integration, persistence, and strengths of representations,'' Percept. Psychophys. **42**(6), 515–525.
Evans, P. M., and Craig, J. C. (**1986**). ''Temporal integration and vibrotactile backward masking,'' J. Exp. Psychol. Hum. Percept. Perform. **12**(2), 160–168.
Evans, P. M., and Craig, J. C. (**1992**). ''Response competition: A major source of interference in a tactile identification task,'' Percept. Psychophys. **51**(2), 199–206.
Felsten, G., and Wasserman, S. S. (**1980**). ''Visual masking: Mechanisms and theories,'' Psychol. Bull. **88**, 329–353.
Garner, W. R. (**1962**). *Uncertainty and Structure as Psychological Concepts* (Wiley, New York).
Gescheider, G. A., Bolanowski, S. J., and Verrillo, R. T. (**1989**). ''Vibrotactile masking: Effects of stimulus onset asynchrony and stimulus frequency,'' J. Acoust. Soc. Am. **85**, 2059–2064.
Horner, D. T. (**2000**). ''Perceptual processing at adjacent locations on a single finger: Masking and response competition,'' Percept. Psychophys. **62**(8), 1545–1555.
Johnson, K. O., Yoshioka, T., and Vega-Bermudez, F. (**2000**). ''Tactile functions of mechanoreceptive afferents innervating the hand,'' J. Clin. Neurophysiol. **17**(6), 539–558.
Kirman, J. H. (**1973**). ''Tactile communication of speech: A review and analysis,'' Psychol. Bull. **80**, 54–74.
Klemmer, E. T., and Muller, P. F. (**1953**). ''The rate of handling information—Key-pressing responses to light patterns,'' Human Factors Operations Research Laboratories Memo Report, No. 34.
Leotta, D. F., Rabinowitz, W. M., Reed, C. M., and Durlach, N. I. (**1988**). ''Preliminary results of speech-reception tests obtained with the synthetic Tadoma system,'' J. Rehabil. Res. Dev. **25**(4), 45–52.
Loomis, J. M., and Lederman, S. J. (**1986**). ''Tactual Perception,'' in *Handbook of Perception and Human Performance: Cognitive Processes and Performance*, edited by K. R. Boff, L. Kaufman, and J. P. Thomas (Wiley, New York), Vol. 2, 31/31–31/41.
Mahar, D. P., and Mackenzie, B. D. (**1993**). ''Masking, information integration, and tactile pattern perception: A comparison of the isolation and integration hypotheses,'' Perception **22**, 483–496.
Reed, C. M., and Durlach, N. I. (**1998**). ''Note on information transfer rates in human communication,'' Presence—Teleoperators Virtual Environ. **7**(5), 509–518.
Reed, C. M., Durlach, N. I., and Braida, L. D. (**1982**). ''Research on Tactile Communication of Speech: A Review,'' ASHA Monographs, No. 20.
Reed, C. M., Durlach, N. I., and Delhorne, L. A. (**1992**). ''Natural methods of tactual communication,'' in *Tactile Aids for the Hearing Impaired*, edited by I. R. Summers (Whurr, London), pp. 218–230.
Reed, C. M., Durlach, N. I., Delhorne, L. A., Rabinowitz, W. M., and Grant, K. W. (**1989**). ''Research on Tactual Communication of Speech: Ideas, Issues, and Findings,'' Volta Review, 91 (Monograph entitled ''Research on the use of sensory aids for hearing-impaired people,'' edited by N. S. McGarr, 65–78.)
Reed, C. M., Rabinowitz, W. M., Durlach, N. I., Braida, L. D., Conway-Fithian, S., and Schultz, M. C. (**1985**). ''Research on the Tadoma method of speech communication,'' J. Acoust. Soc. Am. **77**(1), 247–257.
Rinker, M. A., Craig, J. C., and Bernstein, L. E. (**1998**). ''Amplitude and period discrimination of haptic stimuli,'' J. Acoust. Soc. Am. **104**(1), 453–463.
Tan, H. Z. (**1996**). ''Information transmission with a Multi-Finger tactual display,'' Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.
Tan, H. Z., Durlach, N. I., Rabinowitz, W. M., and Reed, C. M. (**1997**). ''Information transmission with a multi-finger tactual display,'' Scand. Audiol. **26**, 24–28.
Tan, H. Z., Durlach, N. I., Reed, C. M., and Rabinowitz, W. M. (**1999**). ''Information transmission with a multifinger tactual display,'' Percept. Psychophys. **61**(6), 993–1008.
Tan, H. Z., and Rabinowitz, W. M. (**1996**). ''A new multi-finger tactual display,'' in *Proceedings of the International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, American Society of Mechanical Engineers Dynamic Systems and Control Division (The American Society of Mechanical Engineers, New York), Vol. 58, pp. 515–522.
Verrillo, R. T., and Gescheider, G. A. (**1992**). ''Perception via the sense of touch,'' in I. R. Summers *Tactile Aids for the Hearing Impaired*, edited by I. R. Summers (Whurr, London), pp. 1–36.

Factors affecting the duration effect in pitch perception for unresolved complex tones[a)]

Louise J. White[b)]
Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, United Kingdom

Christopher J. Plack[c)]
Department of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom

(Received 3 May 2002; revised 8 August 2003; accepted 8 September 2003)

Previous research has shown that fundamental frequency ($F0$) discrimination thresholds for complex tones containing unresolved harmonics decrease as the duration of the tone increases [White and Plack, J. Acoust. Soc. Am. **103**, 2051–2063 (1998)]. In this paper $F0$ discrimination was measured as a function of duration for complexes with $F0$s of 62.5, 125, and 250 Hz, bandpass filtered into two spectral regions (2750–3750 and 5500–7500 Hz). The harmonics were summed either in sine phase (SINE) or with alternating sine-cosine phase (ALT), which affects the envelope of the waveform and the pitch of the complex. Tone duration was 20, 40, 80, and 160 ms. The improvement in $F0$ discrimination with duration increased with decreasing $F0$. When harmonics where spectrally filtered between 2750 and 3750 Hz, for complexes with an $F0$ of 62.5 Hz, $F0$ discrimination thresholds decreased from approximately 30% for a 20-ms tone to approximately 3% for a 160-ms tone. For complexes with an $F0$ of 250 Hz, thresholds decreased from 3% for a 20-ms tone to 1% for a 160-ms tone: a lower envelope repetition rate led to a larger change in performance with increasing duration. The phase manipulation also affected the size of the duration effect, in that the effect was less for an ALT complex compared to a SINE complex with the same $F0$, consistent with the change in envelope repetition rate. Overall, the results suggest that for unresolved complex tones it is primarily envelope repetition rate, not spectral region, that determines both the $F0$ discrimination threshold and the size of the duration effect.
[DOI: 10.1121/1.1621860]

PACS numbers: 43.66.Fe, 43.66.Hg, 43.66.Mk [NFV] Pages: 3309–3316

I. INTRODUCTION

Many of the sounds in our environment such as vowels or musical notes are complex tones. These consist of a series of sinusoidal components (harmonics) whose frequencies are integer multiples of the repetition rate of the complex, also known as the fundamental frequency ($F0$). The pitch of a complex tone usually corresponds to $F0$, even though there may be many harmonics present.

The lower-numbered harmonics in a complex tone (harmonic numbers less than about 10) are resolved by the peripheral auditory system. That is, the individual harmonics produce separate peaks in the excitation pattern on the basilar membrane (Moore and Glasberg, 1983) and higher activity in those auditory-nerve fibers whose characteristic frequencies match the harmonics. Pitch could be determined from these lower-numbered harmonics by recognition of the pattern of peaks in the excitation pattern (Goldstein, 1973; Srulovicz and Goldstein, 1983; Terhardt, 1974) or by assessing the firing patterns of the auditory-nerve fibers tuned to individual harmonics (Meddis and Hewitt, 1991a, b).

A complex containing only higher-numbered, unresolved harmonics still has a pitch corresponding to $F0$ (Houtsma and Smurzynski, 1990; Moore and Rosen, 1979). There is no information in the shape of the excitation pattern regarding the frequencies of individual harmonics, so this pitch must be derived from the temporal pattern of the interaction between the harmonics on the basilar membrane. The temporal pattern repeats at $F0$ and auditory-nerve fibers will tend to phase lock to this, therefore allowing the $F0$ of the complex to be determined. For example, Joris and Yin (1992) showed that auditory neurons in the cat phase locked to amplitude modulation of pure tones for modulation frequencies of up to 1000 Hz; and Grothe *et al.* (2001) found neurons in the medial superior olive of the big brown bat that phase locked to the envelope of a sinusoidally amplitude modulated sound, although the neurons' best frequency was at the carrier rate.

Plack and Carlyon (1995) showed that $F0$ discrimination thresholds were high for short-duration complexes containing only unresolved harmonics, and that the improvement in $F0$ discrimination with increasing tone duration was greater for unresolved than for resolved harmonics. For example, $F0$ discrimination thresholds for a complex tone with an $F0$ of 250 Hz and a duration of 25 ms were around 1% when the complex had been filtered to contain only resolved harmonics and around 8% when the complex contained only unresolved harmonics. As the signal duration was increased to 200 ms, little improvement in performance was seen for the resolved complex, but at 200 ms thresholds for the unre-

[a)]Some of the data in this paper were presented at the 135th meeting of the Acoustical Society of America in Seattle, Washington, in June 1998 [White and Plack, J. Acoust. Soc. Am. **103**(5), 2767(A) (1998)].
[b)]Electronic mail: l.white@kingston.ac.uk
[c)]Electronic mail: cplack@essex.ac.uk

solved complex were decreased to 2%. Plack and Carlyon argued that that the pitch mechanism uses a *longer integration time* for unresolved than for resolved harmonics.

White and Plack (1998) presented further support for this hypothesis. They measured $F0$ discrimination performance in terms of the detectability index, d', as a function of duration for resolved and unresolved complexes with $F0$s of 62.5 and 250 Hz. Stimuli were either single tones or two 20-ms tones separated by a silent gap of varying duration. In agreement with Plack and Carlyon (1995), White and Plack found that the improvement in performance with increasing tone duration was greater for unresolved than for resolved harmonics. For a single tone burst consisting of unresolved harmonics, increases in duration were associated with improvements in $F0$ discrimination up to durations of 160 ms. The size of the duration effect is possibly related to $F0$: It was larger for the complex tone with the lower $F0$. This could indicate that the integration time is longer for low $F0$s. Wiegrebe (2001) reached a similar conclusion based on the pitch of regular-interval noises. It was suggested in that paper that the size of the temporal window for determining pitch is dependent on pitch itself.

White and Plack (1998) found that for stimuli comprised of two 20-ms presentations separated by a silent gap, performance was constant for gap durations of 5 ms and longer. The d' for two 20-ms bursts separated by a gap was approximately $\sqrt{2}$ (1.414) greater than the d' for one 20-ms burst. This suggests that the auditory system was using a mechanism that took independent samples or looks at the signal on either side of the discontinuity and combined them optimally according to the rules of signal detection theory (Viemeister and Wakefield, 1991). White and Plack hypothesized that the pitch mechanism for resolved harmonics uses a relatively short sampling window of perhaps 20-ms duration for both continuous and gapped presentations, while the mechanism for unresolved harmonics appears to use a more complex strategy: This may involve a long integration time for continuous complexes, which may be reset to produce discrete samples when there is a discontinuity in the complex of 5 ms or more.

One of the aims of the present paper was to examine further the effect of duration on $F0$ discrimination for complexes containing unresolved harmonics. Specifically, the aim was to determine whether it is $F0$ that determines the size of the duration effect (and therefore the size of the integration window for the pitch mechanism) and not spectral region or harmonic density, which in many experiments covary with resolvability and $F0$.

The pitch we hear is generally related to the *repetition rate* of the waveform, or the repetition rate of the *envelope* of the waveform for some complex-tone stimuli (these are not necessarily the same; see Shackleton and Carlyon, 1994). This may reflect a common pitch mechanism across both pure and complex tones that is sensitive to repetition rate. For example, Plack and Carlyon (1995) noted that the duration effect for an unresolved complex with an $F0$ of 250 Hz is similar to the duration effect for a pure tone with a frequency of 250 Hz (Moore, 1973). This experiment tests the differential effects of $F0$ and envelope repetition rate on pitch discrimination performance as a function of duration.

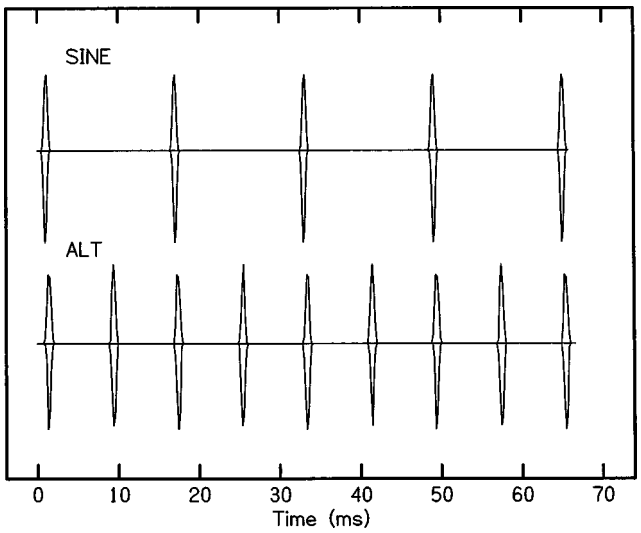


FIG. 1. Envelope of a complex tone with an F0 of 62.5 Hz. Top panel shows all the harmonics added in sine phase (SINE), bottom panel shows alternating harmonics added in cosine phase (ALT).

The experiment made use of a manipulation described by Shackleton and Carlyon (1994). Complexes were comprised of harmonics that were either summed in sine phase (SINE) or in alternating sine-cosine phase (ALT). This manipulation affects the envelope of the waveform (see Fig. 1) but not the harmonic density. Shackleton and Carlyon's pitch-matching experiments revealed that a complex with unresolved harmonics in ALT phase was associated with a pitch of $2*F0$, while a complex with harmonics summed in SINE phase had a pitch at $F0$. The duration of the complex tones used in this study was 400 ms. Shackleton and Carlyon argued that the doubling of pitch with the ALT manipulation only occurs for *unresolved* harmonics. Hence, this stimulus gives one the chance to separate pitch from $F0$ and harmonic density, and the phase manipulation can also be used as a test of resolvability.

II. METHOD

A. Rationale

The experiment tested the hypothesis that it is $F0$ that determines the size of the duration effect and therefore the size of the pitch integration window for tones added in sine phase. $F0$ discrimination thresholds were measured as a function of duration for three $F0$s, two spectral regions, and two phase manipulations: Either all harmonics summed in sine phase (SINE condition) or harmonics alternating between sine and cosine phase (ALT condition).

On the basis of the hypothesis that it is $F0$ and not spectral region which determines the size of the duration effect, it is predicted that there will be very little effect of spectral region on the improvement in performance with increasing tone duration, but that there will be a significant effect of $F0$. The phase manipulation allows pitch and $F0$/harmonic density to be dissociated, so it can be determined which of these is responsible for the size of the duration effect in unresolved complex tones.

B. Stimuli

The $F0$s of the complexes were 62.5, 125, and 250 Hz. The harmonics were bandpass filtered into two distinct spectral regions: 2750 to 3750 Hz and 5500 to 7500 Hz (at the 3-dB down points). Stimuli were filtered using a digital FIR filter with a nominal attenuation of 90 dB per octave. For complexes with $F0$s of 62.5 and 125 Hz, steady-state tone durations of 20, 40, 80, and 160 ms were used. For the 250-Hz $F0$ tone durations were 10, 20, 40, 80, and 160 ms. Stimuli were not ramped, but bandpass filtering occurred after gating, so that wideband spectral splatter was removed. These stimuli were presented either with all the harmonics in sine phase (SINE condition) or alternating harmonics in sine and cosine phase, with the first harmonic in sine phase and the second in cosine phase, etc. (ALT condition). The level of each harmonic component was 50 dB SPL. The interstimulus interval was 500 ms. Figure 1 shows the envelopes of two waveforms with the same 62.5-Hz $F0$, filtered into a 2750–3750-Hz spectral region. The top panel shows the SINE condition and the bottom panel shows the ALT condition.

The stimuli were generated digitally with 32-bit resolution on a Silicon Graphics workstation. The sampling rate was 48 kHz. Stimuli were filtered using a digital FIR filter with a nominal attenuation of 90 dB per octave. Stimuli were played out through a built-in 16-bit DAC and were presented monaurally to the right ear via Beyer DT 100 headphones that were connected directly to the headphone output of the computer.

C. Procedure

A two-interval, two-alternative forced choice task was used throughout with feedback. A two-down, one-up adaptive tracking procedure (Levitt, 1971) was used to measure all the thresholds in the experiment. The $F0$ of the complex in the standard interval was fixed at 62.5, 125, or 250 Hz and the complex in the comparison interval had a higher $F0$ that was varied. The percentage $F0$ difference between the two observation intervals was increased and decreased by a factor of 2 for the first four turnpoints and by a factor of 1.41 thereafter. Sixteen turnpoints were measured and the threshold estimate was taken as the geometric mean of the last 12. Ten such threshold estimates were made and the final estimate was the geometric mean of these ten. The conditions were presented randomly over a number of sessions until each listener had completed the ten threshold estimates. All data for the lower spectral region were collected prior to those for the higher spectral region.

D. Listeners

Five listeners took part in this experiment. One was the first author (S4) and the other four were all experienced in psychoacoustic tasks and were paid an hourly wage. They were aged between 28 and 38 years and had pure-tone thresholds within normal clinical limits: Their audiometric thresholds for pure tones between 250 and 8000 Hz were less than 20 dB SPL. Listeners were given at least 2 h training on the task before data collection began. They were tested individually in an IAC single-walled sound-attenuating booth.

E. Results and analysis

The data are presented as discrimination thresholds ($\Delta F0/F0$ in percent) and are plotted in Fig. 2 for the 2750–3750-Hz spectral region and in Fig. 3 for the 5500–7500-Hz spectral region. Thresholds are shown as a function of tone duration. The data for each subject are in a separate panel and the mean data are in the bottom right panel. For both lower and higher spectral regions, performance improved as the duration of the complex increased. Performance was worse overall for the tones with an $F0$ of 62.5 Hz and improved as the $F0$ of the tone increased. The greatest variation in performance for tones with different $F0$s was at short durations. Notice that the results of the 250-Hz $F0$ ALT condition are not shown. Two of the listeners, S5 and S3, found this condition impossible and the others had very high discrimination thresholds (>20%). The authors felt that performance on this condition was very unreliable and the data should not be included in the analysis. Carlyon and Deeks (2002) have shown recently that repetition rates above 500 Hz are not discriminable for unresolved complex tones. The envelope repetition rate for the 250-Hz ALT condition is 500 Hz.

A repeated measures analysis of variance was conducted on the data combined from Figs. 2 and 3 (excluding the 10-ms data for the 250-Hz $F0$ and the data from the 250-Hz ALT condition). The factors were complex tone type (the five combinations of $F0$ and phase relation), spectral region, and duration. The analysis showed that there was a highly significant effect of the complex tone type [$F(4,16)=26.22$, $p<0.001$] and of duration [$F(3,12)=79.98$, $p<0.001$] on $F0$ discrimination. This is in agreement with previous findings (White and Plack, 1998). There was no overall effect of spectral region [$F(1,4)=1.46$, $p>0.05$]. There was a highly significant interaction between complex tone type and duration, which will be explored in subsequent analyses [$F(4,12)=19.28$, $p<0.001$].

However, there was an unexpected significant interaction between complex tone type and spectral region [$F(4,16)=6.37$, $p<0.05$]. Closer inspection of this effect revealed that in the 250-Hz SINE condition, the $F0$ discrimination thresholds were significantly lower in the 2750–3750-Hz spectral region than in the 5500–7500-Hz spectral region. This appears to be due to two listeners, S2 and S5. As discussed below, this is probably because portions of the 250-Hz SINE stimulus in the lower spectral region are resolved or the stimulus envelope in this region is more "peaky," which would also give listeners additional information upon which to make their pitch discrimination judgment. Similarly, there was also a highly significant interaction between spectral region and duration [$F(3,12)=14.77$, $p<0.01$] which was unexpected. The thresholds for the 20-ms complexes were higher in the lower spectral region. Although all the listeners completed the task for the low spectral region first, the complex tone types and durations were

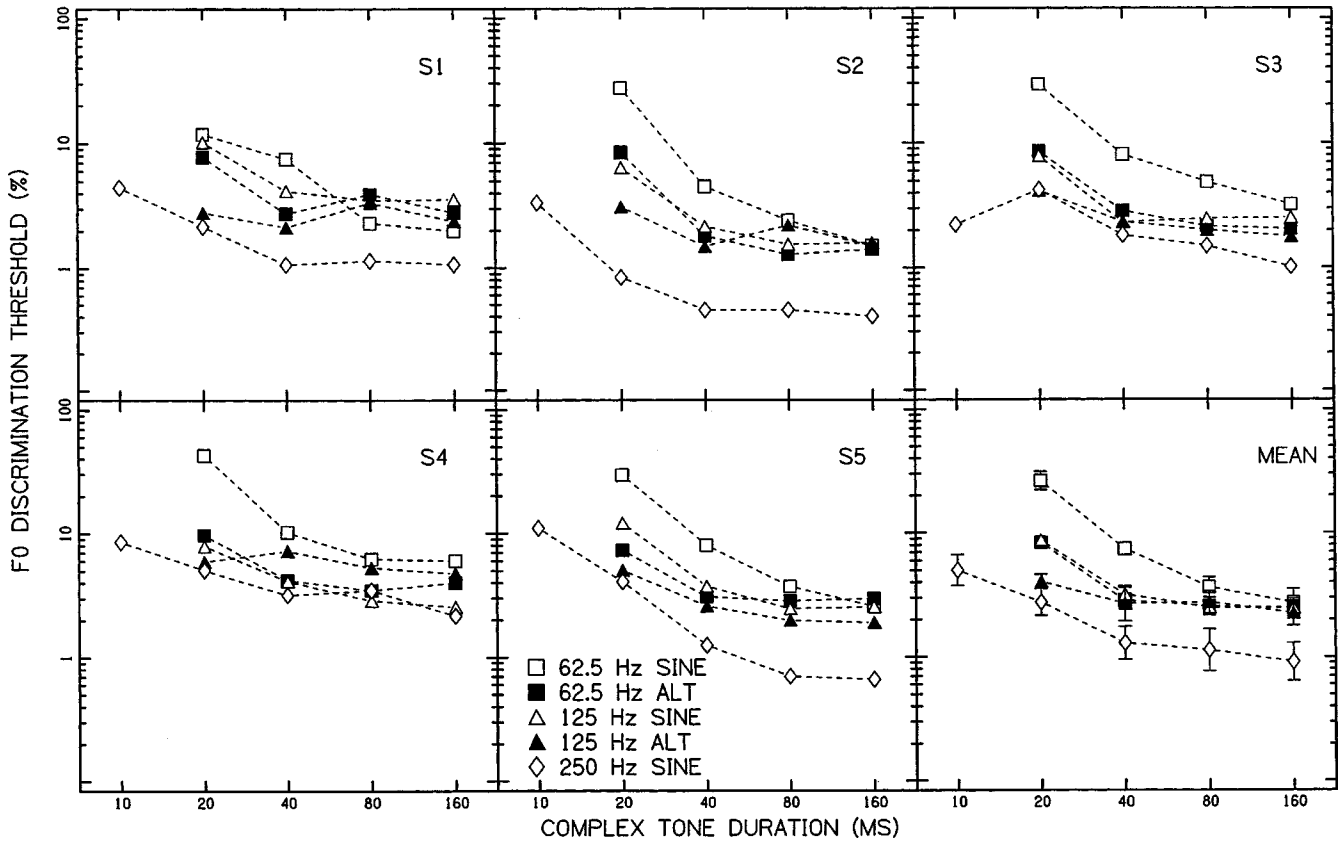


FIG. 2. *F*0 discrimination as a function of duration for complex tones filtered between 2750 and 3750 Hz. Error bars are standard errors.

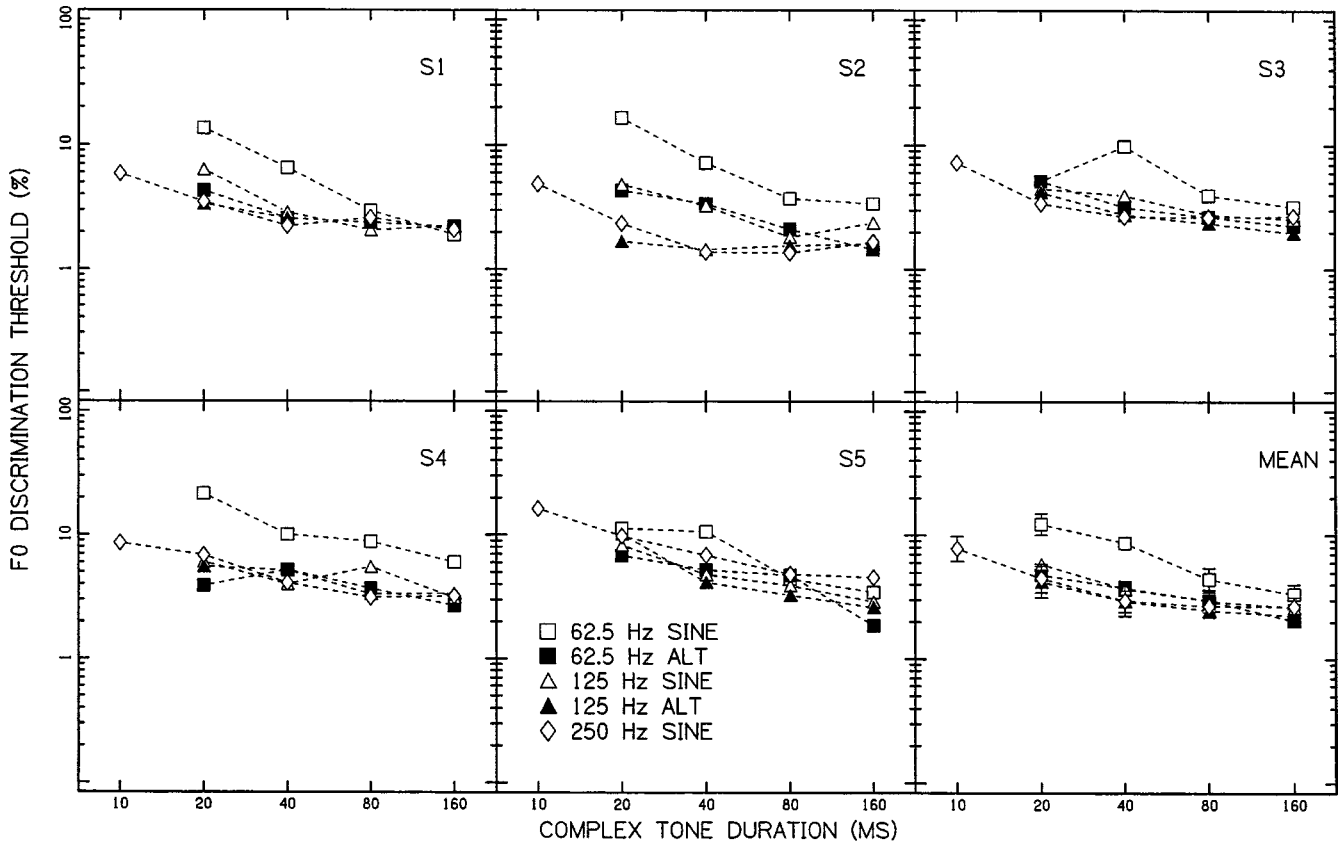


FIG. 3. *F*0 discrimination as a function of duration for complex tones filtered between 5500 and 7500 Hz. Error bars are standard errors.

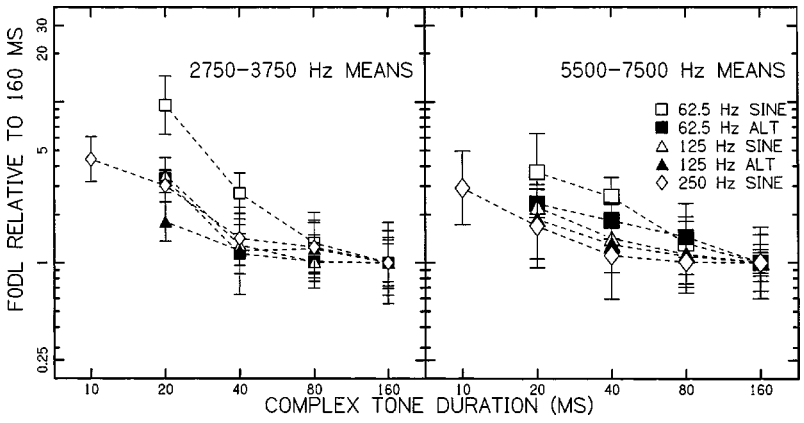


FIG. 4. The results from Figs. 2 and 3 expressed as the ratio between the $F0$ discrimination threshold for a 40-ms tone and the $F0$ discrimination threshold for a 160-ms tone.

randomized so it is not likely to be a practice effect, and similar differences for other tone durations were not seen.

When looking at Figs. 2 and 3 it can be seen that the thresholds in the 2750–3750-Hz spectral region for the 250-Hz SINE condition were much lower than the others. It is possible that, in the 250-Hz SINE condition, some of the harmonics were partially resolved in the low-frequency skirts of the filter, leading to better $F0$ discrimination performance in this spectral region (Plack and Carlyon, 1995; White and Plack, 1998). Also, if one compares the patterns of data for the 62.5-Hz ALT condition and the 125-Hz SINE condition, they are very similar, reflecting the similarity in the envelopes of the waveforms for these conditions. The fact that the 250-Hz SINE condition data are different from the 125-Hz ALT data could be explained by there being some resolved harmonics in the 250-Hz $F0$ condition. The 250-Hz SINE data from the low spectral region were therefore excluded from subsequent analyses.

A further analysis was carried out for complexes of 40 ms or more, excluding the 20-ms data and all the data from complexes with an $F0$ of 250 Hz (both SINE and ALT conditions). Once again there was no overall effect of spectral region [$F(1,4)=1.19$, $p>0.05$] but there were significant effects of complex tone type [$F(3,12)=19.24$, $p<0.01$], and of duration [$F(2,8)=61.12$, $p<0.01$]. There was a significant interaction between complex tone type and duration [$F(6,24)=23.34$, $p<0.01$]. There was no interaction between spectral region and complex tone type [$F(3,12)=1.81$, $p>0.05$], nor was there an interaction between duration and spectral region [$F(2,8)=1.94$, $p>0.05$]. Overall, this analysis demonstrates that for tones of 40 ms or more, the effect of tone duration on $F0$ discrimination for complexes consisting of *unresolved* harmonics is not affected by spectral region.

In order to compare the size of the duration effect for the different conditions and spectral regions, $F0$ discrimination thresholds were plotted relative to the thresholds in the corresponding 160-ms condition (Fig. 4). This was felt to be a simple and satisfactory measure of the size of the duration effect. There is a great deal of variability in the data for the 20-ms duration, with the discrimination thresholds being significantly higher for the complexes with harmonics in the lower spectral region (see the previous analysis). Therefore, it was decided to use the ratios of 40-ms discrimination thresholds to 160-ms discrimination thresholds (hereafter called 40/160 ratios) to explore the duration effect.[1]

A repeated measures analysis of variance was performed on the 40/160 threshold ratios from Fig. 4 and Table I, excluding the complex tones with an $F0$ of 250 Hz. This allowed us to specifically look at the effect of phase (spectral region$\times F0\times$phase). There was no effect of spectral region on the size of the duration effect [$F(1,4)=5.36$, $p>0.05$], but there was, clearly, a significant effect of fundamental frequency [$F(1,4)=18.42$, $p<0.05$] and of phase [$F(1,4)=10.13$, $p<0.05$]. There was no interaction between spectral region and $F0$ [$F(1,4)=1.19$, $p>0.05$] nor between spectral region and phase [$F(1,4)=2.819$, $p>0.05$].

TABLE I. The ratios of 40-ms discrimination thresholds to 160-ms discrimination thresholds.

	S1	S2	S3	S4	S5	Mean	Standard error
2750–3750 Hz							
62.5-Hz SINE	3.77	2.99	2.51	1.70	3.11	**2.72**	0.34
62.5-Hz ALT	0.99	1.29	1.40	1.05	1.04	**1.14**	0.08
125-Hz SINE	1.18	1.36	0.92	1.62	1.47	**1.29**	0.12
125-Hz ALT	0.90	0.99	1.31	1.52	1.36	**1.19**	0.12
250-Hz SINE	1.00	1.12	1.8	1.48	1.90	**1.42**	0.18
5500–7500 Hz							
62.5-Hz SINE	3.42	2.13	3.08	1.66	3.07	**2.58**	0.33
62.5-Hz ALT	1.17	2.33	1.41	1.93	2.79	**1.83**	0.30
125-Hz SINE	1.28	1.37	1.55	1.28	1.65	**1.42**	0.08
125-Hz ALT	1.23	0.90	1.40	1.53	1.58	**1.30**	0.12
250-Hz SINE	1.07	0.83	0.99	1.27	1.51	**1.11**	0.26

F. Effects of phase manipulation on the size of the duration effect

Shackleton and Carlyon (1994) demonstrated that the pitch of an unresolved complex with the ALT phase manipulation was one octave higher than the same complex with all the harmonics summed in sine phase. One of the aims of this study was to determine the effect of this phase manipulation on the size of the duration effect. In order to do this means comparisons were carried out between 40/160 ratios for the conditions collapsed across spectral regions. If the size of the duration effect is determined by the $F0$ of the complex, there will be no effect of the phase manipulation and no significant difference between the ratios with the same $F0$s. If, however, the size of the duration effect is governed by the *pitch* of the complex, then there will be a significant difference between complexes with the same $F0$ but with harmonics summed with different starting phases (ALT vs SINE). The envelope of, say, a 62.5-Hz $F0$ complex with the ALT manipulation is similar to that of a 125-Hz $F0$ complex with all the harmonics summed in sine phase. If the size of the duration effect is governed by the repetition rate of peaks in the envelope, one might expect the size of the effect to be the same for these two complexes.

As stated above, overall the phase manipulation *did* cause a significant difference between the duration effects (measured as the 40/160 threshold ratios) [$F(1,4)=10.13$, $p<0.05$] and a significant interaction between $F0$ and phase [$F(1,4)=8.11$, $p<0.05$]. Means comparisons showed that for the complex with an $F0$ of 62.5 Hz the duration effects were significantly larger in the SINE condition compared with the ALT condition [$t(9)=3.59$, $p<0.05$]. However, for complexes with an $F0$ of 125 Hz there was no significant difference in the size of the duration effect for the ALT and SINE conditions [$t(9)=1.15$, $p>0.05$]. There was also no significant difference between the size of the duration effect for the 62.5-Hz ALT condition and the 125-Hz SINE condition [$t(9)=0.90$, $p>0.05$]. This pattern of results is expected if it is pitch, not $F0$, that determines the size of the duration effect. In order to explore the size of the duration effect for the 250-Hz $F0$ tone, a means comparison was carried out on the 40/160 ratios from the higher spectral region only (filtered between 5500–7500 Hz) for the 125-Hz ALT and 250-Hz SINE complexes. Even though the difference between the means for these conditions was small, a significant difference was found between the size of the duration effect for the 125-Hz ALT and the 250-Hz SINE complex tones [$t(4)=3.01$, $p>0.05$]. According to Shackleton and Carlyon, these complexes should have the same pitch.

Overall, the evidence from the above analyses suggests that it is the pitch, not the $F0$ *per se*, of the complex that determines the size of the duration effect for the lower $F0$ complexes ($F0$s of 62.5 and 125 Hz). However, the pattern of the analyses for complexes whose envelopes repeat at more than 125 Hz is not so clear. It is not so easy to dissociate the effects of $F0$ and pitch on the size of the duration effect for these complexes. If one inspects the individual listeners' data (Figs. 2 and 3), two of the subjects (S2 and S3) *do* show very similar patterns of results for those complexes with similar envelope (and pitch) for the higher $F0$s.

In summary, this experiment replicated earlier findings that $F0$ discrimination performance for complexes containing unresolved harmonics improves with duration. The largest relative improvement in performance with duration was seen for complexes with a low $F0$. The size of this duration effect was mainly dependent on envelope repetition rate or pitch, rather than on spectral region or $F0$ *per se*.

III. GENERAL DISCUSSION

A. Effects of baseline performance

One factor that may underlie the duration effect for these stimuli is the baseline discrimination performance. If performance is poor for short-duration complexes then there may be more possibility for improvement as duration is increased, perhaps before asymptotic performance is reached. In other cases for which performance is relatively better, the size of the duration effect could be limited by ceiling effects. Alternatively, it is conceivable that the auditory system uses a longer integration window when the cues to pitch are not robust. In order to ascertain whether the size of the duration effect was determined by the baseline discrimination ($F0$ discrimination threshold) at 160 ms or the overall repetition rate of the envelope, correlations were calculated between the $F0$ discrimination threshold at 160 ms and the duration effect (40/160 ratio), and between the repetition rate and the duration effect. These correlations were carried out on the mean data, excluding the data from the 250-Hz $F0$ in the lower spectral region. There was no significant correlation between the size of the duration effect and baseline discrimination ($r=-0.94$, df=24, $p>0.05$). There was a significant correlation between the size of the duration effect and the repetition rate ($r=-0.68$, df=24, $p<0.05$). The correlation between the $F0$ discrimination threshold at 160 ms and the repetition rate was also determined, as it was necessary for determining the relative significance of the other two correlations. The strength of the correlations was then compared using Williams's equation (cited in Howell, 1997). The size of the duration effect was significantly more highly correlated with repetition rate than with the $F0$ discrimination threshold at 160 ms [$t(22)=3.47$, $p<0.01$]. The direction of the correlations demonstrates that the larger duration effects are seen for tones with a lower repetition rate. The analysis indicates that overall it is repetition rate that plays a role in determining the size of the duration effect rather than absolute $F0$ discrimination performance, with a larger duration effect for complexes with a lower repetition rate. If the duration effect reflects the size of an integration window, then the results suggest that the auditory system uses a longer integration window for tones with a lower repetition rate.

B. Is discrimination limited by the number of waveform cycles?

The analyses presented here demonstrate that it is the envelope repetition rate of the complex waveform which is the main determinant of the size of the duration effect. Following this, it is possible that discrimination performance is limited by the *number of waveform cycles* in the stimulus, as would be predicted for pure tones by the optimal processing

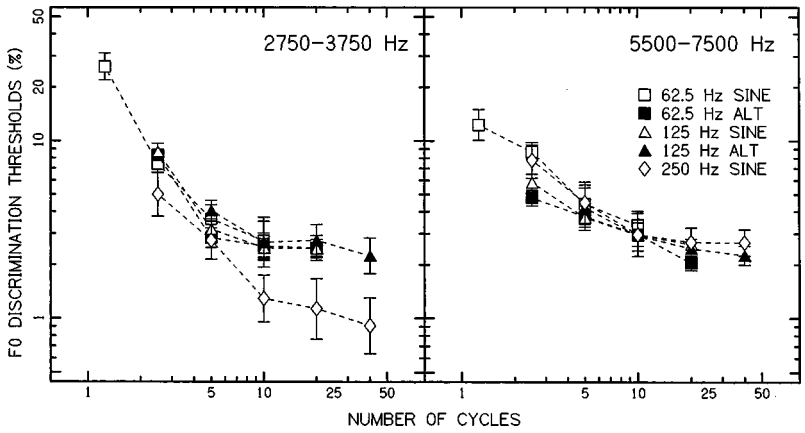


FIG. 5. *F*0 discrimination thresholds for unresolved complexes expressed in terms of the number of envelope cycles in the stimulus.

model of Hanekom and Krüger (2001). Perhaps on account of the variability of phase locking, the auditory system may need a certain number of cycles to make an accurate estimate of the repetition rate, independent of the spectral composition of the stimulus.

In order to illustrate this further, the mean $F0$ discrimination thresholds were plotted as a function of the *number of cycles* of the tone they contained. (For the ALT stimuli, this was calculated on the basis of the repetition rate of the envelope, twice $F0$.) For example, a 40-ms tone with an $F0$ of 62.5 Hz (SINE condition) contains 2.5 repetitions of the envelope. Thus, it was possible to compare the baseline performance and the duration effects for unresolved complex tones in terms of the number of cycles. The results are shown in Fig. 5.

Aside from the data for the 250-Hz SINE condition in the low spectral region, for which some of the harmonics were probably resolved, there is reasonable agreement across the complex-tone conditions when this measure is used. The $F0$ discrimination threshold is approximately constant for a given number of envelope cycles.

IV. CONCLUSIONS

This experiment has extended the findings of Plack and Carlyon (1995) and White and Plack (1998). The main findings of this study can be summarized as follows.

(i) The improvement in discrimination performance with duration for unresolved complex tones with a low $F0$s and durations between 40 and 160 ms is dependent mainly on the repetition rate of the envelope. The spectral region of the tone and the harmonic density of the tone have little influence on the duration effect.

(ii) For these stimuli, the F0DL may be determined primarily by the number of cycles of the stimulus waveform.

ACKNOWLEDGMENTS

The research was supported by a grant from the Medical Research Council (UK), G9630752N. Helpful comments on an earlier version of this paper were provided by Michael Akeroyd and two anonymous reviewers.

[1]A repeated measures analysis of variance was performed on the 20/160ms ratios. There was, as expected, a significant effect of spectral region [$F(1,4)=8.38$, $p<0.05$] and a significant interaction between the effect of spectral region and the effect of duration [$F(4,16)=5.07$, $p<0.01$]. It is not clear why these 20-ms data should be different for the lower spectral region. In light of these results it was decided that it would be more meaningful to use the 40/160-ms ratio to explore the duration effect.

Carlyon, R. P., and Deeks, J. M. (**2002**). ''Limitations on rate discrimination,'' J. Acoust. Soc. Am. **112**, 1009–1025.

Goldstein, J. L. (**1973**). ''An optimum processor theory for the central formation of the pitch of complex tones,'' J. Acoust. Soc. Am. **54**, 1496–1516.

Grothe, B., Covey, E., and Casseday, J. H. (**2001**). ''Medial superior olive of the big brown bat: Neuronal responses to pure tones, amplitude modulations, and pulse trains,'' J. Neurophysiol. **86**, 2219–2230.

Hanekom, J. J., and Krüger, J. J. (**2001**). ''A model of frequency discrimination with optimal processing of auditory nerve spike intervals,'' Hear. Res. **151**, 188–204.

Houtsma, A. J. M., and Smurzynski, J. (**1990**). ''Pitch identification and discrimination for complex tones with many harmonics,'' J. Acoust. Soc. Am. **87**, 304–310.

Howell, D. C. (**1997**). *Statistical Methods for Psychology* (Wadsworth, Belmont).

Joris, P. X., and Yin, T. C. T. (**1992**). ''Responses to amplitude-modulated tones in the auditory nerve of the cat,'' J. Acoust. Soc. Am. **91**, 215–232.

Levitt, H. (**1971**). ''Transformed up–down methods in psychophysics,'' J. Acoust. Soc. Am. **49**, 467–477.

Meddis, R., and Hewitt, M. (**1991a**). ''Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery: Phase sensitivity,'' J. Acoust. Soc. Am. **89**, 2883–2894.

Meddis, R., and Hewitt, M. (**1991b**). ''Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery: Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.

Moore, B. C. J. (**1973**). ''Frequency difference limens for short-duration tones,'' J. Acoust. Soc. Am. **54**, 610–619.

Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, New York).

Moore, B. C. J., and Glasberg, B. R., (**1983**). ''Suggested formulae for calculating auditory-filter bandwidths and excitation patterns,'' J. Acoust. Soc. Am. **74**, 750–753.

Moore, B. C. J., and Rosen, S. M. (**1979**). ''Tune recognition with reduced pitch and interval information,'' Q. J. Exp. Psychol. **31**, 229–240.

Plack, C. J., and Carlyon, R. P. (**1995**). ''Differences in fundamental frequency discrimination and frequency modulation detection between complex tones consisting of resolved and unresolved harmonics,'' J. Acoust. Soc. Am. **98**, 1355–1364.

Schouten, J. F. (**1970**). ''The residue revisited,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Lieden).

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.

Srulovicz, P., and Goldstein, J. L. (**1983**). ''A central spectrum model: A synthesis of auditory-nerve timing and place cues in monaural communi-

cation of frequency spectrum,'' J. Acoust. Soc. Am. **73**, 1266–1276.
Terhardt, E. (**1974**). ''Pitch, consonance, and harmony,'' J. Acoust. Soc. Am. **55**, 1061–1069.
Viemeister, N. F., and Wakefield, G. H. (**1991**). ''Temporal integration and multiple looks,'' J. Acoust. Soc. Am. **90**, 858–865.
Wakefield, G. H., and Nelson, D. A. (**1985**). ''Extension of a temporal model of frequency discrimination: Intensity effects in normal and hearing impaired listeners,'' J. Acoust. Soc. Am. **77**, 613–619.
White, L. J., and Plack, C. J. (**1998**). ''Temporal processing of the pitch of complex tones,'' J. Acoust. Soc. Am. **103**, 2051–2063.
Wiegrebe, L. (**2001**). ''Searching for the time constant of neural pitch extraction,'' J. Acoust. Soc. Am. **109**, 1082–1091.

Binaural models and the strength of dichotic pitches

William M. Hartmann[a] and Peter Xinya Zhang
Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824

(Received 21 April 2003; revised 21 August 2003; accepted 12 September 2003)

Modern physiologically based models of the binaural system incorporate internal delay lines in the pathways from left and right peripheries to central processing nuclei. Different binaural models for the formation of dichotic pitch employ these delay lines in different ways. Consequently, the different models make different predictions for the relative strengths of dichotic pitches made with particular phase conditions. The differences are magnified for dichotic pitches at low frequencies where especially long delay lines may be required. Data from four low-frequency pitch strength experiments on pure-tone-like dichotic pitches (two on Huggins pitch and two on binaural coherence edge pitch) are consistent with models of the equalization-cancellation type and not consistent with the central activity pattern model. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1624072]

PACS numbers: 43.66.Pn, 43.66.Ba, 43.66.Fe [LRB] Pages: 3317–3326

I. INTRODUCTION

Among the diverse phenomena arising in the study of human hearing, dichotic pitch is one of the most fascinating (Cramer and Huggins, 1958; Bilsen and Raatgever, 2000, 2002). A dichotic pitch stimulus consists of a white noise in the left ear and a white noise in the right ear, with nothing in the amplitude or phase spectrum of the noise in either ear alone that is frequency specific. Nevertheless, a listener has the impression of a clear pitch located somewhere within the head. The listener can match that pitch consistently to within a few percent (Hartmann, 1993). The origin of dichotic pitches is not to be found in the individual noises sent to left and right ears. Instead, it is in the relationship between the two noises, particularly their relative (interaural) phase relationship.

Dichotic pitches may be classified into two types, pure-tone-like and complex-tone-like. The former are made with an interaural phase relationship that encodes a single frequency leading to a binaural sensation rather like a sine tone in noise. These include the Huggins pitch (Cramer and Huggins, 1958; Guttman, 1962), the binaural edge pitch (Klein and Hartmann, 1980; Frijns *et al.*, 1986), and the binaural coherence edge pitch (Hartmann and McMillon, 2001). These pitches have been compared for a large number of listeners by Akeroyd *et al.* (2001). The complex-tone-like pitches are made with interaural phase relationships that somehow encode many harmonically related frequencies, and they sound like complex tones in noise. These effects include the Fourcin pitch (Fourcin, 1962, 1970), the dichotic repetition pitch (Bilsen, 1972; Bilsen and Goldstein, 1974), and the multiple phase shift pitch (Bilsen, 1976).

The present article focuses on two pure-tone-like dichotic pitches, the Huggins pitch (HP) and the binaural coherence edge pitch (BICEP). These two effects were studied to compare two well-known binaural models, the equalization-cancellation (EC) model (Durlach, 1960, 1972) and the central activity pattern (CAP) model (Raatgever and Bilsen, 1986). Specifically, we asked which model best accounts for the relative strength of the different phase configurations for these two kinds of dichotic pitch.

According to the EC model, the binaural system *subtracts* the left- and right-ear signals to form a central spectrum.[1] The subtraction process is preceded by an equalization stage in which the amplitudes are adjusted and the phases are shifted so as to optimize the central spectrum for performing the task. For instance, it maximizes the signal-to-noise ratio in a signal detection task. In modern versions of the EC model, such as the modified EC model (e.g., Culling *et al.*, 1998a,b), the phase shifting operation is performed by binaural delay lines, tuned in frequency and in interaural delay. Therefore, different frequency bands may be equalized by different phase shifts. In the EC model, a dichotic pitch is detected as an anomalous point (peak or edge) along the internal frequency axis in the central spectrum, as modified by EC operations (Durlach, 1962).

According to the CAP model, the binaural system *adds* the left- and right-ear signals in channels that are tuned both in frequency and in interaural time delay (ITD) to create a central activity pattern in the frequency-ITD plane. In this model a dichotic pitch is detected as an excitation peak in that plane, with the pitch itself determined by the frequency axis and the lateralization of the pitch image determined by the ITD axis. Although addition is not a popular model for binaural processing, the Appendix shows that it can be regarded as equivalent to binaural cross-correlation in the tonotopic-ITD plane, as introduced by Jeffress (1948).

Both the EC and the CAP models depend upon interaural delay lines. The EC model uses the delay lines to perform the phase equalization operation. The CAP model uses the delay lines as a fundamental part of the binaural display. The models differ in their rules for binaural combination; the EC model employs subtraction but the CAP model uses addition.

The goal of the present article is to distinguish between the EC and CAP models. The key to our approach lies in the different binaural combination rules (subtraction versus addition) and in the character of the population of interaural

[a]Electronic mail: hartmann@pa.msu.edu

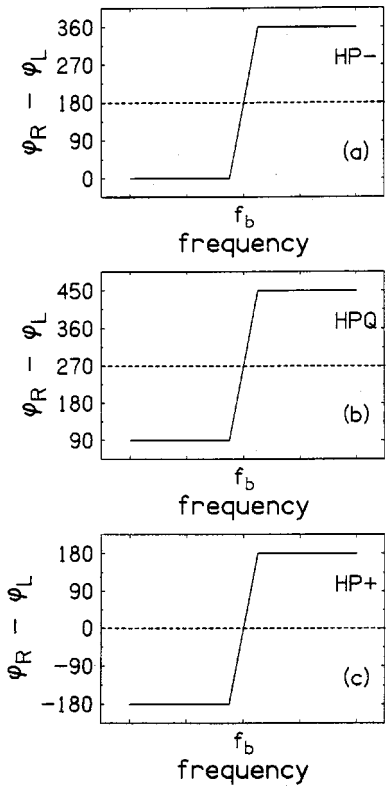


FIG. 1. Although the phases of noise components for Huggins pitch (HP) are random, there is a strict relationship between the phases for left and right ears as a function of component frequency. Parts (a)–(c) show the interaural phase functions for HP−, HPQ, and HP+ respectively. In each case the interaural phases change by 360° in the boundary region. The center of the boundary region is the boundary frequency, f_b. Pitch matching experiments show that the Huggins pitch corresponds to f_b.

delay lines. In contemporary models of the binaural system, delay lines become less numerous, and therefore noisier on the average, as the delay increases to large values (Stern and Trahiotis, 1995). The decrease in fidelity with increasing interaural delay τ is commonly modeled by the central weighting function $p(\tau)$. In the EC model $p(\tau)$ is used to explain the larger masking level difference (MLD) in the $N0S\pi$ configuration compared to the $N\pi S0$ configuration, especially for low signal frequencies (Colburn, 1977). In the CAP model, the central weighting accounts for the unique lateralization of dichotic pitches (Raatgever, 1980; Raatgever and Bilsen, 1986; Bilsen and Raatgever, 2000).[2] The focus of this article is therefore on dichotic pitches at low frequencies where long delay lines may be required. The selective degradation of the binaural processing for long delay lines gives the experiments of this article their special leverage.

II. MODELS FOR THE HUGGINS PITCH

The Huggins pitch is made with a broadband noise that is identical in the two ears except for the interaural phases of the spectral components. Figure 1(a) shows the stimulus known as HP−, where the interaural phase difference is zero (or 360°, equivalent to zero) for all frequencies except for the boundary region where the interaural phase increases monotonically from 0 to 360°. The boundary region is centered on the boundary frequency f_b where the interaural phase is 180°. Figure 1(c) shows the stimulus called HP+, which is the same as HP− except for a global phase shift of 180°, equivalent to the simple inversion of one of the two channels. It is useful to introduce the concept of a *background* phase, 0° for HP− and 180° for HP+.

According to the EC model of Huggins pitch, the binaural system cancels the background components of HP− [Fig. 1(a)] by subtraction, leaving a residual consisting only of the components in the boundary region. The peak of the residual occurs where the interaural phase is 180°, and that peak corresponds to the perceived pitch. By contrast, detecting HP in the HP+ configuration [Fig. 1(c)] requires internal delay lines to adjust the phases of the low- and high-frequency components. These components must be shifted by about 180° (i.e., half the period corresponding to f_b) to make the effective interaural background phase close to zero. Only then can the low and high frequencies be cancelled in the subtraction process.

Because detection of HP+ requires interaural delay, and detection of HP− does not, the EC model predicts that HP− should be stronger. The difference might occur for all boundary frequencies, even for favorable frequencies like 500 Hz, because in any tuned channel it is not possible to completely cancel a constant π phase shift with a single internal delay. As the boundary frequency is reduced, for example, towards 100 Hz, the required delays increase towards 5 ms. Delays this long stretch the limits of any binaural model and the EC model clearly predicts that detection performance will be worse for HP+ compared to HP− for boundary frequencies in this range.

The operation of the CAP model is shown in Fig. 2 where the loci of the phase boundary frequencies are shown in parts (a) and (c) for HP− and HP+, respectively, and filled and open circles correspond to the lowest and highest boundary frequencies used in our experiments. Thus, according to the CAP model, HP+ and HP− are detected at different ITDs in the frequency-ITD plane. Because the interaural phase at the boundary frequency is zero for HP+, the image for HP+ appears directly in the center (ITD=0) so that it is not necessary for any delay lines to be involved. Because the interaural phase at the boundary frequency is π for HP−, the image for HP− appears at the left or right. For HP− with a boundary frequency of 100 Hz, the image appears at an ITD value of plus or minus 5 ms, where delay lines are sparse and the image is weak. In their 1986 paper on the CAP model, Raatgever and Bilsen used the $p(\tau)$ function developed by Colburn (1977). For an ITD of 5 ms, $p(\tau)$ drops to 1% of its small-ITD value. Therefore, the CAP model clearly predicts that detection performance will be worse for HP− compared to HP+ for boundary frequencies in this range.

One more model should be added to the pair of models under consideration, namely Green's 1966 variation on the EC model. In Green's variation, the cancellation process can be done either by addition or subtraction. As it turned out, Green's variation provided a uniquely convenient model for binaural edge pitch (Klein and Hartmann, 1980), and it may find increased support from recent physiology emphasizing the joint role of excitation and inhibition in the medial superior olive (Grothe, 2000; Brand *et al.*, 2002). According to

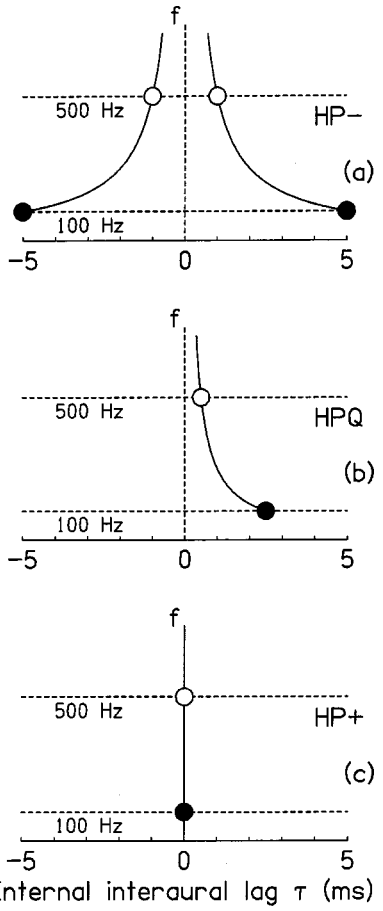


FIG. 2. Peaks of the central spectra according to the CAP model for HP−, HPQ, and HP+. The location of the peaks of the excitation are plotted in the frequency-ITD plane. (a) For boundary frequency f_b, the HP− pitch image appears at an ITD of $1/(2f_b)$, (b) the HPQ image appears at an ITD of $1/(4f_b)$, and (c) the HP+ image appears at an ITD of zero. If f_b is low then HP− ought to be weak because of its extreme laterality. The operation is indicated by the highest and lowest frequencies used in the experiments, indicated by open filled symbols.

Green's variation, both HP+ and HP− can be formed without internal delay lines. For HP+, cancellation except for the boundary region is achieved by addition. For HP−, cancellation except for the boundary region is achieved by subtraction. However, for interaural background phases other than 0 and 180° the model achieves only imperfect cancellation of the background. The most disadvantageous case occurs when the background phases become ±90°, and the component at the boundary frequency is subjected to the same cancellation as the background. Therefore, Green's variation would predict that the worst performance would occur for ±90°, a condition called HPQ (Q for ''quadrature'') as shown in Fig. 1(b).

III. EXPERIMENT 1—HUGGINS PITCH DETECTION

The goal of experiment 1 was to test the predictions of the models for HP for low boundary frequencies. According to the EC model, detection performance should be best for HP− and worst for HP+. According to the CAP model performance should be best for HP+ and worst for HP−. According to Green's version of the EC model, performance should be best for HP+ and HP− and worse for HPQ.

A. Method

The listener heard two noises in succession. One of them had a Huggins-type phase boundary region, the other did not. The ability of a listener to correctly distinguish between these noises was taken to be a measure of the strength of the HP.

1. Stimuli

Each noise stimulus had two channels. The left-ear channel was broadband white noise with 16 384 spectral components of equal amplitude and random phase. To create a HP stimulus, the right-ear channel was generated from the left-ear channel by changing the phases of components. Below the phase boundary region, the interaural phase was 0 for HP−, 90° for HPQ, and −180° for HP+. In the boundary region, the interaural phase increased linearly with frequency by 360° so that the interaural phases for high-frequency components were the same as for the low. Therefore, at the boundary frequency (center frequency of the boundary region) the interaural phases were 180° for HP− (that's why it's called HP−), 270° for HPQ, and 0° for HP+. These phase variations are shown in Fig. 1. The width of the boundary region was 6% of the boundary frequency.

To create the noise without the HP, the interaural phase was the same as for the low-or high-frequency regions of the HP noise. Therefore, the listener could not use a change in the background interaural phases (e.g., $N0$ vs. $N\pi$) to make decisions. The listener had to use information within the boundary region.

Noise stimuli were calculated by an array processor (Tucker Davis AP2) and converted to audio by 16-bit DACs (Tucker Davis DD1). The sample rate per channel was 20 ksps. With a continuously cycled memory buffer of 32 768 words, the period was 1.6384 s and the frequency spacing between adjacent components was 0.61 Hz. The maximum frequency of the broadband noise as converted was 10 kHz, and the output was low-pass filtered at 8 kHz at a rate of −115 dB/oct. The phase integrity of the two channels was tested by adding or subtracting them electronically at a point in the signal chain immediately prior to the headphone power amplifiers. The cancellation was good to at least 40 dB.

2. Listeners

There were three listeners, M (female, age 42), W (male, age 61), and X (male, age 26). All listeners had normal hearing except that W had a bilateral hearing loss above 8 kHz typical of males of this age. Listeners W and X were the authors and had considerable experience in dichotic listening. Listener M had no previous experience.

3. Procedure

On each trial, the listener heard two 500-ms noise intervals with a silent gap of 500 ms between them. One of the intervals included the HP. The listener's task was to decide which interval had the pitch and to press the corresponding button on a response box. There was no feedback.

The phases of the components were randomized on each interval. Each experimental run (block) included 80 trials,

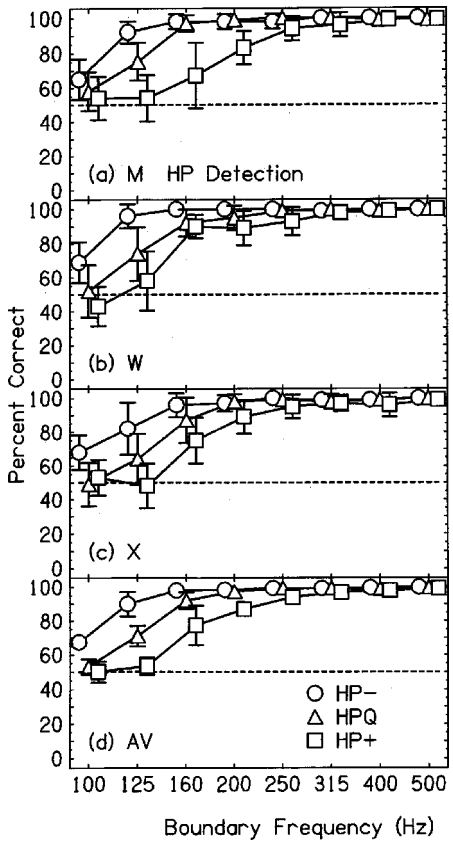


FIG. 3. Results of experiment 1—Huggins pitch detection for three listeners (a,b,c) followed by the average (d) as a function of the boundary frequency. Error bars are two standard deviations in overall length.

ten trials with each of eight nominal boundary frequencies: 100, 125, 160, 200, 250, 315, 400, and 500 Hz. On each trial the boundary frequency was varied by a random amount within a rectangular distribution having an overall width of 6% of the nominal boundary frequency.[3] The order of appearance of the boundary frequencies within a run was random. Every run presented a fixed type of Huggins stimulus—HP−, HPQ, or HP+. Each listener did ten runs of each type, in a haphazard order based on alternation and randomness. The level of the noises was 65 dB, making the spectrum level 26 dB $re\ 10^{-12}$ W per square meter per Hertz. The listener heard the stimuli via Sennheiser HD 480-II headphones while seated in a double-walled sound-treated room.

B. Results

The results of experiment 1 are given in Fig. 3, which shows the percentage of correct identifications of the interval with the HP stimulus as a function of the boundary frequency for the three different phase configurations, or stimulus types.

1. Observations

Figure 3 shows two trends, common among the listeners. First, for each stimulus type (HP−, HPQ, or HP+) performance decreased for decreasing boundary frequency. Second, for boundary frequencies where performance was better than guessing but worse than perfect, there was a clear ordering among the stimulus types. The ordering was the same for all three listeners, and, consequently, for the average shown in the bottom panel. Performance was best on HP−, worst on HP+, and intermediate on HPQ. This is the ordering predicted by the EC model.

2. Statistics

Statistical tests of the results were complicated by both floor (near 50%) and ceiling (near 100%) effects in performance. To make the critical test between HP− and HP+ we included paired comparisons where at least one of the two configurations led to a correct response rate between 55% and 95%. For the three listeners and eight frequencies there were 15 such pairs. The difference in percentages of correct responses had the same sign (sign test) for all 15 pairs, in favor of HP−. A one-tailed *t*-test at the 0.05 level found a significant advantage for HP− on 14 of the 15 pairs.

IV. EXPERIMENT 2—HUGGINS PITCH DISCRIMINATION

Whereas experiment 1 studied HP detection, experiment 2 studied HP discrimination—specifically discrimination between two different boundary frequencies, which lead to two different pitches. An advantage of the discrimination experiment in studying the Huggins pitch effect is that it tests for the musical sense of highness or lowness that is the hallmark of pitch. Other investigators have used a sine-tone matching task (e.g., Guttman, 1962). A discrimination task was actually used in the first reports of dichotic pitch effects (Cramer and Huggins, 1958). Experiment 2 was a boundary-frequency discrimination experiment to compare HP−, HPQ, and HP+.

A. Method

The stimuli, listeners, and protocol were the same as in experiment 1. The difference was that in experiment 2 both intervals included a phase boundary region. The boundary frequency in the second interval was 9% (1.5 semitones) higher, or 9% lower, or the same as in the first interval. In a three-alternative forced-choice task the listener had to compare the pitches on the two intervals and indicate a choice (higher, lower, or the same) by means of push buttons. The advantage of the three-alternative task is that it reduced the guessing limit to 33% correct, affording a greater range for comparing the three HP configurations.

B. Results

The results of the HP discrimination experiment are shown in Fig. 4. For listeners M and W there was a clear ordering. Performance was best on HP−, worst on HP+, and intermediate for HPQ. For listener X, the critical data were ordered in the same way with one exception (200 Hz). Because the data from the three listeners were similar, the average shown in the bottom panel (d) of Fig. 4 exhibits a strong ordering suggesting low-frequency pitch strength in the order: HP−>HPQ>HP+. A statistical test parallel to the test used for experiment 1, in the range 38% (33+5) to 95%, led to 18 comparisons. All differences had the same sign, and

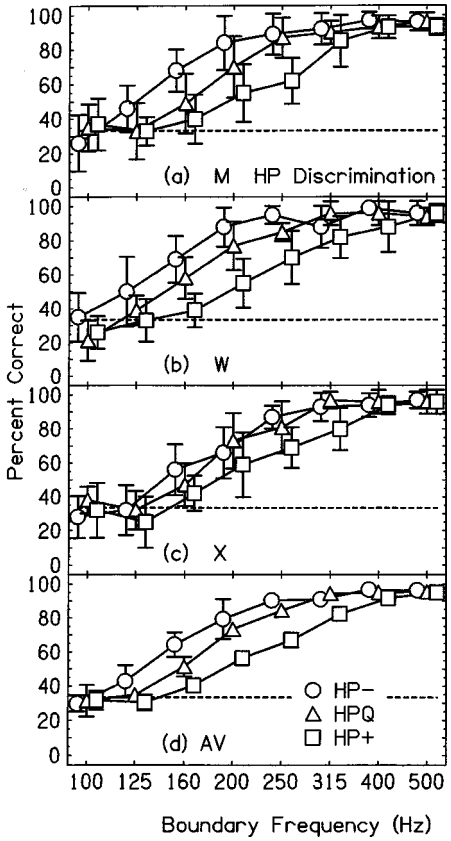


FIG. 4. Results of experiment 2—Huggins pitch discrimination for three listeners (a,b,c) followed by the average (d) as a function of the boundary frequency.

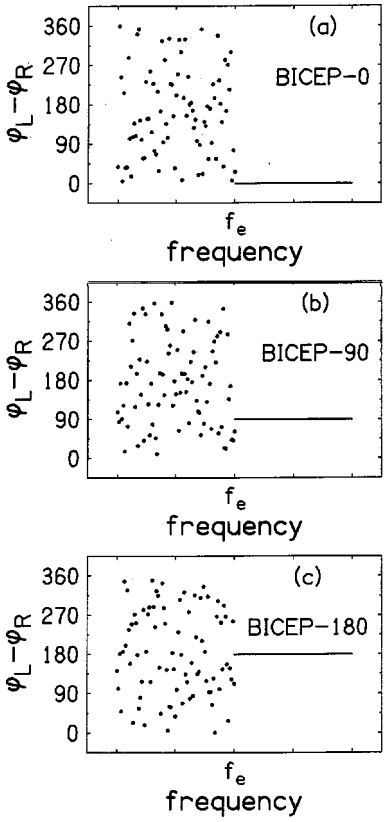


FIG. 5. Parts (a)–(c) show the interaural phases for the binaural coherence edge pitch (BICEP). Above the edge frequency, f_e, the interaural phase is fixed at 0°, 90°, or 180°. Below the edge frequency the interaural phase is completely random because the noises in the two ears are independent.

12 of them showed a significant advantage for HP− over HP+ at the 0.05 level. The ordering and statistical advantage are the same as found in experiment 1 and serve as further evidence in favor of the EC model.

V. EXPERIMENT 3—BICEP DETECTION

The binaural coherence edge pitch (BICEP) is created by a discontinuity in coherence where the binaural cross-correlation function is equal to zero (incoherent) on one side of the edge frequency and finite (coherent) on the other. The edge (or boundary) can take on several different forms. The noise components above the edge frequency can be incoherent while the components below the edge are coherent. Alternatively, the noise above the edge can be coherent while the noise below is incoherent. Hartmann and McMillon (2001) found that for low edge frequencies the BICEP was much stronger when the noise was coherent above the edge. (The interpretation of this result was that the critical element in BICEP perception is the ability to cancel the coherent components to create an edge in the central spectrum. When the noise is coherent only below a low-frequency edge there is not much to cancel and the BICEP is weak. To perceive an edge one must be aware of two different regions.) Therefore, experiment 3 used only the BICEP configuration with coherent noise above the edge, as shown in Fig. 5.

Analogous to HP−, HPQ, and HP+, the interaural phase of the coherent components was either 0°, 90°, or 180°, called BICEP-0, BICEP-90, and BICEP-180, respectively. The EC model predicts that for low edge frequencies, BICEP-0 should be the easiest to hear and BICEP-180 should be the hardest. This prediction is based on the idea that the coherent components of BICEP-0 can immediately be cancelled by a subtraction operation whereas the coherent components of BICEP-180 require a phase shift (presumably mediated by internal delay lines) before cancellation. Again, in parallel with the predictions for HP, the CAP model predicts the reverse order of difficulty, and Green's variation of the EC model predicts that BICEP-90 should be the hardest.

The BICEP matching experiments described in Fig. 2 of the article by Hartmann and McMillon (2001) suggested that it was easiest to hear BICEP-180 (there called ''pi-phase BICEP'') because two of the five listeners were much more successful in matching pitches to that phase condition, at least for edge frequencies between 550 and 750 Hz. Brief experiments at low edge frequencies, however, did not lead to the same conclusion. Low edge frequency experiments were pursued further in the present work.

A. Method

The experiments reported here focused on the critical low-frequency region and use BICEP detection rather than the BICEP matching procedure of Hartmann and McMillon. The protocols and listeners were the same as in experiment

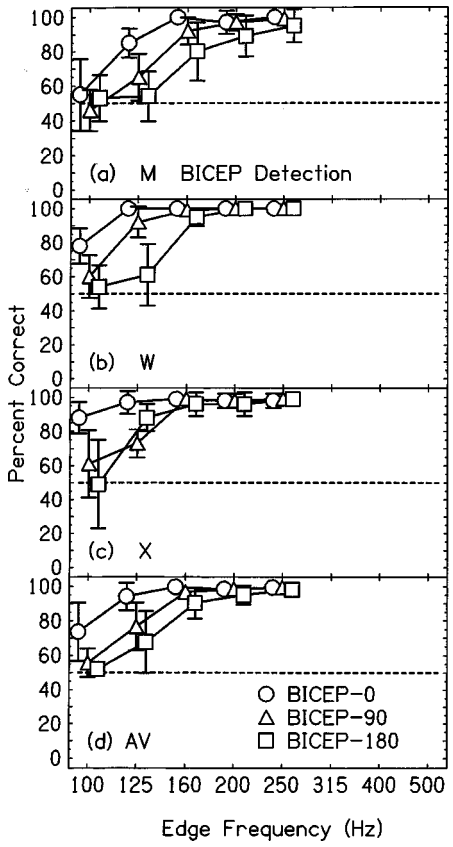


FIG. 6. Results of experiment 3—BICEP detection for three listeners (a,b,c) followed by the average (d) as a function of the boundary frequency.

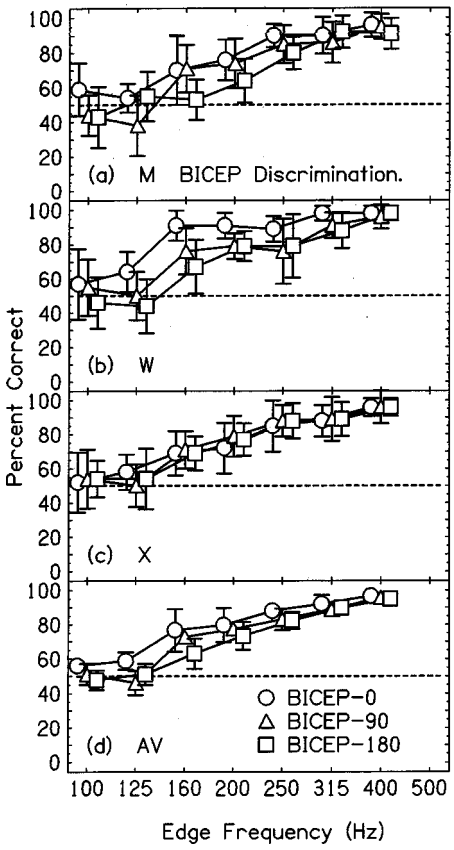


FIG. 7. Results of experiment 4—BICEP discrimination for three listeners (a,b,c) followed by the average (d) as a function of the boundary frequency.

1. The stimuli were also the same except that BICEP noises replaced HP noises. Edge frequencies were distributed randomly over a range of ±5% of the nominal edge frequency. Because preliminary experiments showed that all the listeners could easily detect the BICEP for edge frequencies above 250 Hz, the range of nominal edge frequencies was reduced to the five frequencies less than or equal to 250 Hz. Therefore, a run included only 50 trials.

B. Results

The results of the BICEP detection experiment are shown in Fig. 6. The plots there show nearly perfect performance for an edge frequency of 250 Hz and somewhat better than chance performance at 100 Hz. Below 250 Hz, performance was best for BICEP-0, worst for BICEP-180, and intermediate for BICEP-90. There was only one important exception to this rule, listener X at 125 Hz. The ordering of performance, with this one exception, supports the EC model for BICEP detection. A statistical test at the 0.05 level between BICEP-0 and BICEP-180, parallel to the test for experiment 1, led to seven paired comparisons with BICEP-0 having a significant advantage on all seven, in further support of the EC model.

VI. EXPERIMENT 4—BICEP DISCRIMINATION

Experiment 4 on BICEP discrimination resembled experiment 2 on HP discrimination in the same way that experiment 3 on BICEP detection resembled experiment 1 on HP detection.

A. Method

The method and listeners for experiment 4 were the same as for experiment 2 except that the BICEP stimulus replaced the HP stimulus. Because the BICEP was harder to hear than the HP [as noted also by Akeroyd *et al.* (2001)] only a two-alternative task was used. The edge frequency on the second interval was either 9% higher or 9% lower than on the first.

B. Results

The results of the BICEP discrimination experiment are shown in Fig. 7. The figure shows that there was not much to be learned from the responses at 400, 100, or 125 Hz because of ceiling or floor effects. For the four frequencies in between, the data for M and W suggest that BICEP-0 and BICEP-90 are stronger than BICEP-180. The data for X showed similar pitch strengths for all three phase conditions. Averaging the data for the three listeners led to a plot that tends to agree with the ordering of strength for experiments 1–3. This ordering favors the EC model. The 0.05-level statistical test parallel to experiment 1 showed an advantage for BICEP-0 over BICEP-180 for 8 out of 12 comparisons for listeners M and W and for 0 out of 5 for listener X, overall 8 out of 17. No test was possible for the reverse hypothesis, which failed the sign test on 13 out of 17 pairs.

VII. DISCUSSION

A. Summary of the experiments

Four experiments on the detection and discrimination of Huggins pitch (HP) and binaural coherence edge pitch (BICEP) were performed to test models of binaural pitch formation. The experiments focused on the region of low boundary frequencies. The results of the four experiments all pointed to the same conclusion: the data agree with the predictions of the equalization cancellation (EC) model and do not agree with the predictions of the central activity pattern (CAP) model. The data also do not agree with the predictions of Green's variation on the EC model. These results might be interpreted as peculiar to the low-frequency region, but the low-frequency data in Figs. 3, 4, 6, and 7 fit smoothly onto the intermediate-frequency data, and there would seem to be no reason to postulate that a different mechanism of binaural pitch formation somehow begins to apply when the frequency becomes lower than about 300 Hz. Instead, the results of these experiments, and also the lateralization experiments of Zhang and Hartmann (2002), suggest a single mechanism that becomes progressively less efficient as the frequency decreases. Therefore, the experiments suggest that the EC model correctly describes binaural pitch formation and the other models do not. It is worth investigating the assumptions that led to this conclusion.

B. Assumptions

The central assumption is that the performance differences observed for different binaural stimuli can be attributed to limitations in internal delay lines as the delays become long. The logic of our experiments is based on the idea that different binaural models use those delay lines in different ways and thus make different predictions for the experimental stimuli. Therefore, it is important to the validity of the conclusions that the observed low-frequency performance differences be caused by the delay lines and not by some other aspect of binaural hearing or pitch formation.

In fact, the data as a whole cannot be understood by assuming that the only cause of poor low-frequency performance is the failure of long delay lines because for each model there are some stimuli that do not require delay lines at all, and yet performance is poor for these stimuli at low frequencies, specifically HP− and BICEP-0 in the EC model and HP+ and BICEP-180 in the CAP model. To understand the low-frequency behavior seen for these cases requires a binaural model with a failure characteristic that depends on frequency. It is not uncommon for binaural models to require delay line weighting on both delay and frequency such that the $p(\tau)$ function becomes $p(\tau,f)$, falling off at both high and low frequencies (Raatgever, 1980; Stern *et al.*, 1988; Shackleton *et al.*, 1992). There is experimental evidence for added internal noise at low frequencies from MLD experiments, particularly the strong level dependence of the MLD (Wilbanks and Whitmore, 1967), which decreases binaural coherence and could disrupt the formation of dichotic pitch at low boundary frequencies.

Apart from the general decrease in performance with decreasing boundary frequency, there remain large observed differences between the different phase conditions, and current binaural models would attribute these differences to the use of internal delays. The observed differences seem consistent with that interpretation in that the differences between HP+ and HP− as well as the differences between BICEP-0 and BICEP-180 appear for boundary (or edge) frequencies between 200 and 300 Hz, corresponding to delay lines between 2.5 and 1.7 ms long. Experiments on other binaural phenomena, e.g., the MLD, have been interpreted assuming delay lines that start to fail for comparable delays (van der Heijden and Trahiotis, 1999).

C. The spatial masking interpretation

An alternative account of the experiments reported here appeals to a form of masking or interference and is primarily based on subjective impressions. It can be illustrated for the Huggins pitches HP+ and HP− in the context of the CAP model. According to the CAP model, HP+ is detected without delay lines, and therefore this phase condition should be the most easily detectable at low boundary frequency. This prediction is contradicted by experiments 1 and 2 which show that it is HP− that is the most detectable phase condition. Bilsen (2000) has observed that in comparison with HP− the pitch image in HP+ sounds as though it is masked more, meaning that the listener has the impression that there is greater interference from the background noise. Our listeners agree with that observation.

An interpretation that is consistent with that observation is as follows. Stimulus HP+ includes a background noise that is $N\pi$, a noise that is lateralized to both left and right sides of the head and tends to spread towards the center. Thus it encroaches on the spatial region occupied by the boundary region, which is S0. The background noise is broadband including a great deal of power below the boundary frequency. The encroachment then interferes with the detection of the pitch at the boundary frequency. By contrast, stimulus HP− includes a background noise that is $N0$, which forms a compact image near the center of the head. The background noise image is so compact that it does not interfere with the boundary region, which is $S\pi$ and appears far to the left or right. As a consequence HP− may be more detectable than HP+, consistent with experiment.

Despite its visceral appeal, there are a number of problems with this spatial interpretation: (1) There are no models of dichotic pitch based on contrasts along the spatial dimension; all models of dichotic pitch are based on contrasts along the tonotopic dimension. (2) A model of dichotic pitch based on spatial contrasts would be contrary to evidence indicating the dominance of tonotopic organization over spatial organization (Culling and Summerfield, 1995). (3) It is not enough for spatially based masking merely to exist. In order to preserve the CAP model, the spatial advantage of the more compact background-noise image for HP−, must outweigh the disadvantage of longer delay lines in the formation of the pitch image itself. (4) The subjective impression of greater interfering noise for the HP+ phase condition is not inconsistent with the EC model. Cancelling a noise stimulus having long interaural delays is expected to lead to a large residual noise because of jitter in the cancellation process.

Further, it is impossible for the EC process to cancel all the noise in a tuned channel using a single value of internal delay for HP+, where the background noise has π interaural phase, but it is trivially easy in the case of HP−, where the background noise phase is zero. (5) One would expect the spatial masking effect to make HP− more detectable than HP+ for all boundary frequencies, not just for low frequencies, contrary to our data.

D. Cross-correlation model

The cross-correlation model advanced by Colburn (1977) to account for the results of diverse experiments including those on the MLD makes predictions that are hard to distinguish from those of the EC model. In fact, Domnitz and Colburn (1976) have shown that in the limit of small signal-to-noise ratio, a normalized cross-correlation model is equivalent to an EC model in which the interaural phase difference of the noise is compensated by the equalization process. Like the CAP model, Colburn's model uses a cross-correlation function, but unlike the CAP model it attributes detection to dips, and not peaks, in the central excitation pattern. The model operation can be illustrated with HP−, for which the background noise is $N0$. The cross-correlation function, $\gamma(\tau)$, is maximum $[\gamma(0)=1]$ at zero interaural lag except in the boundary region where there is a dip in the maximum. This dip is very important perceptually because the binaural system is most sensitive to deviations from perfect ($\gamma=1$) cross-correlation (Gabriel and Colburn, 1981). Because the dip for HP− occurs for no internal delay, this model agrees with the EC model in predicting that low-frequency dichotic pitch should be strongest for HP−.

The cross-correlation model is naturally applied to an MLD experiment where a signal with an interaural phase that is different from the masker causes a just-detectable deviation from $\gamma=1$. The model immediately predicts that the greatest MLD should occur for the $N0S\pi$ condition, and with the introduction of the delay weighting, $p(\tau)$, the model predicts the smaller MLD for the $N\pi S0$ condition. In contrast with MLD experiments, the decorrelations presented by dichotic pitch stimuli are larger. The application of the cross-correlation model to dichotic pitches would require that the same principles continue to apply. It is a matter of parsimony to assume that they do.

E. The exponential CAP model

The Appendix introduces a modification to the CAP model including the assumption that the left and right inputs to the central processor are exponentially rectified versions of the input signals. This modification has the effect of passing the mathematical predictions of the CAP model through a nonlinear transformation, essentially the modified Bessel function I_0. As shown in the Appendix, the predicted CAP for tonotopic place ω and interaural delay τ becomes

$$\gamma(\omega,\tau)=I_0\{\{g_L^2+g_R^2+2g_Lg_R \times\cos[\phi_R(\omega)-\phi_L(\omega)-\omega\tau]\}^{0.5}\}.$$

The synchrony coefficients g_L and g_R, for left- and right-ear pathways respectively, can be expected to depend on the interaural delay line.

The argument of the Bessel function, itself a function of tonotopic coordinate ω and ITD τ, is essentially the cross-correlation function from the CAP model. The CAP model predicts dichotic pitch images at peaks in the ω-τ plane. The sharper the peaks, the stronger the dichotic pitch ought to be. The purpose of the calculations described in this section was to discover whether this simple and plausible nonlinear transformation on the model could result in predictions in agreement with our experiments by sharpening the peaks.

In the context of the CAP model, g_L and g_R should be approximately the same in the perception of HP+, which is located at $\tau=0$. But they could be quite different in the perception of HP−, which is located off to the left or right because a longer delay line to one side might well introduce increased jitter which would decrease the synchrony coefficient on that side. *A priori*, it seemed possible that the nonlinear transformation could lead to a peak that was actually sharper for g_L not equal to g_R and hence predict better performance for HP− than for HP+.

Colburn (1973) and Stern and Colburn (1978) took values of g to be less than or equal to $\sqrt{20}$. Therefore, we performed calculations for many pairs of g_L and g_R less than 5. Plots of the calculated cross-correlation functions showed that by any reasonable measure of sharpness, peaks for $g_L \neq g_R$ are never sharper than the peaks for $g_L=g_R$. Therefore, even after the exponential modification and asymmetry, the CAP model continues to predict that HP+ should be stronger than HP−, contrary to experiment.

VIII. CONCLUSIONS

Four experiments on the detection and discrimination of Huggins pitch (HP) and binaural coherence edge pitch (BICEP) were performed to test models of binaural pitch formation. By studying these dichotic pitches at low frequencies, the experiments were able to distinguish among alternative models of pitch formation. The results favored the equalization-cancellation (EC) model or close relatives such as Colburn's cross-correlational model (1977) or the modified EC model of Culling *et al.* (1998a), elaborated by Akeroyd and Summerfield (2000).

The results did not support Green's variation on the EC model. Whereas Green's variation predicts that HPQ and BICEP-90 should be the most difficult of the HP and BICEP stimuli, the observed order of difficulty put these 90° conditions in between the 0° and 180° conditions. This conclusion casts some doubt on the validity of Green's variation as it has been applied to binaural edge pitch.

The results of the four experiments disagreed with the predictions of the central activity pattern (CAP) model. From an esthetic point of view, this conclusion is disappointing because the CAP model is the only model that integrates both the pitch and the lateralization of the pitch image in an economical way. Further, the predictions of the CAP model for the lateralization of HP+ and HP− agree qualitatively with observation. In the end, however, the four experiments

on pitch strength presented in this article all lead to the same conclusion favoring the EC model.

ACKNOWLEDGMENTS

This work was supported by the NIDCD under Grant No. DC 00181. We are grateful to Dr. L. R. Bernstein, Dr. F. A. Bilsen, Dr. H. S. Colburn, Dr. J. F. Culling, and Dr. J. Raatgever for useful discussions.

APPENDIX: CROSS-CORRELATION AND THE CAP MODEL

The CAP model has been remarkably successful in accounting for a variety of binaural effects. The model is based on the concept of a central spectrum, derived by Raatgever and Bilsen (1986) by assuming that the signals to the left and right ears are simply added. The central spectrum is then the power spectrum of that sum for a given delay between the summed waveforms. With unit-amplitude noise signals and infinitely sharp auditory filters, Bilsen and Raatgever found a central spectrum at best frequency ω and internal delay τ given by

$$P(\omega,\tau)=1+\cos[\phi_R(\omega)-\phi_L(\omega)-\omega\tau], \tag{A1}$$

where the difference between signal phases $\phi_R(\omega)$ and $\phi_L(\omega)$ is the interaural phase for frequency ω.

The purpose of this appendix is to show how an equivalent result can be obtained by making the more usual assumption that the central representation is the result of a cross-correlation operation. Cross-correlation, like an AND gate in electronics, depends upon a product and not a sum.

1. Weak synchrony limit

In the limit of weak synchrony the ensemble-average ($\langle\langle\cdots\rangle\rangle$) firing rates in left and right channels can be expanded as a unit spontaneous rate plus a modulation term that is linear in the signal,

$$\langle r_L(\omega,t)\rangle=1+c\cos[\omega t+\phi_L(\omega)], \tag{A2}$$

for the left-ear channel and similarly for the right. This simple form maintains the assumption that the auditory filter with best frequency ω is infinitely sharp. Modulation fraction c must be less than 1 to keep the firing rate positive.

The ensemble average cross-correlation function is

$$\langle\gamma(\omega,\tau)\rangle=\frac{1}{T_D}\int_0^{T_D}dt\langle r_L(\omega,t+\tau)r_R(\omega,t)\rangle, \tag{A3}$$

where T_D is the internal integration time of the neural cross-correlator. With the assumption that ωT_D is considerably larger than one, and that ensemble averages in left and right ears are independent, the cross-correlation function becomes

$$\langle\gamma(\omega,\tau)\rangle=1+\frac{c^2}{2}\cos[\phi_R(\omega)-\phi_L(\omega)-\omega\tau]. \tag{A4}$$

Equation (A4) for the cross-correlation function is the same as Eq. (A1) for the central spectrum except that the peaks and valleys in the ω-τ plane are not as large in the cross-correlation function. They are reduced by a factor of $c^2/2$. However, nothing in the application of the CAP model depends critically on the absolute sizes of peaks and valleys.

2. Strong synchrony—general

If the synchrony is strong, it is helpful to represent the neural firing rate as an exponential rectification of the input signal $x(t)$ (e.g., Stern and Colburn, 1978),

$$r_L(t)=\exp[g_L x_L(t)], \tag{A5}$$

where g_L is a measure of synchrony for the left ear input to a binaural cross correlator, and similarly for the right ear input. Again assuming infinitely sharp auditory filtering,

$$\langle\gamma(\omega,\tau)\rangle=\frac{1}{T_D}\int_0^{T_D}dt\exp\{g_L\cos[\omega(t+\tau)+\phi_L(\omega)]+g_R\cos[\omega t+\phi_R(\omega)]\}. \tag{A6}$$

The integral is a modified Bessel function I_0 for which a series formula is given by Abramowitz and Stegun (1964) (p. 375):

$$\langle\gamma(\omega,\tau)\rangle=I_0(\{g_L^2+g_R^2+2g_Lg_R\times\cos[\phi_R(\omega)-\phi_L(\omega)-\omega\tau]\}^{0.5}). \tag{A7}$$

The argument of the square root is essentially the frequency and delay dependence of the CAP model. Unlike its more famous cousin J_0, which oscillates, the modified Bessel function I_0 is monotonic and smooth. For small argument x, $I_0(x)=1+x^2/4$. Thus, like the low-synchrony limit, the exponential rectifier provides a way to recover the CAP model, or a monotonic transformation of it, from the mathematical operation of cross-correlation. The cross-correlation operation represents the operation of the binaural system better than channel addition.

[1]Durlach's original articles developing the EC model paid little attention to the form of the final displayer. The idea of the central spectrum as a display seems to have originated in a report by Bilsen (1972).

[2]Were it not for the central weighting function a dichotic pitch predicted to be centered could instead be lateralized where the ITD is $\pm 1/f_b$.

[3]The boundary frequency was roved because the nominal boundary frequencies form simple ratios, and these have implications for pitch. By roving the boundary frequencies over the range of a semitone, the experimenters discouraged the listeners from making unwanted musical associations across different trials, especially in the pitch-sensitive discrimination experiments 2 and 4.

Abramowitz, M., and Stegun, I. (**1964**). *Handbook of Mathematical Functions*, National Bureau of Standards Applied Math Series no. 55 (USGPO, WAshington, DC).

Akeroyd, M. A., and Summerfield, A. Q. (**2000**). "The lateralization of simple dichotic pitches," J. Acoust. Soc. Am. **108**, 316–334.

Akeroyd, M. A., Moore, B. C. J., and Moore, G. A. (**2001**). "Melody recognition using three-types of dichotic pitch stimulus," J. Acoust. Soc. Am. **110**, 1498–1504.

Bilsen, F. A. (**1972**). "Pitch of dichotically delayed noise," in *Hearing Theory—1972*, edited by E. de Boer, B. L. Cardozo, and R. Plomp (IPO, Eindhoven, Holland), pp. 5–8.

Bilsen, F. A. (**1976**). "Pronounced binaural pitch phenomenon," J. Acoust. Soc. Am. **59**, 467–468.

Bilsen, F. A. (**2000**). Personal communication.

Bilsen, F. A., and Goldstein, J. L. (**1974**). "Pitch of dichotically delayed noise and its possible spectral basis," J. Acoust. Soc. Am. **55**, 292–296.

Bilsen, F. A., and Raatgever, J. (**2000**). "On the dichotic pitch of simultaneously presented interaurally delayed white noises. Implications for binaural theory," J. Acoust. Soc. Am. **108**, 272–284.

Bilsen, F. A., and Raatgever, J. (**2002**). *Demonstrations of Dichotic Pitch*, compact disc, available from the authors at Perceptual Acoustics Laboratory, Applied Physics Department, Delft University of Technology, Delft, The Netherlands.

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (**2002**). ''Precise inhibition is essential for microsecond interaural time difference coding,'' Nature (London) **417**, 543–547.

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.

Cramer, E. M., and Huggins, W. H. (**1958**). ''Creation of pitch through binaural interaction,'' J. Acoust. Soc. Am. **30**, 413–417.

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.

Culling, J. F., Summerfield, A. Q., and Marshall, D. H. (**1998a**). ''Dichotic pitches as illusions of binaural unmasking I. Huggins pitch and the binaural edge pitch,'' J. Acoust. Soc. Am. **103**, 3509–3526.

Culling, J. F., Marshall, D. H., and Summerfield, A. Q. (**1998b**). ''Dichotic pitches as illusions of binaural unmasking II. The Fourcin pitch and the dichotic repetition pitch,'' J. Acoust. Soc. Am. **103**, 3527–3540.

Domnitz, R. H., and Colburn, H. S. (**1976**). ''Analysis of binaural detection models for dependence on interaural target parameters,'' J. Acoust. Soc. Am. **59**, 598–601.

Durlach, N. I. (**1960**). ''Note on the equalization and cancellation theory of binaural masking level differences,'' J. Acoust. Soc. Am. **32**, 1075–1076.

Durlach, N. I. (**1962**). ''Note on the creation of pitch through binaural interaction,'' J. Acoust. Soc. Am. **34**, 1096–1099.

Durlach, N. I. (**1972**). ''Binaural signal detection—equalization and cancellation theory,'' in *Foundations of Modern Auditory Theory, Volume 2*, edited by J. Tobias (Academic, New York), pp. 369–462.

Frijns, J. H. M., Raatgever, J., and Bilsen, F. A. (**1986**). ''A central spectrum theory of binaural processing. The binaural edge pitch revisited,'' J. Acoust. Soc. Am. **80**, 442–451.

Fourcin, A. J. (**1962**). ''An aspect of the perception of pitch,'' Proc. 4th Int. Congress on Phonetic Sciences, Helsinki (Mouton, The Hague), pp. 355–399.

Fourcin, A. J. (**1970**). ''Central pitch and auditory localization,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sitjhoff, Leiden), pp. 319–328.

Gabriel, K. J., and Colburn, H. S. (**1981**). ''Interaural correlation discrimination I. Bandwidth and level dependence,'' J. Acoust. Soc. Am. **69**, 1394–1401.

Green, D. M. (**1966**). ''Signal-detection analysis of equalization and cancellation model,'' J. Acoust. Soc. Am. **40**, 833–838.

Grothe, B. (**2000**). ''The evolution of temporal processing in the medial superior olive, an auditory brainstem structure,'' Prog. Neurobiol. **61**, 581–610.

Guttman, N. (**1962**). ''Pitch and loudness of a binaural subjective tone,'' J. Acoust. Soc. Am. **34**, 1996(A); Bell Telephone Laboratories memorandum MH-1232-NG-KN, 30 November, 1962.

Hartmann, W. M. (**1993**). ''On the origin of the enlarged melodic octave,'' J. Acoust. Soc. Am. **93**, 3400–3409.

Hartmann, W. M., and McMillon, C. D. (**2001**). ''Binaural coherence edge pitch,'' J. Acoust. Soc. Am. **109**, 294–305.

Jeffress, L. A. (**1948**). ''A place theory of sound localization,'' J. Comp. Physiol. Psychol. **41**, 35–39.

Klein, M. A., and Hartmann, W. M. (**1980**). ''Binaural edge pitch,'' J. Acoust. Soc. Am. **70**, 51–61.

Raatgever, J. (**1980**). ''On the binaural processing of stimuli with different interaural phase relations,'' thesis, Delft (unpublished).

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.

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.

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 *Handbook of Perception and Cognition—Hearing*, edited by B. Moore (Academic, San Diego), pp. 347–386.

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.

van der Heijden, M., and Trahiotis, C. (**1999**). ''Masking with interaurally delayed stimuli: The use of internal delays in binaural detection,'' J. Acoust. Soc. Am. **105**, 388–399.

Wilbanks, W. A., and Whitmore, J. K. (**1967**). ''Detection of monaural signals as a function of interaural noise correlation and signal frequency,'' J. Acoust. Soc. Am. **43**, 785–797.

Zhang, P. X., and Hartmann, W. M. (**2002**). ''On the lateralization of the Huggins pitch,'' J. Acoust. Soc. Am. **112**, 2271(A).

High-rate conditioning pulse trains in cochlear implants: Dynamic range measures with sinusoidal stimuli[a)]

Robert S. Hong and Jay T. Rubinstein[b)]
Department of Otolaryngology-Head and Neck Surgery, University of Iowa Hospitals and Clinics, 200 Hawkins Drive, 21165 PFP, Iowa City, Iowa 52242-1093

(Received 10 December 2002; revised 25 June 2003; accepted 15 September 2003)

The addition of a continuous, unmodulated, high-rate pulse train to the electrical signals of cochlear implant recipients results in statistically significant increases in psychophysical dynamic range (41 out of 46 electrode pairs tested). The observed increases in dynamic range are thought to result from nerve conditioning by appropriate levels of high-rate pulse train. Five dynamic range profiles are characterized, defining the different responses of dynamic range observed with increasing levels of the conditioner. Four of the five profiles demonstrate increases in dynamic range, with three showing behavior consistent with stochastic resonance. One profile depicts evidence of adaptation in response to higher levels of the conditioner, with a recovery period lasting throughout the duration (on the scale of tens of minutes) of experimentation. Dynamic range profiles are shown to be similar across sinusoidal frequencies (202, 515, and 1031 Hz) but potentially different across electrode pairs (electrodes 1-2, 7-8, and 15-16). Correlation analysis does not reveal any predictors of optimal conditioner level or amount of dynamic range increase with the conditioner. [DOI: 10.1121/1.1623785]

PACS numbers: 43.66.Ts, 43.66.Cb, 43.66.Ba [MRL] Pages: 3327–3342

I. INTRODUCTION

Stochastic resonance is a phenomenon whereby the addition of an optimal amount of random noise to a nonlinear system results in an enhancement of input signal detection. This has been demonstrated in a wide variety of biological and physical systems (Wiesenfeld and Moss, 1995), including the neural encoding of information across a number of species (Douglass *et al.*, 1993; Levin and Miller, 1996). The functional role of noise in neural pathways suggests that the incorporation of a fixed level of noise (Collins *et al.*, 1995) into prosthetic systems that provide sensory information to humans may result in an improvement in prosthetic performance. One such prosthetic system is the cochlear implant. The work presented in this paper seeks to take advantage of the potential signal-encoding benefits of adding noise to an electrically stimulated system through the application of a high-rate conditioning pulse train in cochlear implants.

The potential benefits of additive noise to the auditory system have been examined through computational models, animal studies, and human psychophysics. Simulations of single auditory nerve fibers stimulated with a cochlear implant (Morse and Evans, 1999a,b; Morse and Roper, 2000) as well as physiological measurements in the toad sciatic nerve (Morse and Evans, 1996) suggested that noise can enhance the temporal coding of vowels. Addition of Gaussian noise to pulse train stimuli presented to deafened guinea pig auditory nerves was found to result in a more ''normal'' nerve response to electrical stimuli (Matsuoka *et al.*, 2000). Manipulations of interpulse interval of electrical pulse trains, thought to be correlated with manipulations of intrinsic membrane noise, in both a stochastic model of the node of Ranvier and deafened cat auditory nerves demonstrated significant enhancement of dynamic range within an optimal range of interpulse intervals (Matsuoka *et al.*, 2001). The addition of appropriate amounts of noise to the electrical signals presented to human cochlear implant patients generated psychophysical improvements in both signal detection and discrimination (Zeng *et al.*, 2000; Chatterjee and Robert, 2001). Improvements in the delivery of fine-time structure information through the cochlear implant with addition of optimal noise may also result in improved sensitivity to interaural time differences for bilateral implant recipients and pitch perception (Smith *et al.*, 2002).

It has previously been observed that the normal mammalian auditory nerve is spontaneously active in quiet (Kiang and Moxon, 1972); in contrast, the deafened auditory nerve in general lacks this spontaneous activity (Liberman, 1978; Liberman and Dodds, 1984), although exceptions do exist (Shepherd and Javel, 1997). Rubinstein *et al.* (1999) showed through computational auditory nerve models that the application of a continuous, unmodulated high-rate pulse train to a population of deafened fibers could restore spontaneouslike activity to those fibers. At rates of ~5 kpps (kilo pulses per second) where this was demonstrated, different nerve fibers were theorized to be driven into different refractory states and thus sensitive to stimulation at different random moments in time. Subsequently, when a sinusoidal signal was presented to the nerve population with an appropriate level of 5-kpps conditioner, the signal was more faithfully encoded compared to without conditioner due to the desynchronization of the responsiveness of different nerve fibers (Rubinstein and Hong, 2003). This enhancement

[a)]Portions of this work were presented in ''Stochastic resonance in cochlear implant patients,'' Midwinter Research Meeting of the Association for Research in Otolaryngology, St. Petersburg, Florida, January 2002.

[b)]Author to whom correspondence should be addressed. Electronic mail: jay-rubinstein@uiowa.edu

was observed through increases in both the dynamic range (measured with rate-level functions) and temporal resolution of the neural representation of the sinusoidal stimulus (Rubinstein *et al.*, 1998). In effect, the psuedospontaneous activity created by the desynchronizing pulse train served as a source of noise for improved signal encoding through stochastic resonance.

The impact of a high-rate conditioning pulse train on signal encoding has also been explored in studies of auditory nerves in animals. Litvak *et al.* (2001) showed that such a pulse train could produce responses in single nerve fibers of cats that resemble spontaneous activity in acoustically sensitive auditory nerves. The presentation of sinusoidal and vowel stimuli along with the conditioning pulse train (through modulation of the conditioner) resulted in improved dynamic range and temporal resolution (Litvak, 2002). Dynamic range was measured with a sinusoid as the range of modulation depths for which nerve firing responses resemble those of healthy ears. Runge-Samuelson (2002) measured the electrical compound action potential (ECAP) in cats and guinea pigs and found a decrease in the slope throughout most of the ECAP growth function for sinusoids presented with the conditioner, a result consistent with increases in dynamic range due to the conditioner.

This paper examines the potential advantages of adding noise to the electrically stimulated auditory system of human cochlear implant subjects via the addition of a continuous, unmodulated, high-rate (5 kpps) pulse train. Specifically, this study focuses on the impact of the conditioner on the psychophysical dynamic range of sinusoidal stimuli. Dynamic range is chosen because of its relevance to both speech and music perception, as increases in electrical dynamic range could lead to improvements in speech discrimination (Fu and Shannon, 2000; Loizou *et al.*, 2000; Zeng and Galvin, 1999) and sound quality (Neuman *et al.*, 1998; Boike and Souza, 2000). Previously, it has been reported that addition of the most favorable level of the conditioner to sinusoidal stimuli resulted in significant increases in dynamic range across cochlear implant subjects, with a mean largest increase of 6.7 dB per subject (Hong *et al.*, 2003). This paper expands on the previous work by characterizing in greater depth the different dynamic range profiles observed with the addition of the conditioner and by systematically examining the effect of the conditioner on dynamic range across sinusoidal frequencies, electrode pairs, and time.

II. METHODS

A. Subjects

Thirty-two post-lingually deafened cochlear implant recipients, ranging in age from 25 to 89 years old, participated in this study (Table I). All subjects were implanted with the Clarion CII device (Advanced Bionics, Sylmar, CA) between January 2001 and February 2002 at the University of Iowa Hospitals and Clinics. All electrodes were fully inserted into the scala tympani. Data were collected between connection and 13 months post-hookup. Experiments were conducted with the prior approval of the Institutional Review Board at the University of Iowa.

B. Stimuli

All stimuli were delivered to the implanted electrode array using the Clarion research interface for second generation Clarion products (CRI-2). The software for the CRI-2 was custom-designed using MATLAB and Texas Instruments Code Composer. Electrical stimuli were presented to electrode pairs 1-2, 7-8, or 15-16 in bipolar configuration. The stimuli had the following characteristics:

(1) *Conditioner.* The conditioner was defined as the continuous 5-kpps (kilo pulses per second) unmodulated pulse train presented to all subjects at amplitudes ranging from 0 to 900 μA. The conditioner was biphasic, phase duration of 50 μs, with cathodal phase first to the lower numbered electrode of each pair. For each trial, the conditioner was delivered at constant amplitude in a continuous manner.
(2) *Sinusoidal bursts*. A sinusoid served as the stimulus to be judged by the subject. The sinusoid was presented at a constant frequency (202, 515, or 1031 Hz) as 500-ms bursts with anodal phase first to the lower numbered electrode of each pair and with an interstimulus interval of 1 s. The lone exception was subject 4, who was tested with 200-ms bursts.

C. Procedures

Subjects were first presented with the conditioner at the desired amplitude and asked if they could hear the conditioner. Those who could were instructed to indicate when the percept disappeared, after which sinusoidal bursts were added to the electrical stimuli. Seven subjects, however, continued to hear the conditioner at a level between ''just perceptible'' and ''soft'' after 5 min. In these seven subjects, sinusoidal bursts were added after 5 min elapsed. For a random subset (four) of these seven, the sinusoid was turned off after exposure to an audible level of sinusoid for 5 to 10 s and subjects were asked again if they could perceive the conditioner. In all cases, the subjects could not detect the presence of the conditioner despite the fact that the conditioner was on continuously.

Threshold and most comfortable loudness were measured for the sinusoidal probe at different levels of the conditioner. This was performed using the following procedures, each manually controlled by the experimenter. Threshold was measured using an up–down adaptive procedure that converging to the 50% level (Levitt, 1971); this paradigm was used instead of a two-down-one-up paradigm because of the possibility that the conditioner may change the slope of the psychometric function. The number of reversals within each adaptive procedure varied from 4 to 16, with more reversals used for thresholds demonstrating greater variability. Most comfortable loudness (MCL) was measured with a procedure in which subjects heard a pulsing sequence of sinusoidal stimuli that steadily increased at an approximate rate of 1, 2, or 5 μA per second. Subjects were instructed to indicate when the sound became ''most comfortable.'' The dynamic range was defined as the difference between threshold and MCL.

TABLE I. Subject demographics and additional tasks performed. The age, etiology of deafness, and CNC word scores are shown for all subjects. CNC word scores are taken at 3 months post-hookup, unless otherwise indicated ([a] or[b]). HINT sentence test scores are shown for subjects who did not take CNC word tests, due to poor HINT performance ([c] or[d]). Subjects in whom data was collected to characterize dynamic range profiles across sinusoidal frequencies, across electrode pairs, or across multiple testing sessions are also indicated. Subject 9 had 2 different electrode pairs tested at the same sinusoidal frequency across time.

Subject	Age	Etiology of deafness	CNC words (% correct)	Additional tasks performed		
				Frequency	Electrode pairs	Time
1	52	Meniere's	34		X	
2	55	enlarged vestibular aqueducts	52			
3	75	noise-induced	50		X	X
4	66	auto-immune	42			
5	68	noise-induced	0[c]		X	X
6	27	hereditary	47			
7	73	unknown	14		X	X
8	46	unknown	56			X
9	68	Meniere's	44		X	XX
10	67	otosclerosis	58[a]			X
11	80	hereditary	62	X		
12	26	unknown	35[b]			
13	69	unknown	44			X
14	47	hereditary	46			
15	54	unknown	46	X		X
16	48	measles	34	X		X
17	62	unknown	4			
18	33	unknown	80			
19	74	noise-induced	0			
20	75	unknown	48			X
21	37	unknown	56			X
22	78	unknown	11[c]			
23	80	unknown	1[c]			
24	59	hereditary	37[b]			
25	79	otosclerosis	2[d]	X		X
26	64	radiation	48			
27	41	unknown	72[a]	X		X
28	68	unknown	0[d]			
29	68	unknown	56	X		
30	74	Meniere's	12			
31	75	cholesteatoma	59			
32	89	unknown	8[d]	X	X	X

[a]CNC at 6 months.
[b]CNC at 12 months.
[c]HINT at 3 months.
[d]HINT at 6 months.

1. General protocol for characterizing dynamic range profiles

The response of psychophysical threshold and MCL to different levels of the conditioner is referred to as the dynamic range profile. The first data point collected for each subject in characterizing such a profile was for the psychophysical threshold of the sinusoidal probe before addition of the conditioner. Only combinations of electrode pair, sinusoidal frequency, and burst duration that resulted in initial thresholds (without the conditioner) at least 10 μA above the noise floor were tested with the conditioner, to allow for the measurement of potential decreases in threshold with the conditioner of up to at least 10 μA. (The noise floor was 4 μA, the level at which sinusoidal stimuli could not be recognized visually on an oscilloscope with the conditioner off. It was also verified that such stimuli could not be visibly detected with a conditioner.) In 1 out of 32 patients (subject 9), no such combinations were found; therefore, additional testing was not performed on this patient. For the other 31 patients, testing proceeded to measurement of MCL without addition of the conditioner. A conditioner was then added at increasing 100-μA increments, with collection of one data point for threshold and MCL at each level. Upon reaching the highest conditioner level to be tested (i.e., completion of one run), the same procedure was repeated with at least three additional runs. The additional runs were performed, however, only at conditioner levels suggested by the initial run as being important to defining the contour of the dynamic range profile. Specifically, additional runs included threshold and MCL measurements with (1) the conditioner turned off; (2) conditioner levels that potentially demonstrated stochastic resonance; (3) the highest conditioner level tested; and (4) at least one conditioner level found in between any trend of increases or decreases in threshold spanning at least 300 μA (three consecutive 100 μA conditioner increments of the initial run). Furthermore, when threshold data suggested the presence of stochastic resonance, step size was narrowed if

time permitted, about the point of interest to more accurately define the conditioner level of maximal dynamic range enhancement.

2. Modified protocol for characterizing dynamic range profiles suggestive of adaptation

In some subjects, the values measured for threshold and MCL in the first run were lower than those measured in all subsequent runs, suggesting adaptation. To determine if statistically significant adaptation occurred, a modified protocol was used in a different testing session such that in the first run, multiple data points were collected at each conditioner level before increasing the conditioner level. Upon completion of this modified first run, having reached the maximum conditioner to be tested, one of two procedures was undertaken. For most subjects, additional runs were then completed with a single data point collected per run (as described in the general protocol). However, for one subject (subject 27), multiple data points were again collected at each conditioner level before increasing the conditioner, except between each new level, the patient was reexposed to a few minutes of the maximum conditioner tested in the initial run. The reexposure to maximum conditioner was necessary since other data (not shown) suggested that recovery from adaptation could occur in the time frame required to complete this alternative procedure undertaken by subject 27.

3. Modified protocol for characterizing dynamic range profiles across sinusoidal frequencies

Seven subjects were tested on one electrode pair with different sinusoidal frequencies (202, 515, and/or 1031 Hz). For each subject, all experiments were completed on the same day in a 3-h testing session. This was important to avoid confounding factors in the analysis of data across sinusoidal frequencies resulting from changes in threshold and MCL that were observed in some subjects over time (weeks to months) from one testing session to the next. Because subjects varied in their speed in performing the tasks required, the subjects who were tested with this modified protocol were limited to those who could complete the testing in the allotted time. The protocol for these experiments was modified from the general protocol such that multiple data points were collected at a particular conditioner level for all sinusoidal frequencies tested before increasing the conditioner level. This allowed for the observation of any effects across frequencies from exposure to higher conditioner levels. For subjects 15 and 16, the general protocol was used at each frequency tested; data from a previous visit for each subject suggested that shifts in threshold and MCL from one run to the next did not occur in the tested electrode pair. For some frequencies, time limitations allowed only one data point to be collected at each conditioner level; these data are indicated as such in Sec. III.

4. Modified protocol for characterizing dynamic range profiles across electrode pairs

Seven subjects were tested with one sinusoidal frequency on different electrode pairs (1-2, 7-8, and/or 15-16). For each subject, all experiments were completed on the same day in a 3-h testing session. This was important to avoid confounding factors in the analysis of data across electrode pairs resulting from changes in threshold and MCL that were observed in some subjects over time (weeks to months) from one testing session to the next. Because subjects varied in their speed in performing the tasks required, the subjects who were tested with this modified protocol were limited to those who could complete the testing in the allotted time. The experiments were carried out such that multiple data points were taken at a particular conditioner level for all electrode pairs tested before increasing the conditioner level. (Two exceptions were for subjects 1 and 9, for whom experiments were completed on one electrode pair before proceeding to the next electrode pair.) This allowed for the observation across electrode pairs of any effects from exposure to higher conditioner levels. For some electrode pairs, time limitations allowed only one data point to be collected at each conditioner level; these data are indicated as such in Sec. III.

5. Determination of highest conditioner level to be tested

The highest conditioner level tested per electrode pair was determined by a number of factors. For initial subjects ran, software limitations limited testing up to a conditioner level of $\sim$250 μA. This limitation was subsequently modified to enable testing up to a maximum of $\sim$900-μA conditioner. Second, subjects were asked to rate the initial loudness of the conditioner (prior to disappearance of the percept) for each new conditioner level tested on an eight point qualitative scale. When subjects perceived the initial percept as ''loud'' to ''upper limit of comfortable loudness,'' the highest two ratings of the scale, the conditioner level at which this occurred was designated the highest conditioner level to be tested for the electrode pair to avoid presentation of an ''uncomfortably loud'' stimulus. This was true for one electrode pair tested on each of six different subjects (subjects 5, 7, 19, 25, 30, and 32) and two electrode pairs for subject 28, the specifics of which are denoted in Sec. III. Third, the thresholds of some subjects decreased to the noise floor of measuring capabilities with addition of the conditioner. The conditioner in these cases was not increased further because any further decreases in threshold would not be measurable. This was true for one electrode pair at one sinusoidal frequency for each of two subjects (subjects 5 and 28), the specifics of which are denoted in Sec. III. Fourth, when data suggested the presence of stochastic resonance, the conditioner was not increased beyond that necessary to show statistically significant stochastic resonance. Fifth, subjects showed considerable variation in the speed at which they performed the required tasks. As a result, the amount of time available per testing session not only limited the number of data points collected as previously indicated, but also the highest level of conditioner that could be tested in the scheduled session.

III. RESULTS

A. Dynamic range profiles

This paper reports the results of 31 cochlear implant subjects, tested on 46 different electrode pairs (electrodes 1-2, 7-8, and/or 15-16) in bipolar configuration at sinusoidal frequencies of 202, 515, and/or 1031-Hz. For each electrode pair, the psychophysical threshold and most comfortable loudness of the sinusoidal stimuli in response to different amplitudes of a continuous 5-kpps unmodulated pulse train were measured. Analysis of the resulting functions reveals that the psychophysical responses to this 5-kpps conditioner can be categorized into five groups, which are referred to as dynamic range profiles. The first profile, designated DR-0, is characterized by the lack of a statistically significant (*t*-test) increase in dynamic range with addition of the conditioner across the different conditioner levels tested. The second profile, DR-1, demonstrates statistically significant increases in dynamic range without statistically significant evidence of stochastic resonance. The final three profiles—SR-1, SR-2, and SR-3—are distinguished by contours suggesting the presence of statistically significant stochastic resonance.

The first profile, DR-0, is characterized by the lack of a statistically significant increase in dynamic range with addition of the conditioner across the range of conditioner levels tested. This is shown in Fig. 1. Figure 1(a) shows the thresholds and MCLs of electrode pair 7-8 in subject 32, tested with a 515-Hz sinusoid. Conditioner levels at which multiple data points were collected are shown with error bars representing 95% confidence intervals. For this patient, the largest observed increase in dynamic range is 0.3 dB (at 300-μA conditioner) across the range of conditioner levels tested. In all observed cases of DR-0, a decrease in threshold (up to 2 dB) was observed, but as with subject 32, a companion decrease in MCL resulted in observation of little to no change in dynamic range. Figure 1(b) shows the behavior of dynamic range in response to the conditioner of the five electrode pairs tested that demonstrated DR-0. The solid lines are drawn through the mean values of dynamic range for which multiple data points were collected per conditioner level. In the one case shown with the dotted line, time limitations permitted the collection of only one data point per conditioner level.

The second profile, DR-1, represents a pattern of increase in dynamic range without evidence of stochastic resonance. Figure 2(a) depicts two variations of this profile. In the first variation, dynamic range increases up to the highest conditioner level tested. In the second variation, dynamic range increases up to a plateau, beyond which further increases in the conditioner result in no significant changes (in contrast to stochastic resonance, where further increases would result in an increase in threshold or decrease in dynamic range). In both variations, the increase in dynamic range with addition of conditioner is attributable primarily to decreases in threshold. Specific criteria were developed to enable categorization of DR-1. To be categorized as DR-1, a statistically significant (*t*-test) increase in dynamic range upon addition of conditioner must be demonstrated. Furthermore, there must be no statistically significant evidence of

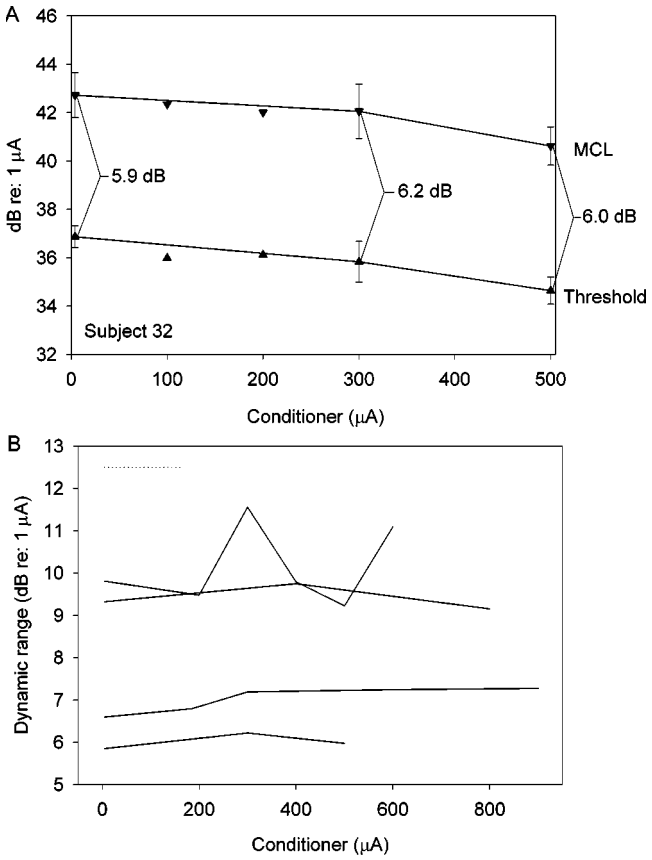


FIG. 1. Dynamic range profile DR-0. Profile is characterized by lack of a statistically significant increase in dynamic range with addition of the conditioner and is exemplified by (a) subject 32, electrode pair 7-8, 515-Hz sinusoid. Error bars represent 95% confidence intervals. (b) All electrode pairs with DR-0 are shown. Solid lines are drawn through mean values of dynamic range for which multiple data points were collected per conditioner level. The dotted line depicts an electrode pair for which one data point was collected per conditioner level.

stochastic resonance in either threshold or dynamic range. Subcategorization into the two variations of DR-1 is facilitated by the following three criteria. First, with increasing conditioner levels, the mean value of dynamic range sometimes decreased following an increase of dynamic range. In these cases, the data is categorized under the first variation if the dynamic range increased again by a statistically significant amount with further increases in conditioner. Otherwise, the data is categorized under the second variation. Second, if the mean value of dynamic range continued to increase at each higher level of conditioner tested, then the data is categorized under the first variation. Finally, for the three cases [denoted by arrowheads in Fig. 2(b)] where only one data point was collected at each conditioner level, categorization is based on visual inspection of the overall trend of the data, as statistics cannot be performed.

Subject 21, electrode pair 1-2 tested with a 515-Hz sinusoid, exemplifies the first variation. The data show the largest dynamic range (18.6 dB) at the highest conditioner level tested (800 μA), resulting in an increase in dynamic range of 11.5 dB. Subject 16, electrode pair 1-2 tested with a 1031-Hz sinusoid, exemplifies the second variation. Dynamic range increased from 6.8 dB without the conditioner to 17.1 dB with 500-μA conditioner. This increased dynamic

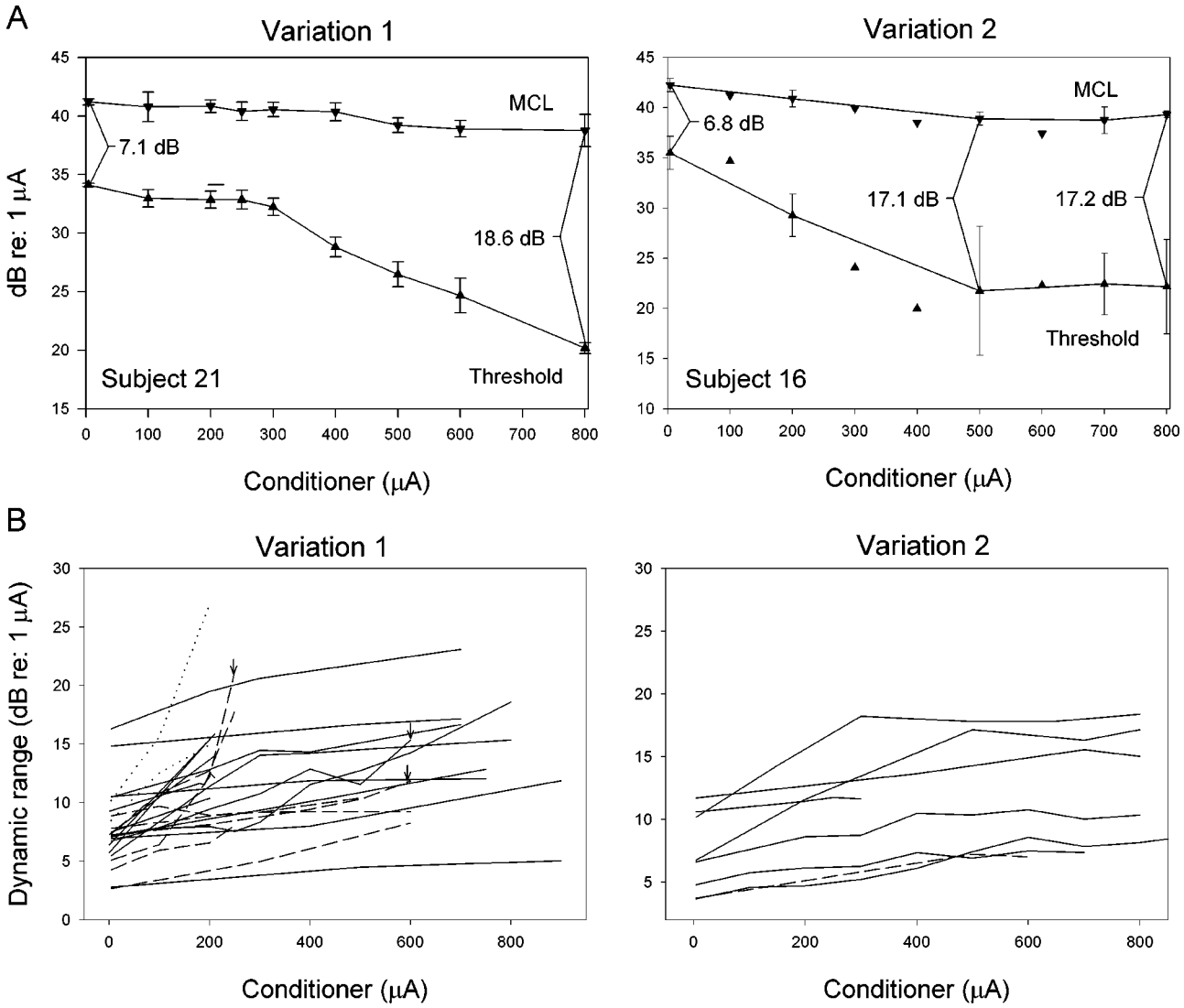


FIG. 2. Dynamic range profile DR-1. Profile is characterized by increases in dynamic range with addition of the conditioner without evidence of stochastic resonance. Two variations of DR-1 were observed: (a) subject 21, electrode pair 1-2, 515-Hz sinusoid, exemplifies the first variation of continual increases in dynamic range with increasing conditioner level. Subject 16, electrode pair 1-2, 1031-Hz sinusoid, exemplifies the second variation of increases in dynamic range up to a point, beyond which no further increases are observed. Conditioner levels at which multiple data points were collected are shown with error bars representing 95% confidence intervals. (b) All electrode pairs that demonstrated DR-1 are shown, segregated into the two variations of DR-1. Lines are drawn through mean values of dynamic range. Dashed lines represent data for which the highest conditioner tested was limited by loudness of the initial conditioner percept. Dotted lines represent data for which the highest conditioner was limited by the threshold falling below the noise floor. Multiple data points were collected per conditioner for all data except the three cases denoted by the arrowheads.

range was then maintained from 500- to 800-μA conditioner. In this second variation, it is possible to define an optimal conditioner level as that which causes the greatest increase in dynamic range above which further increases in the conditioner have little to no effect. For subject 16, this optimal level was 500 μA.

Figure 2(b) shows the dynamic range data of each of the 32 electrode pairs that demonstrates each variation of the DR-1 profile. From the figure, it can be seen that for electrode pairs demonstrating dynamic range increases without stochastic resonance, continued increases in conditioner level result in continued increases in dynamic range for some (variation 1) but not others (variation 2). With respect to the optimal conditioner, relevant only for the second variation of DR-1, the right panel of Fig. 2(b) demonstrates that it occurred between 250- and 700-μA conditioner for the tested electrode pairs.

From Fig. 2(b), it is observed that the highest conditioner level tested for different electrode pairs with the DR-1 profile ranges from 200 to 900 μA. This variability is not ideal for comparing across electrode pairs, but there are a number of practical reasons for it. In eight cases (dashed lines), the subject perceived the conditioner as ''loud'' at the maximum level tested; the conditioner was not increased further to prevent exposing the patient to ''uncomfortable loudness.'' In two other cases (dotted lines), threshold decreased to the noise floor of measuring capabilities at the highest conditioner level tested. Finally, the maximum conditioner level in four other cases was restricted to less than 250 μA by initial limitations in software at the time of testing.

The first profile suggesting the presence of stochastic resonance, SR-1, is demonstrated in Fig. 3. In SR-1, enhancement of dynamic range is observed with increasing conditioner level to an optimal point, beyond which further

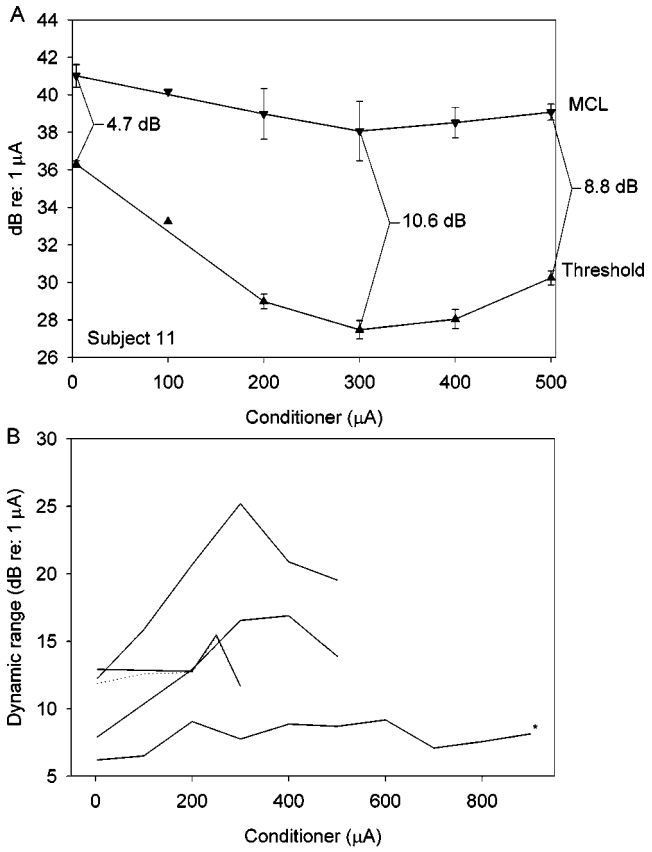


FIG. 3. Dynamic range profile SR-1. Profile is characterized by changes in dynamic range in response to the conditioner consistent with stochastic resonance. This is exemplified by (a) subject 11, electrode pair 1-2, 1031-Hz sinusoid. Error bars represent 95% confidence intervals. (b) All electrode pairs with SR-1 are shown. Solid lines are drawn through mean values of dynamic range for which multiple data points were collected per conditioner level. The dotted line represents an electrode pair for which one data point was collected per conditioner level. In one subject, stochastic resonance is statistically significant for threshold but not dynamic range (indicated by *).

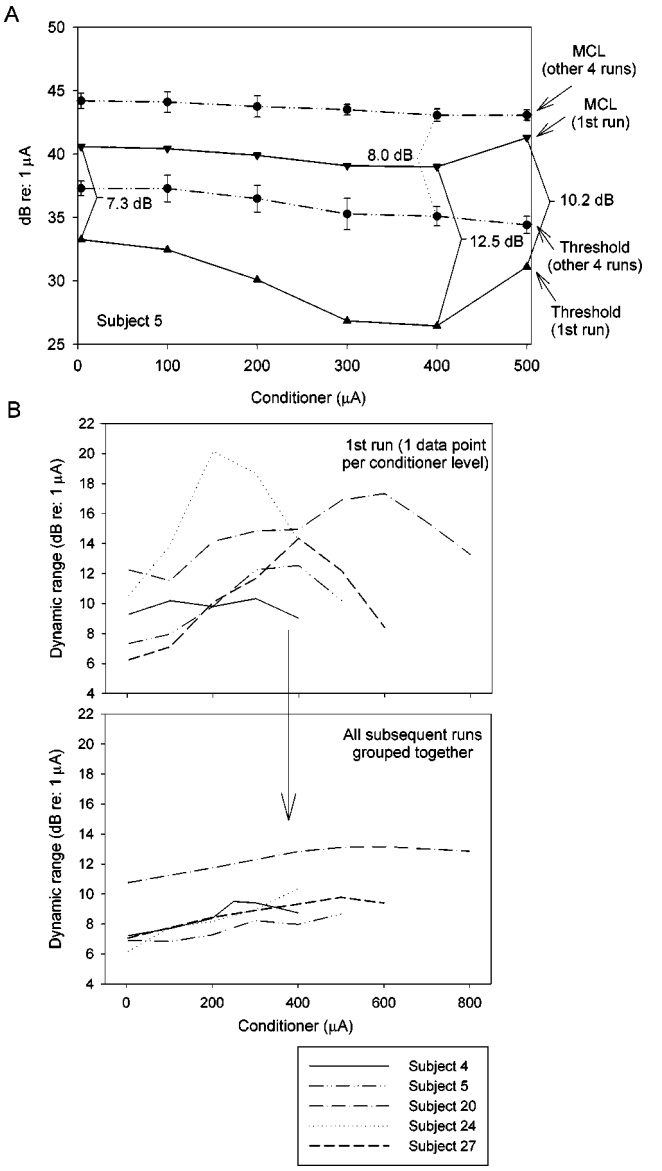


FIG. 4. Dynamic range profile SR-2. Profile is characterized by dynamic range behavior in response to the conditioner consistent with stochastic resonance in the first run and by adaptation in subsequent runs. This is exemplified by (a) subject 5, electrode pair 1-2, 515-Hz sinusoid. The solid lines depict threshold and MCL data from the first run, and the dash-dot-dot lines depict data from all subsequent runs. Error bars represent 95% confidence intervals. (b) The behavior of dynamic range in response to the conditioner for five subjects with SR-2, with only one data point collected per conditioner level in the first run, is shown.

increases in the conditioner result in loss of dynamic range. In all cases observed, this psychophysical behavior is primarily a result of changes in threshold, with a statistically significant (when multiple data points were collected; shown in solid lines) lowest threshold at the optimal point. Furthermore, with the exception of one case [denoted with an asterisk in Fig. 3(b)], statistically significant evidence of stochastic resonance is evident not only in threshold but also dynamic range. Figure 3(a) shows the profile of subject 11, tested on electrode pair 1-2 with a 1031-Hz sinusoid. The dynamic range increased from 4.7-dB without the conditioner to 10.6-dB with the optimal conditioner (300 μA); further increases in the conditioner to 500 μA resulted in loss of dynamic range to 8.8 dB. Threshold with a 300-μA conditioner was significantly lower than that without the conditioner ($t=46.94$; $p<0.001$) and at 500 μA ($t=10.69$; $p<0.001$), consistent with stochastic resonance. Figure 3(b) shows data from the five electrode pairs that demonstrated changes in threshold and dynamic range consistent with SR-1. The optimal conditioner level across this data set ranged from 250 to 600 μA.

The second profile consistent with stochastic resonance is designated SR-2. Like SR-1, an optimal level of the conditioner is observed, about which increases or decreases in conditioner amplitude result in loss of dynamic range. SR-2 is distinct from SR-1 in that adaptation is seen in the threshold and MCL after exposure to the conditioner. Psychophysical data from one subject that exemplifies SR-2, observed using the general protocol for characterizing dynamic range profiles described in Sec. II, is shown in Fig. 4(a). Figure 4(a) shows the SR-2 profile of subject 5 on electrode pair 1-2, tested with a 515-Hz sinusoid. In the first run (solid lines), dynamic range increased from 7.3 dB without the conditioner to 12.5 dB at the optimal 400-μA conditioner, before decreasing to 10.2 dB at 500-μA conditioner. In the next four runs (dash-dot-dot lines), however, both thresholds and

MCLs shifted upward. As a result, the previously observed 12.5-dB dynamic range at 400-μA optimal conditioner decreased to 8.0-dB. Figure 4(b) shows SR-2 in the five electrode pairs with this profile as observed using the general protocol. The optimal conditioner ranged from 200 to 600 μA. In all cases, increases in dynamic range with the optimal conditioner in the initial run were reduced in subsequent runs due to adaptation.

The results shown in Fig. 4 suggest that adaptation may affect the interpretation of dynamic range enhancement and stochastic resonance. For purposes of determining benefit in dynamic range from the conditioner, data collected from the first run, where adaptation is not clearly evident, are the data of interest. This is because for these data, adaptation appears to occur primarily after exposure of the subject to a conditioner level beyond the optimal conditioner, with a return of threshold and MCL to the preadapted state with time (on the scale of hours) when this exposure is removed (unpublished data). Future cochlear implant subjects programmed with a speech processor utilizing conditioning would be exposed on a continuous basis only to as high a conditioner level as necessary to achieve maximum dynamic range benefits, which would be a level prior to triggering the adaptation observed with the SR-2 profile.

To allow a test of the hypothesis by obtaining multiple data points of the first run, testing occurred with the modified protocol for characterizing dynamic range profiles suggestive of adaptation described in Sec. II. Figure 5(a) shows the dynamic range profiles of the same subject (subject 5) from Fig. 4(a), but tested 4 months later using the modified protocol. Figure 5(a) shows psychophysical data from subject 5, electrode pair 1-2, tested with a 515-Hz sinusoid. A dynamic range profile consistent with stochastic resonance occurred for the first run of data collected (solid lines), as the threshold at 300-μA conditioner was significantly lower than that without the conditioner ($t'=18.62$; $p<0.001$) and with 400-μA conditioner ($t=4.74$; $p<0.001$). This translated into an increase in dynamic range of 6.8 dB with the optimal conditioner. After subject 5 was exposed to 400-μA conditioner, adaptation was observed as thresholds and MCLs both shifted upward (dash-dot-dot lines). As a result, increases in dynamic range from addition of the conditioner were reduced: the dynamic range at 300-μA conditioner observed with adaptation was reduced from 13.9 to 10.8 dB, translating into a smaller observed increase in dynamic range (3.7 dB) with the optimal conditioner.

Figure 5(b) depicts SR-2 in the electrode pairs of five subjects observed with this modified protocol. In all five cases, both the threshold and dynamic range demonstrated statistically significant evidence of stochastic resonance at the optimal conditioner level. Again, as with Fig. 4(b), the trend was towards reduced increases in dynamic range with the conditioner following adaptation. For all five subjects, the changes in dynamic range were due to upward shifts in thresholds and MCLs, as exemplified by subject 5 in Fig. 5(a) (threshold and MCL data for other subjects not shown). The optimal conditioner as seen in Fig. 5(b) ranged from 100 to 400 μA. Note that subject 27, like subject 5, was also tested with both protocols, with data found in both Figs. 4 and 5.

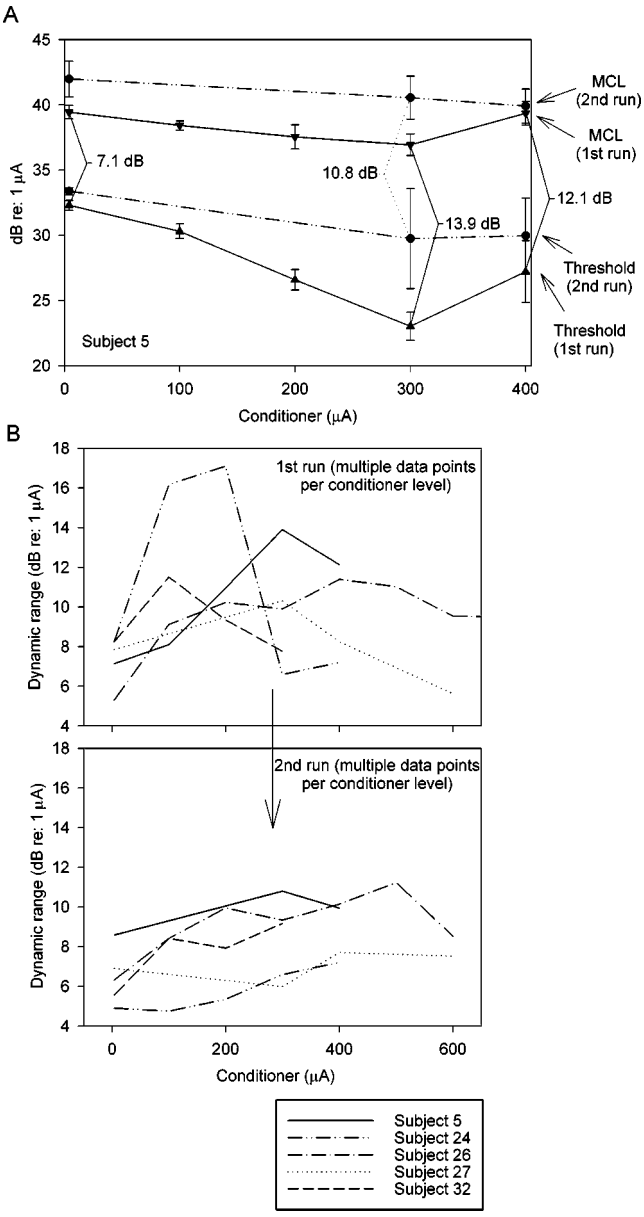


FIG. 5. Dynamic range profile SR-2 (modified protocol for characterizing adaptation). Profile is characterized by dynamic range behavior in response to the conditioner consistent with stochastic resonance in the first run and by adaptation in subsequent runs. This is exemplified by (a) subject 5, electrode pair 1-2, 515-Hz sinusoid. The solid lines depict threshold and MCL data from the first run, and the dash-dot-dot lines depict data from the second run. Error bars represent 95% confidence intervals. (b) Data from five subjects with the SR-2 profile demonstrated with multiple data points in each run are shown.

The third profile consistent with stochastic resonance is SR-3. In this profile, the behavior of the mean values of threshold and dynamic range with increasing conditioner levels are consistent with stochastic resonance, but not at statistically significant values (t-test). However, stochastic resonance is suggested by statistically significant changes in the variance of the threshold with increasing conditioner. SR-3 was found on one electrode pair in one subject. Figure 6 shows this lone example of SR-3 in subject 17 on electrode pair 7-8, tested with 1031-Hz sinusoid. As the conditioner

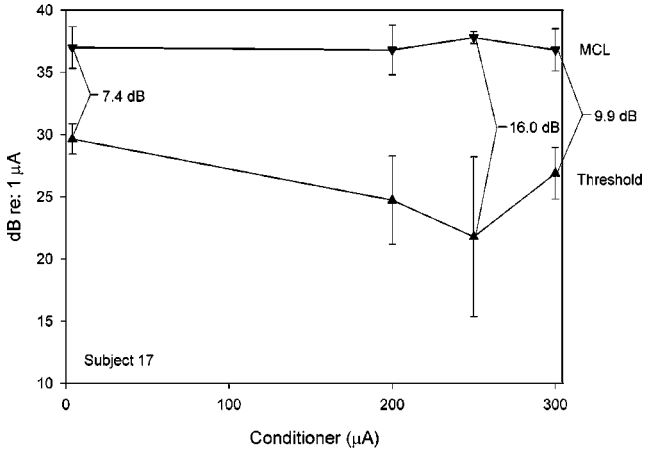


FIG. 6. Dynamic range profile SR-3. Profile is characterized by changes in dynamic range and variance of threshold in response to the conditioner consistent with stochastic resonance. This is demonstrated by subject 17, electrode pair 7-8, tested with a 1031-Hz sinusoid. Error bars represent 1 standard deviation.

was increased to 250 μA, the dynamic range increased to 16 dB (an increase of 8.6 dB) and the variance of the threshold (F statistic$=27.72$; $p<0.01$). Further increases in the conditioner to 300 μA produced a decrease in dynamic range (to 9.9 dB) and variance of threshold (F statistic$=25.91$; $p<0.05$). The variance of the dynamic range did not demonstrate statistically significant differences at different conditioner levels.

B. Dynamic range profiles across sinusoidal frequencies

Seven cochlear implant subjects were tested for response of psychophysical dynamic range to a 5-kpps conditioning pulse train across a range of sinusoidal frequencies (202, 515, and 1031 Hz). For each subject, experiments were carried out on the same day on the same electrode pair (bipolar configuration). The results are shown in Fig. 7. Solid lines are drawn through mean values of dynamic range for which multiple data points were collected per conditioner level. Dotted lines represent electrode pairs for which one data point was collected per conditioner level. From the figure, it is observed that dynamic range profiles are similar across sinusoidal frequencies. This holds true whether the profile is variation 1 of DR-1 (subjects 15, 25, and 29), variation 2 of DR-1 (subjects 16 and 32), SR-1 (subject 11), or SR-2 (subject 27). Additionally, the conditioner level at which the largest increase in dynamic range occurs is relatively consistent across frequencies, varying at most by 200 μA of conditioner (subject 16), with the differences in dynamic range across the variation of conditioner level never reaching statistical significance. For example, for subject 16, the largest dynamic range for the 515-Hz sinusoid was found at 600 μA and for the 1031-Hz sinusoid was found at 800 μA. However, the dynamic range for the 515-Hz sinusoid was not statistically significantly different at 800 μA than at the optimal 600 μA, while the dynamic range for the 1031-Hz sinusoid was not statistically significantly different at 600 μA than at the optimal 800 μA. Subject 16 does not have a profile for a 202-Hz sinusoid because threshold without the conditioner was too near the noise floor for meaningful testing.

C. Dynamic range profiles across electrode pairs

Seven cochlear implant subjects were tested for response of psychophysical dynamic range to a 5-kpps conditioning pulse train across a range of electrode pairs (1-2, 7-8, and/or 15-16) in bipolar configuration. For each subject, experiments were carried out on the same day at the same sinusoidal frequency. Some electrode pairs were tested up to higher conditioner levels than others because there was variation from electrode pair to electrode pair of the conditioner level at which the conditioner was initially perceived as ''loud'' or which caused threshold to decrease below the noise floor. This was relevant in two cases: the conditioner was initially perceived as ''loud'' for electrode pair 15-16 of subject 32 at 250-μA conditioner, and threshold decreased below the noise floor for electrode pair 7-8 of subject 5 beyond 200-μA conditioner. Figure 8 shows the dynamic range data across electrode pairs. In some cases, time limitations allowed only one data point to be collected at each conditioner level (shown in dotted lines in Fig. 8). From the figure, it is clear that dynamic range profiles and resulting enhancement of dynamic range may differ across electrode pairs. Variation in these profiles across electrodes was observed for five out of seven subjects (exceptions are subjects 1 and 30) and was seen across profile types, including DR-0, variation 1 of DR-1, variation 2 of DR-1, SR-1, and SR-2. For example, subject 5 showed a SR-2 profile with a 6.8-dB dynamic range increase on electrode 1-2, a variation 1 of DR-1 profile with a 6.5-dB increase in dynamic range on electrode 7-8, and a variation 1 of DR-1 profile with a 1.5-dB increase in dynamic range on electrode 15-16. (Electrode 7-8, as previously mentioned, was not tested beyond 200 μA because threshold decreased to the noise floor of measuring capabilities with 200-μA conditioner.) Additionally, changes in dynamic range from addition of the same level of the conditioner differed across electrode pairs. In subject 7, a 2.9-dB increase in dynamic range occurred with addition of 200-μA conditioner on electrode pair 7-8, but resulted in relatively little change in dynamic range on electrode 1-2 or 15-16 at the same conditioner level. Furthermore, the level of the conditioner producing the greatest increase in dynamic range may differ across electrode pairs. Subject 5 demonstrated an optimal conditioner level at 300 μA on electrode pair 1-2; in contrast, on electrode pair 15-16, dynamic range was essentially unchanged at 300-μA conditioner and did not increase significantly until 500-μA conditioner. However, it is not always the case that the optimal conditioner differed across electrodes, even if dynamic range profiles differ, as shown by the dynamic range profiles of subject 3, tested with a 202-Hz sinusoid. The minimum conditioner level producing the largest dynamic range increase occurred at 250 μA for electrode pairs 1-2, 7-8, and 15-16, despite the finding of a SR-1 profile on electrode pairs 1-2 and 7-8, and a variation 2 of DR-1 profile on electrodes 15-16.

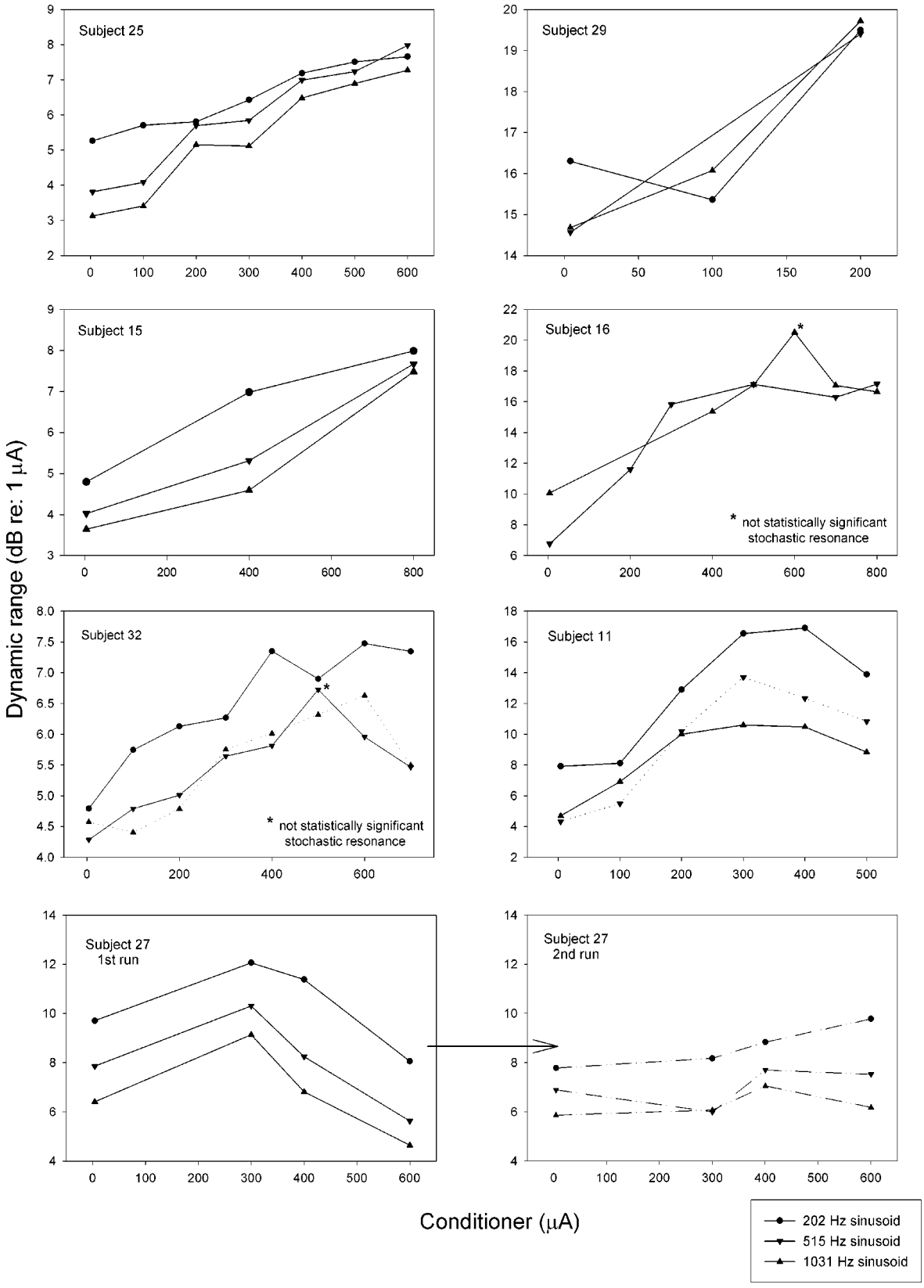


FIG. 7. Dynamic range profiles across sinusoidal frequencies. The response of dynamic range to the conditioner is depicted for seven subjects, tested with 202-, 515-, and/or 1031-Hz sinusoid. All subjects were tested on electrode pair 1-2, except subject 32 (electrode pair 7-8). The arrow denotes changes in dynamic range due to adaptation for subject 27 (SR-2 profile), with dash-dot-dot lines denoting the adapted runs. Solid lines are drawn through mean values of dynamic range for which multiple data points were collected per conditioner level. Dotted lines represent electrode pairs for which one data point was collected per conditioner level.

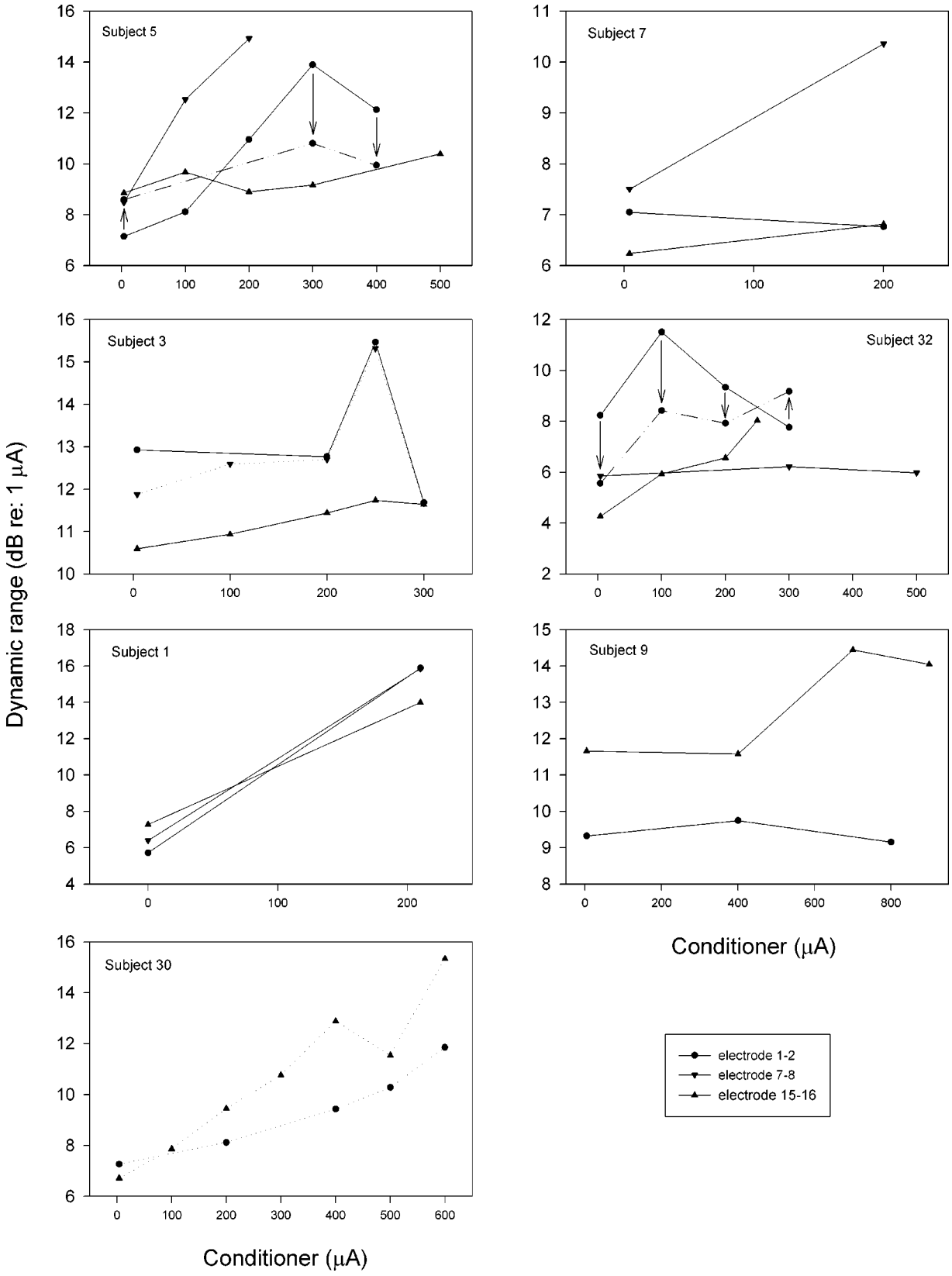


FIG. 8. Dynamic range profiles across electrode pairs. The response of dynamic range to the conditioner is depicted for seven subjects, tested on electrode pairs 1-2, 7-8, and/or 15-16. All subjects were tested with a 515-Hz sinusoid except subject 1 (202 Hz), subject 3 (202 Hz), and subject 7 (1031 Hz). Arrows denote changes in dynamic range due to adaptation for a particular electrode pair (relevant for SR-2 profile), with dash-dot-dot lines denoting the adapted runs. Solid lines are drawn through mean values of dynamic range for which multiple data points were collected per conditioner level. Dotted lines represent electrode pairs for which one data point was collected per conditioner level.

TABLE II. Dynamic range profiles across time. The dynamic range profiles of 15 electrode pairs (bipolar configuration) tested on two or three different test dates is shown. For each test date (denoted by time after initial stimulation), the sinusoidal frequency tested is the same. The highest conditioner level tested, dynamic range profile, optimal conditioner, and largest dynamic range increase found are shown for each test date. All subjects are tested with 500-ms sinusoidal bursts. Boldface signifies dynamic range profiles that differed with testing across time. DR-1 (variation 2) is indicated by denotation of an optimal conditioner for the DR-1 profile. Abbreviations: freq.=frequency; DR=dynamic range; cond.=5 kHz conditioner; d=day; wk=weeks; m=months Opt. cond.=optimal conditioner level=5 kHz conditioning pulse train that produces the largest increase in dynamic range, above which no further increases in dynamic range are observed; relevant for DR-1 (var. 2), SR-1, and SR-2.

Subject	Electrode pair	Sine freq. (Hz)	Test date no. 1	Highest cond. tested (μA)	DR profile	Opt. cond. (μA)	DR increase (dB)	Test date no. 2	Highest cond. tested (μA)	DR profile	Opt. cond. (μA)	DR increase (dB)	Test date no. 3	Highest cond. tested (μA)	DR profile	Opt. cond. (μA)	DR increase (dB)
27	1-2	515	1d	600	SR-2	400	8.1	1m	600	SR-2	300	2.5					
32	7-8	515	**1d**	700	**DR-1**	500	2.4	**1m**	500	**DR-0**	⋯	0.4					
21	1-2	515	**1d**	500	**SR-1**	300	13	**1m**	800	**DR-1**	300	9.3	**3m**	800	**DR-1**	⋯	11.5
7	15-16	515	**1wk**	250	**DR-0**	⋯	0.4	**3m**	250	**DR-1**	⋯	1.3					
25	1-2	515	2wk	600	DR-1	⋯	5.6	1m	600	DR-1	⋯	4.2					
16	1-2	1031	1m	800	DR-1	500	10.4	2m	800	DR-1	600	10.4					
3	1-2	202	1m	220	SR-1	200	4.3	2.5m	500	SR-1[a]	250	3.5	10m	300	SR-1	250	2.5
10	1-2	515	1m	600	DR-1		8.8	7m	800	DR-1		9.9					
5	1-2	515	3m	500	SR-2	400	5.2	7m	400	SR-2	300	6.8					
13	1-2	515	**3.5m**	550	**SR-2**	250	8.9	**9m**	700	**DR-1**	⋯	6.2					
9	1-2	515	5m	800	DR-0	⋯	0.4	7m	700	DR-0	⋯	−0.9					
9	15-16	515	5m	900	DR-1	700	2.8	7m	900	DR-1	700	3.8					
8	1-2	202	5m	750	DR-1	⋯	5.7	13m	600	DR-1	⋯	4.4					
15	1-2	202	8m	900	DR-1	⋯	5	12m	800	DR-1	⋯	3.2					
20	1-2	515	**10.5m**	800	**SR-2**	600	5.1	**11.5m**	750	**DR-0**	⋯	−0.5					

[a]Changes in dynamic range consistent with stochastic resonance, but threshold differences not statistically significant ($p<0.1$).

D. Dynamic range profiles across time

Fourteen cochler implant subjects were tested for response of psychophysical dynamic range to a 5-kpps conditioning pulse train in testing sessions separated by weeks to months. One subject (subject 9) was tested at different times on each of two different electrode pairs. The results are shown in Table II. The table demonstrates that the dynamic range profiles of 5 out of 15 electrode pairs tested changed with time (shown in boldface in Table II). There is no clear trend with respect to months after initial hookup of cochlear implant for those whose profiles changed over time. The profiles of subject 21 changed between 1 day, 1 month, and 3 months after initial stimulation, while those of subject 20 changed between 10.5 and 11.5 months after initial stimulation. In considering those ten electrode pairs that maintained the same dynamic range profiles over time, the amount of increase in dynamic range also stayed relatively constant (within 1–2 dB of each other), with the one exception of subject 27 (8.1- versus 2.5-dB increase in dynamic range over time). For the six electrode pairs demonstrating the presence of an optimal conditioner level (i.e., DR-1 [variation 2], SR-1, or SR-2 profiles) over time, the conditioner level at which the greatest increase in dynamic range was found also stayed relatively constant over time, never varying by more than 100 μA.

E. Overview of dynamic range profiles

The classification of dynamic range responses to different levels of a 5-kpps conditioner from 31 patients is shown in Fig. 9. Patients tested with multiple sinusoidal frequencies on the same electrode pair are only represented once in this count, because no difference in dynamic range profiles is observed between frequencies. In contrast, each unique electrode pair tested from a subject is represented as a data point in Fig. 9, since the data suggests that different profiles may be seen for different electrode pairs within the same subject. As previously described, five electrode pairs tested across time showed a change in dynamic range profile (Table II). For the four electrode pairs that showed two different profiles with repeat testing, each profile is represented as $\frac{1}{2}$ electrode pair in Fig. 9. For the electrode pair 1-2 of subject 21,

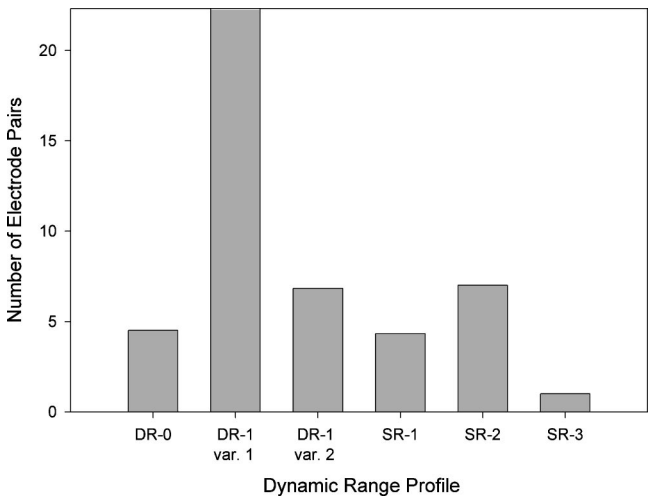


FIG. 9. Dynamic range profiles. The dynamic range profiles of 46 electrode pairs from 31 patients are shown. For electrode pairs tested over time: if tested three times, then each time=$\frac{1}{3}$ of a profile; if tested two times, then each time=$\frac{1}{2}$ of a profile. DR-0=no statistically significant increase in dynamic range; DR-1, variation 1=dynamic range continues to increase across range tested; DR-1, variation 2=dynamic range increases and then plateaus; SR-1=stochastic resonance without adaptation; SR-2=stochastic resonance simultaneous with adaptation; SR-3=stochastic resonance of variance.

which demonstrated three different profiles in each of three testing sessions, each profile is represented as $\frac{1}{3}$ electrode pair.

Evidence of stochastic resonance (SR-1, SR-2, and SR-3) was found in 12 $\frac{1}{3}$ out of 46 electrode pairs. The largest category of dynamic range profiles is variation 1 of DR-1 (22 $\frac{1}{3}$ electrode pairs). The DR-1 profile (variation 2) was observed in 6 $\frac{5}{6}$ electrode pairs. Only 4 $\frac{1}{2}$ out of 46 electrode pairs showed an increase in dynamic range of less than 1 dB with addition of the 5-kpps conditioner. For those profiles where a clear largest increase in dynamic range could be found, beyond which further increases in conditioner level did not result in additional gains in dynamic range (i.e., for DR-1 [variation 2], SR-1, SR-2, and SR-3), regression analysis was performed. This analysis examined potential associations between both threshold and dynamic range without the conditioner versus magnitude of the largest dynamic range increase. Analysis was performed for both individual sinusoidal frequencies and across frequencies. No significant associations were found.

The dynamic range profiles of a small subset of electrodes tested (4 out of 46 electrode pairs) were exceptional in that increases in MCL were found to contribute to increases in dynamic range. All four cases occurred with variation 1 of the DR-1 profile. In all cases, however, decreases in threshold played the more significant if not dominant role in the increases in dynamic range observed with addition of the conditioner.

All electrode pairs tested that showed increases in dynamic range with addition of conditioner when dynamic range was analyzed in dB also showed increases in dynamic range when it was analyzed in microamps. There are isolated cases where the optimal conditioner level differs when the dynamic range data is analyzed in linear versus log units, leading to reclassification of the data into a different profile, but the overall picture is the same no matter which units are used.

F. Optimal conditioner correlations

The optimal conditioner level is defined as the amplitude of the continuous 5-kpps unmodulated pulse train at which the largest increase in dynamic range was found. It is possible to define an optimal conditioner for four of the dynamic range profiles described: DR-1 (variation 2), SR-1, SR-2, and SR-3. The range of optimal conditioner levels found across subjects spans from 100 to 700 μA. Regression analysis of unconditioned threshold versus optimal conditioner level, and for unconditioned dynamic range versus optimal conditioner level, revealed no significant linear associations. The same analyses for data grouped by sinusoidal frequency also revealed no significant associations. Additional analysis of patient performance with their everyday speech processing strategies on CNC monosyllabic word tests (3 months after initial stimulation), increase in dynamic range with the optimal conditioner, and optimal conditioner level also did not reveal any significant correlation amongst any two of the three.

IV. DISCUSSION

A. Response of dynamic range to conditioning pulse trains

The addition of the 5-kpps conditioner to the electrical signals of cochlear implant patients resulted in increases in psychophysical dynamic range for almost every electrode pair tested. For each patient tested, there was at least one electrode pair that received benefit in dynamic range from the conditioner, resulting from statistically significant decreases in threshold. With respect to stochastic resonance, ~25% of electrode pairs tested exhibited behavior (SR-1, SR-2, or SR-3 profile) consistent with this phenomenon: increasing levels of the conditioner correspond to increasing dynamic range up to a point (corresponding to the stochastic resonance point), beyond which further increases in the conditioner led to a deterioration of dynamic range. For the other 75% of electrode pairs tested—those demonstrating DR-0 or DR-1 profiles—no evidence of stochastic resonance was observed. However, it is possible for these electrode pairs that such evidence may have been found if a smaller step size for conditioner level had been used and/or it had been possible to test the conditioner up to higher levels. Also, for the two cases where the conditioner was not increased because thresholds went under the noise floor, it is possible that had testing continued, the threshold may have passed back through the noise floor, thus providing evidence of stochastic resonance.

Increases in dynamic range with high-rate electrical stimulation are not unprecedented. A number of studies have demonstrated that psychophysical dynamic range expands with increases in pulse rate (Shannon, 1985, 1993; Pfingst and Morris, 1993). More recently, Battmer *et al.* (2002) has observed dynamic range increases in patients using the Clarion high resolution strategy, a strategy based on nonsimultaneous pulsatile stimulation at high rates. The relationship of these findings to those of conditioning pulse trains is currently unclear.

The variation in dynamic range profiles observed from patient to patient, as well as from one electrode pair to the next within the same patient, is understandable in light of the single auditory nerve fiber experiments performed by Litvak (2002) with high-rate conditioning pulse trains in cats. This investigator demonstrated that significant variation existed in the response to the 5-kpps conditioner from fiber to fiber. When the conditioner was applied at a fixed level for 10 min to the auditory nerve, some fibers exhibited a sustained response throughout the presentation of the conditioner, while others responded only transiently during the first 1–2 min of stimulation. Furthermore, approximately 20% of fibers that responded to the conditioner exhibited sustained discharge rates beyond that of spontaneous activity in healthy auditory neurons. When the dynamic range of the differently responding fibers was measured with sinusoidal modulation of the conditioner, sustained responders were found to have a significantly larger dynamic range (up to four times higher in dB) than the transient responders.

The experiments in this paper presented a fixed level of the conditioner to all nerve fibers within the electrical field of

a particular electrode pair at each conditioner amplitude tested. If it is assumed that the responses of human auditory nerve fibers to the conditioner are similar to that of cats, then for a particular conditioner level, the population of fibers may consist of sustained responders with spontaneouslike activity, sustained responders without spontaneouslike activity, and/or transient responders. For populations of nerve fibers dominated by transient responders across all conditioner levels tested, the resulting increases in dynamic range observed with the conditioner may be minimal (i.e., DR-0 profile). In contrast, for populations dominated by sustained responders with spontaneouslike activity at a particular conditioner level, significant increases in dynamic range may occur.

It is speculated that different auditory nerve fibers may be sensitive to different levels of optimal conditioner based on individual fiber electrophysiology and electro-anatomical location relative to the stimulating electrode. Thus, at the conditioner level resulting in the largest increase in dynamic range in each profile described, only a subset of the stimulated nerve fibers most sensitive to that conditioner amplitude receives maximum benefit from the conditioner. Consequently, the profiles characterized as consistent with stochastic resonance (i.e., SR-1, SR-2, and SR-3) may represent stochastic resonance for a population of stimulated nerve fibers, but not for each individual fiber—that is, they describe the conditioner level at which the summation of dynamic range increases is greatest across nerve fibers. This interpretation represents one rationale for why the dynamic range increases found with addition of the conditioner are less than the ~20-dB increases predicted by stochastic models of auditory nerve fibers (Rubinstein, 1998).

The potential sensitivity of nerve fibers to different levels of the conditioner also offers an explanation for the different types of dynamic range profiles observed. An electrode pair that stimulates a population of fibers with sharp dynamic-range-versus-conditioner-level functions but a large variance in optimal conditioner levels per fiber may demonstrate little to no benefit from the conditioner (as seen in the DR-0 profile). In contrast, an electrode pair that stimulates a population with a broad dynamic-range-versus-conditioner-level functions and small variance in optimal conditioner per fiber may realize large increases in dynamic range. Because different electrode pairs stimulate different populations of neurons, and because each population of neurons likely differs in the relative numbers of fibers exhibiting stochastic resonance and magnitude of this effect at each conditioner level, it is reasonable that stimulation with different electrode pairs results in different dynamic range profiles.

B. Adaptation

Two parts of the experiments in this paper yield evidence of adaptation. The first occurs after higher levels of conditioner were initially presented to the cochlear implant subject. An initial ''buzzing'' of the conditioner was perceived that faded away to imperceptibility on the order of seconds to minutes. This psychophysical adaptation is consistent with that observed by others in response to continuous high-rate electrical stimulation (Shannon, 1983; Brimacombe and Eisenberg, 1984). This adaptation likely has both peripheral and central components. Adaptation on this time scale with continuous high-rate stimulation has been measured peripherally with both single unit recordings (Moxon, 1971; van den Honert and Stypulkowski, 1987; Litvak, 2002) and electrically evoked compound action potentials (Abbas *et al.*, 2001). Those results suggest that the disappearance of the conditioner's ''buzzing'' with time in this study may be due to a decrease in peripheral neural activity, resulting in subthreshold perception levels. In contrast, the observation that threshold decreases with addition of the conditioner argues that peripheral adaptation effects are limited and central adaptation is the more likely cause.

Effects of adaptation were also observed in electrode pairs demonstrating a SR-2 dynamic range profile. After exposure to the highest tested conditioner level, patients for these electrode pairs demonstrated an elevation in both threshold and MCL for repeated measurements of dynamic range at lower conditioner levels (Figs. 6 and 7). To account for confounding effects from this adaptation, dynamic range increases were calculated by comparing the dynamic range prior to exposure to the conditioner to that of the first exposure to each increasing level of the conditioner (i.e., data from the ''first run''). It is currently unclear for the SR-2 profile whether decreases in dynamic range observed for the conditioner beyond the optimal level are due to adaptation, stochastic resonance, or both. For example, it is possible that the optimal conditioner observed in this profile is also the level at which adaptation first occurs, with the resulting adaptation leading to higher thresholds and MCLs at all conditioner levels subsequently applied. Further studies are needed to distinguish effects of adaptation from stochastic resonance for electrode pairs with the SR-2 profile.

Litvak (2002) observed peripheral adaptation that ''may occur exclusively in response to high-level stimulation'' with the 5-kpps conditioner consistent with those seen in the SR-2 profile. This adaptation occurred on the scale of hours, with longer continuous exposure to the conditioner (measured up to 1-h duration) leading to longer subsequent recovery periods. During testing, patients were exposed to varying levels of the conditioner over 2–3 hs. Time with the conditioner off, when recovery may have occurred, lasted on the scale of tens of minutes. The results of this paper demonstrating reliable measures of threshold and MCL after first exposure to the highest conditioner level tested (Fig. 6) are consistent with an adaptation with a recovery period on the scale of hours. The exact duration of this adaptation in humans remains to be characterized, as do the mechanisms responsible for this adaptation.

The peripheral adaptation observed in these experiments on the scale of hours suggests that a careful consideration of the safety of high-rate electrical stimuli is necessary. For example, one hypothesis for the mechanism of adaptation is that sustained high-rate stimulation leads to a depletion of the electrochemical gradient responsible for auditory nerve conduction, with possible harmful effects to the nerve. However, there is no current evidence that high-rate stimuli produce long-term deleterious effects. First, none of the patients tested, including those demonstrating adaptation, have re-

ported any negative effects on auditory perception following exposure to the conditioner. Second, histopathological data suggests that chronic high-rate electrical stimulation does not adversely affect the auditory nerve (Tykocinski *et al.*, 2002). Third, patients implanted with the house single-channel cochlear implant have received high-rate electrical stimulation (16-kHz sinusoidal carrier) for years (e.g., Brimacombe and Eisenberg, 1984) without clinical deterioration in performance linked to nerve injury. Nevertheless, it is important to proceed with caution in applying the high-rate conditioner to cochlear implant recipients to avoid any potential harm to the patient. Therefore, as studies move towards examining the conditioner in ''take home'' signal processing strategies for speech and music perception, they will exclude those demonstrating the SR-2 profile.

C. Reliability of measurements

The effect of the conditioner on dynamic range was measured during multiple testing sessions separated by many weeks to months and was found to be relatively consistent with respect to dynamic range profile, amount of increase in dynamic range, and optimal conditioner level. The consistency is remarkable given that many subjects were tested in the first few months after hookup of their implants, a time period when patients often demonstrate significant changes in threshold and MCL (Pfingst, 1990; Butts *et al.*, 2000). This consistency across testing sessions supports the test–retest reliability of the measures used in this paper. It is important to emphasize that it does not, however, lend insight into the effect of long-term (weeks to months) stimulation with the 5-kpps conditioner, because the capability to program the conditioner into the ''take-home'' signal processing strategies of cochlear implant users does not currently exist.

D. Implications for signal processing

The results presented in this paper suggest that incorporation of a 5-kpps conditioner into a signal processing strategy may have significant benefits for cochlear implant patients, given the increases in dynamic range observed in this study. These increases in dynamic range should translate into less compression needed by the signal processor when the conditioner is applied, resulting in less compressive distortion (Boike and Souza, 2000). Ideally, for greatest benefit, the conditioner would not only increase dynamic range but also the number of discriminable steps of loudness. However, the effect of conditioning on the intensity difference limen is currently unknown and an issue for future study.

The results of this study also have multiple implications for the design and programming of a conditioned signal processing strategy. They suggest that for optimal benefit from such a strategy, it may be necessary to program each electrode pair with a different level of the conditioner, as different electrode pairs may generate different dynamic range profiles. However, at least for stimuli ranging from 202 to 1031 Hz, it does not appear that the conditioner needs to be tuned to individual stimuli; dynamic range profiles appear to be consistent across sinusoidal frequencies. It will also be important to consider effects of adaptation when determining the appropriate levels of the conditioner for a conditioned strategy. As demonstrated by the SR-2 profile, adaptation may lead to misinterpretation of the conditioner level resulting in the largest increase in dynamic range if it is not properly accounted for in the programming of thresholds and MCLs. Finally, the programming of a conditioned strategy will be easier if the most favorable conditioner level can be predicted without having to test each electrode pair at multiple conditioner amplitudes. Analysis of threshold and dynamic range data does not reveal guidelines for predicting this level psychophysically. This is true for the data both across and grouped by sinusoidal frequencies. The latter grouping is necessary because (1) unconditioned thresholds are higher with higher sinusoidal frequency (Pfingst and Morris, 1993), consistent with the results in this paper, while (2) optimal conditioner level is similar across sinusoidal frequencies (Fig. 7), suggesting that correlations are unlikely between unconditioned threshold and optimal conditioner level for data across frequencies. In future studies, it will be of interest to look at alternatives to psychophysical techniques for determining the optimal conditioner level. One such alternative may be with electrophysiological techniques such as neural response telemetry (Abbas *et al.*, 1999; Brown *et al.*, 2000), which may lead to a less time-consuming programming of the conditioner in signal processing strategies than psychophysics.

ACKNOWLEDGMENTS

This research was supported by the NIH Program Project Grant No. DC00242, Neural Prosthesis Contract No. DC92107, Doris Duke Clinical Research Fellowship Program, Advanced Bionics Corporation, and Texas Instruments. The authors are grateful to Paul Abbas and Chris Turner for helpful comments regarding the manuscript. The authors also thank Dan Wehner for his work in designing the custom interface used for patient testing, and Gang Chen and Haiming Chen for their work in subsequent modifications.

Abbas, P. J., Miller, C. A., Rubinstein, J. T., Robinson, B. K., and Hu, N. (**2001**). ''The neurophysiological effects of simulated auditory prostheses stimulation,'' Eighth Quarterly Progress Report N01-DC-9-2107, NIH Neural Prosthesis Program.

Abbas, P. J., Brown, C. J., Shallop, J. K., Firszt, J. B., Hughes, M. L., Hong, S. H., and Staller, S. J. (**1999**). ''Summary of results using the Nucleus CI24M implant to record the electrically evoked compound action potential,'' Ear Hear. **20**, 45–59.

Battmer, R., Buchner, A., Frohne-Buchner, C., Popp, P., and Lenarz, T. (**2002**). ''The Clarion CII High Resolution mode: experience after 6 months,'' abstract, 7th International Cochlear Implant Conference, Manchester, UK, P081.

Boike, K., and Souza, P. (**2000**). ''Effect of compression ratio on speech recognition and speech-quality ratings with wide dynamic range compression amplification,'' J. Speech Lang. Hear. Res. **43**, 456–468.

Brimacombe, J. A., and Eisenberg, L. S. (**1984**). ''Tone decay in subjects with the single-channel cochlear implant,'' Audiology **23**, 321–332.

Brown, C. J., Hughes, M. L., Luk, B., Abbas, P. J., Wolaver, A., and Gervais, J. (**2000**). ''The relationship between EAP and EABR thresholds and levels used to program the Nucleus 24 speech processor: data from adults,'' Ear Hear. **21**, 151–163.

Butts, S. L., Hodges, A. V., Dolan-Ash, S., and Balkany, T. J. (**2000**). ''Changes in stimulation levels over time in Nucleus 22 cochlear implant users,'' Ann. Otol. Rhinol. Laryngol. Suppl. **185**, 53–56.

Chatterjee, M., and Robert, M. E. (**2001**). ''Noise enhances modulation sensitivity in cochlear implant listeners: stochastic resonance in a prosthetic sensory system?'' J. Assoc. Res. Otol. **2**, 159–171.
Collins, J. J., Chow, C. C., and Imhoff, T. T. (**1995**). ''Stochastic resonance without tuning,'' Nature (London) **376**, 236–238.
Douglass, J. K., Wilkens, L., Pantazelou, E., and Moss, F. (**1993**). ''Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance,'' Nature (London) **365**, 337–340.
Fu, Q-J., and Shannon, R. (**2000**). ''Effects of dynamic range and amplitude mapping on phoneme recognition in Nucleus-22 cochlear implant users,'' Ear Hear. **21**, 227–235.
Hong, R. S., Rubinstein, J. T., Wehner, D., and Horn, D. (**2003**). ''Dynamic range enhancement for cochlear implants,'' Otol. Neurotol. **24**, 590–595
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.
Levin, J. E., and Miller, J. P. (**1996**). ''Broadband neural encoding in the cricket cercal sensory system enhanced by stochastic resonance,'' Nature (London) **380**, 165–168.
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.
Liberman, M. C., and Dodds, L. W. (**1984**). ''Single-neuron labeling and chronic cochlear pathology. II. Stereocilia damage and alterations of spontaneous discharge rates,'' Hear. Res. **16**, 43–53.
Litvak, L. (**2002**). ''Towards a better speech processor for cochlear implants: auditory nerve responses to high rate electric pulse trains,'' Ph.D. thesis, MIT, Cambridge, MA.
Litvak, L., Delgutte, B., and Eddington, D. (**2001**). ''Auditory nerve fiber responses to electric stimulation: Modulated and unmodulated pulse trains,'' J. Acoust. Soc. Am. **110**, 368–379.
Loizou, P., Dorman, M., and Fitzke, J. (**2000**). ''The effect of reduced dynamic range on speech understanding: implications for patients with cochlear implants,'' Ear Hear. **21**, 25–31.
Matsuoka, A. J., Abbas, P. J., Rubinstein, J. T., and Miller, C. A. (**2000**). ''The neuronal response to electrical constant-amplitude pulse train stimulation: additive Gaussian noise,'' Hear. Res. **149**, 129–137.
Matsuoka, A. J., Rubinstein, J. T., Abbas, P. J., and Miller, C. A. (**2001**). ''The effects of interpulse interval on stochastic properties of electrical stimulation: models and measurements,'' IEEE Trans. Biomed. Eng. **48**, 416–424.
Morse, R. P., and Evans, E. F. (**1996**). ''Enhancement of vowel coding for cochlear implants by addition of noise,'' Nat. Med. **2**, 928–932.
Morse, R. P., and Evans, E. F. (**1999a**). ''Additive noise can enhance temporal coding in a computational model analogue cochlear implant stimulation,'' Hear. Res. **133**, 107–119.
Morse, R. P., and Evans, E. F. (**1999b**). ''Preferential and non-preferential transmission of formant information by an analogue cochlear implant using noise: the role of the nerve threshold,'' Hear. Res. **133**, 120–132.
Morse, R. P., and Roper, P. (**2000**). ''Enhanced coding in a cochlear-implant model using additive noise: Aperiodic stochastic resonance with tuning,'' Phys. Rev. E **61**, 5683–5692.
Moxon, E. C. (**1971**). ''Neural and mechanical responses to electrical stimulation of the cat's inner ear,'' Ph.D. thesis, MIT, Cambridge, MA.
Neuman, A., Bakke, M., Mackersie, C., Hellman, S., and Levitt, H. (**1998**). ''The effect of compression ratio and release time on the categorical rating of sound quality,'' J. Acoust. Soc. Am. **103**, 2273–2281.
Pfingst, B. E. (**1990**). ''Changes over time in thresholds for electrical stimulation of the cochlea,'' Hear. Res. **50**, 225–236.
Pfingst, B. E., and Morris, D. J. (**1993**). ''Stimulus features affecting psychophysical detection thresholds for electrical stimulation of the cochlea. II: Frequency and interpulse interval,'' J. Acoust. Soc. Am. **94**, 1287–1294.
Rubinstein, J. T., and Hong, R. S. (**2003**). ''Signal coding in cochlear implants: exploiting stochastic effects of electrical stimulation,'' Ann. Otol. Rhinol. Laryngol. **112**, 14–19.
Rubinstein, J. T., Abbas, P. J., and Miller, C. A. (**1998**). ''The neurophysiological effects of simulated auditory prostheses stimulation,'' Eighth Quarterly Progress Report N01-DC-6-2111, NIH Neural Prosthesis Program.
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.
Runge-Samuelson, C. (**2002**). ''Response of the auditory nerve to sinusoidal electrical stimulation: effects of high-rate pulse trains,'' Ph.D. thesis, University of Iowa, Iowa City, IA.
Shannon, R. V. (**1983**). ''Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics,'' Hear. Res. **11**, 157–189.
Shannon, R. V. (**1985**). ''Threshold and loudness functions for pulsatile stimulation of cochlear implants,'' Hear. Res. **18**, 135–143.
Shannon, R. V. (**1993**). ''Psychophysics of electrical stimulation,'' in *Cochlear Implants: Audiological Foundations*, edited by R. S. Tyler (Singular, San Diego, CA), pp. 357–388.
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.
Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). ''Chimaeric sounds reveal dichotomies in auditory perception,'' Nature (London) **416**, 87–90.
Tykocinski, M., Linahan, N., Shepherd, R. K., and Clark, G. M. (**2002**). ''Chronic monopolar high rate stimulation of the auditory nerve,'' in *Cochlear Implants—an Update*, edited by T. Kubo, Y. Takahashi, and T. Iwaki (Kugler, The Hague, Netherlands), pp. 3–9.
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.
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.
Zeng, F.-G., and Galvin, J. (**1999**). ''Amplitude mapping and phoneme recognition in cochlear implant listeners,'' Ear Hear. **20**, 60–74.
Zeng, F.-G., Fu, Q.-J., and Morse, R. (**2000**). ''Human hearing enhanced by noise,'' Brain Res. **869**, 251–255.

Respiratory constraints on speech production: Starting an utterance

Janet Slifka[a)]
Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307

(Received 3 January 2003; accepted for publication 30 September 2003)

The aim of this research is to model the sequences of respiratory and articulatory events that occur when a speaker starts talking. In this paper, the focus is placed on a set of timing landmarks—utterance onset, phonation onset, and the initial peak in subglottal pressure—in relation to the first sound segment in the utterance and the location of the first syllable in the utterance that is judged as prominent by a listening panel. Simultaneous recordings of the acoustic signal, airflow, lung volume, and subglottal pressure as estimated from esophageal pressure were collected from four native speakers of American English. The timing of the initial alveolar pressure peak generally coincides with a prominent syllable if that syllable is the first or second syllable in the utterance, and precedes later prominences. There is considerable scatter in phonation pressure and phonation onset time across speakers and utterances. Normalizing these parameters to the initial pressure peak provides a more coherent view of the pressure trajectory and the influence of the initial sound segment on phonation parameters. Use of the data in the development of a model of the respiratory system during speech is discussed.
[DOI: 10.1121/1.1627838]

PACS numbers: 43.70.Fq [AL] Pages: 3343–3353

I. INTRODUCTION

Models of speech production, such as articulatory speech synthesis, provide a platform to test our understanding of the speech process. One of the parameters in this process is subglottal pressure, which provides the energy to create sound sources in speech. During the production of an utterance, the subglottal pressure is generally considered to be relatively constant and has been characterized as one which may have local rises and drops associated with sharp changes in the impedance to air flow from the lungs (for example, Ohala, 1990) or as one that allows for the possibility of local pressure rises above the mean background pressure that are linked to prosodic prominences. [See, for example, Draper *et al.* (1959); Ladefoged (1967); and Netsell (1969).] Models of speech production generally specify a time track for subglottal pressure in an utterance that is either a constant or a rather arbitrary rise, plateau, and fall.

A few studies discuss the manner in which the pressure rises to this relatively constant subglottal pressure at the initiation of an utterance. Lieberman (1967) reports that phonation begins 120–200 ms after the start of expiration at a pressure that is most often near 8–9 cm H_2O for data from three speakers and 19 utterances. Atkinson (1973) followed up with a study for a single male speaker, reading 18 utterances, that reported that the subglottal pressure at the onset of phonation was in the range of 6.7–9 cm H_2O on average and was reached in 105–270 ms on average. For a single male speaker reading, in Swedish, a 55-s passage from a novel, Fant and colleagues report that the subglottal pressure at the onset of phonation is in the range of 3–5 cm H_2O and reaches an initial peak in the range of 6–8 cm H_2O in 120–180 ms. (Fant *et al.*, 2000).

These data have a certain degree of variability but still provide a starting point to specify a time track of subglottal pressure in a model of speech production, but in specifying that track it is unclear how such values as the pressure and time at phonation onset might be related to the pressure and time at which the relatively constant pressure for speech is reached. Guidelines to aid the researcher in specifying the details of a specific subglottal pressure trace in relation to a specific utterance are not yet fully described.

Because prosodic prominences within an utterance have been linked to local subglottal pressure changes (e.g., Ladefoged, 1967; Lieberman, 1967; Netsell, 1969; Atkinson, 1973), it may be that the first perceived prominence in the utterance is correlated to the initial pressure peak. Another possibility is that the pressure rise is unrelated to the first prominent syllable and is guided solely by the physical requirement of generating a pressure that is in the proper range for speech as quickly as possible while managing the recoil forces of the respiratory system. One aim of this paper is to address this question. A second aim of this paper is to expand the data regarding the influence of the first sound segment in the utterance on the details of the pressure rise. Pressure is built up behind an impedance to air flow from the lungs. Different impedances are associated with different sound segments, and it is unclear how the resulting variations in the pressure rise are related, if at all, to the initial peak in pressure. This paper reports results across a range of initial sound segments that compare phonation onset pressure and timing to peak pressure and timing.

Data on timing and pressure landmarks for phonation onset and initial pressure peak can provide another step in

a)Electronic mail: slifka@mit.edu

our ability to specify subglottal pressure traces in speech production models. Additionally, these data can improve our understanding of the predictability of prosodic cues at utterance initiation. If an initial portion of the speech signal is produced during rising subglottal pressure, the signal will have amplitude and spectral changes correlated to the pressure rise. These acoustic changes are part of the set of acoustic cues to prosodic boundaries.

II. METHODS

A. Acoustic and physiological data acquisition

Simultaneous recordings of the acoustic signal, airflow, and physiological signals used to estimate lung volume and lung pressure were made at the Voice and Speech Laboratory of the Massachusetts Eye and Ear Infirmary for four native speakers of American English. While a speaker is capable of a wide range of actions with the respiratory system, vocal folds, and articulators, normal muscular habits during speech are the focus of the present research. No special requests were made of the speakers regarding lung volume, syllable or word prominence, or the loudness of the speech. The subset of read speech tasks used for this study consists of simple declarative utterances and short paragraphs.

This study aimed to include two male and two female subjects, but recordings were made from a larger number of subjects. Based on preliminary studies and evidence in the literature, it was expected that some subjects would need to be excluded from the analysis stage. Two subjects were excluded because a number of their utterances were corrupted by evidence of a low amplitude peristaltic wave or spasm occurring near utterance termination. Two other subjects were excluded due to inconsistencies in lung volume estimation parameters across the duration of the acquisition session. Four subjects were obtained with consistent calibrations, and these four subjects range in age from 21 to 28, have normal hearing, have no history of respiratory ailments, are nonsmokers, and speak American English as their first language. In this document, the speakers are referred to as subject 4 through subject 7 for consistency with an existing publication (Slifka, 2000). All subjects signed a consent form and were compensated for their participation.

B. Measurement methods

Airflow estimates were made using a circumferentially vented pneumotachograph mask (Rothenberg, 1973) from Glottal Enterprises, Inc. Mounted to the exterior of this mask was a pressure-sensitive microphone (Sony Co.) located approximately 15 cm from the speaker's mouth and used to record the acoustic signal. Intraoral pressure was measured, during some calibration tasks, with a small open tube placed in the subject's mouth and connected at the other end to a pressure transducer.

Lung volume is controlled by the movements of two essentially independent structures: the ribcage and the abdomen (Konno and Mead, 1967). Most people change lung volume through some combination of movement of the ribcage and movement of the abdomen. The cross-sectional areas of the ribcage and abdomen were estimated using a calibrated respiratory inductive plethysmography system, after accounting for linear drift (Respitrace, Ambulatory Monitoring, Alpsley, NY). These cross-sectional areas can be mapped to lung volume estimates (Konno and Mead, 1967). In order for this method to produce consistent results, any changes in ribcage and diaphragm cross-sectional areas must arise from lung volume changes and not from movement due to postural changes in the subject. Each subject was seated in a chair with foot and head supports and cautioned to maintain the same posture throughout the recording process.

Subglottal pressure, the drive for sound sources in speech, can be considered to be equal to the air pressure in the lungs, alveolar pressure (P_{ALV}), when the airflow rate is not high enough to cause a significant drop across the airway resistance. High rates of flow generally occur during speech only for some aspirated consonants. P_{ALV} is derived in this research using a measurement of pleural pressure (P_{PL}). P_{ALV} is the sum of the pleural pressure (P_{PL}) and the static recoil pressure of the lungs (P_L) as given in Eq. (1):

$$P_{ALV} = P_{PL} + P_L . \tag{1}$$

The volume-dependent recoil pressure of the lungs, P_L, is measured and accounted for in the calibration process in order to estimate P_{ALV} from a measurement of P_{PL} (as in Kunze, 1964).

Pleural pressure is measured using an esophageal balloon (Van den Berg, 1956). A thin latex balloon was passed through the nasal cavity and placed at a level below the trachea and above the diaphragm, and was inflated with at least 0.5 cm^3 of air. Placement with respect to whether the tip of the balloon had crossed the diaphragm was checked by having the subject make inspiratory and expiratory efforts against an occlusion, as in Baydur *et al.* (1982). During the speech tasks, the distal end of the esophageal catheter was connected to a differential pressure transducer (Glottal Enterprises, Inc.) and the other port of the transducer was left open to atmospheric pressure. To characterize the frequency response of the esophageal balloon system, the air pressure in a chamber was varied, using a syringe, for frequencies up to 6 Hz. The air pressure in the chamber was measured using both a pressure transducer directly as well as the esophageal balloon measurement system with its own associated pressure transducer. The magnitude response is flat across the 6 Hz range and the time delay is 2–3 ms.

Calibration maneuvers were performed before and after the speech task segment of the experiment. Each data channel was digitized on-line at a sample rate of 10 kHz using the Axoscope (Axon Instruments, Inc.) recording software and Digidata digitizers. Signals were filtered to prevent aliasing, and were amplified using a Cyberamp 380 Programmable Signal Conditioner (Axon Instruments) prior to digitizing.

C. Calibration procedures and results

Voltage levels from all transducers were correlated to known pressure, flow, and amplitude values. Pressure transducers were calibrated using a water manometer. Airflow was calibrated using a compressed air source and an airflow meter. Sound pressure level (SPL) was calibrated using an electro-larynx (Cooper-Rand), which produces a wideband

signal, held at two different locations from the microphone. An SPL meter (Brüel & Kjær) was held next to the microphone. The recorded voltage values from the microphone were converted to root-mean square (rms) values and correlated to the SPL values.

Lung volume values are expressed relative to the functional residual capacity (FRC). Vital capacity (VC) is the volume change between the maximally stretched and maximally compressed lungs/chest wall system. FRC is generally in the range of 30%–40% VC. Generally, in normal resting breathing in an upright person, exhalation ends at FRC. Lung volume estimation involved the following calibration maneuvers performed by the subject: (1) resting breathing, (2) isovolume maneuvers, and (3) spirometer maneuvers. First, the ribcage and abdomen voltage traces for resting breathing were examined to determine approximate voltage values for FRC under the assumption that, during resting breathing, an exhalation ends at FRC. All abdomen and ribcage voltage traces were scaled such that zero volts corresponded to the FRC estimate. Isovolume maneuvers, performed near FRC, were used to determine the scale factors for ribcage and abdomen voltage traces. The gain of the ribcage signal was adjusted so that the values for the ribcage voltage and the abdomen voltage were the same for that fixed volume. The two signals were then summed and converted to liters using values collected when the subject exhaled into a spirometer. These final calibrated lung volume curves are displayed as volume in liters referenced to FRC, i.e., the difference between the volume and FRC. The forced vital capacity for each subject, as measured with a spirometer, is (in liters) subject 4: 3.5, subject 5: 3.1, subject 6: 4.3, and subject 7: 4.1. Subjects 4 and 7 are male, and subjects 5 and 6 are female.

For additional verification of the calibration process, lung volume estimated from the respitrace signals was compared to the time integral of low-pass filtered airflow. As summed across all isolated read utterances, the average difference between these two methods of estimating lung volume was less than 0.1 l.

To measure the static recoil pressure of the lungs (P_L), one port of the differential pressure transducer measured pleural pressure (P_{PL}) and the other port measured oral pressure through a mouthpiece. Oral pressure measured with the glottis spread is a good estimate of alveolar pressure (P_{ALV}). The mouthpiece has an opening to the atmosphere that could be occluded by the experimenter. The subject was asked to inhale a large breath, and to exhale slowly without closing down the glottis. During the exhalation, the mouthpiece was occluded for brief periods of time. During those occlusions, the differential transducer measured $P_{PL}-P_{ALV}$ or $-P_L$. All calibrated alveolar pressure data were derived from pleural pressure by compensating for the static recoil pressure of the lungs for each speaker, i.e., $P_{ALV}\cong P_{PL}+P_L$.

For verification of the calibration process, the pressure difference was measured between the calibrated alveolar pressure and the intraoral pressure while the subjects produced a continuous series of /pæ/ syllables. In this task, the subject is asked to essentially produce a constant loudness /æ/ with interruptions by /p/ consonants. During the /p/ closure, a short tube connects the air inside the subject's mouth to a pressure transducer. When the subject performs the task well, the intraoral pressure during the closure for a voiceless stop consonant has been shown to be within 1 cm H_2O of the subglottal pressure as measured with a tracheal puncture (Hertegård *et al.*, 1995). In this way, the two separate measures of alveolar pressure can be compared for consistency. A typical example of the comparison between the measured intraoral pressure and the estimated P_{ALV} is shown in Fig. 1. All subjects show close agreement between the intraoral pressure measure and the alveolar pressure estimate during the /p/-closures.

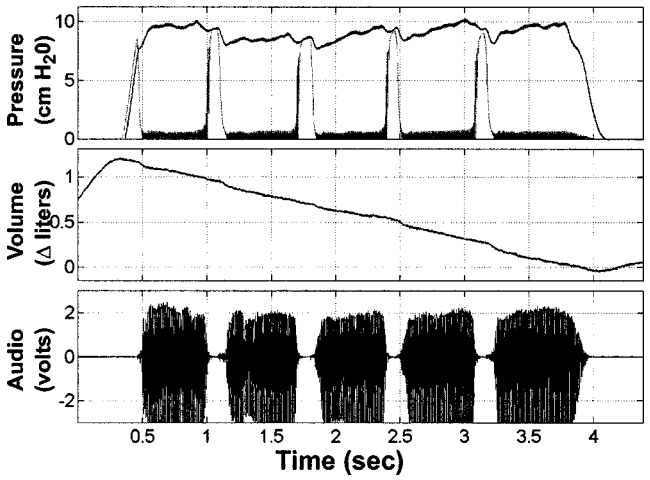


FIG. 1. Example of measured intraoral pressure and estimated alveolar pressure during production of a series of /p æ/ syllables for subject 4. The top panel shows alveolar pressure (dark line) and intraoral pressure (gray line). The middle panel is lung volume and the bottom panel is audio. During the /p/ closure, alveolar pressure as estimated using an esophageal balloon system agrees with intraoral pressure as measured with a pressure transducer.

Of the acquired utterances, only those which were initiated at the start of an exhalation and terminated at the end of an exhalation were used in the analysis, i.e., all analyzed utterances were read on a single breath. Pressure and air flow for each utterance in the data set were checked at the point where flow crossed from negative to positive. At this point, the pressure in the lungs should be zero. Utterances where the magnitude of this difference was greater than 1.5 cm H_2O were excluded from analysis. Evidence of any disfluency during the reading of an utterance caused the utterance to be excluded. Exclusions for all of these reasons across all four subjects amounted to 16%. For the four subjects combined, the total number of utterances is 241.

A subset of the utterances (19.8%) were read as part of a short paragraph, and the remainder of the utterances (80.2%) were not part of a paragraph but were read as an isolated utterance. The distribution of the onset segments is given in Table I. The onset segments are roughly split between vowels (44%) and consonants (56%). In the vowel set, roughly half (45%) are either the reduced vowel schwa as in "*a*ssume" or the lax and often reduced vowel /ɪ/ as in "*i*f." The consonant set is made up of primarily stops (23%), voiceless fricatives (39%), and the voiced fricative /ð/ as in "that" (26%). The remaining 12% are glidelike consonants (/j/ and /h/). The mean utterance duration is 2.27 s. The shortest utterance duration in 0.90 s, and the longest utterance duration

TABLE I. Segment percentages for utterance onset (241 utterances total).

Utterance onset			
Vowel onset	44%	Consonant onset	56%
Segment percentages within vowel onsets		Segment percentages within consonant onsets	
Reduced vowel (schwa as in ''*a*ssume'') or short vowel /ɪ/ as in '*i*f')	45%	Stops (voiced and voiceless)	23%
All other vowels	55%	Voiceless fricatives	39%
		/ð/ as in '*t*hat'	26%
		/h/ as in '*h*at'	7%
		/j/ as in '*y*ou'	5%

is 4.21 s. The number of words in an utterance ranged from 4 to 17.

D. Perceptual ratings of prominences

A prominence rating experiment was conducted with the goal of determining the initial and final perceived prominent syllables in the acquired utterance set. This rating test is a modified version of a rating-based protocol used by Fant and Kruckenberg (1989). It was carried out to gain general and essentially binary information about whether the boundary syllables in the utterances were prominent or not. Listeners were instructed that they would hear a series of utterances and they should make a judgment about the prominence of the syllables in those utterances. They were asked to think of the prominences as being on a 0–30 scale, where prominent syllables are in the upper half of the range and all other syllables are in the lower half of the range. For each utterance, the listeners marked an ''*X*'' through the first syllable in the utterance that would fall in the prominent range and marked an ''*X*'' through the last syllable in the utterance that would fall in the prominent range. They were not asked to assign a numerical rating to the prominences. Each utterance could be replayed as often as the subject needed.

In the utterance set to be rated, 12 utterances were repeated twice (three from each speaker). These utterances and the ratings given to them were used as a consistency check on the stability of each listener's criteria for completing the task. Five subjects participated in this test where all were graduate students or engineering professionals and all reported normal hearing. These subjects had consistency measures ranging from 8/12 to 12/12 with an average value of 10/12.

A simple majority of 3/5 votes was sufficient to label a syllable as prominent. Of the total 241 utterances in the data set, 11 onsets received less than a majority agreement (8 of those were utterances from subject 7) and were removed from the data set. For the remaining 230 utterances, the average number of votes for syllables judged as prominent was 4.35 (0.74 standard deviation). In 84% of the utterances, a 4/5 or a 5/5 vote decided the prominent syllable. In later discussions, these ratings are considered in terms of three groups: (1) utterances where the perceived prominence is on the first syllable, (2) utterances where the perceived prominence is on the second syllable, and (3) utterances where the perceived prominence was on the third, fourth, or fifth syllable. The utterances that received a 3/5 rating made up 9% of the utterances from groups 1 and 2 combined and made up 37% of the utterances in group 3. This result would suggest that there was more uncertainty in the listening panel when placing the initial perceived prominence later into the utterance.

Times were marked at the beginning and end of the first prominent syllable in the utterance. These times were determined through visual inspection of the audio signal and the corresponding spectrogram as well as through listening to the audio signal. These times were marked to allow for analysis regarding possible correlation to the time of the initial peak in alveolar pressure.

E. Utterances and acoustic landmarks

Several timing landmarks in the acoustic and pressure signals were defined for the initiation of utterances. Times were labeled for zero flow, start of the utterance, onset of phonation, and initial alveolar pressure peak, as shown in Fig. 2 for an utterance from subject 5. The vertical lines correspond to (1) start time, start pressure; (2) phonation onset time, phonation pressure; and (3) peak time, peak pressure. The utterance begins with the word ''say.'' The /s/ begins near 1 cm H_2O in pressure, phonation begins at a pressure of 7.8 cm H_2O, and the peak in pressure is 9.4 cm H_2O.

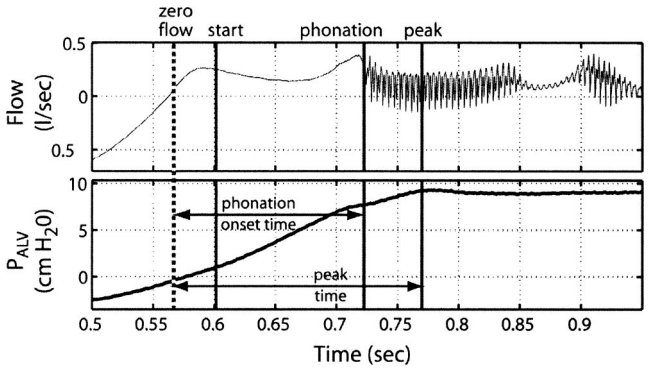


FIG. 2. Initiation landmarks. Vertical bars mark zero flow, start of the utterance, phonation onset, and initial pressure peak. Start of the utterance and phonation onset are timing landmarks determined from the acoustics and flow signals. Start pressure and phonation pressure are the alveolar pressures corresponding to the utterance onset and phonation onset landmarks, respectively. Phonation onset time is defined as the time from the zero flow point until phonation onset. Peak pressure is the initial peak in alveolar pressure. Peak time is defined as the time from zero flow to the first pressure peak. The signals shown are from subject 5. Panels plot airflow (l/s) and alveolar pressure (cm H_2O). The utterance begins with the words ''Say the.''

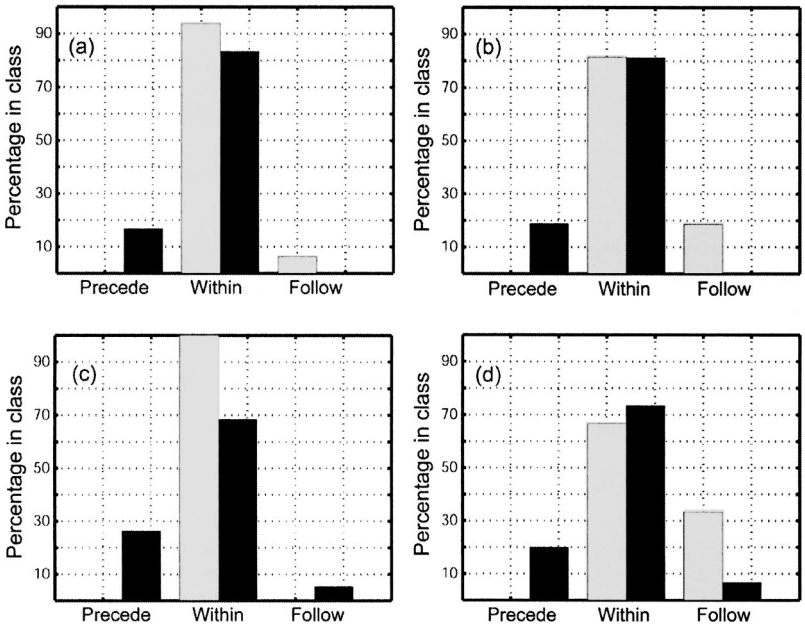


FIG. 3. Comparison of pressure peak time to prominent syllable boundaries. The timing of the initial pressure peak either preceded, fell within, or followed the boundaries of the initial prominent syllable. That alignment is represented on the *x*-axis. Two groupings of perceived prominent syllables are shown for each timing alignment. Light bars represent utterances where the initial rated prominence in on the first syllable. Black bars represent utterances where the initial rated prominence is on the second syllable: (a) subject 4, (b) subject 5, (c) subject 6, and (d) subject 7.

The start of the utterance was determined from the first acoustic evidence of sound—whether that sound was noise-like or periodic. Start pressure refers to the alveolar pressure at the start of the utterance. Onset of phonation was determined from the first acoustic indication of vocal fold vibration. For utterances beginning with a voiced phonetic segment, the start of the utterance is the same as the onset of phonation. For voiced stop onsets such as /b/, the start of the utterance was set as the burst release as there was no evidence of voicing during the /b/ closure. Phonation pressure refers to the alveolar pressure at the onset of phonation, and phonation onset time is the time between the zero flow instance and the onset of phonation. The first peak in pressure was determined automatically from an algorithm which low-pass filtered the pressure to a cutoff of 100 Hz with an FIR filter to reduce measurement variations, and found the first occurrence of two consecutive points where the slope in the pressure curve was less than or equal to zero. The initial pressure peak was set at the first of those two points. The term peak time is used to refer to the time between the zero flow instance and the time of the initial pressure peak.

III. RESULTS

A. Rated prominences

This section provides baseline data for the defined set of acoustic and physiological landmarks in relation to the first perceived prominent syllable. Figure 3 considers two groups of utterances: (1) those utterances where the initial rated prominence falls in the first syllable (light bars, 117 utterances) and (2) those utterances where the initial rated prominence falls in the second syllable of the utterance (black bars, 68 utterances). For each group of utterances, the timing of the pressure peak either preceded, fell within, or followed the limits of the perceptually prominent syllable. The pressure peak occurs in the initial prominent syllable for 100 out of the 117 utterances in group 1 and for 52 out of the 68 utterances in group 2. For these two groups of utterances, the pressure peaked in the initial prominent syllable in an average of 82% of the utterances (152 out of 185 utterances). When the second syllable in the utterance was rated as prominent and the pressure peak did not occur within that syllable, it fell within the first syllable in 14 out of the 16 utterances. For the 45 utterances with initial prominences in the third, fourth, or fifth syllable, the pressure peak preceded that syllable in an average of 91% of the utterances. As stated earlier, these prominences received less consistent ratings from the listening panel.

The overall mean peak pressure is 9.1 cm H_2O with a standard deviation of 2.4 cm H_2O, and the overall mean peak time is 243 ms with a standard deviation of 74 ms. Table II addresses variation in peak time and peak pressure values on a speaker-dependent basis for the utterances as grouped by the location of the initial prominent syllable as well as for each speaker overall. These data are given as mean and standard deviation. The shortest mean peak time is 190 ms with a standard deviation of 60 ms. A peak time at two standard

TABLE II. Peak pressure and timing values for the syllable prominence categories and for each speaker overall. Times are given in milliseconds referenced to the time at which airflow is zero and pressure is given in cm H_2O. Values are given as ''mean (standard deviation).''

	Subject 4	Subject 5	Subject 6	Subject 7
	Group 1: First syllable is prominent			
Peak time	230 (60)	210 (60)	240 (50)	260 (70)
Peak P_{ALV}	8.21 (1.00)	11.51 (1.64)	9.51 (1.14)	6.23 (0.93)
	Group 2: Second syllable is prominent			
Peak time	260 (70)	190 (60)	280 (80)	280 (90)
Peak P_{ALV}	8.01 (0.75)	12.10 (1.96)	10.68 (1.03)	6.98 (0.82)
	Group 3: Later syllable is prominent			
Peak time	240 (50)	210 (60)	260 (60)	310 (90)
Peak P_{ALV}	8.59 (1.50)	12.53 (1.94)	9.64 (0.52)	6.63 (0.85)
	Overall			
Peak time	237 (65)	203 (59)	256 (68)	278 (83)
Peak P_{ALV}	8.25 (1.09)	11.92 (1.82)	9.94 (1.16)	6.52 (0.93)

TABLE III. Pressure and timing values corresponding to syllable prominence ratings for the initial prominent syllable in the utterance. Values are listed as "mean (standard deviation)." Time is measured in milliseconds and pressure in cm H_2O.

Quantity	Subject 4	Subject 5	Subject 6	Subject 7
	Group 1: First syllable is prominent			
Phonation onset time	145 (71)	118 (51)	170 (64)	150 (84)
Start P_{ALV}	3.46 (1.66)	3.72 (2.64)	4.31 (2.57)	2.07 (1.47)
Phonation P_{ALV}	5.83 (2.07)	8.07 (2.84)	7.60 (2.04)	3.84 (1.53)
	Group 2: Second syllable is prominent			
Phonation onset time	93 (24)	65 (21)	151 (71)	120 (55)
Start P_{ALV}	2.82 (0.99)	3.60 (1.63)	5.20 (1.97)	2.99 (1.30)
Phonation P_{ALV}	4.41 (0.76)	6.07 (2.43)	6.76 (1.24)	3.84 (1.13)

deviations below this mean is 70 ms which would correspond to a 3.6-Hz sinusoid. This value falls well within the response limits of the measurement system (flat magnitude response to at least 6 Hz).

In comparing the data for groups 1 and 2 in Table II, the suggestion is that the average peak time for later prominences is slightly delayed, but only on the order of 20–50 ms. Two-sample *t*-tests were run with a null hypothesis that the means are equal and a significance level of 0.05 (alpha). Between groups 1 and 2 for the parameter of peak time, the *t*-test indicates that these means come from the same population for three of the four speakers with the exception, speaker 6, having a *p*-value of 0.03 (subject 4: $p=0.13$, $t=-1.53$, df=48, subject 5: $p=0.23$, $t=1.22$, df=41, subject 6: $p=0.03$, $t=-2.24$, df=45, subject 7: $p=0.53$, $t=-0.63$, df=43). A *t*-test between the two groups for the parameter of peak pressure is less clear. For subjects 4 and 5, the *t*-test indicates that the means come from the same population and for subjects 6 and 7, the *t*-test indicates that the means come from different populations (subject 4: $p=0.45$, $t=0.76$, df=48, subject 5: $p=0.29$, $t=-1.07$, df=41, subject 6: $p<0.01$, $t=-3.58$, df=45, subject 7: $p=0.01$, $t=-2.63$, df=43).

From the overall average given at the bottom of Table II, some interspeaker variability is evident. The speaker with the highest mean peak pressure, subject 5, has the highest standard deviation in peak pressure and the lowest mean peak time. The speaker with the lowest overall mean peak pressure, subject 7, has the smallest standard deviation in peak pressure and the longest mean peak time.

Mean and standard deviation data for the other physiological landmarks in this study, phonation onset time, start P_{ALV} and phonation P_{ALV}, are given in Table III for the utterances in groups 1 and 2. All four subjects appear to show a trend for average phonation onset time to shorten as the initial prominence is moved later into the utterance. However, this trend is only statistically significant for two of the four subjects as given by a two-sample *t*-test (subject 4: $p<0.01$, $t=2.96$, df=48, subject 5: $p<0.01$, $t=3.93$, df=41, subject 6: $p=0.34$, $t=0.96$, df=45, subject 7: $p=0.21$, $t=1.27$, df=43). All subjects also show a trend for average phonation pressure to decrease or remain the same as the initial prominence is moved later into the utterance, but this trend is only statistically significant for the same two speakers that showed a significant trend in phonation onset time. (For a two-sample *t*-test: subject 4: $p<0.01$, $t=2.79$, df=48, subject 5: $p=0.02$, $t=2.35$, df=41, subject 6: $p=0.11$, $t=1.60$, df=45, subject 7: $p=0.98$, $t=0.03$, df=43.) These results indicate that location of the initial perceived prominence is not a primary determinant of these physiological landmarks for all speakers. The nature of the pressure rise and the timing of the acoustic landmarks will also be influenced by the sound segment at the boundary and the corresponding impedance to air flow. The next section presents measurements that focus on this influence.

B. Initial sound segment type

This section provides data for the defined set of acoustic and physiological landmarks in relation to the initial sound segment in the utterance and compares phonation onset to the initial alveolar pressure peak. Measurements at these acoustic and physiological landmarks are given in Table IV, grouped according to the initial sound segment of the utterance into vowels plus /j/, stop consonants, or noiselike consonants. In this task, the noiselike consonants are voiceless fricatives, /ð/, and /h/.

The noiselike consonant onsets have a lower range of average start pressures (on the order of 2–3 cm H_2O) and the stop consonant onsets have a higher range of average start pressures (on the order of 3–5 cm H_2O). For the vowel+/j/ grouping, start pressure is the same as phonation onset pressure (on the order of 3–6 cm H_2O, on average).

In general, the trend is for a longer phonation onset time to coincide with a higher phonation pressure. Mean phonation onset time and mean phonation pressure both increase from vowel/voiced sonorant onsets to stop consonant onsets and to noiselike consonant onsets. Subject 7 is the exception in that mean phonation pressure for stop-consonant and noiselike-consonant onsets are nearly the same. The noiselike consonants include the segment /ð/, which is sometimes produced in a stoplike manner (Glass, 1988). The data from Table IV show that each subject has a different range of lung volumes at utterance initiation. However, for a given subject, there is little difference in mean lung volume at utterance initiation across the three groupings of initial sound segments (<3% VC).

Figure 4(a) is a scatter plot of phonation pressure versus phonation onset time for all four speakers. Data are grouped into three classes: vowel segments and /j/ (circles "○"), stop

TABLE IV. Pressure and timing values corresponding to acoustic and physiological landmarks. Time is given in milliseconds, pressure in cm H_2O, and lung volume (ΔLV) given as the change in volume in %VC from FRC.

Quantity	Subject 4	Subject 5	Subject 6	Subject 7
		Vowels and voiced sonorant onsets		
Phonation onset time	80 (30)	50 (20)	130 (70)	110 (70)
Phonation P_{ALV}	4.11 (1.19)	4.34 (2.31)	5.80 (1.61)	3.26 (1.30)
ΔLV at utterance start	29.4 (3.3)	31.6 (7.4)	15.3 (2.8)	12.9 (6.7)
		Stop consonant onset		
Start P_{ALV}	4.42 (1.51)	3.00 (1.16)	4.92 (1.60)	2.83 (2.46)
Phonation onset time	120 (60)	100 (50)	130 (40)	160 (80)
Phonation P_{ALV}	5.62 (2.31)	6.81 (1.23)	6.69 (1.11)	4.46 (1.86)
ΔLV at utterance start	30.1 (3.1)	29.0 (5.8)	17.0 (2.7)	14.8 (6.6)
		Noiselike consonant onsets		
Start P_{ALV}	2.23 (0.85)	2.67 (1.92)	3.38 (2.32)	1.72 (1.17)
Phonation onset time	160 (60)	120 (50)	180 (60)	180 (70)
Phonation P_{ALV}	5.98 (1.74)	8.69 (2.72)	8.06 (1.88)	4.36 (1.23)
ΔLV at utterance start	29.6 (5.5)	31.7 (8.8)	18.5 (4.0)	13.1 (6.7)

consonants and /ð/ (diamonds "♦"), and voiceless fricatives (crosses "×"). Figure 4(b) models the data points in each segment group as a bivariate Gaussian probability distribution and plots a contour line that contains 70% of the data points in each group. This figure is presented as a tool to visualize the concentrations given in the scatter plot of Fig. 4(a). In these plots, there is considerable spread in the distributions as well as a substantial degree of overlap. The basic trend, as already observed from the discussion of the mean and standard deviation data, is that vowels following voiceless fricatives show the highest phonation pressure and longest phonation onset time, while the stops and /ð/ and the vowels and /j/ groupings have phonation parameters that are predominantly lower in pressure and shorter in time. However, it is difficult at this point to see how to apply these data to a model for speech production. For example, if the pressure peaks to a mean background level of 8 cm H_2O, is there any reason to predict a phonation pressure of 4 cm H_2O rather than one of 7 cm H_2O? The connection to information about the overall pressure trajectory as it rises to the mean background level is not well quantified.

Figures 4(c) and (d) replicate the data from Figs. 4(a) and (b) but the data are normalized by the peak pressure and peak time in each utterance. With this normalization, phonation pressure equals peak pressure at a value of one on the ordinate, and phonation onset time equals peak time at a value of one on the abscissa. There are three points on the normalized graph that the pressure trajectory must pass

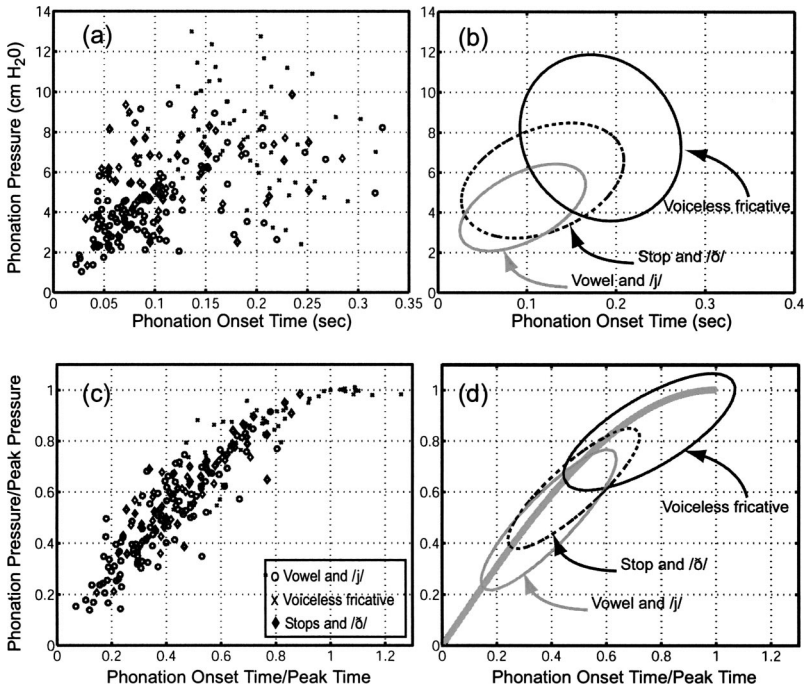


FIG. 4. (a) Scatter plot of the phonation pressure and phonation onset time data for all four speakers. Data are grouped into three classes: vowel segments and /j/ (circles "○"), stop consonants and /ð/ (diamonds "♦"), and voiceless fricatives (crosses "×"). (b) Models the data points in each segment group as a bivariate Gaussian probability distribution as plotted in a contour line that contains 70% of the data points in each group. (c) and (d) replicate the data from (a) and (b) after normalization to peak time and peak pressure for each utterance.

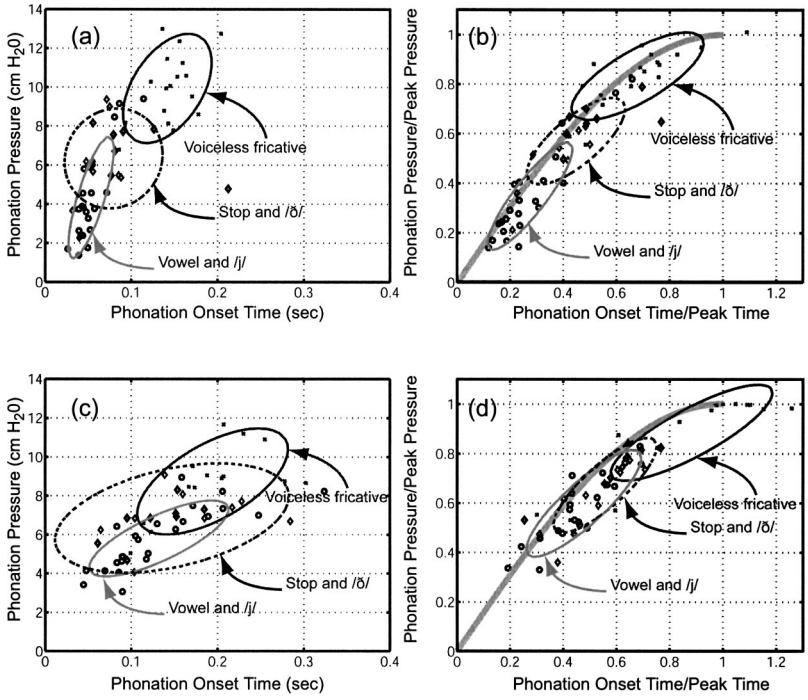


FIG. 5. (a) and (c) are scatter plots of phonation pressure and phonation onset time data for three groupings: vowel segments and /j/ (circles "○"), stop consonants and /ð/ (diamonds "◆"), and voiceless fricatives (crosses "×"). (a) Speaker 5 and (c) speaker 6. (b) and (d) are the results of modeling the normalized data points in each segment group as a bivariate Gaussian probability distribution and plotting a contour line that contains 70% of the data points in each group. (b) Speaker 5 and (d) speaker 6.

through for an utterance. One is the point (0,0). The measured data provide the other two points along the trajectory: the pressure must pass through (phonation onset time/peak time, phonation pressure/peak pressure) and (1,1), which is the initial pressure peak. For the purpose of discussion and visualization, a quarter-period of a sine wave is shaded in gray behind the distributions in Fig. 4(d). The use of this sine-wave does not contain any theoretical implication nor is the normalized phonation pressure used in deriving the curve; it is simply provided for orientation to a general trajectory for a rise in pressure.

In the normalized distributions, phonation onset tends to occur in different regions for different initial sound segment groupings. The vowel and /j/ onsets occur in the lower left quadrant and the voiceless fricative onsets occur primarily in the upper right quadrant with more speaker dependency evident in the grouping of stops and /ð/. Across all speakers, an utterance beginning with a vowel or /j/ has a phonation onset pressure that is generally 30%–70% of the eventual pressure peak. For the voiceless fricative onset segment, the pressure rise occurs predominantly during the frication noise, and the phonation pressure is much closer to the initial peak pressure (60%–100%). The other class of segments (stops+/ð/) falls in the middle range.

The distribution for each group of segments can be modeled on a speaker-dependent basis. Speakers in this study, which focuses on speaker habits, show two types of behavior—substantial overlap of the distributions for the segment groups of vowel+/j/ and stops+/ð/, or relatively larger spacing between the three groups. Examples of each case are given in Figs. 5(a) and (b) for speaker 5 and Figs. 5(c) and (d) for speaker 6. Both are female speakers. Speaker 5, as seen in Fig. 5(b), shows three regions with less overlap between them. The lowest region in normalized time and normalized pressure is the vowels+/j/ group. The highest region in normalized time and pressure is the voiceless fricative group. This region has a distribution that does not exceed 0.9 in normalized phonation time. Two-sample *t*-tests were run with the null hypothesis that the means are the same and a significance level of 0.05 (alpha). Between the vowels+/j/ group and the stop+/ð/ group, the *t*-tests indicate that both the time and pressure values arise from different populations (normalized phonation onset time: $p<0.01$, $t=-3.69$, df=36, normalized phonation pressure: $p<0.01$, $t=-3.98$, df=36). Between the stop+/ð/ group and the voiceless fricative group, the *t*-tests indicate that both the time and pressure values also arise from different populations (normalized phonation onset time: $p<0.01$, $t=-7.66$, df=29, normalized phonation pressure: $p<0.01$, $t=-4.99$, df=29). In contrast, speaker 6 [Fig. 5(d)] has a voiceless fricative group that encompasses the peak point of (1,1). For those cases where the normalized phonation onset time is greater than one, the pressure reached an initial peak prior to the onset of phonation. The vowel+/j/ and stop+/ð/ groupings occupy more of the central region near 0.4–0.6 in normalized time and 0.5–0.7 in normalized pressure. The two-sample *t*-test between the vowel+/j/ group and stop+/ð/ group for this speaker yields a *p*-value of 0.08 ($t=-1.77$, df=39) for normalized phonation onset time and a *p*-value of 0.11 ($t=-1.63$, df=39) for normalized phonation pressure indicating that these populations are not as well separated. When comparing the stop+/ð/ and voiceless fricative groups, the *t*-test indicates that these populations are separate (normalized phonation onset time: $p<0.01$, $t=-7.47$, df=26, normalized phonation onset pressure: $p<0.01$, $t=-6.93$, df=26).

IV. DISCUSSION

The data show a correlation between the timing of the initial pressure peak and the location of the first perceived prominence in the utterance, in that, if the prominence oc-

curred on the first syllable, pressure generally peaked in the first syllable. If the prominence occurred in the second syllable, pressure generally peaked in the second syllable. However, there appear to be limits to the degree to which the peak is delayed into the utterance. If the initial perceived prominence occurs later in the utterance, the pressure peak generally precedes the initial prominent syllable. Additionally, the time to the initial peak is not clearly different for the utterances in groups 1 and 2. When the initial prominence is placed after the first syllable of the utterance, the speaker may vary the peak time to place the peak within the prominent syllable, may change the realization of the initial syllable at the articulators (such as a shortened segment realization or a reduced form of a vowel at utterance onset), or may use some combination of those strategies. It may be that, for initial syllables that are not prominent, the realization of the initial syllable is compressed in some manner. This question would benefit from further study.

Another factor in reaching the initial pressure peak is the dynamics of the respiratory system. In creating the subglottal pressure for speech, a speaker manages the recoil forces of the respiratory system in a fairly rapid transition from inhalation to exhalation where sound generally begins during the transition from net inspiratory force to net expiratory force (Slifka, in press). Variability in such parameters as peak time and peak pressure is also dependent on each speaker's respiratory system characteristics and the manner in which the recoil forces are managed. It was noted that subject 5 has the highest mean peak pressure, the largest standard deviation in peak pressure, and the shortest mean peak time. Essentially, this is a speaker that rapidly reaches a high peak pressure with a good deal of variability in reaching the peak. Subject 7 has the lowest mean peak pressure, the smallest standard deviation in peak pressure, and the longest mean peak time. This type of observed variability is consistent with studies of muscular control which conclude that rapid movement to a target results in more variation around that target than slow movement to a target. This principle is embodied in Fitt's law which states that there is a speed-accuracy trade-off for a wide range of target-directed movements (Fitts, 1954; Fitts and Peterson, 1964).

Data were also collected comparing phonation parameters to the location of the initial perceived prominence. To create a pressure in the lungs, there must be a resistance to the airflow. At the start of the utterance, this resistance changes as the articulators and vocal folds move into position to create the initial sound of the utterance. The different impedances associated with different sound segments lead to variations in the manner in which the pressure rises and phonation begins. The mean phonation pressures for the speakers in this study range between 3.8 and 6.9 cm H_2O and mean phonation onset times range between 89 and 152 ms across all utterances. Each subject appears to have a different preferred range for initiating phonation but all show a standard deviation on the order of 2 cm H_2O for the utterances in group 1. By comparison, for the utterances in group 2, three of the four subjects show a reduction in mean phonation pressure on the order of 1–2 cm H_2O as well as a reduction in the standard deviation. The fourth subject shows little difference between the two conditions, but is also the subject that has the lowest mean phonation pressure (on the order of 4 cm H_2O). As such, there is less room to reduce the pressure for this subject. Half of the subjects showed a statistically significant trend in the decrease in phonation onset time and phonation pressure between groups 1 and 2. The other two subjects did not. These data indicate that such a reduction is a weak trend and subject to interspeaker variability.

The parameters of phonation onset time and phonation pressure were also examined as grouped according to the sound segment at the start of the utterance. The trend for a longer phonation onset time to coincide with a higher phonation onset pressure agrees with the results of Lieberman (1967) for three male subjects and 19 total utterances (p. 95) and those of Atkinson (1973) in a study using a single male speaker (18 utterances). In the Atkinson study, the speech task was slightly different in that the subject was instructed to place emphasis on a particular word in the utterance. In the present study, there was no such instruction. Atkinson reported mean phonation onset times for four different onset segments: /b/: 290 ms, /ð/: 180 ms, /w/: 190 ms, and /h/: 105 ms. The values for /ð/, /w/, and /h/ are on the order of the results in the present study. The value for the stop consonant /b/ is longer, but this may be a result of the task in the Atkinson data collection. All of the utterances that began with a /b/ (eight total) were ''Bev loves Bob'' where two of these began with an explicit emphasis an in ''BEV loves Bob.'' There were no voiceless fricatives in the study.

Phonation pressure depends on vocal fold mass and stiffness, and the degree of adduction of the vocal folds, among other factors [for example, Titze (1988) and Lucero (1998)]. In normal speech, the data show a range of phonation pressures. Consider an example with two repetitions of an utterance that begins with a vowel, where utterance 1 has a peak pressure of 12 cm H_2O and utterance 2 has a peak pressure of 9 cm H_2O. These values are reasonable for a speaker such as subject 5 who uses a peak pressure with a mean of 11.9 cm H_2O and standard deviation of 1.8 cm H_2O in this data set. The normalized groupings for a speaker such as subject 5 would suggest that phonation begins near 30% of the eventual pressure peak. In this example, utterance 1 would begin phonation near 3.6 cm H_2O and utterance 2 near 2.7 cm H_2O. For each utterance, the speaker uses a particular configuration of laryngeal adjustments and respiratory actions to reach the peak pressure. To reach a peak pressure of 11.9 cm H_2O, the vocal folds may be stiffer or more closely adducted. The fact that the phonation pressure is higher for this peak pressure than for that of 9 cm H_2O peak is evidence for the appearance of that particular laryngeal configuration early in the utterance. A speaker may assume a posture for speech in which some parameters that reflect a later target are set early in the process.

An eventual application of the data in this study is the development of a more detailed model of the respiratory system actions at the start of an utterance. From these data on the relation between timing landmarks at utterance initiation, a basic prototype can be proposed. Figure 6 is a prototype for normalized phonation pressure and normalized phonation onset time. This curve was generated as a least-squares fit to

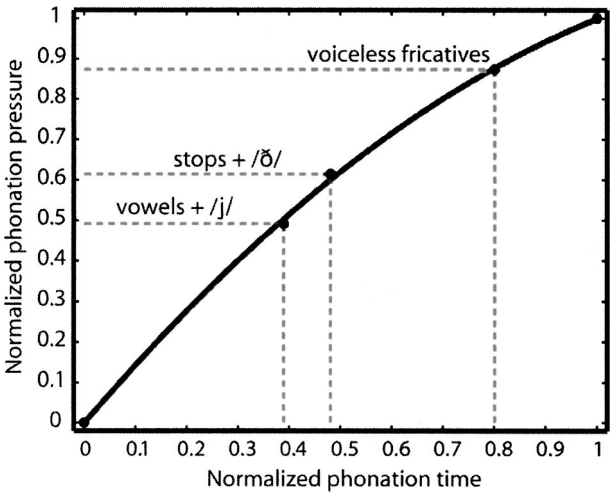


FIG. 6. Prototype for the pressure rise at utterance onset from group averages across all four speakers. Normalized phonation onset time and phonation pressure averages are used for three groups of sound segments at the start of the utterance: vowel and /j/, stops and /ð/, and voiceless fricatives.

a second-order polynomial using the group averages across all four speakers. The resulting polynomial is given in Eq. (2):

$$y = -0.48x^2 + 1.48x, \tag{2}$$

where y represents the normalized phonation pressure and x represents the normalized phonation onset time. In this prototype, utterances initiated with a voiceless fricative have a phonation pressure near 87% of the peak pressure, on average, and the phonation onset time is near 80% of the peak time. Phonation onset for utterances initiated with a voiced segment is near 39% of the peak pressure, on average, and the phonation onset time is in the range of 49% of the peak time. For utterances initiated with a stop or /ð/, the phonation pressure is near 61% of the peak pressure and the phonation onset time is near 48% of the peak time. When such a prototype is coupled with a characterization of a speaker's range of initial peak pressures and knowledge of the location in the utterance of the initial prominent syllable, a more detailed model for utterance initiation is possible. Such a prototype could be used in applications such as articulatory-based text-to-speech synthesis to produce more natural sounding prosodic boundary cues.

As the pressure rises toward the peak, the amplitude of the signal and its spectral characteristics go through changes both due to the rising pressure and the changing nature of the impedance to the air flow from the lungs. Within a short window of time, changes occur in the acoustics that form the coordinated set of acoustic cues to the start of an utterance. For utterances beginning with a vowel or /j/, these changes occur during phonation. For utterances beginning with a voiceless fricative, these changes primarily occur prior to phonation, given that the normalized phonation onset pressure is on the order of 87% of the initial alveolar pressure peak. In addition, spectral changes associated with subglottal coupling may occur as the vocal folds move from an abducted position to an adducted position for phonation (for example, Stevens, 1994).

V. SUMMARY

The eventual aim of this research is to model the sequences of respiratory and articulatory events that occur when a speaker starts talking. This research examines correlations between the actions of the respiratory system and segmental and prosodic properties at speech boundaries and contributes to the specification of the normal range of respiratory actions at utterance onset. The results indicate that, in general, the pressure rises to an initial peak that occurs within the first syllable in the utterance that was judged to be prominent when that prominence occurs early in the utterance and precedes the initial prominent syllable when that prominence occurs later in the utterance. The current research agrees with prior studies (such as Atkinson, 1973) in terms of trends in phonation onset time and phonation pressure. The present study adds to this information by quantifying some aspects of the link between the initial pressure peak and the onset of phonation across a range of boundary sound segments. Normalizing to the time and level of the initial pressure peak provides a more coherent view of the pressure trajectory and supports modeling goals. A vowel or voiced sonorant at the boundary segment has a low phonation pressure (roughly 30%–70% of the peak pressure) and a short phonation onset time (roughly 20%–60% of the peak time). Pressure rises through the voiced segment, leading to changes in signal amplitude. Phonation onset pressure is relatively high (roughly 60%–100% of the peak pressure) when the onset is an unvoiced fricative with a phonation onset time that is roughly 70%–100% of the peak time.

There is variability in the speech signal that arises from each speaker's particular anatomy and muscular habits. Evidence for this variability at the start of the utterance includes (1) differences in the level and timing of the peak pressure as well as the variation around that peak from utterance to utterance, (2) the manner in which phonation onset parameters vary in relation to differences in the initial perceived prominence in the utterance, and (3) the variation in a particular speaker's phonation onset parameters as a function of the sound segment at the boundary. For example, in this data set, there are speakers with relatively separate regions on the normalized curve for three different groupings of initial sound segments and there are speakers who show more overlap in the groupings.

These data on respiratory parameters could be used to assess possible deviant respiratory control, for example, in disorders of neuromotor control or in speech of individuals with hearing disorders, and provide cues for such applications as the detection of speech boundaries in a system for automatic speech recognition. Data for normalized initiation landmarks, a speaker's peak pressure range, and the location of the first prominent syllable in the utterance can be used to aid the development of a model of the respiratory system during speech that is capable of producing more accurate subglottal pressure variations at major speech boundaries and the corresponding prosodic cues.

ACKNOWLEDGMENTS

This work greatly benefited from the advice and support of Professor K. N. Stevens. This work was supported in part

by NIH Grant Nos. 5T32DC0038, 5R01-DC00266-14, and DC00075, the Clarence J. LeBel Chair at MIT, HST Fellowship, and a Dissertation Enhancement grant from the National Science Foundation INT 98-21048. The author would like to thank the reviewers for their time and comments.

Atkinson, J. E. (**1973**). ''Aspects of Intonation in Speech: Implications from an Experimental Study of Fundamental Frequency,'' Ph.D. dissertation, The University of Connecticut, Storrs, CT.

Baydur, A., Behrakis, P. K., Zin, W. A., Jaeger, M., and Milic-Emili, J. (**1982**). ''A simple method for assessing the validity of the esophageal balloon technique,'' Am. Rev. Respir. Dis. **126**(5), 788–791.

Draper, M., Ladefoged, P., and Whitteridge, D. (**1959**). ''Respiratory muscles in speech,'' J. Speech Hear. Res. **2**, 16–27.

Fant, G., and Kruckenberg, A. (**1989**). ''Preliminaries to the study of Swedish prose reading and reading style,'' Royal Institute of Technology, Speech Transmission Laboratory, Quarterly Progress and Status Report, 2/1989, pp. 1–83.

Fant, G., Kruckenberg, A., Liljencrants, J., and Hertegård, S. (**2000**). ''Acoustic phonetic studies of prominence, in Swedish,'' Royal Institute of Technology, Speech Transmission Laboratory, Quarterly Progress and Status Report, 2-2/2000, pp. 1–51.

Glass, J. R. (**1988**). ''Finding acoustic regularities in speech: applications to phonetic recognition,'' Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Fitts, P. M. (**1954**). ''The information capacity of the human motor system in controlling the amplitude of movement,'' J. Exp. Psychol. **47**, 381–391.

Fitts, P. M., and Peterson, J. R. (**1964**). ''Information capacity of discrete motor response,'' J. Exp. Psychol. **67**, 103–112.

Hertegård, S., Gauffin, J., and Lindestad, P.-Å. (**1995**). ''A comparison of subglottal and intraoral pressure measurements during phonation,'' J. Voice **9**, 149–155.

Konno, K., and Mead, J. (**1967**). ''Measurement of the separate volume changes of ribcage and abdomen during breathing,'' J. Appl. Physiol. **22**, 407–422.

Kunze, L. (**1964**). ''Evaluation of methods of estimating sub-glottal air pressure,'' J. Speech Hear. Res. **7**, 151–164.

Ladefoged, P. (**1967**). ''Stress and respiratory activity,'' in *Three Areas of Experimental Phonetics* (Oxford U.P., London).

Lieberman, P. (**1967**). *Intonation, Perception, and Language* (MIT, Cambridge, MA).

Lucero, J. C. (**1998**). ''Optimal glottal configuration for ease of phonation,'' J. Voice **12**(2), 151–158.

Netsell, R. (**1969**). ''A physiological-acoustic-perceptual study of syllable stress,'' Ph.D. dissertation, Univ. of Iowa, Iowa City, IA.

Ohala, J. J. (**1990**). ''Respiratory activity in speech,'' in *Speech Production and Speech Modelling, NATO ASI Series D: Vol. 55*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Boston, MA), pp. 23–51.

Rothenberg, M. (**1973**). ''A new inverse-filtering technique for deriving the glottal airflow wave-form during voicing,'' J. Acoust. Soc. Am. **89**, 1777–1781.

Slifka, J. (**2000**). ''Respiratory constraints on speech production at prosodic boundaries,'' Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Slifka, J. (**in press**). ''Respiratory system pressures at the start of an utterance,'' to appear in *Dynamics of Speech Production and Perception, NATO Advanced Study Institute 2002*, Il Ciocco, Italy.

Stevens, K. N. (**1994**). ''Prosodic influences on glottal waveform: preliminary data,'' in *Proceedings of the International Symposium on Prosody*, Yokohama, Japan, pp. 53–63.

Titze, I. R. (**1988**). ''The physics of small-amplitude oscillation of the vocal folds,'' J. Acoust. Soc. Am. **83**, 1536–1552.

Van den Berg, J. W. (**1956**). ''Direct and indirect determination of mean subglottal pressure,'' Folia Phoniatr. **8**, 1–24.

Influence of collision on the flow through in-vitro rigid models of the vocal folds

M. Deverge
Fluid Dynamics Laboratory, Technical University of Eindhoven, Postbus 513, 5600-MB Eindhoven, The Netherlands

X. Pelorson[a] and C. Vilain
Institut de la Communication Parlée, I.N.P.G, 46 av. F. Viallet, 38031 Grenoble cedex 1, France

P.-Y. Lagrée
Laboratoire de Modélisation Mécanique, Université Paris VI, 4 place Jussieu, Tour 66, case 162, 75252 Paris Cedex 05, France

F. Chentouf, J. Willems, and A. Hirschberg
Fluid Dynamics Laboratory, Technical University of Eindhoven, Postbus 513, 5600-MB Eindhoven, The Netherlands

(Received 25 October 2003; revised 1 August 2003; accepted 22 September 2003)

Measurements of pressure in oscillating rigid replicas of vocal folds are presented. The pressure upstream of the replica is used as input to various theoretical approximations to predict the pressure within the glottis. As the vocal folds collide the classical quasisteady boundary layer theory fails. It appears however that for physiologically reasonable shapes of the replicas, viscous effects are more important than the influence of the flow unsteadiness due to the wall movement. A simple model based on a quasisteady Bernoulli equation corrected for viscous effect, combined with a simple boundary layer separation model does globally predict the observed pressure behavior. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1625933]

PACS numbers: 43.70.Jt, 43.70.Bk [AL] Pages: 3354–3362

I. INTRODUCTION

Voiced sound production, or phonation, is driven by a modulation of the flow passing through the glottis as a result of the oscillation of the vocal folds. Typically, the fundamental oscillation frequency for a male speaker is of order of 10^2 Hz, which is much lower than the frequency range perceptually relevant for speech (of order of 10^3 Hz for most voiced sounds). One can therefore expect that, to simulate this behavior, two different models can be used, one predicting the oscillation of the vocal folds and a second predicting the sound production. As the matter of fact, a simple mechanical model such as a two-mass model[1–6] or a three-mass model[7] combined with a simplified flow model does indeed predict the self-sustained oscillations of the vocal folds. In such lumped parameter models the mechanics of the vocal folds is approximated by rigid masses attached to springs. The oscillations of this mass-spring system are driven by the difference in hydrodynamic force on the vocal folds during opening and closing phases. The contact between the vocal folds during the closure of the glottis is described as a change in stiffness of the springs. The flow is interrupted but the movement of the vocal folds continues as they are allowed to penetrate each other.

In earlier papers[2,8] we have verified that, for conditions typical to those encountered during phonation, a quasi-steady incompressible flow model based on the concept of viscous boundary layers appears to be a reasonable approximation of the glottal flow. However, the accuracy of this flow model during the closure of the glottis was left as an open problem. Considering the simplicity of the two-mass model it was thought, in practice, more reasonable to ignore the deviations from such a flow model which will certainly occur when the vocal folds collide. However, the closing phase of the vocal folds is known to be a very important feature of voiced sound production. This is the abrupt event which is needed to generate the higher harmonics which are perceptually relevant for speech. Many aspects of the voice quality or of the ''naturalness'' of the synthetic sounds can be related to this particular event.[9,10]

During the closure of the vocal folds the quasi-steady, incompressible boundary layer theory will fail because the flow channel, the glottis, becomes too thin to allow a distinction between a frictionless main flow and viscous boundary layers. Eventually the flow becomes dominated by viscous effects. On the other hand, the volume flux induced by the wall displacement becomes locally larger than the flux driven by the *trans*-glottal pressure difference. In such a case the flow becomes essentially unsteady.

The goal of this study is to investigate whether both viscous and unsteady phenomena are equally important as well as whether they appear simultaneously. The answer to this question will obviously depend upon the shape of the glottis. We will therefore consider three different shapes. In addition to two rounded models we will also present results for a channel with a uniform height. Such a straight channel is interesting because we can obtain analytical solutions for

[a]Author to whom correspondence should be addressed; electronic mail: pelorson@icp.inpg.fr

the flow equations in three different limits: the steady viscous flow, the unsteady frictionless flow and the unsteady viscous flow. In the steady viscous flow case an integral formulation of the boundary layer theory is used. As shown by Ishizaka[1] and Van Zon[11] an analytical formulation can be obtained for a uniform channel. The unsteady frictionless flow is based on the Bernoulli equation. Last, the unsteady viscous flow approximation is obtained by assuming an equilibrium between viscous and pressure forces in a quasiparallel flow. This corresponds to the lubrification theory of Reynolds.[12]

Finally we will also compare our data with a commonly used flow model. It is based on a correction of the steady Bernoulli equation for viscous effects based on the assumption of a Poiseuille flow.[1] A more elaborate description, accounting for a nonfixed flow separation point[13,2] will also be considered.

In the first section of this paper the dimensionless parameters relevant for our study are discussed. In the second section three theories are described together with the simplified theory based on the equation of Bernoulli corrected for friction losses. In Sec. III a brief description of the setup and of the experimental method used will be presented. The last part of this paper will be devoted to the analysis of the experimental results. First, only steady flows will be considered, this will allow us to evaluate the effects of viscosity. Second, the results obtained for the oscillating glottis will then be presented and discussed

II. DIMENSIONLESS PARAMETERS AND BASIC ASSUMPTIONS

The experiments presented here have been designed to simulate the conditions typical of voiced sound production. In particular, the pressure differences in the flow are small compared to the atmospheric pressure and the acoustical wave lengths are very large compared to the length L of the glottis. One can therefore assume that the flow is locally incompressible.

We consider the flow through an oscillating rigid model of the vocal folds with length L (in the flow direction) and width W. The minimum aperture of the glottis, h_g occurs at $x=x_g$ (henceforth called the throat of the channel) and varies in time between $h_{\min}$ and $h_{\max}$. An estimate for the flow velocity is the velocity U_B calculated from the pressure difference p_u-p_d across the glottis by means of the Bernoulli equation for steady nonviscous flows:

$$U_B=\sqrt{\frac{2(p_u-p_d)}{\rho}}, \tag{1}$$

where ρ is the air density which we assume to be constant. The pressures p_u and p_d correspond, respectively, to positions just upstream and downstream of the glottis. We thus assume implicitly that p_d is the pressure in the free jet downstream of the glottis. The fact that $h_g/L=O(10^{-1})$ and $h_g/W=O(10^{-1})$ indicates that a quasi-one-dimensional approximation for the flow should be reasonable $\vec{v}\simeq(u(x,y,t),0,0)$. This implies that the pressure is approximatively uniform in a cross section normal to the flow direction $p\simeq p(x,t)$. The ratio h_g/δ_v between the channel height and of the viscous boundary layer thickness $\delta_v=\sqrt{\nu L/U_B}$, where ν is the kinematic viscosity, yields an indication for the importance of viscosity. This ratio is related to the Reynolds number $\mathrm{Re}_h=U_B h_g/\nu$,

$$\left(\frac{h_g}{\delta_v}\right)^2=\mathrm{Re}_h\frac{h_g}{L}. \tag{2}$$

The Reynolds number $\mathrm{Re}_L=\mathrm{Re}_h(L/h_g)=U_BL/\nu$, based on the length L of the channel, provides an indication for the onset of turbulence in the glottis. Using values typical of voiced sound production, one gets $\mathrm{Re}_L=O(10^4)$. A laminar flow within the glottis can therefore be expected but the jet formed by flow separation downstream of the glottis will be turbulent.

A measure for the unsteadiness of the flow is the ratio of the volume flux due to the wall movement $f(h_{\max}-h_{\min})WL$, where f is the fundamental frequency of the motion, and the volume flux U_Bh_0W driven by the pressure difference p_u-p_d across the glottis,

$$\frac{f(h_{\max}-h_{\min})L}{U_Bh_g}=\mathrm{Sr}_L\frac{h_{\max}-h_{\min}}{h_g}, \tag{3}$$

where $\mathrm{Sr}_L=fL/U_B$ is the Strouhal number based on the channel length L. In the case of a uniform straight channel one obviously has $h_g=h_{\min}$ because a collision $h_{\min}\to 0$ implies an essentially unsteady flow. In the case of a more complex geometry however the choice of a relevant length scale for h_g is still an open question.

III. THEORETICAL MODELS

A. Introduction

Although the application to channels of arbitrary shapes will be presented, the theoretical models considered here will mainly focus on the case of straight uniforms channel. In all cases it is considered that the edges of the inlet are always well rounded so that any singular losses at the inlet are negligible. Because $h_g/W\ll 1$ the flow velocity u_u far upstream of the glottis is neglected and a uniform pressure p_u is assumed. In the case of a uniform channel, at the downstream end of the glottis the edges are considered as sharp so that the flow separation occurs at this fixed point. It is assumed that, in all cases, the pressure in the jet formed by this flow separation is equal to the pressure p_d far downstream of the channel. Last, as explained above and because $h_g/L\ll 1$ and $h_g\ll W$ a quasiparallel flow $\vec{v}=(u(x,y,t),0,0)$ is considered.

B. Inviscid unsteady flow

Strictly speaking, a purely inviscid flow theory would ignore flow separation and thus cannot explain the modulation of the flow by the vocal folds. This corresponds to the so-called paradox of d'Alembert.[14,15] Indeed, viscous effects induce a flow separation and the formation of a jet. Turbulent dissipation of the kinetic energy within the jet explains the volume flow control. If one assumes a quasisteady behavior of the jet, this implies that the pressure in the jet is equal to the pressure p_d downstream of the glottis. If other effects of

viscosity are neglected, the velocity $u(0)$ at the inlet of the channel (at $x=0$) can be related to the velocity $u(L)$ at the channel exit by means of the Bernoulli equation for an incompressible flow

$$\rho\frac{d}{dt}[\phi(x)-\phi(0)]+\frac{1}{2}\rho[u(x)]^2+p_d = \frac{1}{2}\rho[u(0)]^2+p(0), \tag{4}$$

where the velocity potential ϕ is given by

$$\phi(x)-\phi(0)=\int_0^x u\,dx \tag{5}$$

which is applied for $x=L$. It is further assumed that

$$p_u=\tfrac{1}{2}\rho[u(0)]^2+p(0) \tag{6}$$

in other words, the unsteadiness of the flow upstream of the inlet $x=0$ is neglected.

For an incompressible flow through a channel of uniform height, h, the mass conservation law yields

$$u(x)-u(0)=-\frac{x}{h}\frac{dh}{dt}. \tag{7}$$

Combining the definition (5) for ϕ and the equation of Bernoulli (4) yields a differential equation for the velocity $u(0)$ at the inlet of the channel,

$$L\frac{du(0)}{dt}=\frac{p_u-p_d}{\rho}+\frac{L^2}{2}\frac{d}{dt}\left[\frac{1}{h}\frac{dh}{dt}\right]-\frac{1}{2}\left[u(0)-\frac{L}{h}\frac{dh}{dt}\right]^2 \tag{8}$$

which for given pressure difference, p_u-p_d and a given h can be integrated as a function of time. For a harmonically oscillating h, the result of this integration converges to a value which is independent from the initial conditions.

In practice, given $u(0)$ one can calculate $u(x)$ and $\phi(x)$ using the mass conservation law (7) and the definition (5). The pressure p_g at $x=x_g$ is then found by applying the equation of Bernoulli (4) between $x=x_g$ and $x=L$.

C. Boundary layer solution for steady flows

For a steady flow through a channel of uniform height, h, and driven by a constant pressure difference, p_u-p_d, the Von Kármán integral formulation of the boundary layer equations can be integrated analytically. This solution was already discussed by Ishizaka[1] but only in the case where the boundary layer approximation remains valid over the full length L of the channel. In such a case, there is always a frictionless core with a uniform velocity $u_e(x)$ in which the Bernoulli equation can be applied. The frictionless core of the jet at the exit of the channel has a velocity $u_e(L)=U_B$ [Eq. (1)]. As a more simple alternative to the method of Thwaites,[8] a method of Pohlhausen of first order is presented here.[12] In general the method of Polhausen assumes that the velocity profile $u(x,y)$ within the viscous boundary layer has a simple shape which can be described by a polynomial of the distance y from the wall. We use here a polynomial of first order. As shown by Van Zon[11] using a linear velocity profile $u(x,y)=u_e(x)y/\delta$, where δ is the thickness of the boundary layers the corresponding volume flux becomes

$$\Phi_V=Wu_e(h-\delta), \tag{9}$$

where δ satisfies the nonlinear equation

$$\frac{4\delta}{h}+9\ln\left(1-\frac{\delta}{h}\right)+\frac{5\delta}{h-\delta}=\frac{6\nu x}{h\Phi_V}. \tag{10}$$

Applying Eqs. (9) and (10) at the exit $x=L$ and using $u_e(L)=U_B$ one has thus a set of two equations from which $\delta_L=\delta(L)$ and Φ_V can be obtained. Once Φ_V has been calculated the viscous boundary layer $\delta(x)$ at any arbitrary position x can be calculated from (10). The corresponding frictionless core velocity $u_e(x)$ is obtained by application of the mass conservation law (9). Finally, the pressure is calculated by using Bernoulli equation

$$p(x)+\tfrac{1}{2}\rho[u_e(x)]^2=p_u. \tag{11}$$

The generalization of this approach for channel of arbitrary length can be made as proposed by Van Zon.[11] When the critical boundary layer thickness

$$\frac{\delta_c}{h}=\frac{4}{9}\left(1-\sqrt{\frac{5}{32}}\right) \tag{12}$$

is reached, the volume flux and momentum flux correspond to those of a fully developed Poiseuille flow. The critical distance $x=l_c$ at which $\delta(l_c)=\delta_c$ can be obtained analytically for a given volume flux Φ_V using Eq. (10). For $L>l_c$ and $x<l_c$ the pressure distribution can be obtained from the Bernoulli equation combined with the equation of mass conservation (9),

$$p_u=p+\frac{1}{2}\rho\frac{\Phi_V^2}{W^2(h-\delta)^2} \tag{13}$$

while for $x>l_c$ the equation of Poiseuille[12] is used

$$p-p_d=\frac{12\rho\nu\Phi_V}{Wh^3}(L-x). \tag{14}$$

Applying Eqs. (13) and (14) at $x=l_c$ and eliminating $p(l_c)$ yields a quadratic equation for Φ_V. Using Eq. (10) in which Eq. (12) is substituted and for $x=l_c$ a linear relationship between Φ_V and l_c: $\Phi_V=(\nu l_c)/(hc)$ can be obtained. By solving the quadratic equation one finally obtains

$$\frac{l_c}{L}=\frac{12c\left(1-\frac{\delta_c}{h}\right)^2}{24c-1}\left[1-\left(1-\frac{h^4(24c-1)(p_u-p_d)}{72\rho\nu^2L^2\left(1-\frac{\delta_c}{h}\right)^2}\right)^{1/2}\right], \tag{15}$$

where the constant c is given by

$$c=\frac{1}{6}\left[\frac{4\delta_c}{h}+9\ln\left(1-\frac{\delta_c}{h}\right)+\frac{5\delta_c}{h-\delta_c}\right]. \tag{16}$$

Depending on x, $p(x)$ can be calculated using Eqs. (13) or (14).

The use of Thwaites's implementation of the integral formulation of Von Kármán equation applied to a channel of arbitrary shape is discussed in detail in Hofman.[8] Similar

results are also discussed by Pelorson[2] using the approach of Pohlhausen. All these approaches involve a numerical resolution which is, in practice, difficult because of the essential nonlinearity of the problem.[12] A direct solution of the equation of Prandtl is also discussed by Lagrée.[16] A systematic comparison showed that in terms of flow separation point prediction, the method of Pohlhausen and Twaithes are equivalent.[17,18] Thwaites method appears more robust numerically and therefore will be used in the following. However, in the case of a straight uniform channel an analytical solution can be obtained as shown previously and there is no need for any numerical resolution. In the following, we will refer to this analytical solution as the solution of van Zon.[11]

D. Lubrification theory of Reynolds

The lubrification theory of Reynolds combines the assumption of a quasiparallel flow together with the assumption that inertial effects are negligible. As the pressure forces balance the viscous ones, the velocity profile in the channel is given by the Poiseuille formula[12]

$$u = -\frac{1}{2\rho\nu}\frac{\partial p}{\partial x}(h-y)y. \tag{17}$$

This velocity profile combined with the mass conservation law

$$W\frac{\partial h}{\partial t} = -\frac{\partial \Phi_V}{\partial x}, \tag{18}$$

where

$$\Phi_V = W\int_0^h u\, dy, \tag{19}$$

yields the equation

$$\frac{1}{12\rho\nu}\frac{\partial}{\partial x}\left(h^3\frac{\partial p}{\partial x}\right) = \frac{\partial h}{\partial t} \tag{20}$$

which for a channel of uniform height can be integrated to give

$$p - p_u = \left(\frac{p_d - p_u}{L}\right)x + \frac{12\rho\nu}{h^3}\left(\frac{dh}{dt}\right)\frac{x(x-L)}{2}. \tag{21}$$

E. Steady Bernoulli corrected for friction

Often one seeks for a simple correction the inviscid theory by adding an extra term to account for viscous pressure losses. This corresponds to the original approach of Ishizaka.[1] A similar approach is proposed by Antunes.[19] For the sake of simplicity, and because it will be shown that the unsteadiness seem to be less important than viscous effects, only the steady flow case will be developed here. This choice is further supported by the fact that if one neglects the effect of the wall movement on the mass conservation law, then the unsteady term in Bernoulli equation should certainly be neglected as well. The wall movement is, indeed, the main cause of unsteadiness. In this sense, the flow model of Ishizaka[1] is not consistent. Such an inconsistent approach might result in some poor behavior of the model as observed by Lous.[5]

The Bernoulli equation corrected for viscous pressure losses Δp_v becomes

$$\frac{1}{2}\rho\left[\frac{\Phi_V}{Wh}\right]^2 + p_d + \Delta p_v = p_u. \tag{22}$$

Using the lubrification theory of Reynolds we find

$$\Delta p_v(x) = \frac{12\rho\nu\Phi_V}{W}\int_0^x \frac{dx}{h^3}. \tag{23}$$

Combined with the modified Bernoulli Equation (22) this equation, when applied at the separation point x_s, yields a quadratic equation for Φ_V which is easily solved for given x_s. In a straight uniform channel one has simply $x_s = L$ and $\Delta p_v = (12\rho\nu\Phi_V L)/(Wh^3)$.

The extension of the above theory to the case of arbitrary shapes can, of course, be done by numerically integrating Eq. (23). However, as the friction losses scale with h^{-3} they depend strongly on the channel height. It can therefore be assumed that the losses are determined locally in a small region close to the throat, x_g of the channel. Using a Taylor expansion, one can therefore use for the channel height h the approximation

$$h \simeq h_g + \frac{(x-x_g)^2}{R}, \tag{24}$$

where R is the (local) radius of curvature of the wall, we find by integrating (23) from $x=-\infty$ to $x=\infty$,[20]

$$\Delta p_v \simeq \frac{9\pi\rho\nu\Phi_V}{2Wh_g^2}\left(\frac{R}{h_g}\right)^{1/2}. \tag{25}$$

In order to solve for Φ_V one now needs an estimation for the height of the channel $h(x_s)$ at the separation point x_s where it is assumed in the first approximation that $p(x_s) = p_d$. The relevance of this later assumption has been discussed in detail by Hofmans.[8] In view of its simplicity the semiempirical criterium of Liljencrants,[13]

$$\frac{h(x_s)}{h_g} = 1 + \epsilon \tag{26}$$

with $\epsilon = 0.1$ seems a reasonable order of magnitude.[8]

In summary, four kinds of theoretical predictions will be considered in the following.

(1) The boundary solution which refers as the (analytical) van Zon solution in the case of a uniform straight channel and as the Thwaites (numerical) solution in the case of a nonuniform channel.
(2) The lubrification theory of Reynolds as developed in Sec. III D.
(3) The steady Bernoulli theory corrected for pressure losses described in Sec. III E.
(4) The unsteady Bernoulli solution presented in Sec. III B.

IV. EXPERIMENTAL SETUP AND PROCEDURE

In order to validate the theoretical models presented in the preceding sections, an experimental setup with oscillating vocal-folds replica is used. The main interest of this approach lies in the better control of the experimental condi-

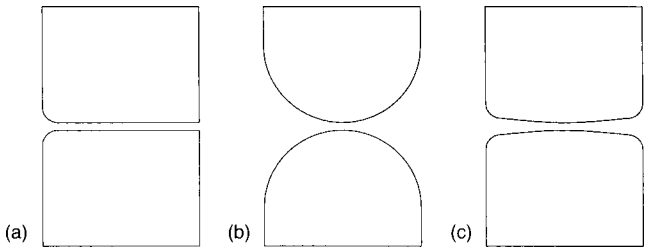


FIG. 1. Mechanical models of the vocal folds: (a) straight uniform channel, (b) rounded vocal folds, and (c) Gaussian vocal folds.

tions compared to in-vivo measurements.[21] While most in-vitro experiments in the literature deal with steady replicas of vocal folds (e.g., Refs. 22–25), or with numerical simulations on steady replicas (e.g., Refs. 26 and 27) very few attempts have been made using moving vocal folds. Much more realistic flow conditions were obtained recently using setups including moving replicas of vocal folds. These experiments were intended either to focus on the onset of phonation[28] or to mimic self-sustained oscillations using forced motion of the vocal-folds replica.[29,30,21] The experimental setup used in this study is in the tradition of these latter studies.

A. Vocal-fold replicas and sensors

Figure 1 shows the three different vocal folds mechanical models (or replicas) used. All these mechanical models have an overall length of 2 cm in the flow direction and a width $W=3$ cm.

Compared with human size, the mechanical model appears thus up-scaled by a factor of 3. To keep the Reynolds number constant, this implies that the velocities in the mechanical model are a factor of 3 times smaller than those expected for humans. At constant Strouhal and Reynolds numbers, the frequency of oscillation of the mechanical replica must be smaller than the one expected for human phonation by a factor of 9.

A pressure tap with a radius of 0.4 mm and 1 mm length is placed 1 cm upstream from the downstream end. This allows measurements of the wall glottal pressure p_g-p_d by means of a Kulite pressure transducer (type XCS-093, diameter 1.6 mm) placed in a cavity below the pressure tap. The pressure gauge was calibrated by using a Betz water micromanometer with a precision of 1 Pa. The response of the gauge was found to be linear within the accuracy of the measurement. The calibration was repeated after each series of measurements and appeared to be stable.

The first replica is the straight uniform channel. On the upstream side, the edges have been rounded with a radius of curvature of 2 mm. On the opposite, the downstream edges are made sharp. The transducer position is located at $x_g=1$ cm.

The second replica, denoted as the rounded vocal folds mechanical model has a length of $L=2$ cm. The walls are half cylinders with a radius of $R=1$ cm. The pressure tap is placed at the throat of the replica $x_g=1$ cm.

The third replica, the Gaussian vocal folds, has a more complex shape. In the region $2\text{ mm}<x<18\text{ mm}$ the vocal folds have a Gaussian shape described by the equation $y=A\exp[-(x-(L/2))^2/\alpha^2]$ with $A=2$ mm and $\alpha=9$ mm. The edges of both the inlet and outlet are rounded with a radius of curvature of 2 mm. The pressure tap is located at $x_g=1$ cm.

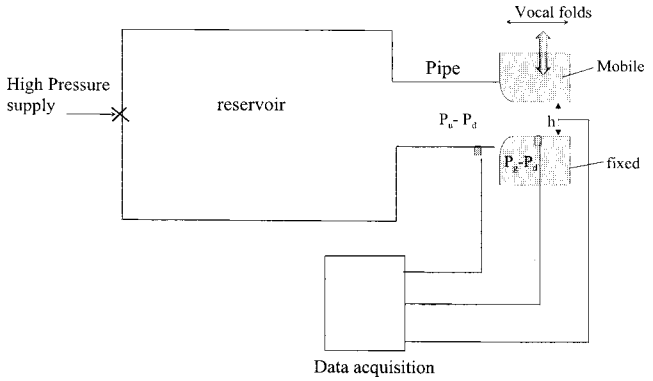


FIG. 2. Global sketch of the setup.

In the case of unsteady flow measurements, the lower fold with the pressure tap is maintained fixed while the upper fold is driven by a piston using an electrical motor and an eccentric wheel. This allowed to simulate self-sustained oscillations with a fundamental frequency of oscillation ranging from 5 Hz up to 35 Hz. Only results concerning the highest frequency are presented here as they correspond to the highest Strouhal numbers achievable using this setup. During the collision the mechanical folds were prevented from bouncing thanks to both to the backlash of the driving mechanism and the strong damping of the piston. The channel height h_g at x_g is measured by means of an optical sensor (type OPB700). The sensor was calibrated by placing gauges of known thickness at the throat of the glottis. The calibration was performed before and after each measurement. The estimated uncertainty in the measurement of the channel height h_g is 10^{-2} mm. The range of variation for h_g was chosen to keep a fairly linear behavior of the sensor.

The presence of an asymmetrical flow (Coanda effect) through rigid nonmoving mechanical replicas of the glottis has been observed in past experiments. Although the most direct way to observe such an asymmetry was to use pressure measurement at both sides of the replica, this phenomenon could be observed in dynamic experiments even using measurements of a single side of the replica due to the presence of an abrupt transition in the pressure signal.[31] This transition corresponds to the time needed by the flow to establish an asymmetrical behavior. As such behavior was never observed here, there is no evidence for the presence of any asymmetry in the flow.

B. Global description of the setup

A global view of the setup is shown in Fig. 2. The vocal folds are mounted at the end of a pipe of 30 cm length and 3 cm diameter connected to a pressure reservoir with a volume of 0.68 m^3 filled with acoustical foam in order to prevent acoustical resonances. The pressure p_u-p_d is measured 2 cm upstream from the inlet of the replica by means of a Kulite pressure transducer (type XCS-093) mounted flush with the pipe wall. The pressure reservoir is filled with

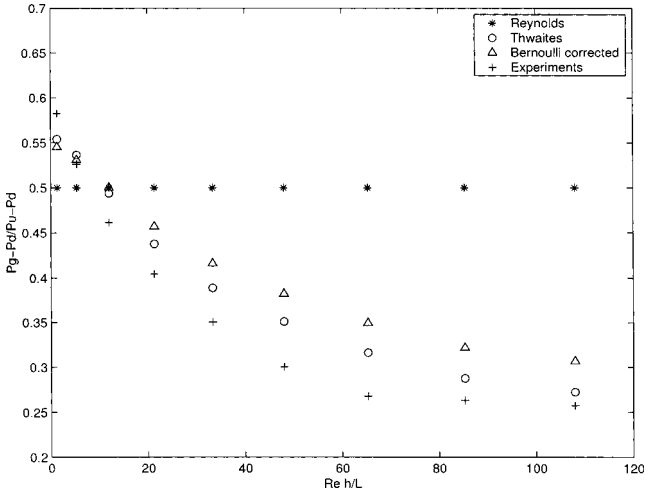


FIG. 3. Steady flow measurements for a straight channel ($p_u - p_d = 1$ kPa).

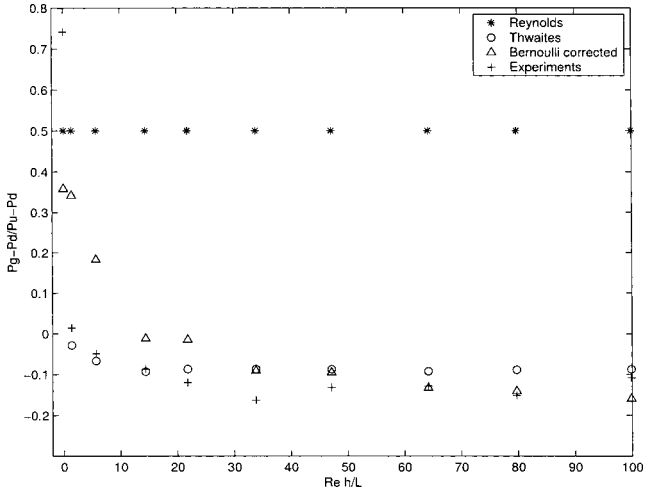


FIG. 4. Steady flow measurements for the rounded replica ($p_u - p_d = 1$ kPa).

acoustical foam to avoid acoustical resonances. The air flows from a 8 bar pressure supply through a choked valve into the reservoir. Downstream of the replica the flow exits into a large room. The pressure p_d in this room is used as a reference by the pressure gauges. The signals are recorded with a sample frequency of 1 kHz by means of a data acquisition card (NI PCI-MIO 16XE10) in a PC.

V. RESULTS

A. Steady flow measurements

In Fig. 3 the measured dimensionless pressure $(p_g - p_d)/(p_u - p_d)$ inside the straight uniform replica is plotted as a function of $\mathrm{Re}_h(h/L)$. The measured data are compared with the predictions from the lubrification theory of Reynolds (Sec. III C), the boundary layer theory (Sec. III B) and the Bernoulli theory corrected for friction (Sec. III D). These measurements have been carried out at a fixed pressure of $p_u - p_d = 1$ kPa. It can be observed that the boundary layer approximation provides quite reasonable agreement while the theory of Reynolds overestimates $(p_g - p_d)/(p_u - p_d)$ for $\mathrm{Re}_h(h/L) > 10$. The inviscid approximation based on Bernoulli (Sec. III B) would just predict $(p_g - p_d)/(p_u - p_d) = 0$. However, the addition of a correction based on a Poiseuille flow profile allows for a reasonable agreement although the predictions are systematically higher than the measured data.

In Fig. 4 are presented the measured $(p_g - p_d)/(p_u - p_d)$ for a steady flow within the rounded replica. These data are compared with the predictions obtained by the boundary layer approximation of Thwaites,[8] the approximation of Reynolds and the Bernoulli theory corrected for friction. It can be observed that the theory of Reynolds (Sec. III D) always predicts $(p_g - p_d)/(p_u - p_d) = 1/2$ due to the symmetry of the replica. This theory cannot predict the negative values of $p_g - p_d$ which are essentially due to inertial effects. The Thwaites theory does predict the order of magnitude of those negative values for $\mathrm{Re}_h(h/L) > 10$ mm. When the glottis is more and more closed $h_g \to 0$ one measures a pressure which is approaching the value predicted by the theory of Reynolds. The use of the Bernoulli theory corrected for friction (Sec. III E) explains the measured data quite well. The separation criterion chosen in Eq. (26) thus seems relevant for this geometry.

The corresponding results for the Gaussian replica are shown in Fig. 5. The conclusions that can be drawn are similar to those expressed for the rounded geometry.

We conclude from those data that for $\mathrm{Re}_h(h/L) > 10$ the boundary layer theory seems reasonable while the data approach the prediction of the lubrification theory of Reynolds for smaller values.

B. Unsteady flow measurements

In Fig. 6(a) are presented the measurements of the glottal pressure $p_g - p_d$, the upstream pressure $p_u - p_d$ and the channel height h. In this experiment, the vocal folds were not allowed to collide, the minimum channel throat was fixed as $h_{\min} = 0.10$ mm. The measured glottal pressure is compared with the prediction obtained by means of the boundary layer theory (Sec. III C), the lubrification theory of Reynolds (Sec. III D) and the inviscid unsteady solution (Sec. III B). The

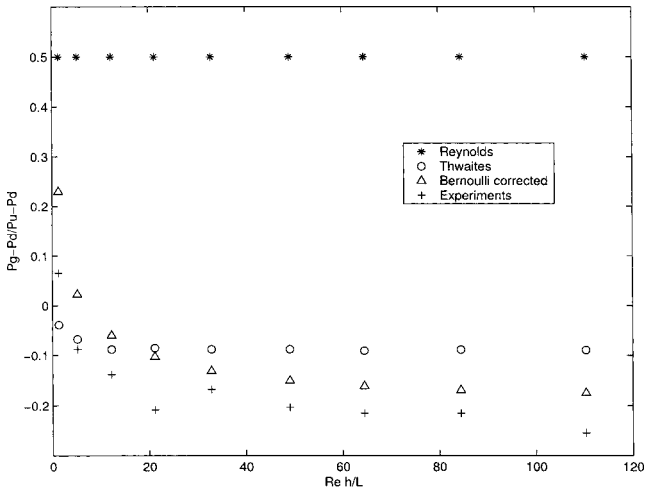


FIG. 5. Steady flow measurements for the Gaussian replica ($p_u - p_d = 1$ kPa).

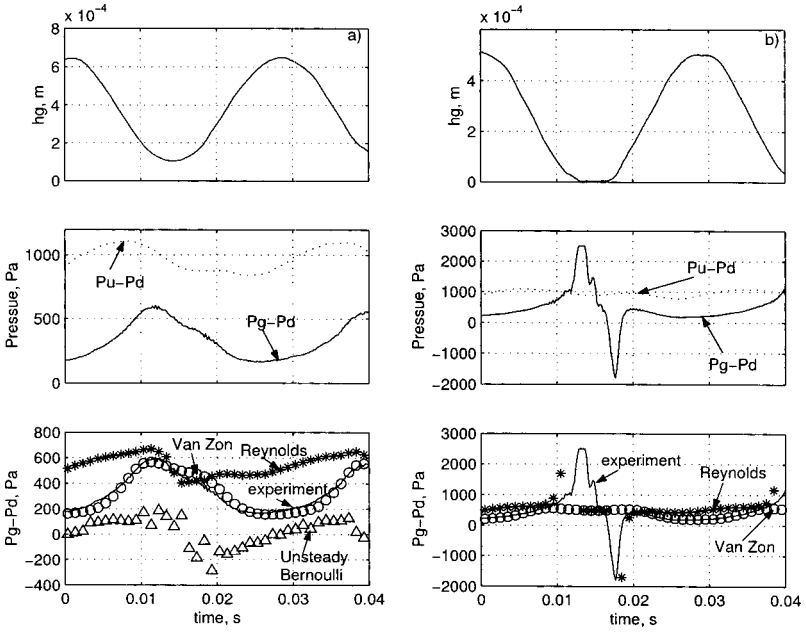


FIG. 6. Experimental and theoretical results for the straight uniform replica (a) $h_{\min}=0.10$ mm and (b) $h_{\min}=0$.

good agreement between the boundary layer prediction and the measured data tends to show that unsteady flow effects are not very important. The comparison with the lubrification theory of Reynolds (Sec. III E) shows that inertial effects at the channel inlet are important except for $h_g \simeq h_{\min}$. For a short time interval (when $h_g \simeq h_{\min}$) the theory of Reynolds predicts the experimental data. We further observe that the inviscid unsteady approximation (Sec. III B) provides quite poor results.

In the case where a collision is allowed [Fig. 6(b)], it can be observed that a finite glottal pressure could be measured even when the glottis is closed. This surprising result is due to the fact that owing to the surface roughness of the mechanical folds (of order of 10^{-6} m) a complete closure cannot be achieved.

It can also be observed that the glottal pressure $p_g - p_d$ can become larger than the transglottal pressure $p_u - p_d$ which can only be explained by the flow unsteadiness due to the movement of the walls. This effect is indeed predicted by the lubrification theory of Reynolds. Deviation between theory and experiments could be partially due to the effect of errors in the measurement of the channel height h_g when $h_g \to 0$. The results of the lubrification theory are indeed very sensitive to such errors, because the viscous losses are proportional to h^{-3}. Last, the inviscid unsteady approximation, which is not shown here, fails to explain the measured data.

In Figs. 7(a) and 7(b) are presented results for the rounded replica for $h_{\min}=0.08$ mm and $h_{\min}=0$ mm, respectively. It can be observed that as h approaches $h_{\min}$ the glottal pressure changes from a negative pressure predicted by the theory of Thwaites towards a positive pressure which approaches $(p_g - p_d)/(p_u - p_d) = 0.5$ as predicted by the steady lubrification theory of Reynolds. When there is a collision, as shown in Fig. 7(b) this value is indeed reached. The fact that

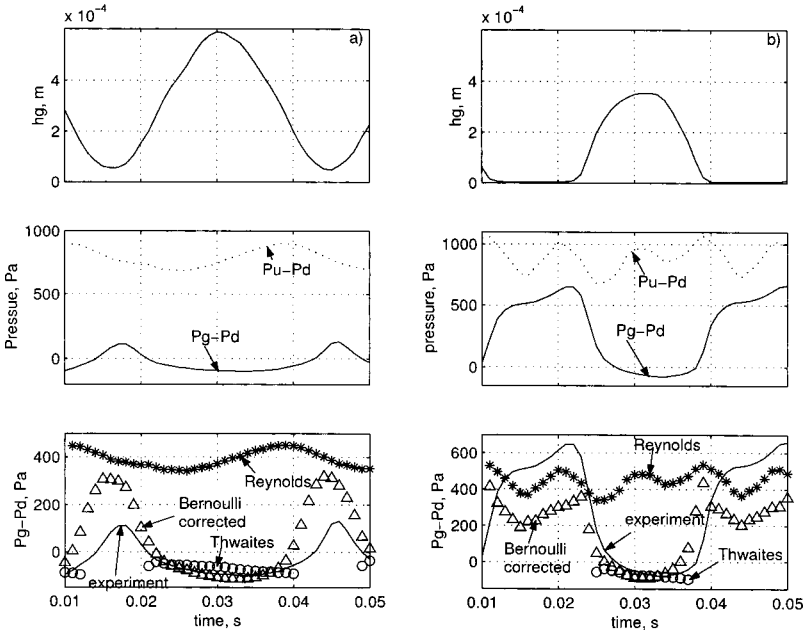


FIG. 7. Experimental and theoretical results for the rounded replica (a) $h_{\min}=0.08$ mm and (b) $h_{\min}=0$.

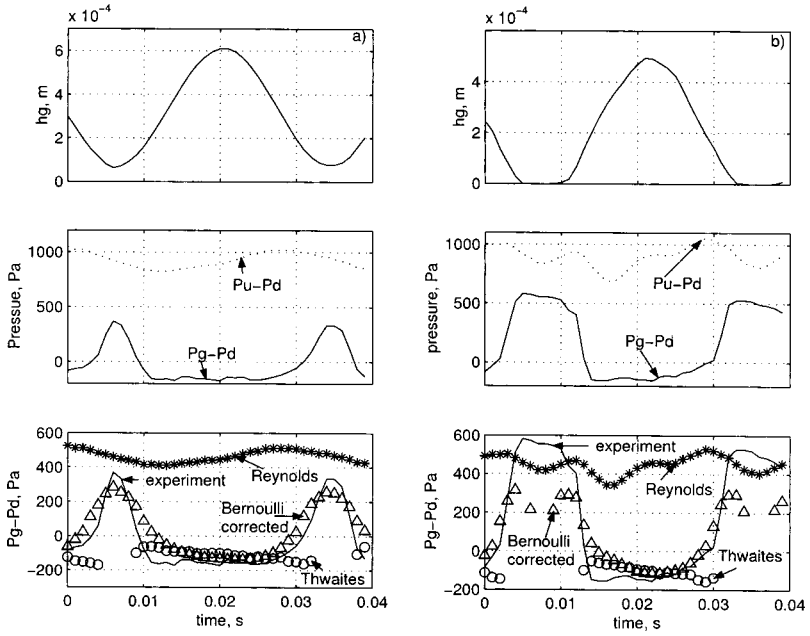


FIG. 8. Experimental and theoretical results for the Gaussian replica (a) $h_{min}=0.05$ mm and (b) $h_{min}=0$.

we do not find glottal pressures significantly higher than this limit tends to indicate that, for the rounded replica, even in the case of a collision the unsteady flow induced by the wall movement remains negligible compared to the viscous effects.

In Fig. 8 it can be seen that the results obtained using a Gaussian replica are very similar to those obtained with the rounded one. The main difference is that the prediction of the steady boundary layer theory of Thwaites as well as the Bernoulli theory modified for friction appear less accurate than for rounded replica experiments.

VI. CONCLUSION

From this study it was observed that upon collision the flow unsteadiness due to the wall movement appears to be only significant in the case of the straight uniform replica. In such a case the unsteady theory of Reynolds (Sec. III E) predicts qualitatively the measurements. In the case of more physiological vocal folds shapes, a transition between a boundary-layer behavior towards a friction dominated behavior was clearly observed without significant effect of the flow unsteadiness.

It was shown that the boundary layer solution is quite reasonable for $\mathrm{Re}_h(h/L)>10$. In practice such a theory is still quite complex and one may prefer a more simple model to predict flow separation. The models of Liljencrants[13] or the model of Pelorson[2] are easy to implement. When such a simple model is used, the Bernoulli equation corrected for a pressure loss term calculated on the basis of the theory of Reynolds provides a surprisingly accurate approximation. Addition of unsteady terms in the equation of Bernoulli will not improve the model. Such an unsteady term should certainly not be used if the mass conservation law does account for wall movement.

It must be noted that in our experiments a complete closure of the glottis could not be simulated due to the surface roughness of the replicas. In the case of human phonation one could expect a complete closure due to both the elastic deformation of the tissues and the presence of moisture or of mucus on the folds. While the dynamics of the tissues during collision can be studied by means of finite element method as shown recently by Gunter,[32] the possible presence of moisture seems another potentially important aspect. The movement of water driven by the colliding folds is expected to involve strong inertial effects due to the relatively large density of water. Therefore a complete and accurate physical model for the closure of the vocal folds might involve more than an interaction between an airflow and dry elastic tissues.

ACKNOWLEDGMENTS

Part of this work was supported by grant EURODOC from the Rhone-Alpes region (SAFIR No. 99009385), by grant VAN GOGH from the French Ministry of National Education, Research and Technology (No. 01381XE) and by grant from French and Dutch national centers for scientific reasearch (CNRS and NWO) (No. DRI-EUR 99/NWO/NLD). The authors thank Dr. Anna Barney for her helpful comments and suggestions. The authors wish also to thank H. Koolmes and F. Van Uittert from Technical University of Eindhoven for their help in the design of the experimental setup.

[1] K. Ishizaka and M. Matsudaira, "Fluid mechanical considerations of vocal-cord vibrations," Speech Commun. Res. Lab. Monograph, **8**, Santa Barbara, CA (1972).
[2] X. Pelorson, A. Hirschberg, R. van Hassel, A. Wijnands, and Y. Auregan, "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model," J. Acoust. Soc. Am. **96**, 3416–3431 (1994).
[3] J. Lucero, "Dynamics of the two-mass model of the vocal folds: Equilibria, bifurcations, and oscillation region," J. Acoust. Soc. Am. **94**, 3104–3111 (1993).
[4] I. Steinecke and H. Herzel, "Bifurcations in an asymmetric vocal fold model," J. Acoust. Soc. Am. **97**, 1874–1884 (1995).
[5] N. Lous, G. Hofmans, R. Veldhuis, and A. Hirschberg, "A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with appli-

cation to prosthesis design,'' Acta Acust. (Beijing) **84**, 1135–1150 (1998).
[6] M. DeVries, H. Schutte, and G. Verkerke, ''Determination of parameters for lumped parameter models of the vocal folds using a finite-element approach,'' J. Acoust. Soc. Am. **106**, 3620–3628 (1999).
[7] B. Story and I. Titze, ''Voice simulation with a body-cover model of the vocal folds,'' J. Acoust. Soc. Am. **97**, 1249–1260 (1995).
[8] G. Hofmans, ''Vortex sound in confined flows,'' Ph.D. thesis, Technical University Eindhoven, 1998.
[9] J. Gauffin and J. Sundberg, ''Spectral correlates of glottal voice source waveform characteristics,'' J. Speech Hear. Res. **32**, 556–565 (1989).
[10] D. Childers and C. Lee, ''Vocal quality factors: Analysis, synthesis, and perception,'' J. Acoust. Soc. Am. **90**, 2394–2410 (1991).
[11] J. van Zon, A. Hirschberg, J. Gilbert, and A. Wijnands, ''Flow through the reed channel of a single reed instrument,'' in Proceedings of the 1st French Congress of Acoustics (Editions de Physique, Lyon, France, 1990), Vol. C2, pp. 821–824.
[12] H. Schlichting and K. Gersten, *Boundary Layer theory*, 8th ed. (Springer Verlag, Berlin, 2000).
[13] J. Liljencrants (private communication).
[14] H. Teager and S. Teager, ''Active fluid dynamics voice production, or there is a unicorn in the garden,'' in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, edited by I. Titze and R. Scherer (The Denver Center For The Performing Arts, Denver, Colorado, 1983), pp. 387–401.
[15] A. Hirschberg, ''Some fluid dynamic aspects of speech,'' Bull. Commun. Parlée **2**, 7–30 (1992).
[16] P. Lagrée and K. Goorman, Self oscillation of a ''one mass'' 2D elastic stenosis submitted to J. Fluids Struct.
[17] G. Hofmans, G. Groot, M. Ranucci, G. Graziani, and A. Hirschberg, ''Unsteady flow through in-vitro models of the glottis,'' J. Acoust. Soc. Am. **113**, 1658–1675 (2003).
[18] C. Vilain, ''Contribution à la synthèse de parole par modèle physique. application à l'étude des voix pathologiques,'' Ph.D. thesis, Institut National Polytechnique de Grenoble, 2002.
[19] P. P. Antunes, and P. Piteau, ''A nonlinear model for squeeze-film dynamics under axial flow,'' in *ASME's Pressure Vessels and Piping Symposium*, edited by M. Petitgrew (ASME Atlanta, 2001).
[20] L. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products*, 6th ed. (Academic, San Diego, CA, 2000).
[21] F. Alipour and R. Scherer, ''Effects of oscillation of a mechanical hemilarynx model on mean transglottal pressures and flows,'' J. Acoust. Soc. Am. **110**, 1562–1569 (2001).
[22] J. van den Berg, J. T. Zantema, and P. Doornenbal, ''On the air resistance and the bernoulli effect of the human larynx,'' J. Acoust. Soc. Am. **29**, 626–631 (1957).
[23] R. Scherer, I. Titze, and J. Curtis, ''Pressure-flow relationships in two models of the larynx having rectangular glottal shapes,'' J. Acoust. Soc. Am. **73**, 668–676 (1983).
[24] N. Binh and J. Gauffin, Aerodynamic measurements in an enlarged static laryngeal model, Technical Phoniatrics, STL-QPSR, Department of Gas Dynamics, Royal Institute of Technology, Stockholm, 1983.
[25] R. Scherer, D. Shinwari, K. de Witt, C. Zhang, B. Kucinschi, and A. Afjeh, ''Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees,'' J. Acoust. Soc. Am. **109**, 1616–1630 (2001).
[26] J. Liljencrants, ''Numerical simulations of glottal flow,'' in *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*, edited by J. Gauffin and B. Hammarberg (Singular Publishing Group, Inc., San Diego, California, 1991), pp. 99–105.
[27] H. Iijima, N. Miki, and N. Nagai, ''Glottal impedance based on a finite element analysis of two-dimensional unsteady viscous flow in a static glottis,'' IEEE Trans. Signal Process. **40**, 2125–2135 (1992).
[28] I. Titze, ''Phonation threshold pressure in a physical model of the vocal fold mucosa,'' J. Acoust. Soc. Am. **97**, 3080–3084 (1995).
[29] L. Mongeau, N. Franchek, C. Coker, and R. Kubil, ''Characteristics of a pulsationg jet through a small modulated orifice, with application to voice production,'' J. Acoust. Soc. Am. **102**, 1121–1133 (1997).
[30] A. Barney, C. Shadle, and P. Davis, ''Fluid flow in a dynamic mechanical model of the vocal folds and tract. 1-measurements and theory,'' J. Acoust. Soc. Am. **105**, 444–455 (1999).
[31] X. Pelorson, A. Hirschberg, A. Wijnands, and H. Bailliet, ''Description of the flow through in-vitro models of the glottis during phonation,'' Acta Acust. (Beijing) **3**, 191–202 (1995).
[32] H. E. Gunter, ''A mechanical model of vocal-fold collision with high-spatial and temporal resolution,'' J. Acoust. Soc. Am. **113**, 994–1000 (2003).

Schlieren imaging of shock waves from a trumpet

Brian H. Pandya, Gary S. Settles,[a] and James D. Miller
Gas Dynamics Laboratory, Mechanical and Nuclear Engineering Department, 301D Reber Building, Penn State University, University Park, Pennsylvania 16802

(Received 2 August 2002; revised 19 September 2003; accepted 29 September 2003)

A sensitive, large-aperture schlieren optical instrument is applied to observe gas-dynamic phenomena at the exit of a trumpet. Shock waves are seen, especially for loud, high-pitched trumpet notes, and several illustrations are given. Microphone waveforms are given for representative examples. These shock waves arise from the shock-tube-like effect of the performer's intermittent breath pressure driving the cylindrical duct of the trumpet, and are the result of cumulative nonlinear acoustic propagation inside the trumpet bore. They are, however, very weak, traveling only marginally above the acoustic speed. In the 118–124 peak dB(A) range, they are near the weak limit of shock wave visibility by schlieren optics. The schlieren evidence confirms that the frequency of the emitted shock waves corresponds to the frequency of the note being played. Ancillary laminar and turbulent jet phenomena associated with the performer's breath are also visible in the images. [DOI: 10.1121/1.1628682]

PACS numbers: 43.75.Fg, 43.75.Wx, 43.75.Yy [NHF] Pages: 3363–3367

I. INTRODUCTION

In this paper, optical flow visualization is applied to image shock waves radiated by a trumpet. Hirschberg *et al.*[1] first discovered that shock waves are generated when playing bright, brassy fortissimo notes on a trombone. They state that the ''brightness of the sound generated by trombones is expected to be due to the essential non-linearity of the wave propagation in the pipe. At fortissimo levels this leads to shock wave formation.''

The physics of brass instruments like the trumpet and trombone is reviewed by Fletcher and Rossing,[2] who note that these instruments have complicated harmonic sound mode frequencies. The nonlinear combination of the buzzing-lip acoustic generator, the mouthpiece resonant frequency, and the compound shape of the instrument all contribute to this complication. They further note that higher harmonics play an increasing role as the loudness of the played note rises. These conclusions are based in part on the work of Backus and Hundley[3] and early work by Beauchamp.[4] Fletcher and Tarnopolsky[5] also measured trumpet sound spectra, revealing many peaks out to at least 10 kHz, even for softly-played notes. The recent work of Gilbert and Petiot[6] is also cited, and Msallam *et al.*[7] and Thompson and Strong[8] have presented nonlinear models of wave propagation in a trombone.

Shock waves are detectable by the sharp rise in static pressure, density, and temperature that occurs across them.[9] While fast sensors can record these changes, the schlieren optical method renders the shock itself visible by way of the gradient it causes in the refractive index of the air.[10]

We have studied shock wave propagation from the horn flare of a *B*-flat trumpet played by the first author of this paper, using a large-aperture, high-sensitivity schlieren system.[10] We further recorded schlieren images, sound pressure levels, mouthpiece pressure waveforms, and microphone spectra for a range of trumpet notes and volume levels. From these data, shock wave strength and propagation, additional flow phenomena occurring at the exit of the trumpet, and the relationship between visible shock radiation and sound spectrum data are addressed.

II. EXPERIMENTAL METHODS

The schlieren system[10] is based on a 1-m aperture parabolic telescope mirror, shown schematically in Fig. 1. Briefly, a conical light beam is projected onto the mirror and refocused by it to form a precise image of the light source slit. Small changes in the refractive index of the air along the 8-m optical path cause light rays to shift at the focus, where a vertical ''knife-edge'' is used as a detector. There results a grayscale image—the schlieren image—depicting refractive-index gradients in the otherwise-transparent air. Shock waves are very thin and have sharp gradients that show up as distinct lines in such an image. For present purposes digital videotape, a CCD still camera, or 35 mm film are used to record the images. A microsecond xenon arc lamp provides schlieren illumination brief enough to ''freeze'' the shock motion.

The schlieren optical response to a shock wave is approximated[10] by $\varepsilon = 200\,(M_s - 1)$, where M_s is the shock Mach number and ε is the resulting refraction angle in arcseconds. The maximum sensitivity of the present schlieren system was earlier estimated to be $\varepsilon = 0.03$ arcseconds, but this already remarkably high sensitivity is augmented further by the sharp visual contrast that occurs across a shock. Thus we expect to see very weak but nonetheless finite-amplitude shocks in the nonlinear acoustic range, e.g., those due to power tools, engines, the bursting of a toy balloon, or even the clapping of hands. At this schlieren sensitivity level, slight concentric-ring anomalies in the mirror figure and room air currents are also seen.

[a]Author to whom correspondence should be addressed. Electronic mail: gss2@psu.edu

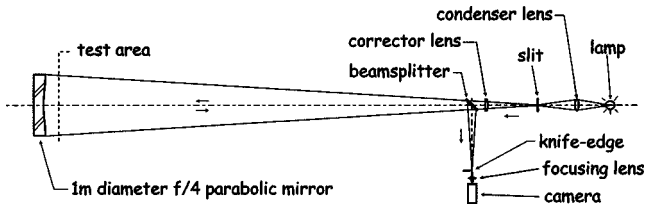


FIG. 1. Schematic of 1-m aperture schlieren optical system, from Ref. 10.

Trumpet rms sound pressure levels in dB(A) are read by a Norsonic Sound Analyzer Nor-110 located 1.23 m forward of the trumpet flare exit and offset at a 5° angle from the axis of the trumpet. (Note, for comparison, that shock waves are visible in the schlieren image over distances between 0 and about 0.75 m beyond the exit of the trumpet flare.) Microphone data are also taken at the 1.23-m location, though the nature of this experiment—being performed directly in front of a large, almost-flat mirror—precludes the usual anechoic-chamber treatment to avoid reverberation. Quantitative data are also obtained from a pressure tap at the trumpet mouthpiece constriction connected to a pressure transducer. No attempt is made to alter the mouthpiece or trumpet-bore configurations in these experiments.

III. RESULTS AND DISCUSSION

A parametric test program revealed many combinations of trumpet note and loudness for which shock waves were visible by schlieren observation. While high-pitched notes (above 700 Hz or the concert F_5 note) at a fortissimo level are the most visible, the fortissimo ''concert *B*-flat-4'' (trumpet middle *C*, 466 Hz) also produces a visible shock wave, as does the high *C* (trumpet C_6 or concert *B*-flat-5, 932 Hz) at mezzo-forte levels. Even high-pitched notes played at mezzo-piano levels reveal visible shocks.

Taken together, these results are summarized as follows: The threshold rms sound pressure level required for shock wave visibility in the present experimental apparatus is approximately 100 dB(A), measured at 1.23-m distance as described earlier. Near the center of the schlieren image, or half the distance from the trumpet horn flare to the sound level meter, threshold waves are expected to have 106 dB(A) peak overpressures, i.e., about 4 Pa above ambient pressure. (This threshold, of course, pertains to our rather-unique schlieren system and would be higher for smaller, less-sensitive optics.) Unfortunately only rms and not peak pressures were recorded, but from digitized microphone data the peaks due to shock waves are found to be about four times the rms signal level. Thus we might expect a 16-Pa peak overpressure or 118 dB(A) due to shock waves that are just visible to our optics. The corresponding shock wave Mach number is only 1.000 07 and the effective threshold refraction angle ε is a remarkably-small 0.014 arcsecond.

It is further observed that the most visible shock waves result from the combination of high blowing strength (resulting in a loud note) and lip tightness (yielding a high pitch). The optimum combination of these factors was found in the G_5 (trumpet high *G* or concert F_5, 698 Hz) note, which also yields the highest measured rms sound pressure level [106 dB(A) at the meter, or 112 dB(A) in the schlieren image]. This is as expected from precedent research, since the sound radiation efficiency of the horn flare increases with higher note frequency.[1–3]

Still further, within the current experimental range several different note frequency and volume combinations yield similar sound pressure levels and similarly visible shock waves. A fortissimo middle *C* and a mezzo-forte high *G* both produce roughly a measured 100 dB(A) rms level at 1.23 m, and are both near the shock wave visibility threshold. Examples of the weakest, intermediate, and strongest visible shock waves are given in the schlieren images of Fig. 2.

These schlieren images reveal spherical moving shock waves (not standing waves, nor the constant-strength waves produced in a shock tube). Thus the shock strength is higher inside the trumpet than at the present external observing and measuring stations, and the radiated shock waves rapidly decay toward linear sound waves with increasing distance from the trumpet.

The quantitative data, not shown here for brevity, reveal excellent agreement between the dominant frequency of the mouthpiece pressure and the fundamental upon which the microphone spectra are based, which in turn equals the nominal frequency of the note being played within an error of only 0.6%. Likewise the microphone pressure waveforms show strong leading spikes near the same fundamental frequency.

Mouthpiece pressure waveforms corresponding to the schlieren images of Fig. 2 also show features similar to those given by previous authors.[1,3–5] The microphone spectra corresponding to Figs. 2(a) and (c) (both high-G notes) differ mostly in amplitude and its redistribution to higher frequencies in the louder case, which is somewhat richer in spectral peaks due to nonlinear harmonic coupling in the bore of the instrument.[1,2,7] The gradual decrease in harmonic peak level above 1 kHz, observed by Fletcher and Tarnopolsky,[5] is seen here only for the softer note.

In Fig. 3 are shown microphone waveforms for fortissimo middle C and high G notes. The mezzo-forte high G waveform is similar to the latter, but with less pronounced spikes. These spikes, occurring at the leading edge of each cycle, suggest shock-wave-like behavior. The waveforms also contain a curious amplitude-modulation tremulant at 50–100 Hz, which may be peculiar to our trumpet player.

In some of the present schlieren images, two shock waves appear in the same image (e.g., Fig. 4). This provides an opportunity to estimate a frequency of shock wave motion radiated by the trumpet. For the fortissimo high G note, several such schlieren frames show two shocks separated by a distance of about $\frac{1}{2}$ m. Given the room temperature of 296 K, elementary gas dynamics[9] yields an acoustic speed of 345 m/s. Traveling essentially at the acoustic speed, the observed weak shocks thus have a frequency around 690 Hz, which is quite close to the 698-Hz nominal frequency of the note. Invoking the trumpet/shock tube analogy first suggested by Hirschberg,[1] a shock tube driven by a constant diaphragm or ''piston'' frequency is expected to expel shock waves at this same frequency. We thus expect that the shock waves observed in the schlieren results are driven directly by the per-

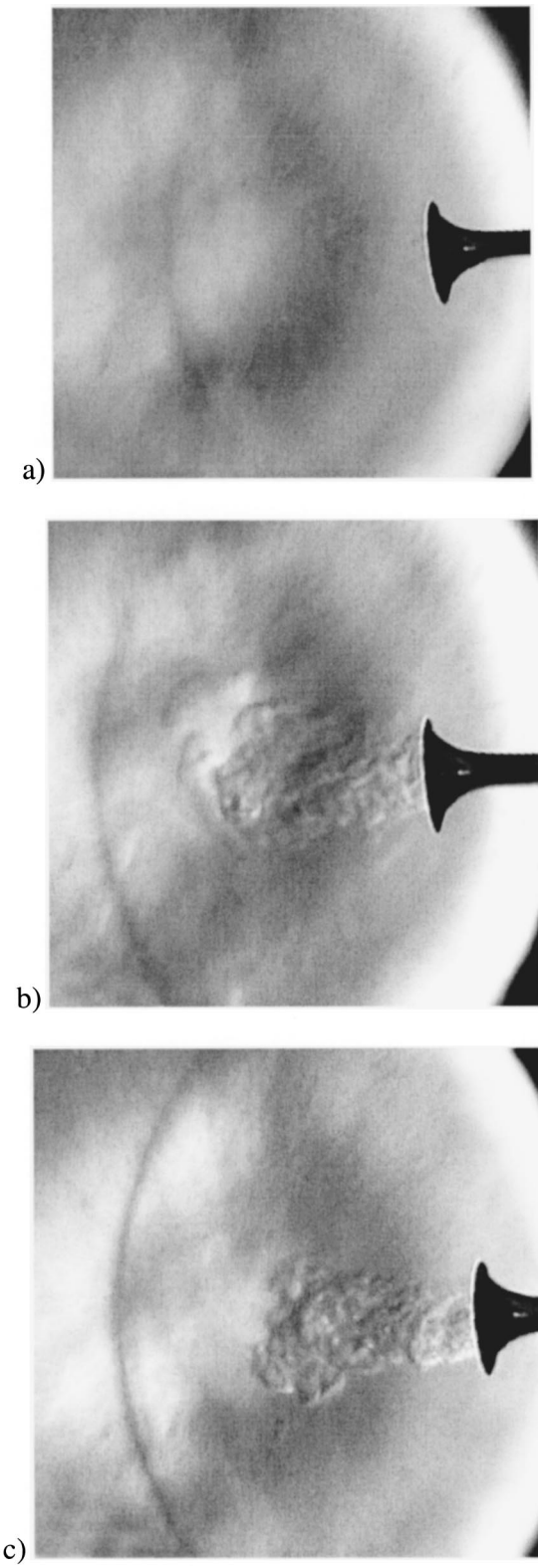


FIG. 2. Schlieren images of shock waves observed while playing trumpet notes (a) mezzo-forte high G, 99 dB(A) rms measured, (b) fortissimo middle C, 105 dB(A) rms measured, and (c) fortissimo high G, 106 dB(A) rms measured. For dimensional reference, the outer diameter of the trumpet flare is 12.2 cm.

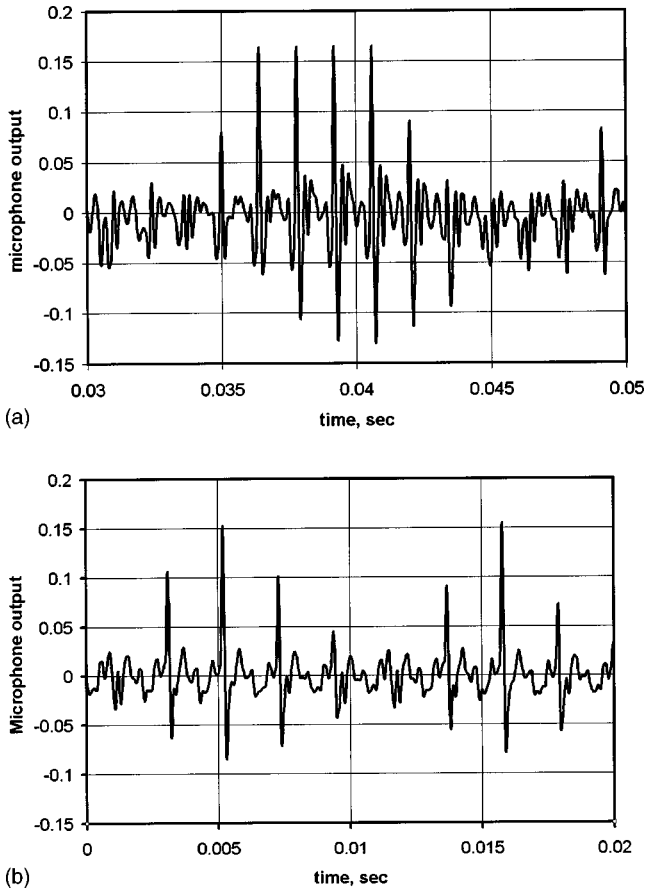


FIG. 3. Microphone waveforms for (a) fortissimo middle C and (b) fortissimo high G.

former's lip buzzing frequency and its interaction with the mouthpiece and the cylindrical section of the trumpet.

The spherical shocks in our schlieren images fade out of visibility within a half-angle of 25°–60° from the trumpet axis, being most visible on-axis. This is due, at least in part, to the vertical schlieren knife-edge cutoff, which renders most visible those disturbances that refract light with a strong horizontal component. Thus all spherical shock waves, no matter how strong, must disappear from the schlieren image as the angle from the trumpet axis ap-

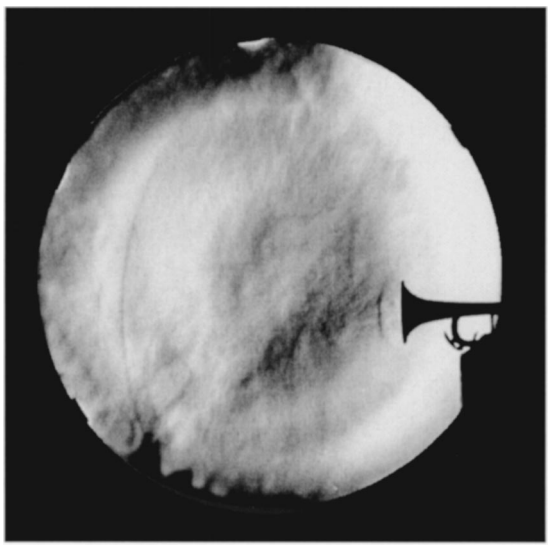


FIG. 4. Schlieren image of two shock waves, captured while playing a fortissimo high G note at a measured rms sound pressure level of 106.4 dB(A).

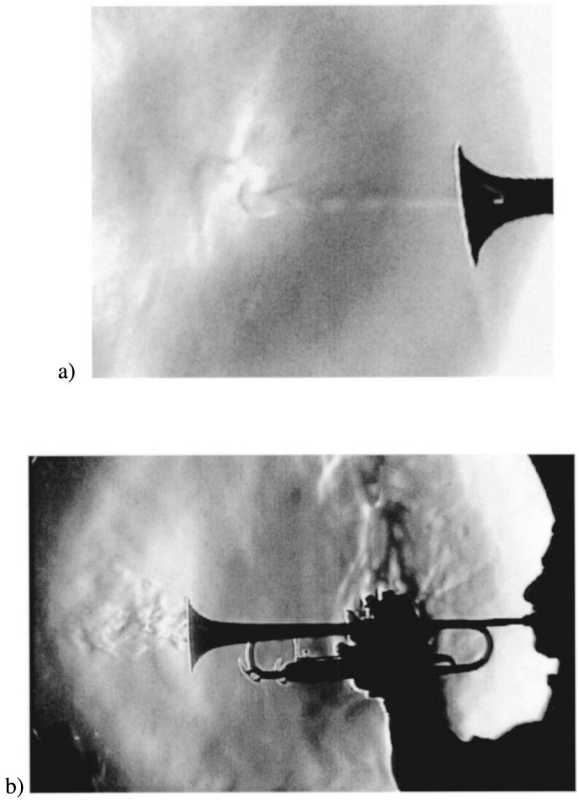


FIG. 5. Schlieren images of trumpet exit flow while playing (a) pianissimo middle C, 80 dB(A) measured and (b) fortissimo high C with small visible shock wave.

proaches 90°. In addition to this lateral loss of schlieren sensitivity, however, shock wave strength itself is likely to decrease off the trumpet axis, since brass instruments are highly directional.

Beyond shock waves, the present schlieren images also reveal low-speed fluid-dynamic phenomena at the trumpet flare exit because the performer's breath, exiting there, is warmer than ambient temperature. While soft trumpet notes produce a laminar jet of performer's breath from the trumpet [Fig. 5(a)], the louder notes always produce turbulent jet flow [Figs. 2(b) and (c), 4, and 5(b)]. Equally interesting is the difference in diameter of the laminar and turbulent jets in these images: the laminar jet is narrow and comparable to the diameter of the straight trumpet section, i.e., about 2.1 cm, while the turbulent jet is at least 4 cm and sometimes 10 cm in diameter.

These observations suggest that the performer's breath undergoes transition from laminar to turbulent pipe flow in the cylindrical bore of the trumpet when loud notes are played. The trumpet horn flare is designed to radiate acoustic energy, but also acts as a wide-angle diffuser to the exiting stream of performer's breath. For soft notes, the speed of the breath through the instrument is low and the flow remains laminar. When this ''laminar pipe flow'' encounters the rapid horn flare divergence, it separates from the duct wall due to the resulting adverse pressure gradient (such flow separation in wide-angle diffusers is treated in introductory fluid mechanics texts, e.g., Ref. 11). Thus a laminar jet of diameter comparable to the trumpet bore is seen emerging from the instrument in Fig. 5(a). Approximating the jet speed by analyzing the corresponding video footage, we find a value of about 0.35 m/s. The corresponding Reynolds number is only 490, well below the traditional criterion for pipe-flow transition.[11]

Loud trumpet notes are another matter, though, for the jet of performer's breath is now seen to exit the trumpet as turbulent flow with a larger jet diameter [Figs. 2(b) and (c), 4, and 5(b)]. Here the jet speed of about $\frac{1}{2}$ m/s corresponds to a Reynolds number of 2600 in the cylindrical trumpet bore, comparable to the expected transition value for simple pipe flow.

Turbulent flow is much more resistant to adverse pressure gradients than laminar flow, thus the exiting stream of performer's breath remains internally attached to trumpet flare until a diameter twice that of the cylindrical bore or larger before finally separating. This explains the observation that the visible turbulent jet is larger in diameter than is the laminar case.

None of these observations of laminar versus turbulent exit flow from the instrument affect the acoustics directly, but they do reveal how the acoustic design of the instrument influences its low-speed fluid-dynamic behavior.

IV. CONCLUSIONS

It is shown that weak-but-visible shock waves occur outside a trumpet when a bright, brassy note is played, whether it be a low-to-mid-pitched fortissimo note or a high-pitched note played at as low as a mezzo-forte level. Though visible as shock waves beyond the exit of the instrument with sensitive schlieren imaging, these waves no longer qualify as shock waves upon reaching an audience at a distance. Indeed, with Mach numbers as low as 1.000 07, the observed shock waves are barely supersonic and represent the loud limit of the sound continuum produced by the trumpet. Quantitative results for representative examples all show behavior dominated by the fundamental frequency of the note being played. Schlieren evidence further suggests that the note frequency is also the frequency of the emitted shock waves. Finally, interesting patterns of laminar and turbulent breath flow exiting the trumpet are also observed by way of schlieren imaging.

ACKNOWLEDGMENTS

The help of L. J. Dreibelbis, L. E. Fisher, M. Doty, Dr. W.-C. Chen, and Dr. S. R. Turns of Penn State, and several useful discussions with Dr. A. Hirschberg, Dr. J. Gilbert, and Dr. J. F. Petiot are gratefully acknowledged. We also acknowledge the suggestions of two anonymous reviewers and an important critique by Dr. N. H. Fletcher.

[1] A. Hirschberg, J. Gilbert, R. Msallam, and A. P. J. Wijnands, ''Shock waves in trombones,'' J. Acoust. Soc. Am. **99**, 1754–1758 (1996).
[2] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer, New York, 1998), Chaps. 8, 13, and 14.
[3] J. Backus and T. C. Hundley, ''Harmonic generation in the trumpet,'' J. Acoust. Soc. Am. **49**, 509–518 (1971).
[4] J. W. Beauchamp, ''Analysis of simultaneous mouthpiece and output waveforms,'' Audio Eng. Soc. Preprint 1626 (1980).
[5] N. H. Fletcher and A. Tarnopolsky, ''Blowing pressure, power, and spectrum in trumpet playing,'' J. Acoust. Soc. Am. **105**, 874–881 (1999).

[6] J. Gilbert and J. F. Petiot, ''Brass instruments, some theoretical and experimental results,'' in *Proc. Inst. Acoustics*, ISMA'97 (1997), Vol. **19**, Book 2, 391–400.
[7] R. Msallam, S. Dequidt, R. Causse, and S. Tassart, ''Physical model of the trombone including nonlinear effects. Application to the sound synthesis of loud tones,'' Acustica **86**, 725–736 (2000).
[8] M. W. Thompson and W. J. Strong, ''Inclusion of wave steepening in a frequency-domain model of trombone sound production,'' J. Acoust. Soc. Am. **110**, 556–562 (2001).
[9] J. E. A. John, *Gas Dynamics*, 2nd ed. (Prentice–Hall, Englewood Cliffs, NJ, 1984).
[10] G. S. Settles, *Schlieren and Shadowgraph Techniques* (Springer, Heidelburg, 2001).
[11] F. M. White, *Fluid Mechanics*, 4th ed (McGraw-Hill, Boston, 1999).

Time-domain simulation of a guitar: Model and method[a)]

Grégoire Derveaux[b)]
INRIA Roquencourt, Projet Ondes, 78153 Le Chesnay Cedex, France

Antoine Chaigne[c)]
ENSTA, Chemin de la Hunière, 91761 Palaiseau Cedex, France

Patrick Joly and Eliane Bécache
INRIA Roquencourt, Projet Ondes, 78153 Le Chesnay Cedex, France

(Received 8 June 2003; revised 29 September 2003; accepted 3 October 2003)

This paper presents a three-dimensional time-domain numerical model of the vibration and acoustic radiation from a guitar. The model involves the transverse displacement of the string excited by a force pulse, the flexural motion of the soundboard, and the sound radiation. A specific spectral method is used for solving the Kirchhoff–Love's dynamic top plate model for a damped, heterogeneous orthotropic material. The air–plate interaction is solved with a fictitious domain method, and a conservative scheme is used for the time discretization. Frequency analysis is performed on the simulated sound pressure and plate velocity waveforms in order to evaluate quantitatively the transfer of energy through the various components of the coupled system: from the string to the soundboard and from the soundboard to the air. The effects of some structural changes in soundboard thickness and cavity volume on the produced sounds are presented and discussed. Simulations of the same guitar in three different cases are also performed: "*in vacuo*," in air with a perfectly rigid top plate, and in air with an elastic top plate. This allows comparisons between structural, acoustic, and structural–acoustic modes of the instrument. Finally, attention is paid to the evolution with time of the spatial pressure field. This shows, in particular, the complex evolution of the directivity pattern in the near field of the instrument, especially during the attack.
 [DOI: 10.1121/1.1629302]

PACS numbers: 43.75.Gh, 43.40.Rj [NHF] Pages: 3368–3383

I. INTRODUCTION

The scope of the reported study is the simulation of the three-dimensional sound-pressure field of a guitar in the time domain. In a guitar, the elastic waves in strings and soundboard, and the acoustic waves in both external air and cavity, are continuously evolving with time. Both fluid and structural waves interact continuously so that the system composed by the instrument, the cavity, and the surrounding air should be considered as a whole. This approach, which is essential for a good understanding of the time-varying vibroacoustical phenomena involved in a guitar, has been made possible by the use of advanced numerical methods.

The method used consists of modeling the vibratory and acoustical phenomena involved in the instrument from the initial pluck to the radiation in a three-dimensional space. Since we wish to focus on the modeling of the soundboard and on the fluid–structure interaction, the model used for the other parts of the instrument is intentionally kept simple. Thus, an idealized plucking force is acting on a one-dimensional damped string model. The motion of the string is assumed to be perpendicular to the top plate. The soundboard itself is more sophisticated, in order to account for its crucial role in the sound of a guitar. It is modeled as an orthotropic heterogeneous damped Kirchhoff–Love plate, with a hole, and clamped at its boundaries. The other constitutive parts of the guitar (back, sides, neck) are assumed to be rigid. These latter assumptions could be eventually revisited in future studies. The plate radiates both inside the cavity and in the external free field. The cavity communicates with the external field through the hole.

In its present state, the purpose of the work is to show the feasibility and the main principles of a complete guitar modeling. No comparison with a particular instrument has been made. The values of the various physical and geometrical parameters used in the simulations correspond to realistic values extracted from the existing literature on the guitar.

The complexity of the system makes it necessary to use different numerical techniques for solving the problem. The plate equation is solved with a specific spectral method. A fictitious domain method, similar to the one previously used by Rhaouti *et al.* for timpani, is used for solving the fluid–structure interaction.[1] Finally, a conservative scheme is used for the time discretization.

The paper starts with the presentation of the guitar model, which is followed by the numerical formulation of the problem. The fourth section is devoted to the presentation of selected preliminary results whose aim is to show the

[a)]Preliminary results to the present study were first reported at: Forum Acusticum, Séville, Spain (*Time-domain simulation of a guitar: Preliminary results*, Forum Acusticum 2002, Special Issue of the Revista de Acustica, Vol. 33) and 144th Acoustical Society of America Meeting, Cancun, Mexico ["Time-domain simulation of a guitar," J. Acoust. Soc. Am. **112**, 2409 (2002)]. The numerical aspects of the method were presented at the Second M.I.T. Conference on Computational Fluid and Solid Mechanics (June, 2003).

[b)]Present address: ENSTA-UMA, 32 Bd Victor, 75015 Paris, France.
[c)]Electronic mail: chaigne@ensta.fr

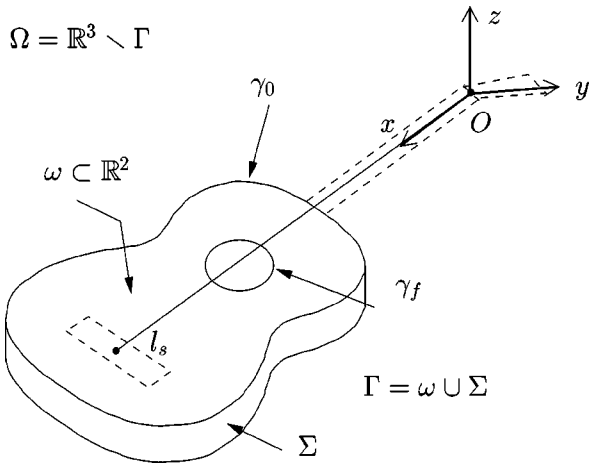


FIG. 1. Geometrical description of the guitar.

main attributes of the method. The detailed presentation of the application of the model to important features in the context of guitar acoustics, such as sound power and efficiency, is left for future studies.

With regard to previous work on guitars, one can summarize the originality of the present work as follows: First, the cavity model and the air–structural coupling is more general than in the boundary element method, developed by Richardson *et al.*[2] It is also a step forward compared to the spatial Fourier method developed by Le Pichon *et al.*, where the velocity distribution on the surface of the body was measured, and not calculated.[3] Second, the results obtained in the time domain with the present model should be compared to the results obtained in the frequency domain by Elejabarrieta *et al.*, who used commercial software.[4] In particular, the bracing used in our model is less accurate than in this latter paper. However, more emphasis is put here on the description of the air sound field and on the effects of the pressure jump on the top plate.

Finally, as far as the numerical technique is concerned, the present modeling is a direct extension to plates of the timpani model developed by Rhaouti *et al.*[1] The modeling of damping in the guitar top plate makes use also of results obtained by Chaigne and Lambourg.[5]

II. PRESENTATION OF THE MODEL

A. Description

The body of the guitar is delimited by a surface denoted Γ which is divided into two parts: $\Gamma = \omega \cup \Sigma$, where ω is the top plate of the instrument and Σ is the rest of the surface (i.e., sides and back). The boundary of ω itself is divided into two parts: γ_0 is the outer boundary of the top plate and γ_f is the inner boundary, along the hole. The surrounding air occupies the domain $\Omega = \mathbb{R}^3 \setminus \Gamma$. Ω corresponds to the internal cavity and the external domain, which communicate via the sound hole. The string of length l_s is rigidly fixed to the neck at a point denoted O, chosen as the origin of the coordinate system (see Fig. 1).

B. The string

The string is supposed to have uniform density ρ_s, uniform tension T, and no stiffness. In the following, only the vertical flexural motion of the soundboard will be considered. Therefore, it is also natural to consider only the vertical polarization of the string. The motion of the string is thus described by its transverse displacement (perpendicular to the soundboard) denoted $u_s(x,t)$, $x \in]0,l_s[$. Assuming that the amplitudes of vibration are sufficiently small, the string equation is a classical 1D damped wave equation

$$\rho_s \frac{\partial^2 u_s}{\partial t^2} - T\left(1 + \eta_s \frac{\partial}{\partial t}\right)\frac{\partial^2 u_s}{\partial x^2} + \rho_s R_s \frac{\partial u_s}{\partial t} = f_s(x,t), \quad \text{in }]0,l_s[. \tag{1}$$

The internal damping is modeled by two terms: a viscoelastic term with coefficient η_s whose essential purpose is to introduce a frequency-dependent damping in the string, and a fluid term with coefficient R_s whose purpose is to introduce a constant internal damping for low frequencies. The string is fixed at one end to the neck of the guitar, which is supposed to be perfectly rigid

$$u_s(0,t) = 0, \quad \forall t > 0. \tag{2}$$

At the bridge, the displacement of the string is assumed to be given by the displacement of the plate at the contact point between string and bridge, denoted (x_0, y_0)

$$u_s(l_s,t) = u_p(x_0,y_0,t), \quad \forall t > 0. \tag{3}$$

The plucking of the string is assumed to be an imposed force (see Fig. 2 and Table I)

$$f_s(x,t) = g(x)h(t), \tag{4}$$

where $h(t)$ represents a simple idealized version of the "stick–slip" mechanism that governs the interaction between string and finger. One pulls on the string during nearly

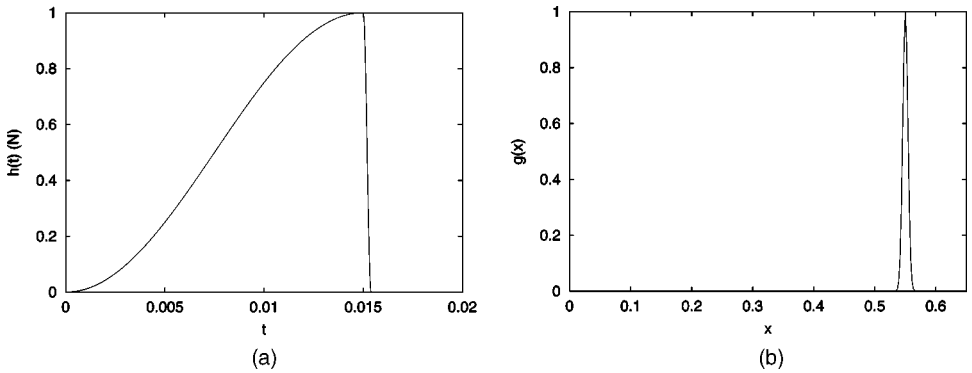


FIG. 2. Representation of the idealized plucking force. (a) Time dependence. (b) Space dependence.

TABLE I. Typical values of physical and numerical values used for the simulations.

String: $\rho_s = 0.005\ 25\ \mathrm{kg\cdot m^{-1}}$, $T = 60$ N, $l_s = 0.65$ m,
$R_s = 0.75\ \mathrm{s^{-1}}$, $\eta_s = 9\cdot 10^{-8}$ s, $(x_0, y_0) = (65\ \mathrm{cm}, 4\ \mathrm{cm})$

Soundboard:

	D_1 (MPa)	D_2 (MPa)	D_3 (MPa)	D_4 (MPa)	ρ_p ($\mathrm{kg\cdot m^{-1}}$)	a (mm)	R_p ($\mathrm{s^{-1}}$)	η_p (s)
Plate	850	50	75	200	350	2.9	7	0.005
Bridge	80	50	900	270	400	6	7	0.005
Struts	100	60	1250	300	400	14	7	0.005

Air: $c_a = 344\ \mathrm{m\cdot s^{-1}}$, $\rho_a = 1.21\ \mathrm{kg\cdot m^{-3}}$

Numerical information:

String:	308 nodes	$h_c = 2.1$ mm
Plate:	3230 nodes	$h_p = 1.2$ cm
Pressure jump:	1260 nodes	$h_\lambda = 1.2$ cm
Acoustic field:	≈400 000 cubes	$h_a = 1.1$ cm
Time step:	$\Delta t = 2.10^{-5}$ s	
Sampling frequency:	$f_e = 50\ 000$ Hz	

Plucking force: $f_s(x,t) = g(x)h(t)$.

$$g(x) = \frac{\exp(-(x-x_0/\delta_s)^2)}{\int_0^{l_s}\exp(-(x-x_0/\delta_s)^2)} \qquad h(t) = \begin{cases} (1-\cos(\pi t/t_1)), & 0 \leqslant t \leqslant t_1, \\ (1+\cos(\pi(t-t_1)/t_2)), & t_1 \leqslant t \leqslant t_2, \\ 0, & t > t_2. \end{cases}$$

$x_0 = 55$ cm, $\delta_c = 0.006$ m, $t_1 = 0.015$ s, $t_2 = 0.000\ 4$ s.

15 ms and releases it during 0.4 ms. This force is distributed over a small segment of the string by means of the smooth positive function g, normalized so that $\int_0^{l_s} g(x)dx = 1$. Despite its simplicity, this excitation is in fairly good agreement with experiments.[6]

C. The soundboard

It is assumed that the only vibrating part of the guitar is the soundboard. The other parts of the guitar body (back, sides, neck) are assumed to be perfectly rigid. If necessary, the motion of the back could be introduced straightaway in this model since it is of exactly the same nature as the top plate. The soundboard is a thin, orthotropic wooden layer, with a soundhole. The thickness is relatively small (about 3 mm) compared to the dimensions of the soundboard (about 50 cm long). In terms of spectral content, our study is limited to the bandwidth 0–3000 Hz. For both these reasons, it is not necessary to use a thick plate model, such as the Reissner–Mindlin one. The flexural Kirchhoff–Love's plate model is thus appropriate to describe the vibrations of the soundboard.

The strutting system and the bridge are considered as heterogeneities of the soundboard. As a consequence, the following physical parameters which characterize the top plate are variable functions of space: density $\rho_p(x,y)$, thickness $a(x,y)$, and rigidity tensor for orthotropic material $\mathbf{C}(x,y)$. In what follows, the space dependence of these parameters will be omitted, in order to simplify the notations.

Finally, the in-plane motion of the soundboard is neglected compared to its flexural motion. The motion of the strutted soundboard is thus completely described by the transverse displacement of the top plate, denoted $u_p(x,y,t)$, $(x,y) \in \omega$. This motion is governed by the following equation:[7–10]

$$a\rho_p \frac{\partial^2 u_p}{\partial t^2} + \left(1 + \eta_p \frac{\partial}{\partial t}\right) \mathrm{div}\,\underline{\mathrm{Div}}\ a^3 \mathbf{C}\underline{\underline{\varepsilon}}(\underline{\nabla} u_p) + a\rho_p R_p \frac{\partial u_p}{\partial t} = \mathcal{F} - [p]_\omega, \quad \text{in } \omega, \tag{5}$$

where $\underline{\mathrm{Div}}$ is the divergence operator for tensors: $(\underline{\mathrm{Div}}.\tau) = \partial_j \tau_{ij}$, $\underline{\underline{\varepsilon}}$ is the plane linearized strain tensor $[\varepsilon_{\alpha\beta}(\underline{\theta}) = \frac{1}{2}(\partial_\beta \theta_\alpha + \partial_\alpha \theta_\beta)]$, div and $\underline{\nabla}$ denote the usual divergence operator and gradient operator for vectors, respectively (see Appendix A for details and more explicit notations).

As for the string, the internal damping in the plate is modeled by two terms: a viscoelastic term with coefficient η_p and a fluid term with coefficient R_p. Notice, however, that since a spectral method is used for the numerical resolution of this equation, it becomes possible to introduce other laws of damping, which do not need to have so simple a counterpart in the time domain. This point will be developed in Sec. III C.

The top plate is clamped on γ_0

$$u_p = 0 \ \text{ and } \ \partial_n u_p = 0, \quad \text{on} \quad \gamma_0, \tag{6}$$

whereas the boundary is free along the hole γ_f (see Appendix A for explicit notations in particular cases)

$$\begin{aligned} &(\underline{\underline{\mathcal{M}}}\,\underline{n}) \cdot \underline{n} = 0, \quad \text{on} \quad \gamma_f, \\ &(\underline{\mathrm{Div}}\,\underline{\underline{\mathcal{M}}}) \cdot \underline{n} + \partial_\tau [(\underline{\underline{\mathcal{M}}}\,\underline{n}) \cdot \underline{\tau}] = 0, \quad \text{on} \quad \gamma_f, \end{aligned} \tag{7}$$

where $\underline{\underline{\mathcal{M}}} = a^3 \mathbf{C}\underline{\underline{\varepsilon}}(\underline{\nabla} u_p)$ is the bending moment. $\underline{n}$ denotes the outer normal and $\underline{\tau}$ the tangential vector along the boundary γ_f. The two conditions in Eq. (7) express approximately that force and moment vanish at the free boundary.

The force density term exerted on the plate by the surrounding air is $[p]_\omega$, the pressure jump across the soundboard. $\mathcal{F}$ denotes the force density exerted by the string on the bridge. It is assumed to be the normal component of the

tension of the string at this point. This force is applied at the contact point (x_0, y_0) between string and bridge

$$\mathcal{F}(x,y,t) \approx -T\partial_x u_s(l_s,t)\,\delta_{x_0,y_0}(x,y). \tag{8}$$

D. The acoustic field

The acoustic field is governed by the linearized Euler's equations

$$\frac{\partial p}{\partial t} = -c_a^2 \rho_a \,\mathrm{div}(\underline{\mathbf{v}}_a) \quad \text{in } \Omega, \tag{9a}$$

$$\rho_a \frac{\partial \underline{\mathbf{v}}_a}{\partial t} = -\underline{\nabla} p \quad \text{in } \Omega, \tag{9b}$$

where c_a is the speed of the sound in air, ρ_a is the density of air, p is the sound pressure in Ω, and $\underline{\mathbf{v}}_a$ the acoustic velocity in Ω. These equations are complemented by a condition of continuity for the normal component of the velocity at the surface ω of the plate

$$\underline{\mathbf{v}}_a(x,y,0,t)\cdot \mathbf{e}_z = \partial_t u_p(x,y,t), \quad \forall (x,y)\in\omega, \quad \forall t>0. \tag{10}$$

In addition, as the body of the guitar is assumed to be perfectly rigid, one has

$$\underline{\mathbf{v}}_a(x,y,z,t)\cdot \mathbf{N}_\Gamma = 0, \quad \forall (x,y,z)\in\Gamma, \quad \forall t>0, \tag{11}$$

where $\mathbf{N}_\Gamma$ denotes the outer normal to the boundary Σ.

E. Energy decay

The system of coupled equations (1) to (11) fulfills a property of energy decay in the case of free oscillations. The respective energies of the string E^s, plate E^p, and sound field E^a, are given by

$$\begin{aligned}
E^s(t) &= \frac{1}{2}\int_0^{l_s} \rho_s \left|\frac{du_s}{dt}\right|^2 dx + \frac{1}{2}\int_0^{l_s} T\left|\frac{du_s}{dx}\right|^2 dx,\\
E^p(t) &= \frac{1}{2}\int\int_\omega \rho_p a \left|\frac{du_p}{dt}\right|^2 dx\,dy\\
&\quad + \frac{1}{2}\int\int_\omega a^3 \mathbf{C}\underline{\underline{\varepsilon}}(\underline{\nabla} u_p):\underline{\underline{\varepsilon}}(\underline{\nabla} u_p)dx\,dy,\\
E^a(t) &= \frac{1}{2}\int\int\int_\Omega \rho_a |\underline{\mathbf{v}}_a|^2 dx\,dy\,dz\\
&\quad + \frac{1}{2}\int\int\int_\Omega \frac{1}{c_a^2\rho_a}|p|^2 dx\,dy\,dz.
\end{aligned} \tag{12}$$

It is easy to show that the total energy of the system $E = E^s + E^p + E^a$ fulfills the following identity property:

$$\begin{aligned}
\frac{dE}{dt}(t) &= \int_0^{l_s} f_s \frac{du_s}{dt}dx - \int_0^{l_s} \eta_s T|\partial_{x,t} u_s|^2 dx\\
&\quad - \int_0^{l_s} \rho_s R_s |\partial_t u_s|^2 dx\\
&\quad - \int\int_\omega \eta_p a^3 \mathbf{C}\underline{\underline{\varepsilon}}(\underline{\nabla}\partial_t u_p):\underline{\underline{\varepsilon}}(\underline{\nabla}\partial_t u_p)dx\,dy\\
&\quad - \int\int_\omega a\rho_p R_p |\partial_t u_p|^2 dx\,dy.
\end{aligned} \tag{13}$$

In addition to its physical meaning, this property guarantees the wellposedness of the continuous model and will be exploited for ensuring the stability of the numerical scheme, as it will be shown in the next section.

III. NUMERICAL RESOLUTION

Let us begin with a brief description of the main difficulties we have to face in order to solve these equations numerically. The 3D nature of the computational domain for the sound radiation yields a large size problem. This domain is unbounded, so that it is necessary to reduce it to a finite size problem, in order to make it computable. The geometry of the instrument is complex and includes third-order boundary conditions [Eq. (7)]. Furthermore, the space- and time discretization of the Kirchhoff–Love equation requires an adapted resolution method, because this equation includes a fourth-order space operator and is intrinsically dispersive. Last but not least, one has to derive a stable resolution scheme.

The problem is first rewritten as an equivalent variational formulation, which is the first step of the finite element method. A variational formulation corresponds to the virtual work principle which is naturally related to the continuous energy identity [Eq. (13)]. The conforming space approximation of the variational formulation of the problem leads to a discrete energy identity, which is a convenient approach to derive the stability of the resolution scheme. The main aspects of the numerical resolution are the following:

(i) The plate equation is solved by a modal decomposition method (Secs. III A 1 and III C).
(ii) The fluid–structure interaction problem is solved with the use of the fictitious domain method (Sec. III A 2).
(iii) One of the main originalities of the method is a stable coupling between a continuous time resolution and a discrete one (Sec. III C).

In what follows, for the sake of simplicity, the damping in the plate is described by a single fluid damping term, and there is no damping in the string. It does not add formal difficulty to take into account the other damping terms presented in Eqs. (1) and (5).

A. Variational formulation

1. The plate

A well-known difficulty for the resolution of the Kirchhoff–Love's problem is due to the presence of a fourth-order space operator. This operator would require *a priori* the use of sophisticated finite elements of class C^1, with continuity of the normal derivative along each edge of the plate triangulation. To circumvent this difficulty, the space operator is split into two second-order operators. We introduce the velocity $v_p = \partial_t u_p$ and the bending moment $\underline{\underline{\mathcal{M}}} = \underline{\underline{\varepsilon}}(\underline{\nabla} u_p)$, which leads to the following equivalent problem of second order in space and first order in time:

$$a\rho_p \partial_t v_p + \operatorname{div} \underline{\operatorname{Div}} \underline{\underline{\mathcal{M}}} + a\rho_p R_p v_p = \mathcal{F} - [p]_\omega, \quad \text{in } \omega, \tag{14a}$$

$$\partial_t \underline{\underline{\mathcal{M}}} - a^3 \mathbf{C} \underline{\underline{\varepsilon}}(\underline{\nabla} v_p) = 0, \quad \text{in } \omega, \tag{14b}$$

$$v_p = \partial_n v_p = 0, \quad \text{on } \gamma_0, \tag{14c}$$

to which the conditions on the free boundary given by Eq. (7) are added. The variational formulation of Eq. (14) is obtained by multiplying Eq. (14a) by a test function $v_p^* : \omega \to \mathbb{R}$ and by multiplying equation Eq. (14b) by a test function $\underline{\underline{\mathcal{M}}}^* : \omega \to \mathbb{R}^4$. Using then an integration by parts on ω leads to the following mixed formulation:

Find $v_p(t) : \omega \to \mathbb{R}$ and $\underline{\underline{\mathcal{M}}}(t) : \omega \to \mathbb{R}^4$ such that

$$\frac{d}{dt} \int_\omega a\rho_p v_p v_p^* - \int_\omega \underline{\operatorname{Div}} \underline{\underline{\mathcal{M}}} \cdot \underline{\nabla} v_p^* - \int_{\gamma_f} \partial_\tau [(\underline{\underline{\mathcal{M}}} \underline{n}) \cdot \underline{\tau}] v_p^* + \int_\omega a\rho_p R_p v_p v_p^* = -T \partial_x u_s(l_s, t) v_p^*(x_0, y_0) - \int_\omega [p]_\omega v_p^*, \quad \forall v_p^*, \tag{15a}$$

$$\frac{d}{dt} \int_\omega a^{-3} \mathbf{C}^{-1} \underline{\underline{\mathcal{M}}} : \underline{\underline{\mathcal{M}}}^* + \int_\omega \underline{\operatorname{Div}} \underline{\underline{\mathcal{M}}}^* \cdot \underline{\nabla} v_p + \int_{\gamma_f} \partial_\tau [(\underline{\underline{\mathcal{M}}}^* \underline{n}) \cdot \underline{\tau}] v_p = 0, \quad \forall \underline{\underline{\mathcal{M}}}^*. \tag{15b}$$

The main interest of this formulation is that it can be approximated using standard Lagrange finite elements (see Sec. III B).

2. The fluid–structure interaction

With regards to the large size of the problem, the finite difference method would certainly be an efficient approach for solving the acoustic equation in terms of computational time, since it relies on the use of a regular mesh made of small cubes. Unfortunately, this choice leads to a staircase approximation of the complex geometry of the guitar which generates spurious diffractions.[11] Furthermore, the construction of a coupling scheme with the finite element method used for the plate problem is not straightforward because of the incompatibility of the meshes. As an alternative, the finite element method allows one to take into account the shape of the instrument accurately. However, this method requires the construction of a tetrahedric mesh and leads to a quite expensive resolution scheme, with regard to computing time and memory.

In practice, we used a fictitious domain method, first introduced by Glowinsky,[12] and which has been successfully applied to kettledrums by Rhaouti *et al.*[1] The fictitious domain method preserves the advantages of both previously mentioned methods: the efficiency of the finite difference method in terms of computing time, and the good approximation of the shape of the guitar. Its leading idea is to reformulate the problem in a domain of simple geometry which ignores the instrument. The fluid–structure interaction is then taken into account via the introduction of the new unknown variable $\lambda = [p]_\Gamma = p_e - p_i$, defined on Γ which denotes the pressure jump across the boundary of the instrument.

Multiplying Eq. (9a) by a test function $p^* : \mathbb{R}^3 \to \mathbb{R}$, and Eq. (9b) by a test function $\underline{\mathbf{v}}_a^* : \mathbb{R}^3 \to \mathbb{R}^3$, and using the Green formula leads to Eqs. (16a) and (16b). It remains to take into account the boundary conditions given by Eqs. (10) and (11) on the surface of the instrument, which is done in a weak sense [Eq. 16(c)]. This leads to the following problem:

Find $p(t) : \mathbb{R}^3 \to \mathbb{R}$, $\underline{\mathbf{v}}_a(t) :: \mathbb{R}^3 \to \mathbb{R}^3$, and $\lambda(t) : \Gamma \to \mathbb{R}$ such that

$$\frac{d}{dt} \int_{\mathbb{R}^3} \frac{1}{\rho_a c_a^2} p p^* + \int_{\mathbb{R}^3} p^* \operatorname{div} \underline{\mathbf{v}}_a = 0, \quad \forall p^*, \tag{16a}$$

$$\frac{d}{dt} \int_{\mathbb{R}^3} \rho_a \underline{\mathbf{v}}_a \cdot \underline{\mathbf{v}}_a^* - \int_{\mathbb{R}^3} p \operatorname{div} \underline{\mathbf{v}}_a^* - \int_\Gamma (\underline{\mathbf{v}}_a^* \cdot \underline{N}) \lambda = 0, \quad \forall \underline{\mathbf{v}}_a^*, \tag{16b}$$

$$\int_\omega v_p \lambda^* - \int_\Gamma (\underline{\mathbf{v}}_a \cdot \underline{N}) \lambda^* = 0, \quad \forall \lambda^*. \tag{16c}$$

In Eq. (16), $\underline{\mathbf{v}}_a$ and p are defined in the complete domain $\mathbb{R}^3$, and not only in the domain $\Omega = \mathbb{R}^3 \backslash \Gamma$. It will then be possible to use a regular mesh for the approximation of the acoustic field. It turns out that this choice leads in fact to a finite difference scheme, which is the major interest of the fictitious domain method. In addition, the geometry of the instrument is well fitted, since it relies on a triangular mesh of its surface for the approximation of λ, which is easy to construct.

3. The string

Because the plate velocity v_p has been introduced in order to solve the Kirchhoff–Love's problem [Eq. (14)], one has to differentiate in time the boundary condition at the bridge given by Eq. (3). This leads naturally to introduce the string velocity $v_s = \partial_t u_s$. It can be shown that in order to preserve the energy identity [Eq. (13)], one also has to introduce the constraint $q = T \partial_x u_s$, which appears, in particular, in the expression of the force exerted by the string on the plate [Eq. (8)]. The following problem of first order in space and first order in time is obtained:

$$\begin{aligned} &\rho_s \partial_t v_s - \partial_x q = f_s, \quad \text{in }]0, l_s[\\ &\partial_t q - T \partial_x v_s = 0, \quad \text{in } 0, l_s[\\ &v_s(0,t) = 0, \quad \forall t \\ &v_s(l_s, t) = v_p(x_0, y_0), \quad \forall t. \end{aligned} \tag{17}$$

The variational formulation of Eq. (17) is obtained by multiplying the first equation by a test function $v_s^* :]0, l_s[\to \mathbb{R}$, and the second by a test function $q^* :]0, l_s[\to \mathbb{R}$, and using integration by part. This leads to the following problem:

Find $v_s : [0, l_s[\to \mathbb{R}$ and $q :]0, l_s[\to \mathbb{R}$ such that

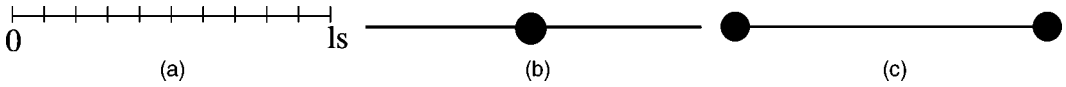


FIG. 3. Mesh for (v_{s_h}, q_h). (a) Mesh of $]0, l_s[$. (b) Degrees of freedom for v_s. (c) Degrees of freedom for q.

$$\frac{d}{dt}\int_0^{l_s} \rho_s v_s v_s^* - \int_0^{l_s} \partial_x q v_s^* = \int_0^{l_s} f_s v_s^*, \quad \forall v_s^*, \tag{18a}$$

$$\frac{d}{dt}\int_0^{l_s} \frac{1}{T} q q^* + \int_0^{l_s} \partial_x q^* v_s - q^*(l_s,t) v_p(x_0,y_0) = 0, \quad \forall q^* \tag{18b}$$

4. Variational formulation of the complete problem

Considering the additional unknowns λ and q, which were introduced in order to solve the acoustic equation and the string equation, respectively, one has to replace $T\partial_x u_s(l_s,t)$ by $q(l_s,t)$ and $[p]_\omega$ by $\lambda_{|\omega}$ in Eq. (15a). Equation (15a) is thus replaced by

$$\frac{d}{dt}\int_\omega \rho_p v_p v_p^* - \int_\omega \underline{\mathrm{Div}}\underline{\underline{\mathcal{M}}} \cdot \underline{\nabla} v_p^* - \int_{\gamma_f} \partial_\tau[(\underline{\underline{\mathcal{M}}}\underline{n}) \cdot \underline{\tau}] v_p^* + \int_\omega \rho_p R_p v_p v_p^* = -q(l_s,t) v_p^*(x_0,y_0) - \int_\omega \lambda_{|\omega} v_p^*, \quad \forall v_p^* : \omega \to \mathbb{R}. \tag{19}$$

Finally, the variational formulation of the complete problem is given by the set of equations (15b), (16), (18), (19).

B. Space discretization

In this variational formulation, the unknowns $(v_s, q, v_p, \underline{\underline{\mathcal{M}}}, p, \underline{\mathbf{v}}_a, \lambda)$ and the test functions $(v_s^*, q^*, v_p^*, \underline{\underline{\mathcal{M}}}^*, p^*, \underline{\mathbf{v}}_a^*, \lambda^*)$ are chosen in appropriate spaces so that all integrals are defined. q has to be chosen in a space of continuous functions because of the term $\partial_x q$ in Eq. (18). For similar reasons, v_p, $\underline{\underline{\mathcal{M}}}$, and λ will also be chosen in a space of continuous functions. On the other hand, the spaces for v_s and for p do not require any condition of continuity. Finally, $\underline{\mathbf{v}}_a$ is in a space of vector field with continuity of the normal component across the surface, because of the term $\mathrm{div}\,\underline{\mathbf{v}}_a$ in Eq. (16).

The finite dimensional spaces chosen for the approximation of the variational formulation (15b), (16), (18), (19) and satisfying the adequate continuity conditions are described below. In the following P_k denotes the polynomial functions of one or more variables of degree less than or equal to k.

The continuous variables v_s and q are approximated by discrete variables in space denoted v_{s_h} and q_h, respectively. A regular mesh of the string, made of small segments of length h_c is constructed [Fig. 3(a)]. v_{s_h} is a piecewise constant and is thus entirely determined by its values at the center of each segment, which are called the degrees of freedom (dof). In other words v_{s_h} is discretized using P_0 discontinuous finite elements. q_h is a continuous piecewise linear function and it is thus fully determined by its values at each node of the mesh (P_1 continuous finite elements). The degrees of freedom for v_{s_h} and q_h are represented in Figs. 3(b) and (c), respectively.

Given a triangular mesh of the soundboard ω, with the smallest edge length denoted h_p [Fig. 4(a)], the variables v_p and $\underline{\underline{\mathcal{M}}}$ are discretized by means of P_2 alike-continuous finite elements.[13] This particular choice reduces the numerical dispersion and leads to the required estimation of the eigenfrequencies of the soundboard.[13] With regard to sound synthesis it is an important aim to simulate the eigenfrequencies of the soundboard with sufficient accuracy, in view of the ability of the human ear to detect small frequency variations. In this context, an order of magnitude of 1% seems a reasonable target. An analysis of the performance of finite element approximations of different orders for the plate equation is presented in Appendix B. Another advantage of using P_2 elements lies in the elimination of the unknown $\underline{\underline{\mathcal{M}}}_h$ in the final scheme, as it will be shown in Sec. III C.

The discrete variables v_{p_h} and $\underline{\underline{\mathcal{M}}}_h$ are determined by their values at the vertex, at the gravity center of each triangle, and at the center of each edge. The degrees of freedom for v_{p_h} and $\underline{\underline{\mathcal{M}}}_h$ are represented in Figs. 4(b) and (c), respectively.

To simulate the free space, the computations are restricted to a box of finite size denoted $\mathcal{C}$, with the help of higher-order absorbing boundary conditions defined on the faces of this cube.[1,14] For the sake of conciseness, these additional equations $\mathcal{C}$ are not given here. Given a regular mesh of $\mathcal{C}$, made of small cubes with edge length denoted h_a [Fig. 5(a)], p is discretized by means of P_0-discontinuous finite

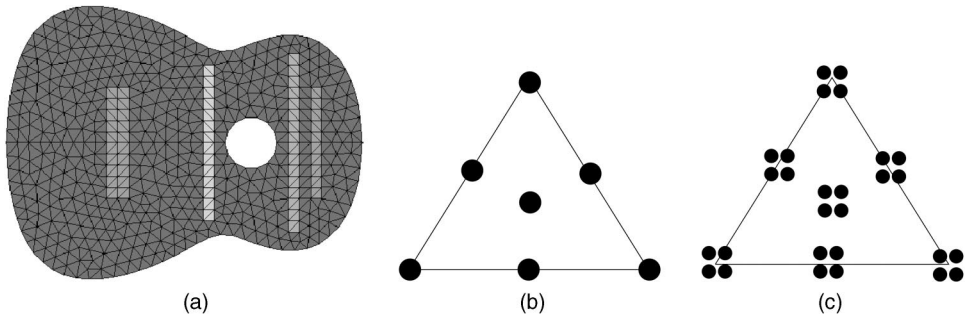


FIG. 4. Mesh for $(v_{p_h}, \underline{\underline{\mathcal{M}}}_h)$. (a) Mesh of ω. (b) Degrees of freedom for v_{p_h}. (c) Degrees of freedom for $\underline{\underline{\mathcal{M}}}_h$.

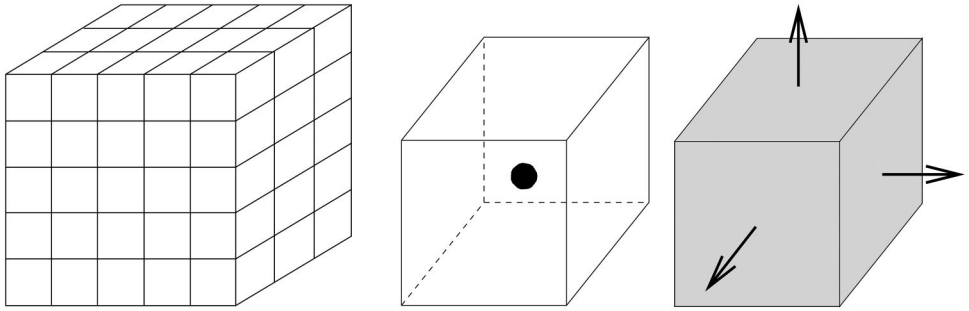


FIG. 5. Mesh for $(p_h, \underline{\mathbf{v}}_{a_h})$. (a) Mesh of $\mathcal{C}$. (b) Degrees of freedom for p_h. (c) Degrees of freedom for $\underline{\mathbf{v}}_{a_h}$.

elements. The field p_h is thus determined by its values at the center of each cube. The approximation of $\underline{\mathbf{v}}_a$ is made with the first-order Raviart–Thomas finite element space.[15] Each component of $\underline{\mathbf{v}}_{a_h}$ is linear in one direction and constant in the other two directions. $\underline{\mathbf{v}}_{a_h}$ is entirely determined by the value of its normal component across each face of the cubes. The degrees of freedom for p_h and $\underline{\mathbf{v}}_{a_h}$ are represented in Figs. 5(b) and (c), respectively.

Finally, a triangular mesh is given on the surface of the guitar Γ, with the smallest edge length denoted h_λ [see Fig. 6(a)]. The pressure jump λ is discretized by means of P_1 continuous finite elements. Its approximation λ_h is determined by the values at the vertex of each triangle. The degrees of freedom of λ_h are represented in Fig. 6(b).

The space discretization of the problem is then obtained by substituting $(v_{s_h}, q_h, v_{p_h}, \underline{\underline{\mathcal{M}}}_h, p_h, \underline{\mathbf{v}}_{a_h}, \lambda_h)$ for $(v_s, q, v_p, \underline{\underline{\mathcal{M}}}, p, \underline{\mathbf{v}}_a, \lambda)$ and $(v_{s_h}^*, q_h^*, v_{p_h}^*, \underline{\underline{\mathcal{M}}}_h^*, p_h^*, \underline{\mathbf{v}}_{a_h}^*, \lambda_h^*)$ for $(v_s^*, q^*, v_p^*, \underline{\underline{\mathcal{M}}}^*, p^*, \underline{\mathbf{v}}_a^*, \lambda^*)$ in the variational formulation (15b), (16), (18), (19). The resulting differential system can be written in the following matrix form:

$$M_h^p \frac{dv_{p_h}}{dt} - H_h^{\mathsf{T}} \underline{\underline{\mathcal{M}}}_h + R_p M_h^p v_{p_h} = -J_h q_h - (B_{\omega_h})^{\mathsf{T}} \lambda_h, \tag{20a}$$

$$M_h^{\mathcal{M}} \frac{d\underline{\underline{\mathcal{M}}}_h}{dt} + H_h v_{p_h} = 0, \tag{20b}$$

$$M_h^s \frac{dv_{s_h}}{dt} - D_h q_h = f_{s_h}, \tag{20c}$$

$$M_h^q \frac{dq_h}{dt} + D_h^{\mathsf{T}} v_{s_h} - J_h^{\mathsf{T}} v_{p_h} = 0, \tag{20d}$$

$$M_h^a \frac{d\underline{\mathbf{v}}_{a_h}}{dt} - G_h p_h - (B_{\Gamma_h})^{\mathsf{T}} \lambda_h = 0, \tag{20e}$$

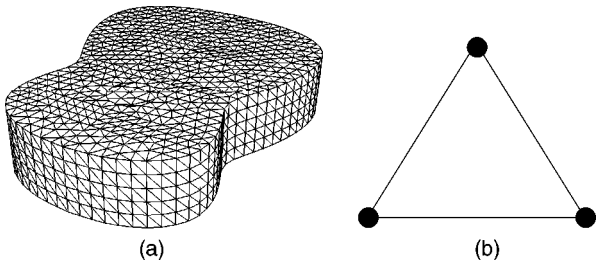


FIG. 6. Mesh for λ_h. (a) Mesh of Γ. (b) Degrees of freedom for λ_h.

$$M_h^{p_a} \frac{dp_h}{dt} + G_h^{\mathsf{T}} \underline{\mathbf{v}}_{a_h} = 0, \tag{20f}$$

$$B_{\omega_h} v_{p_h} - B_{\Gamma_h} \underline{\mathbf{v}}_{a_h} = 0, \tag{20g}$$

where A_h^{T} denotes the transpose of a matrix A_h. M_h^s, M_h^q, $M_h^{\mathcal{M}}$, M_h^a, and $M_h^{p_a}$ are mass matrices. f_{s_h} denotes a vector which approximates the plucking force exerted on the string. D_h, J_h, H_h, C_h, B_{ω_h}, G_h, and B_{Γ_h} are matrices which represent the discrete operators approximating the continuous operators of the variational formulation (15b), (16), (18), (19). G_h, for example, is the approximation of the three-dimensional gradient operator $\underline{\nabla}$.

C. Time discretization

For the time discretization, a classical explicit centered finite difference scheme of second order is used for the string and for the acoustic field. For the plate equation, it appears that the choice of an explicit finite difference scheme in time imposes the selection of a small time step Δt because of a restricting stability condition of the form $\Delta t \leqslant C h_p^2$, where C is a constant depending on the physical parameters of the plate. A natural idea would be to choose an implicit, unconditionally stable scheme, allowing a larger Δt. However, in this case the analysis of the error made on the eigenfrequencies of the soundboard shows that, for accuracy considerations, one has to choose Δt proportional to h_p^2 as for the explicit scheme.[16] For this reason, it has been finally decided to solve the plate equation analytically in time, which allows one to choose Δt without degrading the accuracy of the semidiscretized spatial scheme.

In a first step, the size of the plate system is reduced by eliminating the bending moment $\underline{\underline{\mathcal{M}}}_h$ in Eqs. (20a) and (20b), which leads to

$$M_h^p \frac{d^2 v_{p_h}}{dt^2} + R_p M_h^p \frac{dv_{p_h}}{dt} + K_h v_{p_h} = -J_h \frac{dq_h}{dt} - B_{\omega_h}^{\mathsf{T}} \frac{d\lambda_h}{dt}, \quad \forall t \geqslant 0, \tag{21}$$

where K_h is the matrix defined by

$$K_h = H_h (M_h^{\mathcal{M}})^{-1} H_h^{\mathsf{T}}. \tag{22}$$

The matrix $M_h^{\mathcal{M}}$ is computed using quadrature formula so that it reduces to a 4×4 block diagonal matrix, without loss of accuracy. This technique, called mass lumping, permits

one to invert $M_h^{\mathcal{M}}$, and, consequently, permits one to compute K_h easily.[13]

For the resolution of Eq. (21), one has to sample the solution. Given a time step Δt, v_{p_h} and $\dot{v}_{p_h}=dv_{ph}/dt$ are calculated at the successive instants of time $t^{n+(1/2)}=(n+\frac{1}{2})\Delta t$. Equation (21) is solved on each interval $[t^{n-(1/2)},t^{n+(1/2)}]$. Since the string and acoustic equations of the problem are solved by means of finite differences with time step Δt, the right-hand side is constant on this interval. The following equation is thus solved at each time step:

$$M_h^p\frac{d^2v_{p_h}}{dt^2}+R_pM_h^p\frac{dv_{p_h}}{dt}+K_hv_{p_h}=-J_h\frac{q_h^{n+1}-q^{n-1}}{2\Delta t}-B_{\omega_h}^{\mathsf{T}}\frac{\lambda_h^{n+(1/2)}-\lambda_h^{n-(1/2)}}{\Delta t},\quad \forall t\in\,]t^{n-(1/2)},t^{n+(1/2)}[,$$

$$v_{p_h}(t^{n-(1/2)})=v_{p_h}^{n-(1/2)}\ \text{ and }\ \dot{v}_{p_h}(t^{n-(1/2)})=\dot{v}_{p_h}^{n-(1/2)}. \tag{23}$$

The resolution of Eq. (23) relies on the computation of the eigenmodes of the positive definite matrix K_h, and on the ordinary differential equation theory. In the absence of damping, the solution is given by

$$v_{p_h}^{n+(1/2)}=\left[\cos(\sqrt{K_h}\Delta t)v_{p_h}^{n-(1/2)}+\frac{\sin(\sqrt{K_h}\Delta t)}{\sqrt{K_h}}\dot{v}_{p_h}^{n-(1/2)}\right]+\frac{I-\cos(\sqrt{K_h}\Delta t)}{\sqrt{K_h}}\left(-J_h\frac{q_h^{n+1}-q^{n-1}}{2\Delta t}-B_{\omega_h}^{\mathsf{T}}\frac{\lambda_h^{n+(1/2)}-\lambda_h^{n-(1/2)}}{\Delta t}\right), \tag{24}$$

with a similar equation for $\dot{v}_{p_h}^{n+(1/2)}$. For our computations, only the first 50 modes of the plate were retained, which corresponds to a cutoff frequency nearly equal to 3 kHz.

The introduction of the damping terms leads in fact to exponentially damped sinusoidal solutions for each modal component. It is a well-known fact that the fluid and viscoelastic terms introduced in the model [R_p and η_p in Eq. (5)] lead to a damping factor $\alpha_n=(R_p+\eta_p\omega_n^2)/2$ for the nth eigenmode of K_h with pulsation ω_n. At this stage, notice that it is possible to choose any other frequency dependence for the damping factors, without requiring that it has a local expression in the time domain. This is one interest of the spectral method which allows one to model the internal damping in the plate with such a flexibility. Considering the damping ratio values for wooden plate given in Ref. 5 in the range [0–3 kHz] corresponding to our case, a linear law has been chosen, of the form

$$\alpha_n=(R_p+\eta_p\omega_n)/2. \tag{25}$$

One must now write a finite difference scheme for the time discretization of Eqs. (20c)–(20g). The crucial point is to propose a stable coupling between two radically different time-resolution techniques: a continuous one for the plate equation and a discrete one for the string and air equations. The scheme is chosen in order to obtain a discrete energy identity similar to Eq. (13). In addition, we look for a scheme in which almost all computations are explicit (i.e., without matrix inversion), in particular for the 3D part. These two reasons lead to computing the variables v_{s_h}, p_h, and λ_h at time $t^{n+(1/2)}$, while q_h and $\underline{\mathbf{v}}_{a_h}$ are computed at time t^n, and to discretize Eq. (20g) which is now differentiated in time. The following system is obtained:

$$\begin{aligned}
&M_h^s\frac{v_{s_h}^{n+(1/2)}-v_{s_h}^{n-(1/2)}}{\Delta t}-D_hq_h^n=f_{s_h}^n,\\
&M_h^q\frac{q_h^{n+1}-q_h^n}{\Delta t}+D_h^{\mathsf{T}}v_{s_h}^{n+(1/2)}-J_h^{\mathsf{T}}v_{p_h}^{n+(1/2)}=0,\\
&M_h^a\frac{\underline{\mathbf{v}}_{a_h}^{n+1}-\underline{\mathbf{v}}_{a_h}^n}{\Delta t}-G_hp_h^{n+(1/2)}-B_{\Gamma_h}^{\mathsf{T}}\lambda_h^{n+(1/2)}=0,\\
&M_h^{p_a}\frac{p_h^{n+(1/2)}-p_h^{n-(1/2)}}{\Delta t}+G_h\underline{\mathbf{v}}_{a_h}^n=0,\\
&B_{\omega_h}\frac{v_{p_h}^{n+(1/2)}-v_{p_h}^{n-(1/2)}}{\Delta t}-B_{\Gamma_h}\frac{\underline{\mathbf{v}}_{a_h}^{n+1}-\underline{\mathbf{v}}_{a_h}^{n-1}}{2\Delta t}=0.
\end{aligned} \tag{26}$$

The resolution of the linear system consisting of Eqs. (24) and (26) imposes, in particular, to invert the mass matrices M_h^s, M_h^q, M_h^a, and $M_h^{p_a}$. To avoid this inversion at each time step, these matrices are reduced to diagonal matrices using the mass-lumping technique [see Eq. (22)].

We seek for $(v_{s_h}^{n+(1/2)},q_h^{n+1},v_{p_h}^{n+(1/2)},\dot{v}_{p_h}^{n+(1/2)},\underline{\mathbf{v}}_{a_h}^{n+1},p_h^{n+(1/2)},\lambda_h^{n+(1/2)})$, with all other terms being known. It is easy to show that if $\lambda^{n+(1/2)}$ is known, the rest of the computation is entirely explicit. For the computation of $\lambda^{n+(1/2)}$, one has to invert the matrix C_λ, where

$$C_\lambda=\frac{2B_{\omega_h}}{\Delta t}\left[I+J_h\left(\frac{1-\cos(\sqrt{K_h}\Delta t)}{\sqrt{K_h}}\right)J_h^{\mathsf{T}}\right]^{-1}B_{\omega_h}^{\mathsf{T}}+\Delta tB_{\Gamma_h}(M_h^a)^{-1}B_{\Gamma_h}^{\mathsf{T}}. \tag{27}$$

C_λ is symmetric and positive without any condition on the time step. Furthermore, this matrix is definite, provided that the following compatibility between the mesh of $\mathcal{C}$ and the mesh of Γ is satisfied:

$$\frac{h_\lambda}{h_a}\geqslant\alpha, \tag{28}$$

where α is a constant. This compatibility condition, which is not demonstrated here, is a consequence of the consistency of the fictious domain method.[17] In practice, α is approximately equal to 1.1.

The stability of the scheme given by Eqs. (24)–(26) is guaranteed with the help of an energy identity similar to Eq. (13). The details of the proof are beyond the scope of this paper. This leads to the following conditions which have to be satisfied for ensuring the stability:

$$\text{(i)}\quad \sqrt{\frac{T}{\rho_s}}\frac{\Delta t}{h_c}\leqslant 1, \tag{29}$$

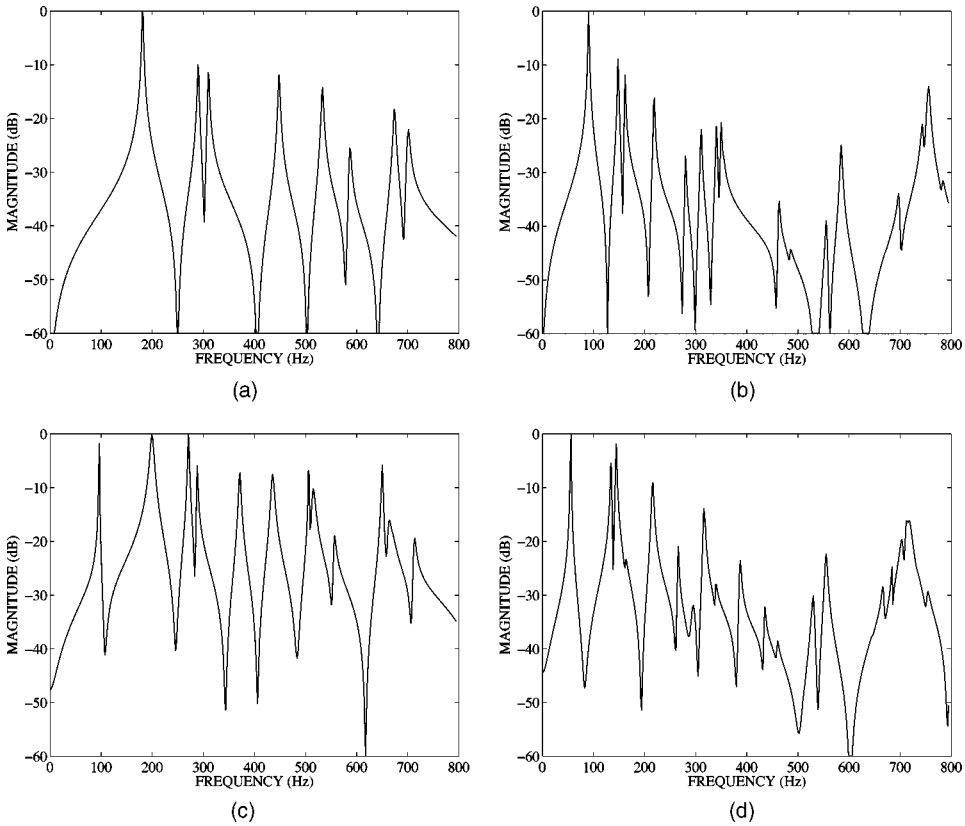


FIG. 7. Admittance curves at the bridge. A pulse force is applied to the lowest guitar string. The strings are damped. (a) Soundboard *in vacuo*, plate thickness 2.9 mm. (b) Soundboard *in vacuo*, plate thickness 1.45 mm. (c) Guitar coupled to the air, plate thickness 2.9 mm. (d) Guitar coupled to the air, plate thickness 1.45 mm.

$$\text{(ii)} \quad c_a \frac{\Delta t}{h_a} \leqslant \frac{1}{\sqrt{3}}.$$

One should notice that these two conditions are optimal, in the sense that they are identical to the usual stability conditions obtained for the standard finite difference discretization of the uncoupled 1D and 3D wave equation. This result shows the robustness of the coupling scheme.

The numerical parameters chosen for the numerical simulations presented in the following section and which fulfill the conditions Eqs. (28) and (29) are given in Table I.

IV. PRELIMINARY RESULTS AND DISCUSSION

The prime and most visible interest of our model is the computation of the complete 3D sound field radiated by the guitar (see Fig. 15). However, it is also necessary to validate our method through comparisons with existing results of the literature on guitars. In what follows, we start by presenting simulation results in terms of admittance, a quantity which is widely used for characterizing the essential coupling properties between strings and plate loaded by air in stringed instruments. In addition to the validation purpose, the specific interest of the presented results on admittances is to show the possible use of our model for comparing two different soundboards mounted on the same instrument. In our example, only the thickness is changed, all other parameters remaining identical. Such a modification would certainly be hard to realize on a real instrument. Finally, we show the differences, in terms of admittance and decay times, for a guitar body *in vacuo* compared to a guitar body in air. This clearly shows the various effects of both radiation and cavity field on the soundboard vibration.

Another interest of our model is the use of a variational formulation from which energetic quantities can be very easily derived. We use that property for gaining important information on the energy balance between strings, soundboard, and air.

The simulated sound pressure is observed in the time domain. Comparisons are made for two different values of box height and plate thickness, all other material and geometrical parameters remaining unchanged. The results show a clear difference between the waveforms. A spectral analysis is performed on the initial part of these simulated sounds. This analysis shows the role of both the string–body and air–body coupling in the spectral content and decay pattern of a guitar sound.

Finally, the spatial properties of the sound field are displayed in terms of instantaneous directivity patterns.

A. Admittance at the bridge

The admittance is defined as the ratio between velocity and force. In order to compute such a quantity, we use our model to simulate the velocity waveform consecutive to a force impulse $f(x,y,t)$ at one particular point at the bridge. In practice, the strings are damped and we use $f(x,y,t) = h(t)\delta_{x_0,y_0}(x,y)$, where $h(t)$ is given in Table I with $t_1 = t_2 = 5 \cdot 10^{-3}$ s. The selected point (x_0,y_0) corresponds to the attachment point of the lowest guitar string (6th string, note E_2). Using standard FFT tools, the admittance is simply

TABLE II. Analysis of simulated guitar sound pressure note E_2 (open 6th string) Damping factors and frequencies. The string partials are in bold. (a) Reference (plate thickness 2.9 mm, cavity height 10.4 cm). (b) Large cavity (plate thickness 2.9 mm, cavity height 21 cm). (c) Thin plate (plate thickness 1.45 mm, cavity height 10.4 cm).

(a) Reference		(b) Large cavity		(c) Thin plate	
F (Hz)	α (s^{-1})	F (Hz)	α (s^{-1})	F (Hz)	α (s^{-1})
82.4	**0.4**	76.7	1.6	55.5	1.6
96.6	2.1	**82.6**	**0.4**	**82.6**	**0.4**
164.7	**0.6**	**164.2**	**1.0**	135.0	4.8
200.6	20.8	183.3	14.9	144.4	6.5
247.8	**0.8**	**247.8**	**0.7**	**163.0**	**11.2**
272.3	7.9	274.7	8.3	166.4	1.7
289.0	7.0	293.0	7.7	216.4	11.9
330.6	**0.7**	**330.6**	**0.7**	**248.4**	**1.0**
372.9	12.9	355.6	7.2	266.7	7.4
413.0	**0.9**	**412.9**	**0.9**	296.1	30.0
437.6	19.0	436.2	15.0	317.0	8.7
495.6	**1.3**	**495.5**	**1.4**	319.0	8.3
508.7	1.6	505.6	1.1	**331.3**	**1.0**
517.1	21.9	515.0	19.7	340.4	8.7
558.3	15.6	**578.6**	**1.1**	388.5	12.0
578.5	**1.1**	656.5	11.4	**413.4**	**0.9**
652.7	9.3	**662.0**	**4.0**	436.6	9.7
661.1	**3.1**	672.1	24.4	461.6	12.4
672.7	26.7	707.4	7.6	**495.8**	**1.0**
716.0	13.0	**743.7**	**1.4**	531.2	8.2
743.7	**1.5**			**532.3**	**20.4**
				557.4	15.7
				578.6	**1.1**
				648.8	16.1
				660.6	**1.6**
				670.4	13.7
				693.4	39.4
				708.6	23.1
				720.5	29.1
				735.2	23.1
				744.4	**2.4**

obtained from the computation of the frequency response between velocity and force.

In order to illustrate the capabilities of our model, Fig. 7 shows typical results obtained in four different situations. The case (a) corresponds to the reference guitar "*in vacuo*." The reference guitar has a soundboard of thickness equal to 2.9 mm and a cavity height of 10.4 cm. In Fig. 7(b), the thickness is equal to 1.45 mm, all other parameters of the guitar remaining unchanged. Figures 7(c) and (d) correspond to the same configurations as (a) and (b), respectively, when the guitar vibrates in air.

As expected, the modal density is higher for case (b) than for case (a), and all frequencies are lowered (see also Table II). Figure 7(c) clearly shows additional peaks in the admittance curve, variation of the bandwidth for some peaks, and slight variations of the peak frequencies, compared to case (a). These phenomena are due to the coupling of the soundboard with cavity and external air. The frequency change of the two lowest peaks, in particular, follow the well-known rule[18]

$$f_+^2 + f_-^2 = f_1^2 + f_2^2, \tag{30}$$

where f_+ and f_- are the two lowest frequencies of the coupled system, whereas f_1 is the lowest acoustic eigenfrequency of the rigid cavity and f_2 is the lowest structural eigenfrequency of the soundboard. Here, we have $f_+ = 200.6$ Hz, $f_- = 96.6$ Hz, $f_1 = 117$ Hz, and $f_2 = 181.9$ Hz, which means that Eq. (30) is verified to within $\pm 5\%$. This discrepancy is mainly due to the fact that we do not have here an isolated system with two degrees of freedom, but a truncated system extracted from a continuous system. Frequency f_2 is obtained from the simulation of the complete guitar "*in vacuo*" (see Table II). Frequency f_1, like the other acoustic modes of the uncoupled system, is obtained from the simulation of the guitar coupled with external air and cavity while keeping the top plate rigidly fixed. Frequencies f_+ and f_- are derived from a spectral analysis performed either on the pressure or on the plate vibration, for the complete guitar.

The other lowest frequencies of the uncoupled acoustic modes are: 355, 699, 905, and 1046 Hz. It is interesting to notice that Eq. (30) also applies to other subsets of modes, provided that these subsets are reasonably well separated from the other modes. For the reference guitar, the coupled (called structural–acoustic or SA) modes at 373 and 437.6 Hz, for example, are due to the coupling between the acoustic mode at 355 Hz and the structural mode at 450 Hz.

Similar conclusions can be drawn from the comparison between Figs. 7(b) and (d), corresponding to the thin plate, with stronger variations of frequencies due to coupling in case (d) compared to case (c). In this latter case, notice that the modal density of the fluid–structure coupled system is high so that the maxima of the admittance curve cannot be well separated. This property also appears in case (b), though for higher frequencies than for case (d).

Finally, one should highlight the fact that the model yields admittance curves up to 3 kHz, as shown in Fig. 8. This is an improvement compared to previously published results on simulated guitars which are generally limited to less than 1 kHz.[19] These figures show, among other things, that the average slope for the mobility of the thin plate is relatively higher than the one of the reference soundboard. This suggests that the guitar with the thin soundboard is relatively more efficient in low frequencies compared to the high-frequency domain.

Detailed analysis of the simulated plate velocity at the junction between string and bridge, using the matrix-pencil method,[20] yields accurate values of the damping factors. Figures 9(a) and (b) show the effect of air loading and cavity on these factors for the soundboard vibration. For clarity, only the lowest modes, between 0 and 800 Hz, are presented. In this bandwidth, the admittance [Fig. 7(a)] shows 8 structural modes (*in vacuo* modes) from 182 Hz (1,1)-mode to 704 Hz (1,4)-mode. For these modes, the damping factors α (in s^{-1}) are only due to the material losses, which are assumed here to increase linearly with frequency [see Eq. (25)].

For the plate coupled with air [case 7(c)], 12 structural–acoustic modes (SA modes) were found. At this stage, notice that the matrix-pencil method yields accurate values of the eigenfrequencies and damping factors for the *in vacuo* soundboard, which validates the analysis. The damping factors for the air–body system are compared to those obtained from the analysis of the plate admittance *in vacuo*. These

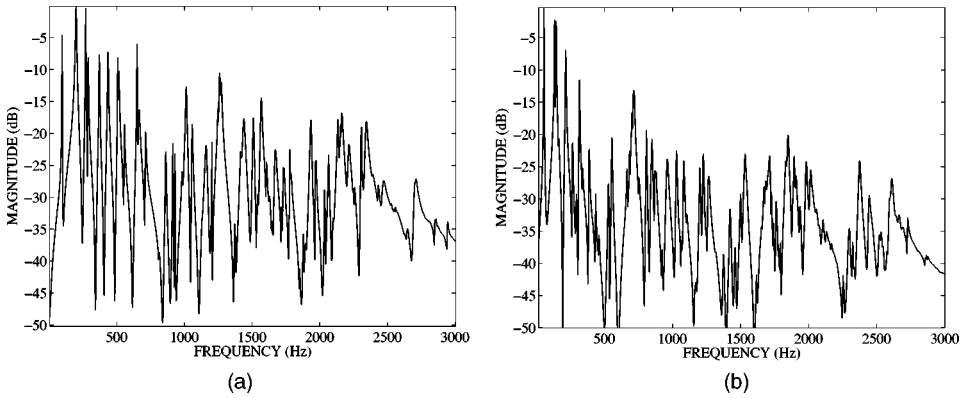


FIG. 8. Admittance curve at the bridge, in the interval [0–3000] Hz. (a) Reference guitar in air. (b) Guitar with thin plate in air.

factors are proportional to the sound power radiated by the instrument: high values of the damping coefficients correspond to strongly radiating modes.[21] For the reference case presented here, it can be seen, for example, that the SA modes at 201, 438, 517, and 673 Hz are efficient radiators.

For the reference guitar, it turns out that the Q-values, related to the damping factors α by the relation $Q=\pi f/\alpha$, are slightly higher in our simulation than those observed experimentally in real guitars (30–130 compared to 20–80), for the SA modes.[22] This should incite us to slightly increase the plate damping in future simulations in order to obtain sounds closer to real guitar sounds. Also, the fluid damping and the visco-thermal losses at the boundaries are not taken into account in our simulations. As a consequence, the simulated Q-values of some SA modes are high compared to real guitars, as it can be observed, for example, on the lowest structural–acoustic mode at 96.6 Hz.

In conclusion, the admittance analysis shows that, despite some crude aspects of the modeling (bracing, rigid back, air damping), our system is able to reproduce the main properties of a real guitar. In addition, the separate modeling of structural, acoustics, and structural–acoustic modes, respectively, not only validates the fictitious domain method, but also yields a powerful tool for the estimation and prediction of changes in radiated sound power consecutive to structural modifications of the instrument.

B. Energy exchange

Figure 10 shows the evolution of various energy quantities vs time for the plucked open string E_2 (fundamental 82.6 Hz).

As the string is in contact with the finger, most of the mechanical energy due to the plucking is transferred to the string. During this period of time, only a small amount of energy is transformed into quasistatic deformation of the plate and motion of the air. As the finger leaves the string, both plate and air energy increase almost instantaneously. The air-energy history is then dominated by almost periodic exchange with fundamental frequency equal to the fundamental frequency of the string. The plate-energy history is also periodic with the same fundamental frequency. For the case presented here, plate and air vibrations are in antiphase. The plate-energy history is also characterized by a modulation at a frequency nearly equal to the 5th harmonic of the fundamental. This property is due to the fact that the 5th harmonic of the string (around 413 Hz) is strongly coupled to an efficiently radiating structural–acoustic mode at 438 Hz.

After the pluck, the average magnitude of both plate and air energies corresponds approximately to 1 percent of the string energy. This proportion varies of course with the coupling properties between string and plate, and with the radiation efficiency of the instrument. However, the figure shown here corresponds to a typical order of magnitude for the lowest open string of a guitar, and this explains why guitar notes usual last for several seconds. As shown in Fig. 11, simulations made with the thickness of the top plate reduced by a factor of 2, all other parameters remaining unchanged, show a significant increase in the radiated sound power. In this case the average value of the air energy is equal to 10% of the string energy. As a consequence, the string vibrations are damped more rapidly.

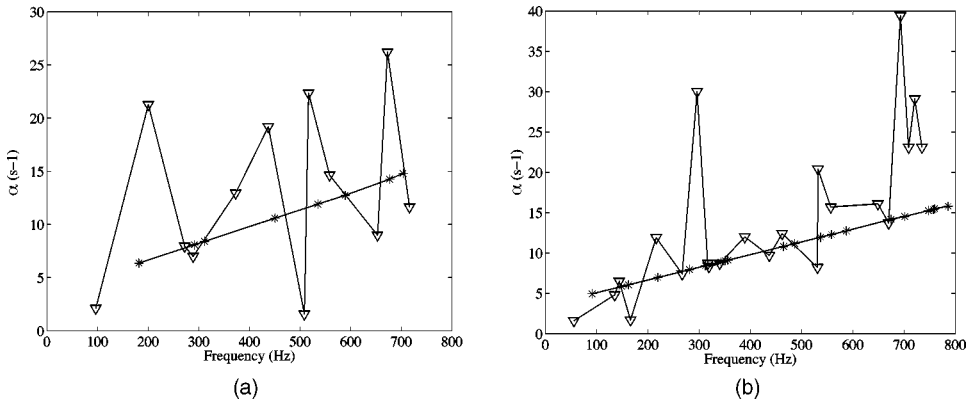


FIG. 9. Damping coefficients α (s^{-1}) for the lowest modes of the plate–air system (0–800 Hz). Comparison between structural and structural–acoustic modes. (a) Plate thickness 2.9 mm. Structural modes (*); structural–acoustic modes (∇). (b) Plate thickness 1.45 mm. Structural modes (*); structural–acoustic modes (∇).

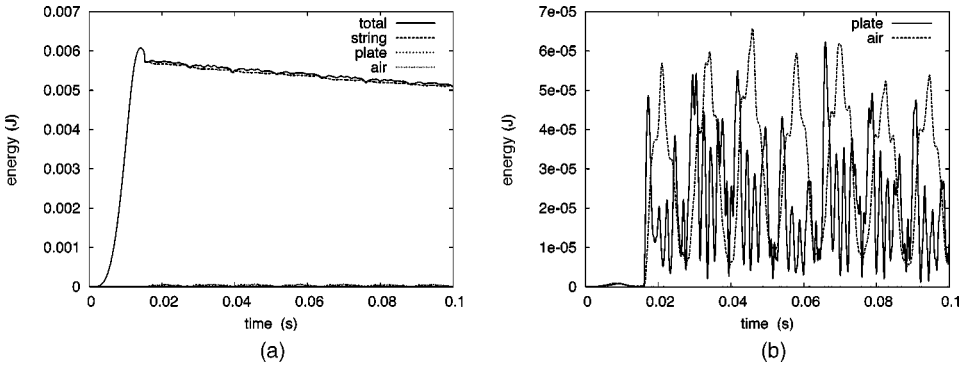


FIG. 10. (a) Evolution of the energy of string, plate, air, and total energy vs time for the plucked open string E_2 (fundamental 82.6 Hz) during the first 100 ms of the tone. Reference guitar. (b) Plate and air energy represented with appropriate scale.

C. The sound of the guitar in time

Examples of guitar pressure waveforms are simulated in three different situations, for a pluck on the open E_2 (6th) string. The first waveform shown in Fig. 12(a) corresponds to the reference guitar (see Table I). The waveform shown in Fig. 12(b) is obtained for a doubling of the cavity height, all other parameters remaining unchanged. Finally, the waveform shown in Fig. 12(c) corresponds to a top plate with thickness divided by a factor of 2, compared to the reference case, all other parameters remaining unchanged. For these three simulations, the tone duration is equal to 3 s. The pressure is simulated at a specific point in free space whose coordinates are $x=40$ cm, $y=5$ cm, and $z=60$ cm (see Fig. 1 for the definition of the axes).

The waveforms look clearly different. An accurate frequency analysis of these waveforms has been carried out with the help of the matrix-pencil method.[20] This analysis yields the frequencies and damping factors shown in Table II.

As expected, the pressure spectra are composed of string modes and of the SA modes of the guitar body coupled with air. All three spectra exhibit the nine lowest string modes and a variable number of SA modes, between 0 and 800 Hz. It can be seen that the coupling does not affect the frequencies and damping factors of the string modes significantly. As expected, the reference spectrum shows, in addition, the 12 SA modes observed in the admittance analysis. The damping factors of these modes are all significantly higher than those of the string modes, so that the well-known ''box sound'' in the guitar transient is damped more rapidly than the string components. The spectrum corresponding to the simulation with a large cavity shows 11 SA modes, in the same frequency range. Due to the doubling of the height some SA modes are strongly modified (such as the lowest one at 96.6 Hz, which becomes equal to 76.7 Hz). Other SA modes, which are mainly imposed by the lateral dimensions of the cavity, remain almost unchanged (272.3 and 289 Hz). For the simulated guitar sound with a thin plate, the pressure spectrum shows 22 SA modes between 0 and 800 Hz. This is, again, a result in accordance with the admittance analysis. As a consequence of the air–plate coupling, the lowest SA mode, in particular, is significantly reduced (55.5 Hz), compared to the reference case (96.6 Hz).

Figures 13 and 14 are aimed at summarizing the effects of the various coupling in the guitar through careful examination of the damping factors of the string's partials. Compared to Fig. 9, which illustrated the air–body coupling only, these data show that the sound of the guitar is governed by the combination of both air–body and string–body coupling.

The solid line in Fig. 13 shows the frequency-dependent damping on the isolated string: the quadratic function is a consequence of the damping model used (see Sec. II). The coupling of the string with the plate *in vacuo* (dashed line) contributes to increase its damping factors, with a more pronounced effect for the 2nd, 4th, 7th, and 8th partials, for which the degree of coupling with the plate modes is higher than for the other partials. The coupling of the plate with air modifies these factors substantially. In general, the damping factors increase due to radiation into the free space. However, since the air coupling also modifies the plate's modes, this can in turn modify the string–plate coupling and thus lead to a decrease of the damping factors for some string's components. Figure 13 shows that it is the case here for the 2nd, 4th, and 7th partials of the note E_2 (open 6th string) coupled to the reference guitar. Similarly, Fig. 14 shows the effect of reducing the plate thickness on the radiation of the

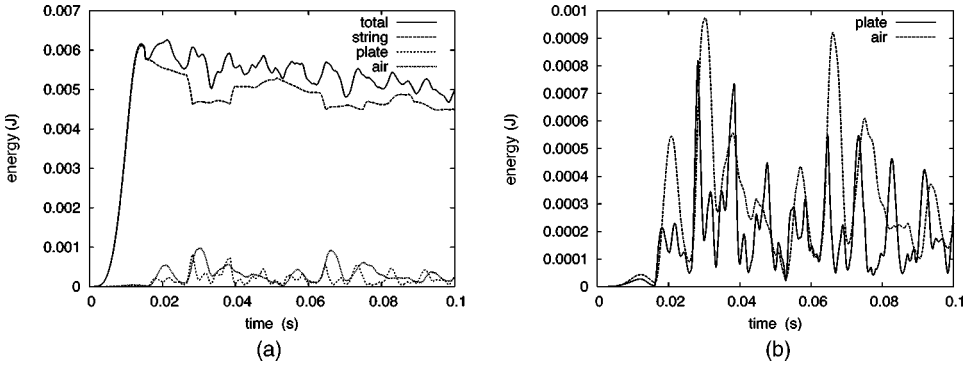


FIG. 11. (a) Evolution of the energy of string, plate, air, and total energy vs time for a plucked open string E_2 (fundamental 82.6 Hz) during the first 100 ms of the tone. Soundboard thickness 1.45 mm. (b) Plate and air energy represented with appropriate scale.

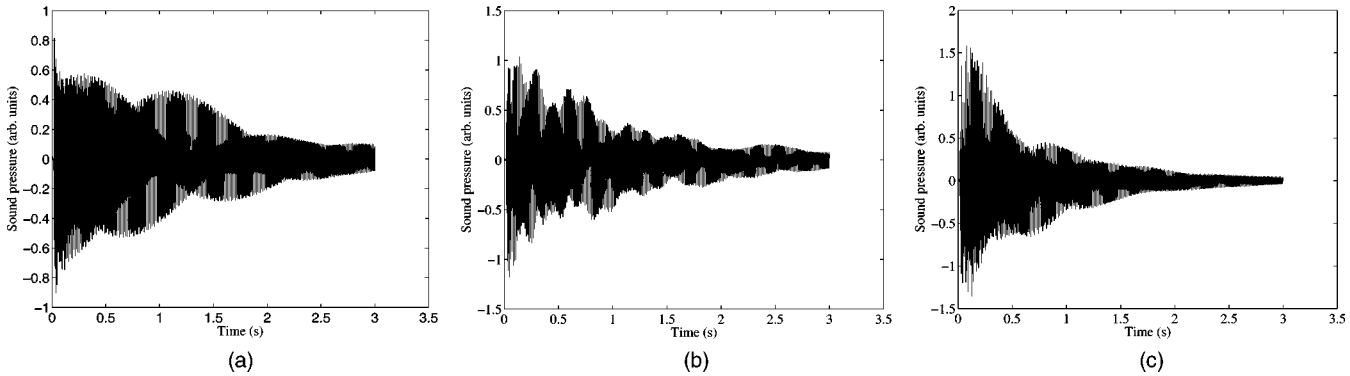


FIG. 12. Waveforms of the pressure recorded at the point ($x=40$ cm, $y=5$ cm, $z=60$ cm) (see Fig. 1 for the definition of the axes). (a) Reference (plate thickness 2.9 mm, cavity height 10.4 cm). (b) Large cavity (plate thickness 2.9 mm, cavity height 21 cm). (c) Thin plate (plate thickness 1.45 mm, cavity height 10.4 cm).

string's components. In the frequency range under examination, it can be seen that the thin plate is more efficient for allowing the guitar to radiate below 400 Hz, whereas the reference plate is more efficient between 400 and 800 Hz. In the low-frequency range, this result is due to the fact that the fluid–structure coupling is more pronounced for the thin plate than for the reference plate. However, as the frequency increases, it becomes hazardous to find a general interpretation for the changes in damping, since the values of the factors depend on the particular degrees of coupling between the string's modes and SA modes. A small change of thickness may alter this coupling substantially, as is experienced daily by guitar makers. More work is needed here for summarizing the efficiency of the instrument in the medium range, and in attempting to express the result in terms of global radiated power.

D. The sound of the guitar in space

Figure 15 shows an example of the spatial distribution of sound pressure in a plane perpendicular to the guitar, in the direction of the strings. For obvious reasons, only a small number of pictures which ''sample'' the acoustic field with a rather large period of time (typically a few milliseconds), can be inserted in this paper. The temporal evolution described below is better seen on a videotape and/or in animated pictures.[23]

The selected pictures are aimed at illustrating some of the basic physical phenomena relative to the acoustic field generated by a guitar, during the first 45 ms of the sound. For this reason, the time interval between consecutive pictures is not imposed to be kept constant. The black color corresponds to a negative acoustic pressure, whereas the white color corresponds to a positive pressure (see the scale). Figure 15(a) (at time 16.4 ms) shows the acoustic field right after the release of the string: two waves of opposite signs propagate inside and outside the cavity, respectively. At time 18.4 ms [Fig. 15(b)], the cavity is filled with a strong, positive pressure field. Plate and rose contributions are in phase, which yields a rather omnidirectional and homogeneous field in front of the instrument. At time 19.2 ms, Fig. 15(c) shows an example where the cavity field is not homogeneous. The external field is not symmetrical, which is a consequence of the fact that the various radiating components of the instruments are no longer in phase. Figure 15(d) (at time 21.1 ms) shows the opposite situation compared to case (b): the internal field is negative, like the near field of the instrument which shows

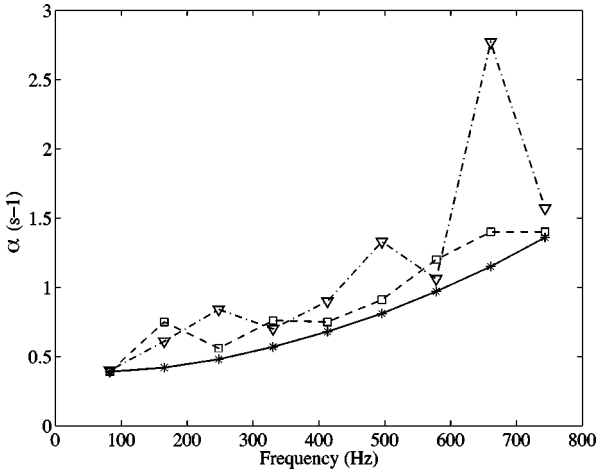


FIG. 13. Damping factors of the string partials. (*———) uncoupled string, (□---) string coupled to the soundboard, (▽-·-) string coupled to soundboard and air.

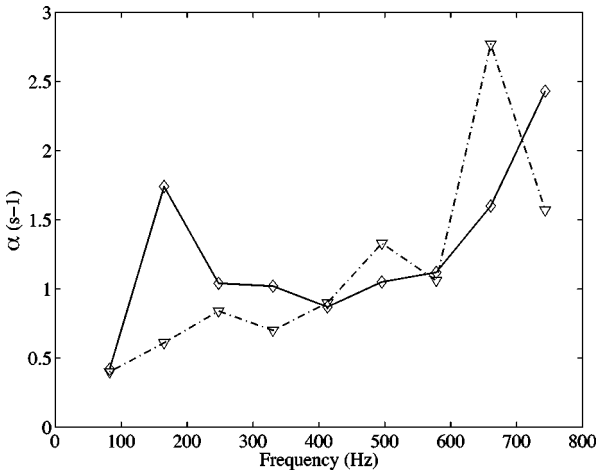


FIG. 14. Damping factors of the string partials coupled to soundboard and air (▽-·-) string coupled to the reference soundboard (plate thickness 2.9 mm), (◇———) string coupled to the thin soundboard (plate thickness 1.45 mm).

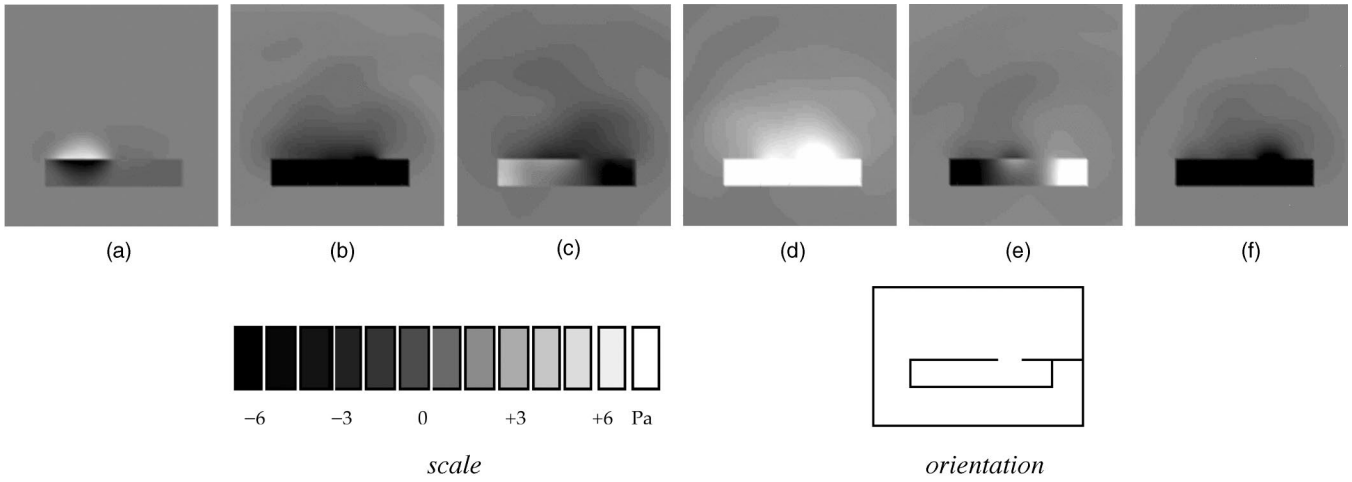


FIG. 15. Sound field radiated by the guitar in the plane $y=0$ for a plucked open string E_2 (fundamental 82.6 Hz) during the 45 first ms. Linear scale (Pa). (a) 16.4 ms. (b) 18.4 ms. (c) 19.2 ms. (d) 21.1 ms. (e) 23.1 ms. (f) 23.9 ms.

again a rather omnidirectional pattern. Similarly, Fig. 15(e) (at time 23.1 ms) shows the situation comparable to case (c), though with opposite signs. This leads to strong heterogeneities both in the internal and external field. The phenomena repeat themselves first pseudoperiodically during the transient where the ''box sound'' is present, and then become progressively periodic. It can be seen, for example, that case (f) (at time 23.9 ms) is comparable to case (b). Animated pictures[23] show, in addition, the propagation of the acoustic waves along the boundaries of the instruments. As a consequence, the acoustic load on the structure varies continuously with time.

The previous results suggest that the directivity of the instrument is not a stationary pattern. This feature is better shown on a logarithmic scale, as seen in Fig. 16, which represents instantaneous directivity patterns. The acoustic pressure is represented in decibels (dB). At time t the following function is plotted: $20\log_{10}|p(x,t)|/p_e$, where $p_e=2\cdot10^{-5}$ Pa is the reference. This figure has the property of enhancing the contributions of the various parts of the instruments to the radiated field while reducing the local differences in pressure, compared to the linear plot. The black color corresponds to 0 dB, while the white color corresponds to 120 dB (see the scale). Figure 16(a) (at time 9.6 ms) shows an example of almost-perfect omnidirectional pattern in front of the instrument. Figure 16(b) (at time 17.3 ms), on the contrary, denotes a multipolar radiation. In addition, the diffraction of the sound field on the back side of the guitar is clearly seen. Other examples of irregular directivity pattern can be also observed in Fig. 16(c) (at time 28.3 ms) and Fig. 16(d) (at time 32.4 ms), while Fig. 16(e) (at time 36.1 ms) shows again an almost constant directivity in the plane under study. In most of the cases, the sound field inside the cavity is significantly louder than the external field.

It could be of interest to validate the pressure patterns obtained with the present model with noncontacting TV-holography techniques, such as those developed by Runnemalm and Molin.[24] This should be done not only for stationary waves, but also for transients.

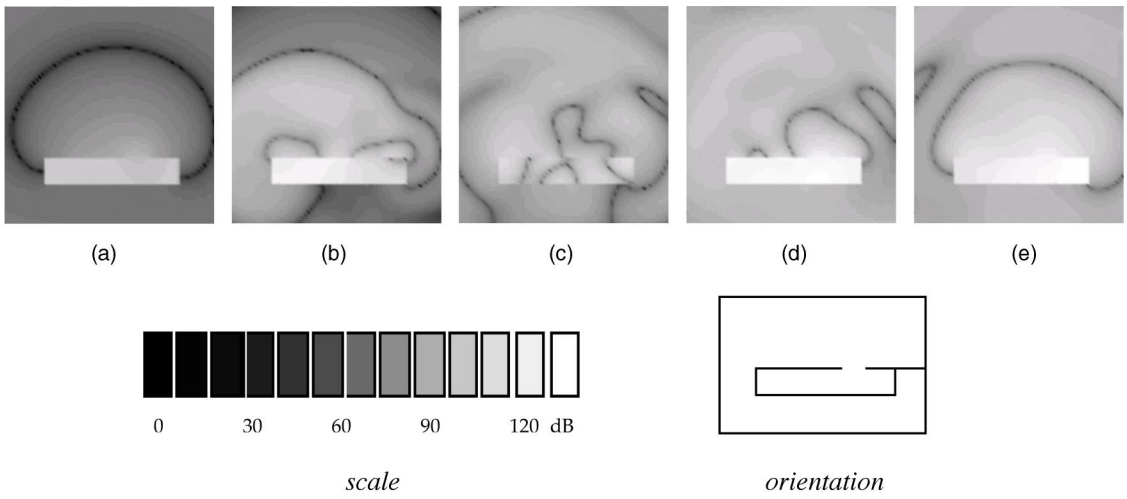


FIG. 16. Instantaneous directivity patterns: sound field radiated by the guitar in the plane $y=0$ for a plucked open string E_2 (fundamental 82.6 Hz) during the 45 first ms. dB scale. At time t, function $20\log_{10}|p(x,t)|/p_e$ is plotted where $p_e=2.10^{-5}$ Pa. (a) 9.6 ms. (b) 17.3 ms. (c) 28.3 ms. (d) 32.4 ms. (e) 36.1 ms.

V. CONCLUSION

The novel aspects of the present work arise from the demanding specificities of the model: complex geometry for the guitar, complex material for the top plate, coupling of the top plate with the partially opened cavity and the external free space. This relatively exhaustive model requires the association of highly elaborated numerical techniques in order to solve it with enough accuracy. The resolution has been made possible here through the use of a spectral method for the top plate, a fictitious domain method for the fluid–structure interaction problem and a conservative scheme for the time discretization. This method is applicable to other radiating structures composed of plates, holes, and cavities, and, in particular, to other stringed musical instruments.

Solving the vibroacoustical system in the time domain yields, among other things, the 3D temporal evolution of the acoustic field. From these results, important new features of the sound radiated by the guitar can be derived. The complexity and the variation in time of the directivity of the instrument, in particular, has been clearly shown.

The present method allows the separate computation of the acoustic modes, assuming a completely rigid body, of the structural modes, assuming that the instrument vibrates *in vacuo*, and of the coupled structural–acoustic modes for the complete air–structure model. This tool is thus applicable to, among other things, the evaluation of the relative contributions of structural losses and radiation losses in the sounds generated by the guitar. It is well-known that such a separation is hard to conduct experimentally.

In its present state, the guitar model is such that the effects of structural modifications on the produced sounds, such as plate thickness and cavity height, can be tested in a straightforward manner. Such simulations were systematically validated in this paper through analysis of admittance at the bridge, energy histories, and pressure signals. A good coherence was found between the results of this analysis, in terms of modal frequencies and damping factors. These results are compatible with experimental results published in the literature devoted to the acoustics of guitars. Because we have control here on the parameters fed into the model, it has been made possible to evaluate in detail the effects of the various sources of coupling (string–plate, plate–air) in terms of frequencies and decay rates.

Despite these interesting features, the present version of the model suffers from some numerical and physical limitations. As seen in Fig. 4, the main limitation due to numerical reasons is the bracing, which is rather crude compared to real guitars. This is due to the necessity of dealing with a more refined mesh if the purpose is to reproduce a real guitar adequately. From a physical point of view, the stronger limitation is probably due to the underestimation of the damping in the fluid, which might be mainly located at the rose and in the boundary layer in the box. More work is needed here for modeling and quantifying these effects convincingly. Other improvements should be investigated, such as the double polarization of the string and the vibrations of the neck. Although this part of the instrument does not participate directly in the radiation, its vibrations have effects on the motion of the other parts (strings, soundboard) to which it is coupled. This feature has been highlighted by Jansson.[25]

The present paper was mainly focused on the guitar model and on the numerical method to solve it, with only a few selected examples showing its capabilities in terms of guitar acoustics. At present, since the air equations are expressed in terms of pressure and acoustic velocity, it becomes possible to investigate other quantities of interest such as sound intensity and sound power. As an example, recent experimental work by Wang and Burroughs showed the interest of representing the complete acoustic intensity field around a stringed instrument in order to characterize its radiating properties.[26] With regard to such problems, the model developed here can be viewed as an interesting tool for validating the experiments theoretically.

APPENDIX A: THE KIRCHHOFF-LOVE'S PLATE EQUATION

1. Orthotropic plate equation

We present here the Kirchhoff–Love's plate equation in some particular Cartesian cases. $\underline{\underline{\mathcal{M}}}=a^3\mathbf{C}\underline{\underline{\varepsilon}}(\underline{\nabla} u_p)$ denotes the bending moment and $\mathbf{C}$ is the rigidity tensor for plates. One can write Eq. (5) in the following form (without damping terms):[7–10]

$$\underline{\underline{\mathcal{M}}}=\begin{pmatrix}\mathcal{M}_{xx}\\ \mathcal{M}_{yy}\\ \mathcal{M}_{xy},\end{pmatrix}=a^3\begin{pmatrix}D_1 & D_2/2 & 0\\ D_2/2 & D_3 & 0\\ 0 & 0 & D_4/2\end{pmatrix}\begin{pmatrix}\partial_{xx}u_p\\ \partial_{yy}u_p\\ \partial_{xy}u_p\end{pmatrix},$$

$$a\rho_p\frac{\partial^2 u_p}{\partial t^2}+\frac{\partial^2\mathcal{M}_{xx}}{\partial x^2}+\frac{\partial^2\mathcal{M}_{yy}}{\partial y^2}+2\frac{\partial^2\mathcal{M}_{xy}}{\partial x\partial y}$$
$$=\mathcal{F}-[p]_\omega \quad \text{in } \omega,$$

where D_1, D_2, D_3, and D_4, are the four rigidity constants of the plate.

2. Homogeneous isotropic plate equation

In the homogeneous isotropic case, one has: $D_1=D_3=(D_2/2+D_4/2)=E/[12(1-\nu^2)]$ where E and ν are the Young's modulus and Poisson's ratio, respectively. The Kirchhoff–Love's plate operator then becomes a usual bi-Laplacian. Equation (5) reduces to

$$a\rho_p\frac{\partial^2 u_p}{\partial t^2}+D_1a^3\left(1+\eta_p\frac{\partial}{\partial t}\right)\Delta^2 u_p+a\rho_pR_p\frac{\partial u_p}{\partial t}$$
$$=\mathcal{F}-[p]_\omega \quad \text{in } \omega.$$

3. Free-boundary condition

We give now the expression of the free-boundary conditions in the case of a plate with a straight edge. For example, if the y axis is a free boundary of the plate, the two conditions given by Eq. (7) can be written

$$\mathcal{M}_{xx}=0 \quad \text{and} \quad \frac{\partial\mathcal{M}_{xx}}{\partial x}+2\frac{\partial\mathcal{M}_{xy}}{\partial y}=0,$$
$$\text{along the edge } x=0.$$

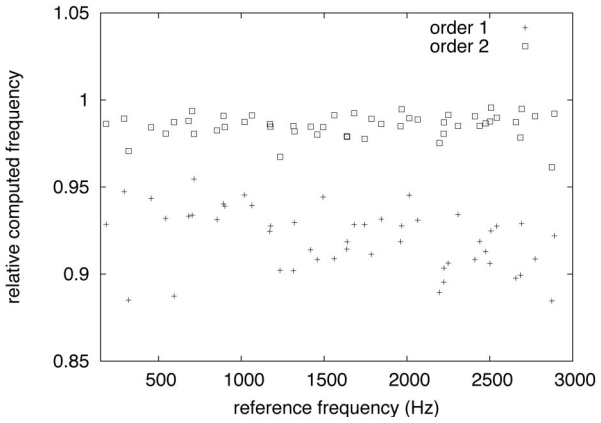


FIG. 17. Comparison between reference and computed eigenfrequencies of the soundboard. (×) relative frequency $f^1_{\text{com}}/f_{\text{ref}}$ computed using $P1$ finite elements. (□) relative frequency $f^1_{\text{com}}/f_{\text{ref}}$ computed using $P2$-alike finite elements.

APPENDIX B: DISPERSION ANALYSIS OF FINITE ELEMENT OF ORDER 1 AND 2 FOR THE PLATE

In order to show the interest of using of second-order finite element for the approximation of the plate equation, we compare here the values of the first 50 eigenfrequencies of the soundboard obtained using $P1$ or $P2$-alike continuous finite elements for the discretization of the variables v_p and $\underline{\underline{\mathcal{M}}}$, with the same computational effort.

Since the actual eigenvalues are not known analytically, references values denoted f_{ref} are first computed using a refined mesh of the soundboard with space step $h_p/2$ and using $P2$-alike continuous finite elements. The results are represented in Fig. 17. It appears immediately that the error made on the estimation of the eigenfrequencies is larger with the use of P_1 finite elements, since the error is greater than 5% for all frequencies. It is thus natural to choose second-order finite elements, which leads to a better approximation of the eigenfrequencies of the soundboard.

[1] L. Rhaouti, A. Chaigne, and P. Joly, "Time-domain simulation and numerical modeling of the kettledrum," J. Acoust. Soc. Am. **105**(6), 3545–3562 (1999).
[2] B. E. Richardson, "Numerical modelling of stringed musical instruments," in *SMAC 93* (Publication of the Royal Swedish Academy of Music, Stockholm, 1993), Vol. 79, pp. 457–462.
[3] A. Le Pichon, S. Berge, and A. Chaigne, "Comparison between experimental and predicted radiation of a guitar," Acustica **84**(1), 136–145 (1998).
[4] M. J. Elejabarrieta, A. Ezcurra, and C. Santamaria, "Coupled modes of the resonance box of the guitar," J. Acoust. Soc. Am. **111**(5), 2283–2292 (2002).
[5] A. Chaigne and C. Lambourg, "Time-domain simulation of damped impacted plates. I. Theory and experiments," J. Acoust. Soc. Am. **109**(4), 1422–1432 (2001).
[6] A. Chaigne, "On the use of finite differences for musical synthesis. Application to plucked string instruments," J. Acoust. **5**, 181–211 (1992).
[7] A. W. Leissa, *Vibrations of Plates* (NASA SP, Washington, D.C., 1969), Vol. 160.
[8] R. F. S. Hearmon, *An Introduction to Applied Anisotropic Elasticity* (Oxford University Press, London, 1961).
[9] S. Ambartsumyan, *Theory of Anisotropic Plates* (Technomic, Stanford, 1970).
[10] G. Caldersmith, "Vibration of orthotropic rectangular plates," Acustica **56**, 144–151 (1984).
[11] F. Millot, F. Collino, and P. Joly, "Fictitious domain method for unsteady problems: Application to electromagnetic scattering," *Mathematical and Numerical Aspects of Wave Propagation* (Mandelieu-La Napoule, 1995) (SIAM, Philadelphia, 1995), pp. 260–269.
[12] R. Glowinski, T. Pan, and J. Periaux, "A fictitious domain method for Dirichlet problem and applications," Comput. Methods Appl. Mech. Eng. **11**1(3–4), 283–304 (1994).
[13] G. Cohen, P. Joly, J. E. Roberts, and N. Tordjman, "Higher order triangular finite elements with mass lumping for the wave equation," SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **38**(6), 2047–2078 (2001).
[14] F. Collino, "Conditions absorbantes d'ordre élevé pour des modèles de propagation d'onde dans des domaines rectangulaires," Technical Report No. 1790, INRIA, Rocquencourt.
[15] J. Nédélec, "Mixed finite elements in $\mathbb{R}^3$," Numer. Math. **35**, 315–341 (1980).
[16] E. Bécache, A. Chaigne, G. Derveaux, and P. Joly, "An explicit finite element scheme for time-dependent Kirchhoff–Love equations," in *Fifth International Conference on Mathematical and Numerical Aspects of Wave Propagation* (SIAM, Santiago, Spain, 2000), pp. 735–740.
[17] P. Joly and L. Rhaouti, "Domaines fictifs, éléments finis H(div) et condition de Neumann: Le problème de la condition inf-sup," C. R. Acad. Sci., Ser. I: Math. **328**(12), 1225–1230 (1999).
[18] O. Christensen and R. Vistisen, "A simple model for low-frequency guitar function," J. Acoust. Soc. Am. **68**, 758–766 (1980).
[19] B. Richardson, T. J. W. Hill, and S. Richardson, "Input admittance and sound field measurements of ten classical guitars," in *Proceedings of the Institute of Acoustics* (IOA, London, 2002), Vol. 24(2), pp. 1–10.
[20] J. Laroche, "The use of the matrix pencil method for the spectrum analysis of musical signals," J. Acoust. Soc. Am. **94**(4), 1958–1965 (1993).
[21] B. David and X. Boutillon, "Using vacuum to measure the acoustical efficiency," in *ISMA 95* (Société Française d'acoustique, Dourdan, France, 1995), pp. 380–385.
[22] B. E. Richardson, "Simple models as a basis for guitar design," Catgut Acoust. Soc. J. **4**(5), 30–36 (2002).
[23] E. Bécache, A. Chaigne, G. Derveaux, and P. Joly, "Numerical simulation of the acoustic guitar," DVD, VHS, and RealPlayer document, English, 2003, http://www.inria.fr/multimedia/Videotheque-fra.html
[24] A. Runnemalm and N.-E. Molin, "Operating deflexion shapes of the plates and standing aerial waves in a violin and a guitar model," Acustica **86**(5), 883–890 (2000).
[25] E. Jansson, *Acoustics for Violin and Guitar Makers* (Royal Institute of Technology, Stockholm, Stockholm, 2002).
[26] L. Wang and C. Burroughs, "Acoustic radiation from bowed violins," J. Acoust. Soc. Am. **110**(1), 543–555 (2001).

Impedance measurements of *ex vivo* rat lung at different volumes of inflation

Michael L. Oelze,[a] Rita J. Miller, and James P. Blue, Jr.
Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois, 405 North Mathews, Urbana, Illinois 61801

James F. Zachary
Bioengineering Program, University of Illinois, 1406 West Green Street, Urbana, Illinois 61801

William D. O'Brien, Jr.
Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois, 405 North Mathews, Urbana, Illinois 61801 and Bioengineering Program, University of Illinois, 1406 West Green Street, Urbana, Illinois 61801

(Received 7 February 2003; revised 5 September 2003; accepted 15 September 2003)

A previous study [J. Acoust. Soc. Am. **111**, 1102–1109 (2002)] showed that the occurrence of ultrasonically induced lung hemorrhage in rats was directly correlated to the level of lung inflation. In that study, it was hypothesized that the lung could be modeled as two components consisting of air and parenchyma (contiguous tissue [pleura and septa]). The speed of sound and lung impedance would then depend on the fractional volume of air in the lung. According to that model, an inflated lung should act like a pressure-release surface for sound incident from tissue onto a tissue-lung boundary. A deflated lung containing less air should allow more acoustic energy into the lung tissue because the impedance was more closely matched to the contiguous tissues. In the study reported herein, a measurement technique was devised to calculate the impedance of seven rat lungs, *ex vivo*, under deflation (atmospheric pressure) and three volumes of inflation pressure (7-cm H_2O, 10-cm H_2O, and 15-cm H_2O). Lungs were dissected from rats and immediately scanned in a tank of degassed 37 °C water. The frequency-dependent acoustic pressure reflection coefficient was measured over a frequency range of 3.5 to 10 MHz. From the reflection coefficient, the frequency-dependent lung impedance was calculated with values ranging from an average of 1 Mrayls in deflated lungs to 0.2 Mrayls for fully inflated lungs. Lung impedance calculations showed that deflated lungs had an impedance closer to water (1.52 Mrayls) than inflated lungs. At all volumes of inflation, the lungs acted as pressure-release surfaces relative to the water. The average of the four lung impedance values (deflated, 7-cm H_2O, 10-cm H_2O, and 15-cm H_2O) at each level of inflation was statistically different ($p<0.0001$). © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1624069]

PACS numbers: 43.80.Cs, 43.80.Gx, 43.80.Jz [FD] Pages: 3384–3393

I. INTRODUCTION

Studies have shown that ultrasound causes lung hemorrhage in animals.[1–17] The mechanism is unknown and under considerable debate; however, the interaction of ultrasound with gas bodies is the primary hypothesis.[18–23] Gas bodies include ''theoretical'' microbubbles within the surfactant layer of each alveolus (≈ 5 μm in diameter) and air in the alveolar architecture of the lung (≈ 40 μm in diameter for each murine alveolus).[24–26]

In vivo experiments in rats demonstrated that deflated lung (lower gas volume) increased the occurrence and degree of hemorrhage when compared to a fully inflated lung.[27] This study hypothesized that the amount of ultrasonic energy transmitted into the lung was inversely related to the volume of inflation. The lung was modeled as bulk material consisting of tissue and air. The less air in the lung, the more the impedance of the lung matched the impedance of adjacent intercostal tissues. The more air in the lung, the greater was the mismatch in impedance between the lung and the surrounding tissues. When the impedance of the lung more closely matched the impedance of the intervening tissues, more ultrasonic energy was transmitted into the lung itself. Thus, a higher occurrence and severity of lung damage resulted in a deflated lung as opposed to an inflated lung.

Previous experiments supported the bulk model of the lung.[28] In these experiments, canine lungs of varying volumes of inflation were fixed and then sliced for measurement of the attenuation and reflection loss. The results showed that the reflection losses were less for inflated lungs than for deflated lungs. Furthermore, the experiments on the canine lungs showed the reflection loss had a frequency dependence.

Several other studies have looked at the propagation of ultrasound in the lung. Dunn and Fry studied the attenuation of ultrasound in freshly excised lungs of dogs at frequencies of 1, 2.25, and 5 MHz and showed that the attenuation had a logarithmic dependence.[29,30] Measurements of sound speed in the lungs showed that the speed of sound was slightly

[a]Electronic mail: oelze@brl.uiuc.edu

linear with frequency.[31] Further studies on lungs with different volumes of inflation yielded increasing attenuation and decreasing sound speed with increased inflation.[31] Studies by Pederson and Ozcan involving cow lungs over the frequency range of 10 to 800 kHz showed measured group velocities had a strong dependence on inflation level, and attenuation exhibited a strong frequency dependence that was weakly dependent on inflation level.[32] A more recent study of ultrasound in pig lungs showed attenuation was large even in deflated lungs.[33] Therefore, in summary, the cumulative interpretation of results from these studies suggest that the amount of ultrasonic energy transmitting into and propagating within a deflated lung is much greater than in an inflated lung.

In the study reported herein, a technique was developed to measure the reflection coefficient and calculate impedance value of intact, fresh lungs, *ex vivo*, from rats for different volumes of inflation. It was hypothesized that fresh lung would have an impedance value dependent on the frequency and volume of inflation. It was previously hypothesized that the impedance value of the lung due to inflation was related to the occurrence of lung damage in rats.[27] Section II describes the theory of lung impedance and its relation to the pressure reflection coefficient. Section III details a new procedure to measure impedance values from fresh, whole lungs. Section IV presents the results of the measurements and the final section (V) gives some conclusions about the study.

II. THEORY

The acoustic pressure reflection coefficient (for plane waves) for sound reflecting at normal incidence from a smooth boundary of two different media (medium I, II) is defined as[34]

$$R=\frac{P_R}{P_I}, \tag{1}$$

where P_R is the amplitude of the reflected sound, and P_I is the amplitude of the incident pressure wave. The sign of R was theorized to be negative because the impedance of the lung (medium II composed of tissue and air) was presumed less than water (medium I). Subsequent measurements from the lung verified that the lung acted as a pressure relief surface relative to water. The pressure reflection coefficient can be measured by comparing the signal reflected, P_R, from the surface of medium II with the sound incident, P_I, on the surface. If a transducer is operated in pulse-echo mode to measure the pressure pulse reflected from the surface, then the pressure amplitude of the reflected signal is found by

$$P_R=S_{\mathrm{Re}}V_R, \tag{2}$$

where V_R is the voltage signal recorded from the received signal and S_{Re} is the receive sensitivity of the pulse-echo transducer.

The incident pressure and the sensitivity of the transducer to receive can be factored out in the measurement of the reflection coefficient by using a reference pulse. The reference pulse is found by using the same settings, equipment, and transducer to reflect a signal from a planar surface of known reflectivity, R_{ref}. Dividing the unknown reflection coefficient, R, by the known reflection coefficient yields

$$\frac{R}{R_{\mathrm{ref}}}=\frac{S_{\mathrm{Re}}V_R/P_I}{S_{\mathrm{Re}}V_{\mathrm{ref}}/P_I}, \quad \frac{R}{R_{\mathrm{ref}}}=\frac{V_R}{V_{\mathrm{ref}}}. \tag{3}$$

Solving for R gives

$$R=\frac{V_R}{V_{\mathrm{ref}}}R_{\mathrm{ref}}, \tag{4}$$

where V_{ref} is the voltage measured from the signal reflected from the surface of the known reflector.

For many surfaces, the reflection coefficient is not constant with frequency. The frequency dependence of the reflection coefficient can be found by taking the magnitude of the Fourier transform of the voltage signals, V_R and V_{ref}. Assuming that the reflection coefficient from the plastic plate is constant with frequency, then

$$|R(f)|=\frac{|FT\{V_R(t)\}|}{|FT\{V_{\mathrm{ref}}(t)\}|}R_{\mathrm{ref}}. \tag{5}$$

The reflection coefficient will be positive or negative depending on the impedance change from one medium to another. If the second medium has a smaller impedance value than the first medium (a pressure release surface), then the pressure reflection coefficient will be negative. If the second medium has larger impedance than the first medium, then the pressure reflection coefficient will be positive.

The pressure reflection coefficient is related to the impedance of the two media by[34]

$$R(f)=\frac{Z_{II}(f)-Z_I(f)}{Z_{II}(f)+Z_I(f)}, \tag{6}$$

where $Z_I(f)$ and $Z_{II}(f)$ represent the impedance of media I and II, respectively. Solving Eq. (6) for the impedance of medium II yields

$$Z_{II}(f)=Z_I(f)\frac{1+R(f)}{1-R(f)}. \tag{7}$$

If the impedance of medium I is known (i.e., water) and the pressure reflection coefficient at the boundary of media I and II is measured, the impedance of medium II can be determined through Eq. (7).

In our previous study, the lung was modeled as two components consisting of air and tissue.[27] The lung's density was modeled as a composite of the fractional volume of air, x_{air}, in the lung and the fractional volume of tissue, $x_{\mathrm{parenchyma}}$, giving

$$\rho_{\mathrm{lung}}=x_{\mathrm{air}}\rho_{\mathrm{air}}+x_{\mathrm{parenchyma}}\rho_{\mathrm{parenchyma}}, \tag{8}$$

where $x_{\mathrm{air}}+x_{\mathrm{parenchyma}}=1$. The adiabatic bulk modulus of the lung was also modeled as a composite of air and tissue

$$B_{\mathrm{lung}}=x_{\mathrm{air}}B_{\mathrm{air}}+x_{\mathrm{parenchyma}}^{3.5}B_{\mathrm{parenchyma}}. \tag{9}$$

The propagation speed and acoustic impedance could then be determined according to

$$c_{\mathrm{lung}}=\sqrt{B_{\mathrm{lung}}/\rho_{\mathrm{lung}}} \tag{10}$$

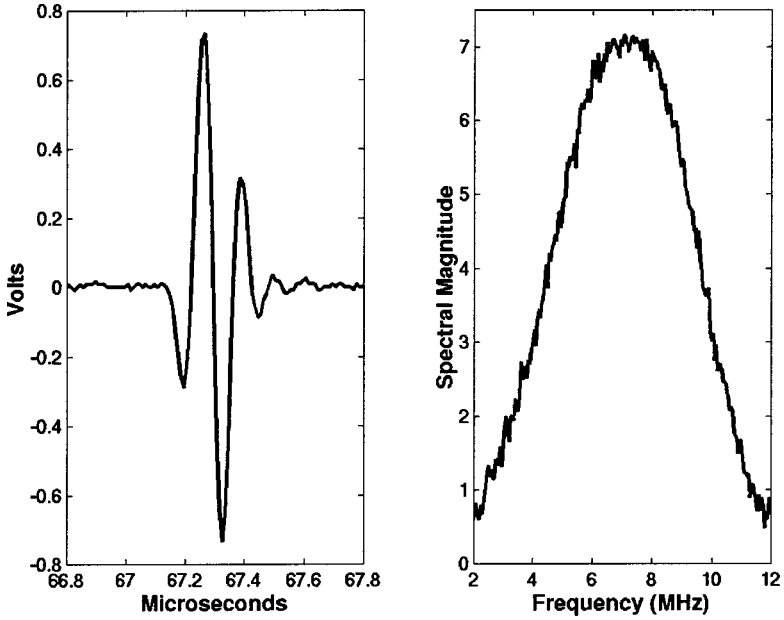


FIG. 1. Acoustic pulse (left) reflected from the plastic plate and the magnitude of the Fourier spectrum (right) of the pulse reflected from the plastic plate.

and

$$z_{\text{lung}} = \sqrt{\rho_{\text{lung}} B_{\text{lung}}}. \tag{11}$$

Previously published values for measurements of the density and bulk properties of air and lung yielded $\rho_{\text{air}} = 1.21\ \text{kg/m}^3$, $\rho_{\text{parenchyma}} = 600\ \text{kg/m}^3$, $B_{\text{air}} = 142\ \text{kPa}$, and $B_{\text{parenchyma}} = 1\ \text{GPa}$.[27] If the volume fraction of air is large (i.e., fully inflated lung), then the lung's adiabatic bulk modulus and the lung's density both will be closer to the values for air. The impedance of the fully inflated lung will be closer to the impedance of air and will act as a pressure release boundary relative to water. On the other hand, a lung with a small amount of air (deflated lung) will have an impedance closer to water or tissue. According to the model, less ultrasonic energy will be transmitted into an inflated lung than into a deflated lung because the impedance of a deflated lung is more closely matched to the impedance of tissue or water. The model has been used to explain the greater percentage of lesion occurrence in the deflated lungs of rats as opposed to inflated lungs.[27]

III. MEASUREMENT METHODS

A. Ultrasound measurement device

A single-element weakly focused ultrasonic transducer (GE Panametrics Inc., Waltham, MA) was used to make measurements of the reflection coefficient from planar reflectors and the lung tissues. The transducer had an aperture diameter of 19 mm and a focal length of 51 mm measured from a planar reflector. The center frequency of the transducer was 7.1 MHz with a −6-dB pulse-echo frequency bandwidth of 5.5 MHz and a −6-dB pulse-echo beamwidth of ~550 μm. The analysis bandwidth ranged from 3.5 to 10 MHz (−9-dB pulse-echo bandwidth). Figure 1 shows the reflected pulse from a flat plastic plate at the focus of the transducer and the corresponding Fourier spectrum (magnitude) of the reflected pulse. The transducer was operated in pulse-echo mode through a Panametrics 5800 pulser/receiver (GE Panametrics, Inc., Waltham, MA). The amplitude of the pulse used for the reflection coefficient study was 1.4 MPa. The signals were recorded with a calibrated PVDF membrane hydrophone (Marconi Model Y-34-6543, Chelmsford, UK) and digitized on an oscilloscope (LeCroy 9354TM, Chestnut Ridge, NY) that had a dynamic range of 48 dB and downloaded to a PC computer for post-processing. The sampling rate was 50 MHz. Reflection measurements were taken near the focus of the transducer where the sound field could be approximated as plane waves.

While the planar reflector is a flat surface, the lung is not a flat surface. One of the difficulties in the experimental setup was to align the lung with the transducer so that the sound reflected from the lung was at normal incidence. Failure to align the lung properly or local curvature of the lung surface would lead to a reduced reflected signal relative to sound reflected at normal incidence. To quantify the possible alignment error, the plastic plate was used to record the relative reflected signal at several angles of incidence. A smooth plastic plate was placed normally incident to the transducer at the focus and rotated axially several degrees. Figure 2 shows a diagram of the plastic plate and transducer setup. The plastic plate was rotated about the z-axis and the magnitude of the signal decrease was then recorded. Figure 3 shows the relative pulse intensity integral (PII) of the reflected signal as the angle of incidence changes from normal. The angle of incidence could be off normal by 2.5° and still have a PII value within 10% error of true normal.

The lung surface was measured similarly to the reflection from the plastic plate. In order to obtain a smooth, flat surface on the lung that would be near normal incidence, the lung was gently pressed up against a plastic holder that was aligned perpendicular to the beam axis. The holder had a small circle cut out in the middle with a diameter of 18 mm that allowed for a scanning window onto the lung surface; the lung surface was in direct contact with the water. The large size of the scanning window (18 mm) compared to the beamwidth (~550 μm) ensured negligible diffraction effects.

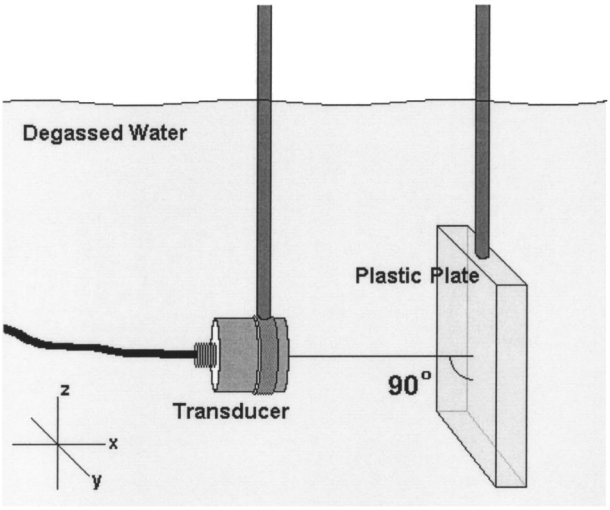


FIG. 2. Experimental setup of the transducer reflecting a pulse at normal incidence from a plastic plate of known reflectivity through a water path.

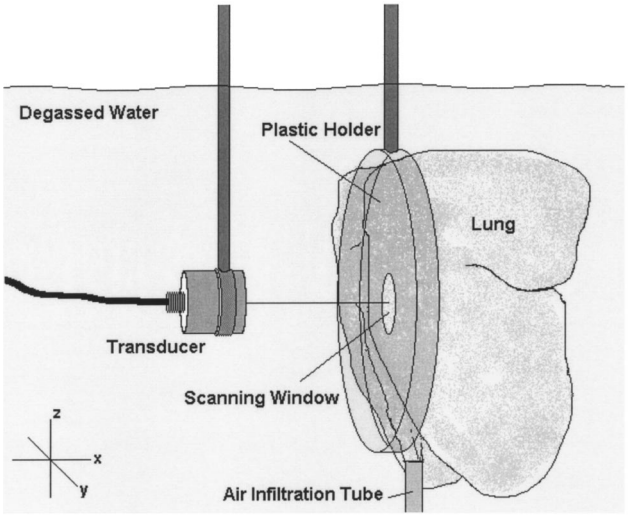


FIG. 4. Experimental setup of the transducer reflecting a pulse from a lung surface through a water path.

Figure 4 shows a diagram of the lung and transducer setup. The lung was set up so that its surface would be near the focus of the transducer. A 5 mm×5 mm square perpendicular to the beam axis of the transducer (*y*, *z* plane) was scanned on the surface of the lung. The rf data waveforms were acquired at intervals of 100 μm in the *y* and *z* directions. Figure 5 shows a pulse reflected from the surface of a lung along with the accompanying Fourier spectrum (magnitude) of the reflected pulse. Scanned results were downloaded to a computer for postprocessing.

Reference pulses were measured from a smooth plastic plate of known reflectivity and a smooth air–water interface. The transducer was aligned perpendicular to the reflecting surface. The same equipment and settings were used for the pulse/echo measurement of the reference pulse as used for the lung reflection measurements. Reflection measurements from the planar surfaces were made at different depths relative to the transducer by scanning axially along the transducer beam. The scan measured reflected values over an axial length of 1 cm centered about the focus with a step size of 100 μm (approximately half the wavelength).

To test the validity of the approach, measurements were taken from a water–air interface and a water–Dow Corning 710 interface. The speed of Dow Corning 710 (1340 m/s) is published and the density (1.11×10^3 kg/m^3) can be found from Dow Corning Corporation (Midland, MI).[35] The water–Dow Corning 710 interface measurements were taken in a water bath temperature of 22.2 °C, making the impedance of the water larger than the impedance of the Dow Corning 710 oil. Figure 6 shows the percent error in the measured impedance of the Dow Corning 710 oil from the published values using the measurement technique described in this study. Figure 7 shows the percent error between the measured and theoretical pressure reflection coefficients for the water–air interface. The small difference between the theoretical and measured values show the measurement technique is valid.

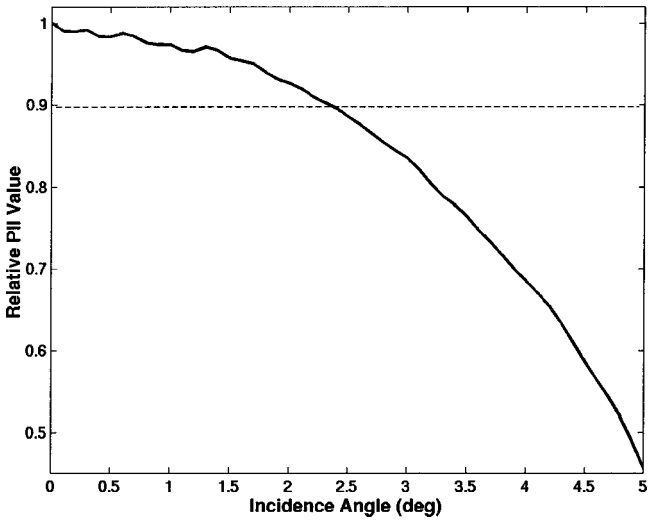


FIG. 3. Relative loss in magnitude of the PII value of reflected signal as the angle of incidence moves away from normal.

B. Animal procedures

The experimental protocol was approved by the Laboratory Animal Care Advisory Committee at the University of Illinois at Urbana—Champaign and satisfied all campus and NIH (National Institutes of Health) rules for the humane use of laboratory animals. Rats were housed in an Association for Assessment and Accreditation of Laboratory Animal Care, Rockville, MD (AAALAC)-approved animal facility, placed in groups of 1–3 in polycarbonate cages with beta-chip bedding and wire bar lids, and provided food and water *ad libitum*. The AAALAC is a private, nonprofit organization that promotes the humane treatment of animals in science through a voluntary accreditation program.

Seven 16-to-20-week-old (240-to-280-g) female Sprague-Dawley rats (Harlan, Indianapolis, IN) were weighed and anesthetized with ketamine hydrochloride (87.0 mg/kg) and xylazine (13.0 mg/kg) administered intraperitoneally. Rats were immediately euthanized while under anesthesia by cervical dislocation. The lungs (with heart still at-

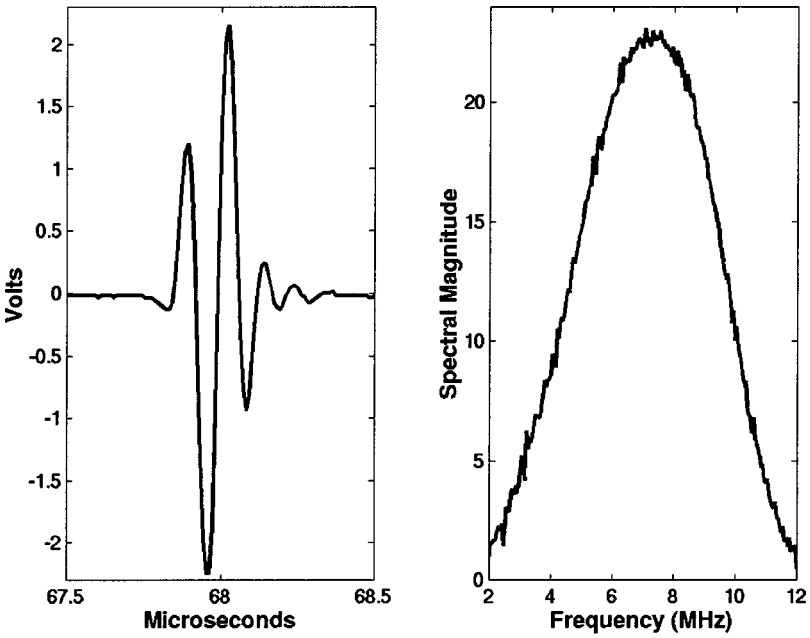


FIG. 5. Acoustic pulse (left) reflected from an inflated lung surface and the magnitude of the Fourier spectrum (right) of the pulse reflected from the lung surface.

tached) and trachea were removed by opening the thorax with a midline incision. Following removal, the lungs were rinsed with 0.9% saline solution. Saline was not allowed to enter the trachea. All lungs were used immediately post harvest.

A 20-gauge needle, with the tip blunted, was inserted into the trachea, which was sealed airtight around the needle hub. Plastic hosing was then used to attach the needle to the lung inflation apparatus. The lung inflation apparatus consisted of a plastic tube connected to two pressure regulators, a tank of compressed air, and a ruler to determine the pressure level (Fig. 8). Pressure in the lungs was measured in terms of cm of H_2O ($1\ \text{cm}\ H_2O = 98\ \text{Pa}$). When a pressure level of 30-cm H_2O in the *ex vivo* lungs was reached, the lung was considered fully inflated or inflated to total lung capacity.[36]

All measurements were taken in degassed water at $37 \pm 0.5\ ^\circ\text{C}$. Water temperature was controlled by a proportional temperature controller (Yellow Springs Instrument Co., Inc., Yellow Springs, OH). A digital thermometer (Control Company, Friendswood, TX), with an accuracy of $0.2\ ^\circ\text{C}$, in the water tank served as the temperature standard. During the time frame of the experimental procedures, the lung did not undergo significant degradation in the degassed water because the visceral pleura of the lung consists of structural proteins (i.e., fibers).

Measurements of the ultrasound reflection coefficient were taken at four different volumes of inflation for each rat lung. The first measurements of the ultrasound reflection coefficient were taken on the deflated lungs (0-cm H_2O). After the deflated measurement was taken, the lung was slowly inflated, at a rate of 1-cm H_2O per minute, to 20-cm H_2O. The lung was held at that pressure for a minimum of 5 min to allow the pressure and air inside the lung to equilibrate. The pressure was then reduced to 15-cm H_2O, at 2-cm H_2O per minute, and held for 15 min to ensure pressure equilibrium. Measurements of the ultrasound reflection coefficient were taken from the inflated lung at 15-cm H_2O. For a normal

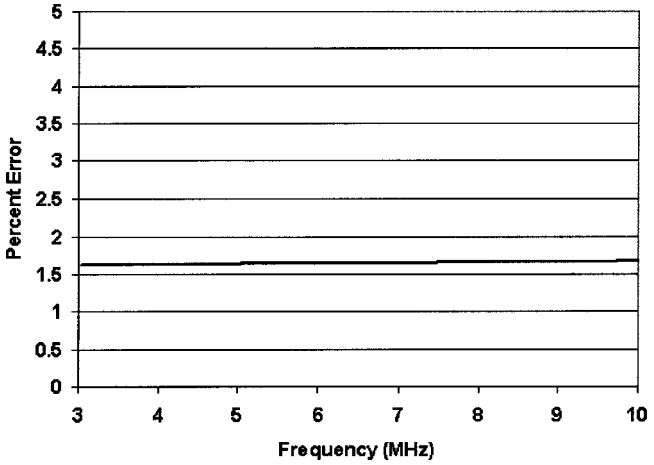


FIG. 6. Plot of the percent difference between the measured and theoretical impedance values versus frequency using the reference spectrum technique for Dow Corning 710.

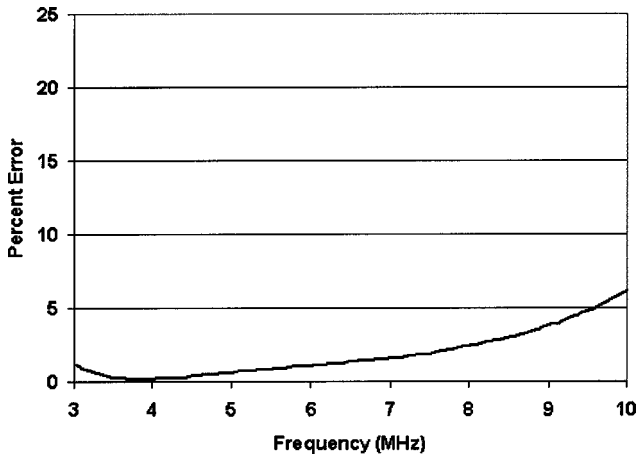


FIG. 7. Plot of the percent difference between the measured and theoretical pressure reflection coefficient using the reference spectrum technique for the water/air interface.

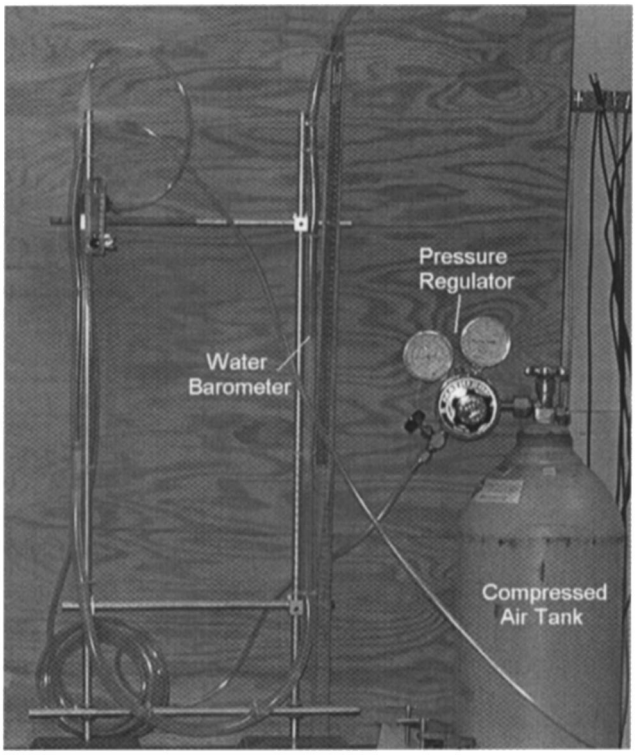


FIG. 8. Picture of the inflation apparatus including compressed air tank, pressure gauge regulators and water barometer.

animal, 15- to 20-cm H_2O pressure is required for adequate but not full inflation.[37] The next two pressure volumes were used to observe lung impedance as the pressure level was reduced below the level necessary to inflate the lung. The pressure was reduced to 10-cm H_2O and 7-cm H_2O, at 2-cm H_2O per minute, and held for 15 min at each pressure level to ensure pressure equilibrium. Measurements of the ultrasound reflection coefficient were taken from the inflated lung at 10-cm and 7-cm H_2O. The 7-cm H_2O was the lowest pressure that could be maintained without the lung visibly deflating during the measurement procedure.

C. Data analysis

Waveforms acquired by the transducer were downloaded to a computer for postprocessing. Matlab® (The Mathworks Inc, Natick, MA) was used to perform the waveform analysis. In order to reduce measurement error of reflections from off-normal incidence, waveform sections were used in the calculations that coincided with flat areas at near-normal incidence on the lung surface. A surface plot of the lung over the 5×5 mm^2 scanned area was constructed from the waveforms. The surface plot was constructed by finding the distance from the source to the peak with the maximum magnitude in each reflected waveform. Figure 9 shows a graph of a surface plot from one of the rat lungs. The flat regions shown by the surface plots were used to determine smooth regions at near-normal incidence from which the reflection coefficient calculations were made. The angle of incidence at a particular point in the scan region was calculated from the average of the arctangents of the slopes between the particular scan point and four immediately surrounding scan points. Only reflected pulses from scan points that showed angles of incidence less than 0.5° from normal were used.

The frequency dependence of each waveform in the selected region was calculated by taking the Fourier transform. Each reflection waveform was made up of 2000 points. A Hanning window was constructed based on the width of the pulse reflected from the reference surface. The Hanning window was applied to the pulse reflected from the lung by centering the window at the point in the waveform where the peak had its maximum magnitude. The size of the deflated rat lung was more than a centimeter in diameter while the length of the pulse was around 300 μm, enabling the gating out of multiple surface echoes. The Hanning window was also applied in the same manner to the pulse reflected from the planar (reference) surface. The absolute values of the Fourier transforms from the waveforms were taken and all the Fourier spectra were averaged together from the selected region to obtain an overall representation of the reflection from the region specified. The average of the Fourier spectra

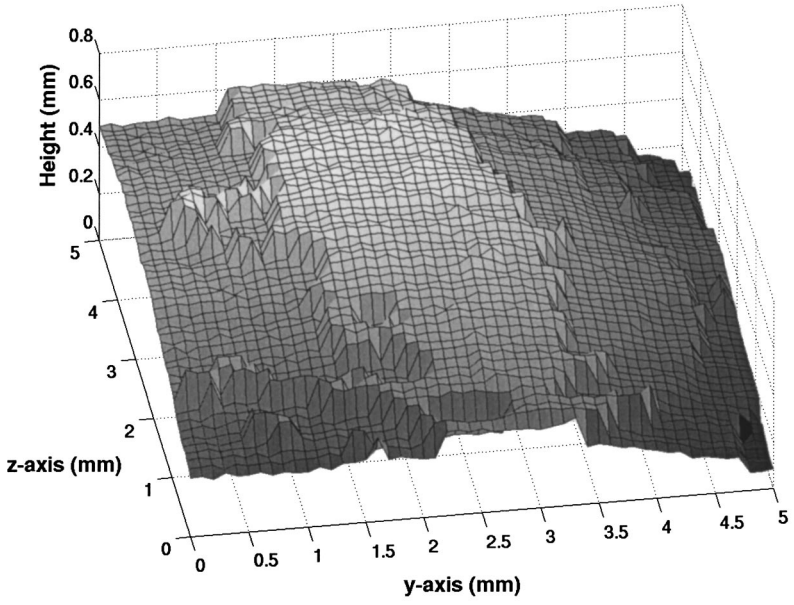


FIG. 9. Surface plot of the lung recorded from the maximum peak magnitude of waveforms scanned over a 5×5 mm^2 window.

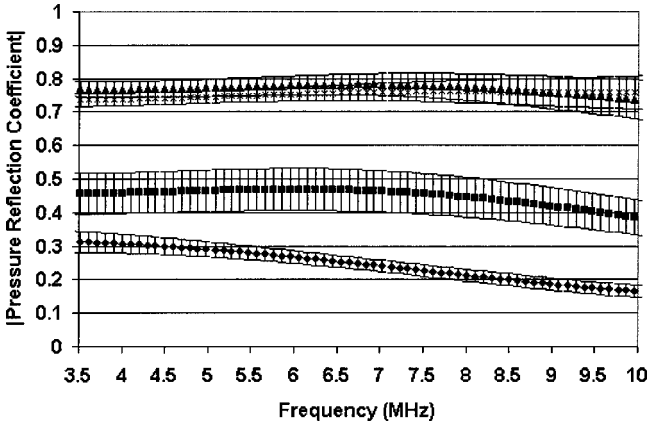


FIG. 10. Plot of the average magnitude of the pressure reflection coefficient from 7 rats versus frequency for four inflation volumes, ♦, pressure of 0.0-cm H_2O (deflated case); ■, pressure of 7-cm H_2O; ▲, pressure of 10-cm H_2O; ×, pressure of 15-cm H_2O. Error bars represent $\frac{1}{2}$ standard deviation about the mean for seven rats.

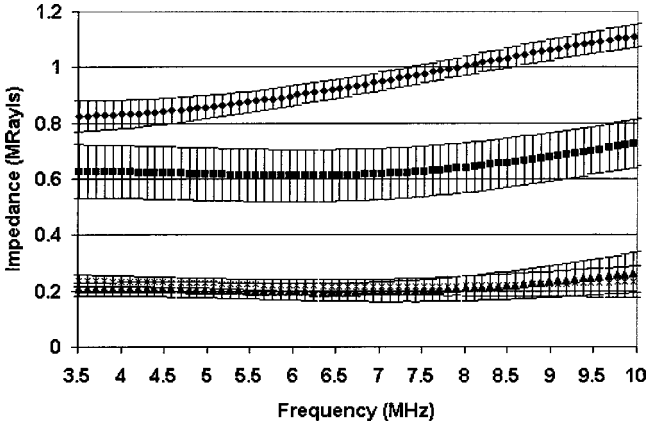


FIG. 11. Plot of the average impedance value from seven rats versus frequency for four inflation volumes, ♦, pressure of 0.0-cm H_2O (deflated case); ■, pressure of 7-cm H_2O; ▲, pressure of 10-cm H_2O; ×, pressure of 15-cm H_2O. Error bars represent $\frac{1}{2}$ standard deviation about the mean for seven rats.

was then divided by the reference spectrum and multiplied by the appropriate reflection coefficient for the reference surface according to Eq. (5).

The impedance was then calculated over the frequency range of 3.5–10 MHz according to Eq. (7) where $Z_I(f)$ represented the characteristic acoustic impedance of water at 37 °C, $Z_I(f)=\rho_{H_2O}c=1.52$ MRayls (ρ_{H_2O} is the density of water, 998 kg m^{-3}, and c is the speed of sound in water at 37 °C, 1525 m s^{-1}). The sign of the reflection coefficient was determined by comparing the reflected pulses from the plastic plate and the lung surface. A comparison of the reflected pulses from Figs. 1 and 5 shows that the pulse reflected from the lung surface was 180° out of phase with the pulse reflected from the plastic plate. Because the plastic plate's impedance was greater than the surrounding water, the opposite sign from the pulse reflected from the lung meant the lung surface represented a pressure relief type of surface. The sign change of the reflected pulse from the plastic plate to the lung was evident at all inflation volumes. In calculating the impedance of the lung surface from Eq. (7), a minus sign was multiplied times the magnitude of the reflection coefficient yielding impedance estimates according to

$$Z_{II}(f)=\rho_{H_2O}c\,\frac{1-|R(f)|}{1+|R(f)|}. \tag{12}$$

IV. RESULTS

The procedure was applied to the four different inflation volumes and graphed for comparison. Figure 10 shows the graph of the average measured reflection coefficient magnitudes versus frequency for the four different inflation volumes. Each data point represents the average of measurements taken from seven different rat lungs. The smallest reflections came from the deflated cases and the largest overall reflections came from the most inflated cases.

Figure 11 shows the average impedance values versus frequency calculated for the seven rat lungs at different inflation volumes. The data show that the more deflated the lung, the closer the impedance value was to water (1.52 MRayls). The closer the impedance of the lung was to water, the more energy was transmitted into the lung itself. Likewise, for the most inflated level of the lung, the impedance had the greatest disparity from water meaning that relatively less energy was being transmitted into the lung. Volumes of inflation between fully inflated and deflated appeared to have impedance values between the deflated and fully inflated cases.

The linear regression analysis was performed on the impedance values for each of the four volumes of inflation (Fig. 12).[38] The linear regression analysis yielded the p-value for the slope, the ±95% confidence band for the regression line and the coefficient of determination, r^2. As a function of frequency (range: 2.15–12.0 MHz), the linear regression equations for each the four volumes of inflation from the seven rat lungs were

$$Z_{\text{deflated}}=0.040f+0.68\ \ (n=714;p<0.0001;r^2=0.34), \tag{13a}$$

$$Z_{7\text{-cm}}=0.019f+0.53\ \ (n=714;p<0.0001;r^2=0.026), \tag{13b}$$

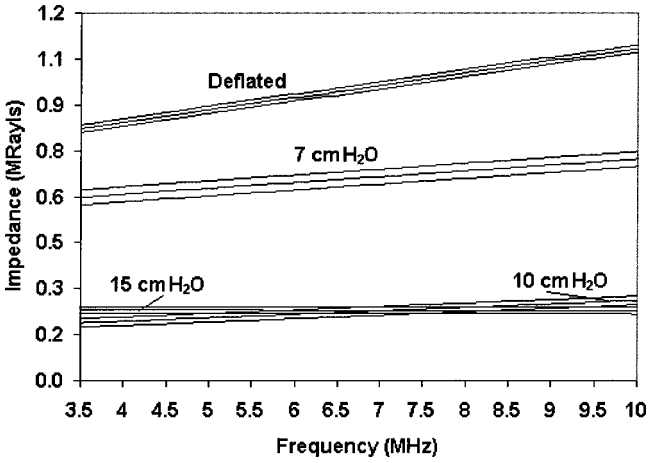


FIG. 12. Linear fits to the impedance data with 95% confidence intervals.

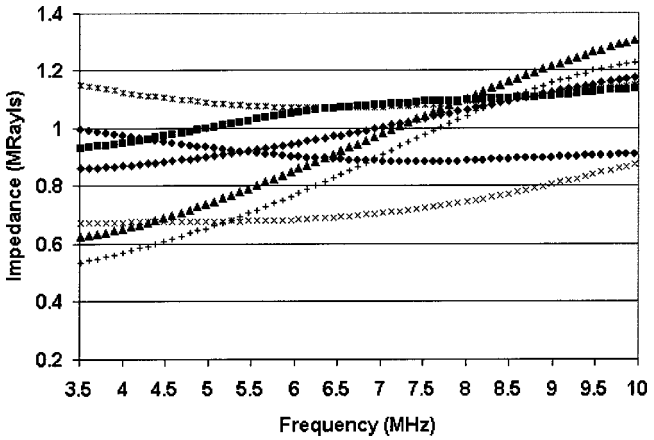


FIG. 13. Impedance values for the seven rats measured from deflated lungs.

$$Z_{10\text{-cm}} = 0.011f + 0.15 \quad (n = 714; p < 0.0001; r^2 = 0.027), \tag{13c}$$

$$Z_{15\text{-cm}} = -0.00027f + 0.23 \quad (n = 714; p = 0.89; r^2 < 0.001), \tag{13d}$$

where the acoustic impedance, Z, is in MRayl and the ultrasonic frequency, f, is in MHz.

To test for equality of the four regression lines for each of the four volumes of inflation an analysis of covariance F-test was performed.[38] The null model constrains the intercepts and slopes to be equal across different volumes of inflation. The general model allows for unequal intercepts and slopes for the volumes of inflation. If not rejected, then the null model is used. If it is rejected, then the general model is used. Comparison of the acoustic impedances in the general model for the four volumes of inflation indicates that the four linear regressions are different ($p < 0.0001$). Likewise, comparison of the acoustic impedances in the general model for the two most inflated cases (10-cm H_2O and 15-cm H_2O) indicates that the two linear regressions are different ($p < 0.0001$).

The previous graphs of the impedance (Figs. 11 and 12) showed general trends for the lungs from the seven rats. The deflated lungs tended to transmit more ultrasound energy into the lung than the inflated lungs. In a case by case examination, each lung will be different with different ultrasonic properties. Figure 13 shows the frequency-dependent impedance values for the seven rat lungs under deflation. The graph shows a definite spread of the measured impedance values from one rat to another. In the previous study, the deflated lungs had the greatest occurrence of ultrasound-induced lesions, though not all of the rat lungs developed lesions.[27] Figure 13 may help explain the fact that under the same exposure conditions some rats will develop lesions while others do not. In general, the deflated lungs will admit more acoustic energy than inflated lungs, but individual lungs will vary in the magnitude of ultrasonic energy transmitted into the lung under deflation. A similar spread in the data was also seen for the inflated lungs showing the biological variability from one animal to the next.

The frequency dependence for the two most inflated cases (10-cm H_2O and 15-cm H_2O) appeared to be minimal but increased as the level of inflation decreased. The cause of the frequency dependence may be due to increasing roughness of the lung surface with deflation. When the lung was inflated, the visceral pleura was stretched taught and appeared to be smooth. In a deflated lung, the loose visceral pleura may have bunched together causing the surface to be less smooth. The morphologic structure of the surface of the lung (visceral pleura) is very uniform in all species but varies in thickness and fiber composition (i.e., density and type) between species. In the inflated state, the visceral pleura of the lung is stretched very taught such that the topographical features of the surface are smooth no matter where the lung is acoustically examined. In the deflated state, the visceral pleura is no longer stretched and is likely contracted such that the topographical features of the surface are ''roughened'' or irregular from area to area examined acoustically. If the roughness is small relative to the wavelength of sound, most of the acoustic energy will be reflected back to the source. As the wavelength becomes smaller relative to the roughness, more acoustic energy is scattered or reflected off in different directions, reducing the amount of acoustic energy reflected back to the source. Figure 10 shows that the higher frequencies (smaller wavelengths) reflected back less sound than the lower frequencies (larger wavelengths). In the case of the deflated lung, the impedance measurement of the lung surface may be a combination of the true lung impedance and surface roughness properties of the lung.

The frequency dependence of the reflection coefficient from the deflated lung may also be due to the amount of energy transmitted into the deflated lung. With the inflated lungs, a single pulse is seen reflected from the lung surface because almost no acoustic energy is transmitted into the lung. In the case of the deflated lung, the return echo tends to have a train of pulses returning (speckle). The train of pulses in the echo suggests that sound is returning from inside the lung as well as from the surface of the lung. It is possible that the pulses corresponding from within the lung are interfering with the sound reflecting from the surface. The interference of the pulses can cause apparent frequency dependence in the windowed waveform used to calculate the reflection coefficient. However, the effects of interference are minimal because on average the backscatter pulses trailing the surface reflection are more than 13 dB less than the surface reflection.

Comparison of the results obtained with the technique used in this study with the results from previous techniques and studies show similar trends. At frequencies of 5.0 and 7.4 MHz, Bauld and Schwan showed a decrease in the reflection coefficient from the inflated to deflated lung of 13 and 25 dB, respectively.[28] Results from Fig. 10 show a decrease from the inflated case to the deflated case of approximately 12 dB for the rat lungs for both 5.0 and 7.4 MHz. Both studies showed that the reflection loss increased with decreasing inflation. The frequency dependence observed by Bauld and Schwan was markedly greater than the frequency dependence observed in the present study.

V. CONCLUSIONS

Measurements were made at four volumes of inflation of the impedance of rat lungs. The measurements showed that the greater the volume of the inflation, the smaller the impedance value of the lung. Measurements also showed that the impedance value of the lung was always less than the characteristic impedance of water at 37 °C (pressure release surfaces relative to water). The fully inflated lungs were more similar to a water–air interface.

The impedance of the lungs at deflation was closer to the characteristic impedance value of water than for the inflated cases. More ultrasonic energy was transferred into the deflated lungs than into the inflated lungs. The same will also be true for ultrasonic incident on the lung from adjacent intercostal tissues. Most of the sound is reflected from the lung surface with just a little inflation. If increased ultrasonic energy transmitted into the lung causes an increase in the incidence of lung damage, then the increased ultrasonic energy transmitted into the deflated lung, due to a more closely matched impedance, supports earlier measurements of higher probability of lung lesions in rats with deflated lungs. However, further study needs to be conducted to show the exact cause of ultrasound induced lung damage.

For the inflated lungs, almost no frequency dependence was observed. Noticeable frequency dependence was seen for the deflated lung. The amount of ultrasonic energy reflected from the deflated lung appeared to decrease with higher frequencies of ultrasound. The more energy that was transmitted into the lung (deflated), the larger the frequency dependence was observed. Several factors may have contributed to the frequency dependence seen in the deflated lungs. Future experiments will examine whether the frequency dependence is due to lung impedance effects, surface roughness of the lung, interference of the surface reflection with pulses returning from within the lung or some other cause.

ACKNOWLEDGMENTS

The authors thank Steven Fay for his invaluable assistance in building the lung inflation apparatus. This work was supported by NIH Grant No. EB02641 (formerly HL58218) awarded to WDO and JFZ.

[1] S. Z. Child, C. L. Hartman, L. A. Schery, and E. L. Carstensen, "Lung damage from exposure to pulsed ultrasound," Ultrasound Med. Biol. **16**, 817–825 (1990).
[2] C. Hartman, S. Z. Child, R. Mayer, E. Schenk, and E. L. Carstensen, "Lung damage from exposure to fields of an electrohydraulic lithotriptor," Ultrasound Med. Biol. **16**, 675–679 (1990).
[3] D. P. Penney, E. A. Schenk, K. Maltby, C. Hartman-Raeman, S. Z. Child, and E. L. Carstensen, "Morphologic effects of pulsed ultrasound in the lung," Ultrasound Med. Biol. **19**, 127–135 (1993).
[4] C. H. Raeman, S. Z. Child, and E. L. Carstensen, "Timing of exposures in ultrasonic hemorrhage of murine lung," Ultrasound Med. Biol. **19**, 507–517 (1993).
[5] L. A. Frizzell, E. Chen, and C. Lee, "Effects of pulsed ultrasound on the mouse neonate: Hind limb paralysis and lung hemorrhage," Ultrasound Med. Biol. **20**, 53–63 (1994).
[6] J. F. Zachary and W. D. O'Brien, Jr., "Lung lesions induced by continuous- and pulsed-wave (diagnostic) ultrasound in mice, rabbits, and pigs," Vet. Pathol. **32**, 43–54 (1995).
[7] C. H. Raeman, S. Z. Child, D. Dalecki, C. Cox, and E. L. Carstensen, "Exposure-time dependence of the threshold for ultrasonically induced murine lung hemorrhage," Ultrasound Med. Biol. **22**, 139–141 (1996).
[8] W. D. O'Brien, Jr. and J. F. Zachary, "Lung damage assessment from exposure to pulsed-wave ultrasound in the rabbit, mouse, and pig," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 473–485 (1997).
[9] D. Dalecki, S. Z. Child, C. H. Raeman, D. P. Penney, C. Cox, and E. L. Carstensen, "Age dependence of ultrasonically induced lung hemorrhage in mice," Ultrasound Med. Biol. **23**, 767–776 (1997).
[10] C. K. Holland, K. Sandstrom, X. Zheng, J. Rodriguey, and R. A. Roy, "The acoustic field of a pulsed Doppler diagnostic ultrasound system near a pressure release surface," J. Acoust. Soc. Am. **95**, 2855 (1994).
[11] W. D. O'Brien, Jr., L. A. Frizzell, D. J. Schaeffer, and J. F. Zachary, "Superthreshold behavior of ultrasound-induced lung hemorrhage in adult mice and rats: Role of pulse repetition frequency and exposure duration," Ultrasound Med. Biol. **27**, 267–277 (2001).
[12] J. F. Zachary, J. M. Sempsrott, L. A. Frizzell, D. G. Simpson, and W. D. O'Brien, Jr., "Superthreshold behavior and threshold estimation of ultrasound-induced lung hemorrhage in adult mice and rats," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 581–592 (2001).
[13] J. F. Zachary, L. A. Frizzell, K. S. Norrell, J. P. Blue, Jr., R. J. Miller, and W. D. O'Brien, Jr., "Temporal and spatial evaluation of lesion reparative responses following superthreshold exposure of rat lung to pulsed ultrasound," Ultrasound Med. Biol. **27**, 829–839 (2001).
[14] W. D. O'Brien, Jr., D. G. Simpson, L. A. Frizzell, and J. F. Zachary, "Superthreshold behavior and threshold estimation of ultrasound-induced lung hemorrhage in adult rats: Role of beamwidth," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 1695–1705 (2001).
[15] A. F. Tarantal and D. R. Canfield, "Ultrasound-induced lung hemorrhage in the monkey," Ultrasound Med. Biol. **20**, 65–72 (1994).
[16] R. Baggs, D. P. Penney, C. Cox, S. Z. Child, C. H. Raeman, D. Dalecki, and E. L. Carstensen, "Thresholds for ultrasonically induced lung hemorrhage in neonatal swine," Ultrasound Med. Biol. **22**, 119–128 (1996).
[17] D. Dalecki, S. Z. Child, C. H. Raeman, C. Cox, and E. L. Carstensen, "Ultrasonically induced lung hemorrhage in young swine," Ultrasound Med. Biol. **23**, 777–781 (1997).
[18] R. E. Apfel, "Comment on 'Ultrasound-induced lung hemorrhage is not caused by inertial cavitation' [J. Acoust. Soc. Am. **108**, 1290–1297 (2001)]," J. Acoust. Soc. Am. **110**, 1737 (2001).
[19] L. A. Frizzell, J. M. Kramer, J. F. Zachary, and W. D. O'Brien, Jr., "Response to "Comment on 'Ultrasound-induced lung hemorrhage is not caused by inertial cavitation'" [J. Acoust. Soc. Am. **110**, 1737 (2001)]," J. Acoust. Soc. Am. **110**, 1738–1739 (2001).
[20] R. E. Apfel, "Reply to Frizzell *et al.*'s comment to our comment," J. Acoust. Soc. Am. **110**, 1740–1741 (2001).
[21] L. A. Frizzell, J. M. Kramer, J. F. Zachary, and W. D. O'Brien, Jr., "Comment on Apfel's second comment," J. Acoust. Soc. Am. **110**, 1742 (2001).
[22] C. K. Holland, C. X. Deng, R. E. Apfel, J. L. Alderman, L. A. Fernandez, and K. J. Taylor, "Direct evidence of cavitation in vivo from diagnostic ultrasound," Ultrasound Med. Biol. **22**, 917–925 (1996).
[23] W. D. O'Brien, Jr., L. A. Frizzell, R. M. Weigel, and J. F. Zachary, "Ultrasound-induced lung hemorrhage is not caused by inertial cavitation," J. Acoust. Soc. Am. **108**, 1290–1297 (2000).
[24] M. L. Crosfill and J. G. Widdicombe, "Physical characteristics of the chest and lungs and the work of breathing in different mammalian species," J. Physiol. (London) **158**, 1–14 (1961).
[25] S. M. Tenney and J. E. Remmers, "Comparative quantitative morphology of the mammalian lung: Diffusing area," Nature (London) **197**, 54–56 (1963).
[26] E. R. Weibel, "Dimensions of the tracheobronchial tree and alveoli," in *Biological Handbooks: Respiration and Circulation*, edited by P. L. Altman and D. S. Dittmer (Federation of American Societies for Experimental Biology, Bethesda, MD, 1971), Chap. 51.
[27] W. D. O'Brien, Jr., J. M. Kramer, T. G. Waldrop, L. A. Frizzell, R. J. Miller, J. P. Blue, and J. F. Zachary, "Ultrasound-induced lung hemorrhage: Role of acoustic boundary conditions at the pleural surface," J. Acoust. Soc. Am. **111**, 1102–1109 (2002).
[28] T. J. Bauld and H. P. Schwan, "Attenuation and reflection of ultrasound in canine lung tissue," J. Acoust. Soc. Am. **56**, 1630–1637 (1974).
[29] F. Dunn and W. J. Fry, "Ultrasonic absorption and reflection by lung tissue," Phys. Med. Biol. **5**, 401–410 (1961).
[30] F. Dunn, "Attenuation and speed of ultrasound in lung," J. Acoust. Soc. Am. **56**, 1638–1639 (1974).
[31] F. Dunn, "Attenuation and speed of ultrasound in lung: Dependence upon frequency and inflation," J. Acoust. Soc. Am. **80**, 1248–1250 (1986).
[32] P. C. Pederson and H. S. Ozcan, "Ultrasound properties of lung tissue and

their measurements,'' Ultrasound Med. Biol. **12**, 483–499 (1986).
[33] Z. Mikhak and P. C. Pederson, ''Acoustic attenuation properties of the lung: An open question,'' Ultrasound Med. Biol. **28**, 1209–1216 (2002).
[34] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982).
[35] F. Dunn, P. D. Edmonds, and W. J. Fry, ''Absorption and dispersion of ultrasound in biological media,'' in *Biological Engineering*, edited by H. P. Schwan (McGraw Hill, New York, 1969), pp. 205–332.
[36] H. Bachofen, A. Amman, D. Wangensteen, and E. R. Weibel, ''Perfusion fixation of lungs for structure-function analysis: Credits and limitations,'' J. Appl. Physiol.: Respir., Environ. Exercise Physiol. **53**(2), 528–533 (1982).
[37] G. Hayatdavoudi, J. D. Crapo, F. J. Miller, and J. J. O'Neil, ''Factors determining degree of inflation in intratracheally fixed rat lungs,'' J. Appl. Physiol.: Respir., Environ. Exercise Physiol. **48**(2), 389–393 (1980).
[38] J. Neter, M. H. Kutner, C. J. Nachtsheim, and W. Wasserman, *Applied Linear Statistical Models*, 4th ed. (Irwin, Chicago, 1996), Chaps. 1–2, pp. 3–94.

Modulation spectra of natural sounds and ethological theories of auditory processing

Nandini C. Singh[a] and Frédéric E. Theunissen[b]
Department of Psychology and Neuroscience Institute, University of California, Berkeley, 3210 Tolman Hall, Berkeley, California 94720-1650

(Received 23 April 2003; revised 28 August 2003; accepted 15 September 2003)

The modulation statistics of natural sound ensembles were analyzed by calculating the probability distributions of the amplitude envelope of the sounds and their time-frequency correlations given by the modulation spectra. These modulation spectra were obtained by calculating the two-dimensional Fourier transform of the autocorrelation matrix of the sound stimulus in its spectrographic representation. Since temporal bandwidth and spectral bandwidth are conjugate variables, it is shown that the joint modulation spectrum of sound occupies a restricted space: sounds cannot have rapid temporal and spectral modulations simultaneously. Within this restricted space, it is shown that natural sounds have a characteristic signature. Natural sounds, in general, are low-passed, showing most of their modulation energy for low temporal and spectral modulations. Animal vocalizations and human speech are further characterized by the fact that most of the spectral modulation power is found only for low temporal modulation. Similarly, the distribution of the amplitude envelopes also exhibits characteristic shapes for natural sounds, reflecting the high probability of epochs with no sound, systematic differences across frequencies, and a relatively uniform distribution for the log of the amplitudes for vocalizations. It is postulated that the auditory system as well as engineering applications may exploit these statistical properties to obtain an efficient representation of behaviorally relevant sounds. To test such a hypothesis we show how to create synthetic sounds with first and second order envelope statistics identical to those found in natural sounds. [DOI: 10.1121/1.1624067]

PACS numbers: 43.80.Ka, 43.64.Bt, 43.64.Qh [WA] Pages: 3394–3411

I. INTRODUCTION

Natural sounds span a restricted range of all possible sounds just as natural scenes only represent a small subset of all possible images (Attneave, 1954; Field, 1987). This phenomenology can be quantified by calculating the degree of statistical redundancy found in natural sounds. The use of this redundancy is clearly demonstrated by the multiple forms of compression that are available for the digital storage of music and that result in relatively little perceptual degradation (Painter and Spanias, 2000). We will argue that the characterization of the statistics of natural sounds is also potentially important for understanding acoustical perception and its underlying neuro-physiological basis. A theory of neural representation and neural computation in sensory systems that takes into account the natural environment, as originally proposed by Attneave (1954) and Barlow (1961), has been fruitful in advancing our understanding of the visual system and we propose that a similar approach will lead to insights in auditory science. This theoretical framework leads to a series of predictions and experiments that have now demonstrated how neural computations and representations in the early stages of the visual system are adapted to the processing of natural scenes (reviewed in Simoncelli and Olshausen, 2001). For example, the spatio-temporal receptive fields of visual neurons have been shown to perform optimal filtering operations on natural images (van Hateren, 1992a; Dan *et al.*, 1996).

The use of natural sounds for understanding auditory processing has, for the most part, followed a different path. On one hand, auditory neuroethologists were pioneers in the use of behaviorally relevant stimuli to probe the physiology of the sensory systems. This approach led to the classic discoveries of pulse-echo tuned neurons in the bat (Suga *et al.*, 1978), song selective neurons in songbirds (Margoliash, 1983) and call selective neurons in the primate (Newman and Wollberg, 1978). In this respect, the auditory system appears to be at least as ''selective'' for specific natural sounds as the visual system is for specific natural images. On the other hand, a systematic study of the statistical structure that characterizes these natural vocalizations and then would yield theoretical predictions for the response properties of single or network of auditory neurons has not been pursued to the same degree as in the visual modality.

Two studies have taken this systematic approach by analyzing the statistics of the sound pressure waveform. In an initial study, Rieke *et al.* (1995) demonstrated that auditory nerve fibers in the frog transmitted information more efficiently when the power spectrum of broadband sounds matched the power spectrum of the natural frog call. More recently, by examining the higher-order statistics of natural sounds, Lewicki found that the basis set that best represented the independent components of vocalizations was obtained by a Fourier decomposition whereas the basis set that best

[a]Current affiliation: National Brain Research Center; Sector-15, Part II; Gurgaon, 122 001, Haryana, India.
[b]Author to whom correspondence should be addressed. Electronic mail: fet@socrates.berkeley.edu

represented the independent components of environmental sounds was obtained by a wavelet decomposition. The biological basis set generated by the filtering properties of the cochlea and the hair cells fell in the middle of these two solutions, suggesting that the initial stage of auditory processing could have evolved to be optimized to the different statistics of these two important groups of natural sounds (Lewicki, 2002).

Here we extend this theoretical approach to the auditory computations and representations not of the sound pressure waveform but of the spectro-temporal amplitude envelopes that are obtained by the decomposition of sound into frequency channels. This decomposition is performed in biological systems by the cochlea and in engineering applications, such as speech recognition, by the use of filter-banks. The importance of the spectro-temporal amplitude envelopes of sound in capturing the significant statistics of the natural sounds as well as in predicting the neural response of higher-level auditory neurons is very well documented. First, spectrograms are used extensively in the analysis of animal vocalizations, not only because they provide a clear pictorial representation of the different types of vocal gestures, but also because the spectrographic representation is a better pictorial match of our perception of the sound than any plot of the sound pressure waveform. For the same reasons, time-frequency representations are used extensively in preprocessing stages of speech recognition or sound compression algorithms (Painter and Spanias, 2000). Second, the importance of the statistical structure of these envelopes for speech perception has clearly been demonstrated. Degradation of this structure along either the spectral or temporal dimension results in loss of intelligibility (Drullman *et al.*, 1994; Drullman, 1995; Shannon *et al.*, 1995). Similarly, psychophysical studies have shown that humans are particularly sensitive to either temporal modulations alone (Viemeister, 1979), spectral modulations alone (Green, 1986) or the joint spectro-temporal modulations (Chi *et al.*, 1999) of these amplitude envelopes, and that this sensitivity is restricted to relatively low modulations rates. Finally, whereas auditory neurons in the auditory midbrain and forebrain are not sensitive to the phase of the sound pressure waveform, they do acquire novel temporal and spectral amplitude modulation tuning that is not observed at the lower levels of auditory processing stream (Popper and Fay, 1992). For these reasons, the characterization of the response properties of higher level auditory neurons has included their response to amplitude modulated tones (Phillips and Hall, 1987; Eggermont, 2002), spectrally modulated sounds (Schreiner and Calhoun, 1994; Calhoun and Schreiner, 1998) and more recently to complex spectro-temporal stimuli which are used to extract the joint spectral-temporal receptive fields (STRFs) of the neurons (Eggermont *et al.*, 1983; deCharms *et al.*, 1998; Theunissen *et al.*, 2000; Depireux *et al.*, 2001; Sen *et al.*, 2001; Escabi and Schreiner, 2002; Miller *et al.*, 2002).

Since both auditory perception and the responses of auditory neurons seem to be particularly sensitive to the structure in sound amplitude envelopes it becomes crucial to describe the statistical nature of this structure in natural sounds. Attias and Schreiner (1997) have begun to study the second order statistics of amplitude envelopes along the temporal dimension but very little is known about the joint statistics of the spectro-temporal modulations of natural sounds. For this reason, we investigated the lower order joint statistics of three different ensembles of natural sounds: human speech, zebra finch song and environmental sounds. As was done in Attias and Schreiner (1997), we calculated and fitted the probability distributions of amplitudes for such envelopes. We then calculated the joint second order statistics of the amplitude envelopes, which we call the modulation spectrum. We found that natural sounds have a characteristic modulation spectrum and discuss the implications of our results for an ethologically based theory of auditory processing.

II. METHODS

A. Estimating modulation spectra

Figure 1 illustrates how the modulation spectrum of a sound ensemble is defined. First, a specific time-frequency representation of the sound is calculated. For example, in Fig. 1(a), a spectrographic representation is chosen to display the spectral and temporal structure present in a zebra finch song. This time-frequency representation can be expressed in its Fourier domain. On the right panel, the 2-D image made by the spectrogram is shown as a weighted sum of sinusoidal gratings of variable period, orientation and phase. Each spectrographic ''grating'' corresponds to a particular broadband sound called a ripple sound. The ripple sounds are characterized by their sinusoidal amplitude modulations in time and in frequency. The function describing the amplitude envelope for each frequency band f of a particular ripple sound is written as

$$S_i(t,f) = A_i \cos(2\pi\omega_{t,i}t + 2\pi\omega_{f,i}f + \phi_i). \tag{1}$$

The spectrogram for the sound of interest can then be written as a sum of such ripple components (or the equivalent integral in a continuous formulation):

$$S(t,f) = A_0 + \sum_i S_i(t,f).$$

A_i determines the relative strength of the modulation depth (relative to the dc term A_0) for that particular ripple sound component. The parameter ω_t describes the modulation frequency of the amplitude envelope along the temporal dimension has units of Hz. In this report, it is referred to as the temporal modulation frequency or simply the modulation frequency. In other reports it has also been named ripple velocity or drifting velocity. Since ripple velocity has been used to describe the number of frequency units spanned per second by the ripple (ω_t/ω_f), we will avoid the use of that term. The parameter ω_f describes the modulation frequency of the amplitude envelope along the spectral dimension and has units of 1/Hz or for wavelet time-frequency representations 1/oct. In this report, it is referred to as the spectral modulation frequency. It has also been called ripple density or ripple peak density (Chi *et al.*, 1999; Klein *et al.*, 2000; Depireux *et al.*, 2001; Escabi and Schreiner, 2002). φ is the initial phase of the ripple. Note that although we use the

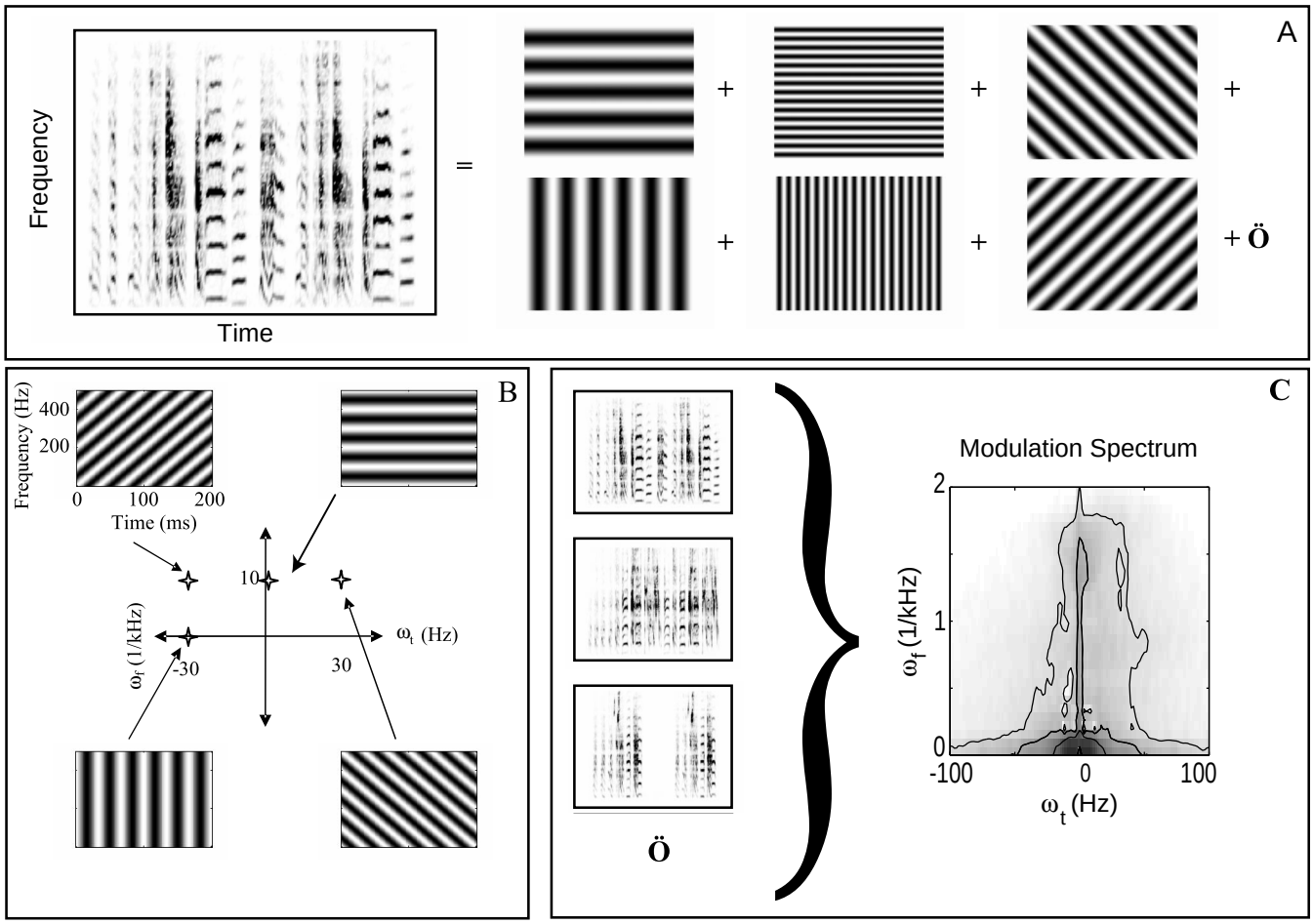


FIG. 1. *Definition of the modulation spectrum.* Panel (a) illustrates the decomposition of a sound represented by its spectrogram into its Fourier components: a sum of ripple sounds, which can be thought of as the acoustic analog of visual gratings. Each subfigure on the right side of the equation is the spectrogram of a single ripple sound component. The sound shown is a song from a zebra finch. As illustrated in (b), ripples are characterized by their temporal modulation, ω_t (Hz), their spectral modulations, ω_f (1/Hz or 1/oct), and their phase. A single point in a two-dimensional Cartesian plot can be used to represent the ripple sound components of a given spectral and temporal modulation, irrespective of phase. (c) To calculate the modulation spectrum, a representative group of sounds is decomposed into its ripple components and the power density of each ripple is estimated and plotted with gray-scale on the two-dimensional Cartesian plot. For the modulation spectrum shown in (c), we used 20 zebra finch songs of approximately 2 s each. The details of the calculations are illustrated in Fig. 2.

symbol ω, the modulation frequencies in Eq. (1) and in the rest of the paper are specified in units of oscillation frequencies and not in angular frequencies.

The modulation spectrum then shows the density distribution of amplitudes A_i of the component ripple sounds for an ensemble of sounds as a function of ω_t and ω_f. Figure 1(c) shows the modulation spectrum for an ensemble of zebra finch song. For time-frequency representations that yield a real valued amplitude envelope, the modulation spectrum is symmetric along the origin and therefore can be shown in two quadrants. Ripple sounds where $\omega_f = 0$ are broadband noise that are sinusoidally modulated in amplitude at a frequency given by ω_t. Ripple sounds where $\omega_t = 0$ are constant sounds with a sinusoidal frequency spectrum where the distance between peaks in the spectrum is given by $1/\omega_f$. Ripple sounds where $\omega_t \cdot \omega_f \geqslant 0$ are down-sweeps and are shown in the upper right quadrant. Ripple sounds where $\omega_t \cdot \omega_f \leqslant 0$ are up-sweeps and are shown in the upper left quadrant [see Fig. 1(b)]. As we will explain below, the range of possible values for ω_t and ω_f is restricted because of the mathematical nature of time-frequency representations.

The calculation of the modulation spectrum is similar to that of the standard frequency spectrum except that it requires the additional preprocessing step of calculating the spectrogram of the sound (or any other time-frequency representation) before calculating the modulus square of the 2-D Fourier transform of the ensemble of spectrograms [Fig. 1(c)]. As is the case for the frequency spectrum, the same result can be obtained by first estimating the auto-correlation function and then calculating the real valued 2-D Fourier transform. Figure 2 illustrates the entire calculation process using this second approach. A spectrogram is first obtained by decomposing the sound into an ensemble of narrow-band signals obtained from the output of a filter bank. The amplitude envelope of each narrow-band signal is obtained from the analytical signal (Flanagan, 1980; Cohen, 1995; Theunissen and Doupe, 1998). The value of the amplitude envelope calculated in that fashion is identical to the amplitude obtained in a short-time Fourier transform of a segment of sound centered at t and windowed with a function given by the Fourier transform of the gain function of the particular filter in the filter bank (Flanagan, 1980). As described in

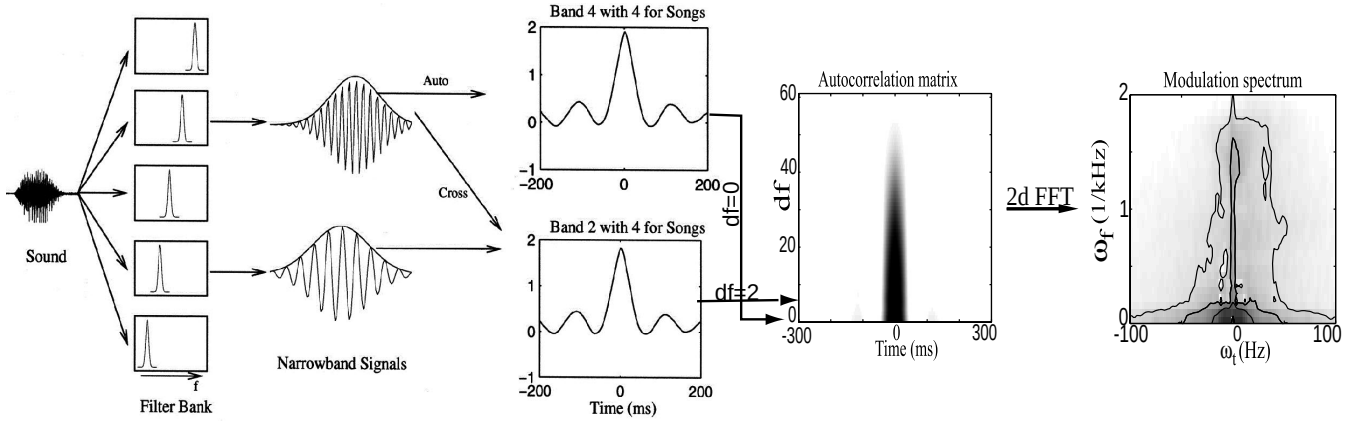


FIG. 2. *Calculation of the modulation spectrum*. A spectrographic representation of the sounds in the ensemble is first obtained by a decomposition into frequency bands using a bank of Gaussian filters. This decomposition results in a set of narrow-band signals with time-varying amplitude envelopes. The spectrogram is a pictorial representation of this time-varying envelope. The stimulus auto-correlation matrix is then obtained by cross-correlating the amplitude envelope of a particular band with the amplitude envelope of all the other bands, including itself. These cross-correlation functions are then averaged for all functions with equal frequency band offsets (df) and collapsed into an auto-correlation matrix, which shows the correlations as a function of time delay (X-axis) and frequency band offset (Y-axis). The two-dimensional Fourier transformation of this auto-correlation matrix is calculated to obtain the modulation spectrum of the sound ensemble.

more detail below, further transformations can then be applied to the amplitude of the envelopes before calculating the modulation spectrum. The amplitude envelopes (or its transformed value) in each band are then used to estimate an autocorrelation matrix, which shows the average product of the amplitude at frequency f and time t with the amplitude at frequency $f+df$ and time $t+dt$. The average is taken over all times t and frequencies f. The 2-D Fourier transform of this auto-correlation matrix yields the modulation spectrum: $P_{\mathrm{MS}}(\omega_t,\omega_f)$, where ω_t are the temporal frequencies corresponding to dt and ω_f are the spectral frequencies corresponding to df. For a modulation spectrum based on a wavelet decomposition, a similar calculation is done but with a filter bank of logarithmically spaced filters and fixed octave widths.

In our studies, we used three separate filter banks. In all three cases, the filters had Gaussian shapes and each filter bank was characterized by the fixed bandwidth of the filters. We used widths of 62.5, 125 and 250 Hz measured as the standard deviation parameter of the Gaussian function describing the gain of each filter. The filters were equally spaced on the frequency axis and separated from each other by one standard deviation. The auto-correlation matrix was calculated for time delays of ± 300 ms. Before taking the 2-D Fourier transform, the auto-correlation matrix was multiplied by a Hanning window. For most of our analyses, we calculated the modulation spectra using a log transformation on the amplitude values and we subtracted the mean log amplitude before windowing. The log transformation was used because, as shown here and previously (Attias and Schreiner, 1997), the distribution of envelope amplitude of natural sounds has a strong exponential component. The last plot in Fig. 2 shows a graphical representation of the modulation spectrum of a zebra finch song obtained in our calculations with the 125-Hz bandwidth filter bank and the log transform. As seen in the figure, the modulation spectrum has a low-pass characteristic both in temporal and spectral modulations and is slightly asymmetric with more energy for down-sweeps than up-sweeps.

B. Time-frequency scale and the estimation of modulation spectra

The bandwidth of the filters in the filter bank has a direct effect on the band occupancy of the amplitude envelope. In each frequency band, the temporal modulation spectrum of the amplitude square of the envelope is restricted to frequencies below the bandwidth of the filter (Flanagan, 1980). Therefore, high frequency temporal modulations can only be observed with wide bandwidth filters. Similarly, the spectral amplitude modulations for a given temporal window along frequency space are restricted to the modulation frequencies below the bandwidth given by the temporal window. Therefore, high frequency spectral modulations can only be measured with wide temporal windows. Since the temporal window is given by the Fourier transform of the gain function of the filter in the filter bank (Flanagan, 1980), high frequency spectral modulations can only be measured with narrow band-pass filters. These properties are another form of the well-known compromise between time and frequency resolution in time-frequency representations (Cohen, 1995).

Because of the time-frequency trade-off in resolution, one cannot generate a spectrographic representation that exhibits both high spectral and high temporal frequency modulations. Since spectrographic representations can be designed to be invertible (up to a single absolute phase) physical sounds that have simultaneously high spectral and high temporal amplitude modulations in the spectrographic time-frequency representation do not exist. More specifically, the uncertainty principle tells us that the product of the bandwidth, σ_f, and duration, σ_t, of the sound sample (the windowed signal used in the spectrographic decomposition) must satisfy the following inequality (Cohen, 1995):

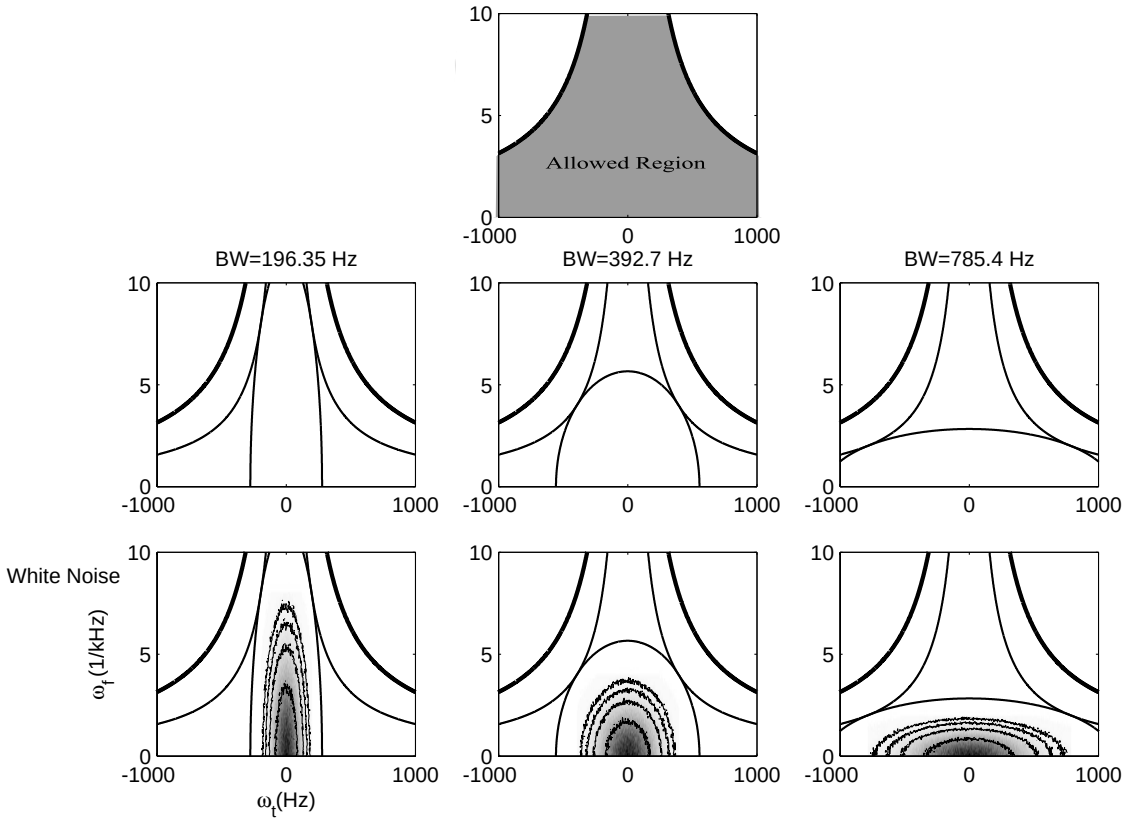


FIG. 3. *Schematics showing the physically plausible region of power for modulation spectrum and the region sampled by choosing a particular time-frequency scale*. The figure in the top row shows the region of physically plausible power for modulation spectra obtained from a spectrographic representation of sound. Since time and frequency are conjugate variables, the uncertainty principle imposes restrictions on the simultaneous sampling of fine temporal and spectral modulation frequencies (see text). The hyperbolic curves given by $\omega_t\ \omega_f = \pi$ enclose the allowed region shaded in gray. The *middle row* shows the actual region sampled by choosing a particular bandwidth for Gaussian filters in the filter bank: only the region within an ellipse can be sampled at a time. For spectrograms, the bounds of the ellipse are given by a second set of hyperbolic curves defined by $\omega_t\ \omega_f = \pi/2$. Choosing narrow bands in the spectrogram (narrow filters in the filter bank) results in a relatively low value for the highest temporal modulation that can be measured and a relatively high value for the highest spectral modulation that can be measured and vice versa. This principle is illustrated by estimating the modulation spectrum of white-noise shown in the *bottom row*. The modulation spectrum of white noise should fill the entire allowed region but, depending on the time-frequency scale chosen, we find power only within the region defined by the ellipse shown in bold. The contour lines show the areas that encompass 50%, 80%, 90% and 95% of the total power.

$$\sigma_f \sigma_t \geqslant \frac{1}{4\pi}.$$

Since we are measuring modulation frequencies (spectral and temporal) by estimating the correlations across multiple measurements of such samples, we can rewrite this inequality, by specifying the upper frequency limit of the temporal and spectral modulations which are given by $\max(\omega_t) = 1/2\sigma_t$, and $\max(\omega_f) = 1/2\sigma_f$. Sounds are therefore restricted to the range of modulation frequencies given by $|\omega_t \cdot \omega_f| \leqslant \pi$.

We provide an illustration of these properties by calculating the modulation spectrum of white noise for the three different bandwidths that we used in our filter bank (see Fig. 3). White noise sound includes all modulation frequencies and should therefore fill uniformly the entire area given by $|\omega_t \cdot \omega_f| \leqslant \pi$ as shown on the top panel (gray area labeled ''Allowed Region''). However, when we choose a particular time-frequency scale for our spectrographic representation, we are effectively only measuring modulation frequencies that are found in a subarea within this allowed region. As mentioned above, the maximum temporal and spectral modulation frequencies are given by the bandwidth of the filter in the filter bank. For Gaussian-shaped filters, the bandwidth of nonzero power is theoretically infinite but the power in the higher frequency modulations quickly decreases. The effective bandwidth of the filter, defined mathematically as the square root of the average deviations square from the center frequency, is simply the standard deviation parameter of the Gaussian filter in the filter bank, σ_{filt}. The bandwidth of the spectral modulation is then given by the maximum spectral modulation: $\text{BW}(\omega_f) = \max(\omega_f) = 1/2\,\sigma_{\text{filt}}$. The time window that corresponds to the Gaussian filter in the filter bank is also a Gaussian function with standard deviation parameter (also the effective duration of the window) given by $\sigma_t = 1/2\,\pi\sigma_f$. The bandwidth of the temporal modulations is then given by the maximum temporal modulation: $\text{BW}(\omega_t) = \max(\omega_t) = 1/2\,\sigma_t = \pi\sigma_f$. Thus when a spectrogram is obtained with Gaussian filters the product of the temporal modulation bandwidth and spectral modulation bandwidth is $\text{BW}(\omega_f) \cdot \text{BW}(\omega_t) = \pi/2 \leqslant \pi$, as required by the uncertainty principle. The hyperbola describing this function within the allowed region is shown in the middle plots of Fig. 3. In addition, for a given time-frequency scale set by the parameter describing the width of the filters in the filter bank (or

equivalently the width of the time window), the modulation spectrum will be restricted to frequencies given by the specific BW(ω_t) and BW(ω_f) corresponding to σ_f. For a Gaussian filter, the regions sampled are not rectangular but, as shown in Fig. 3, ellipsoid with major and minor axes given by $\sqrt{2}$BW(ω_t) and $\sqrt{2}$BW(ω_f).

Figure 3 shows the modulation spectrum for white noise calculated for the three values of the time-frequency scale parameter used in the spectrograms in our analyses, which was determined by the width of the filter in the filter bank (σ_{filt}=62.5,125,250 Hz). The corresponding spectral-temporal modulation ellipse was then determined by BW(ω_t) (196.35, 392.7, 785.4 Hz) and BW(ω_f) (8,4,2 kHz^{-1}). All of the significant estimated energy in the modulation spectrum fell within this ellipse. For white noise, a large area at the center of the ellipse was uniformly sampled, illustrating the fact that white noise also has white modulation spectra. Note, however, that the power in the modulation does decrease significantly before reaching the edge of the ellipsoid area. The modulation spectra for white noise obtained at different time-frequency scales has the same geometric shape but occupies a different modulation frequency area. Clearly, for white noise, there is no appropriate time-frequency scale since the entire range would be required to properly describe the spectral and temporal modulations actually present in the sound. In this case, the modulation spectra obtained from a spectrographic representation are uniquely a reflection of the shape and bandwidth of the filters in the filter bank. For other sound ensembles, however, there might be time-frequency scales that include most of the energy in the spectral and temporal modulations present in the sounds. For those sound ensembles, one could therefore estimate the modulation spectrum from a single spectrographic representation with the realization that modulation frequencies that would fall outside the sampled ellipse would be filtered out.

We attempted to estimate a measure of the optimal time-frequency scale of the spectrogram by measuring the entropy of the power density function given by the modulation spectrum. We reasoned that the modulation spectrum with the highest entropy would be the one that included most of the temporal and spectral modulation structure found in the song. To calculate the entropy, we transformed the modulation spectra into a discrete probability function that specified the probability of the occupancy of a discrete subdivision of the modulation spectrum defined into small rectangles defined by $(\omega_t - d\omega_t, \omega_x - d\omega_x)$ and $(\omega_t + d\omega_t, \omega_x + d\omega_x)$, $p(\omega_t, \omega_f)$. The limits for ω_t and ω_f were chosen to cover the space given by the corresponding sample ellipse for each bandwidth and $d\omega_t$ and $d\omega_f$ were set for all three bandwidths at $d\omega_t$=1.66 Hz and $d\omega_f$=0.065 cycles/kHz. The entropy of the probability distribution is then obtained with

$$H(P) = \sum_{\omega_t} \sum_{\omega_f} -p(\omega_t, \omega_f) \log_2(p(\omega_t, \omega_f)).$$

This entropy measure has units of bits but its absolute value is not interpretable since it is dependent on the size of $d\omega_t$ and $d\omega_f$ and on the choice of units. We are using the measure solely in a relative manner to compare the variability in the modulation spectrum obtained at different time-frequency scales.

C. Descriptive quantifiers of modulation spectra

To quantitatively describe some of the structure observed in the modulation spectra of our sound ensembles, we used a small set of simple measures that estimated the separability, symmetry, low-pass quality and shape. For these four measures, we used the modulation spectra on the log amplitude with the mean level subtracted as explained in Sec. II.A. We also calculated separately the relative power of the dc component of the linear amplitudes to yield a measure of modulation depth. Similar measures have been also used to quantify the modulation spectrum of the spectral-temporal receptive field of auditory neurons (Depireux *et al.*, 2001).

1. Separability

A fully separable modulation spectrum is one that will factorize into a function of ω_t and ω_f over all quadrants. $P_{\text{MS}}(\omega_t, \omega_f) = G(\omega_t).H(\omega_f)$. A separable modulation spectrum signifies that the probability of occurrence of joint spectral-temporal modulations (down-sweeps of up-sweeps) is expected from the average probability of the spectral or temporal modulations measured separately. To quantify the separability, we calculated the singular value decomposition of the modulation spectrum,

$$P_{\text{MS}}(\omega_t, \omega_f) = \sum_{i=1}^{n} \lambda_i g_i(\omega_t) \cdot h_i(\omega_f), \lambda_1 > \lambda_2 > \cdots > \lambda_n,$$

and calculated the ratio of the first singular value relative to the overall power given by the sum of all singular values:

$$\alpha_{\text{sep}} = \frac{\lambda_1}{\Sigma_{i=1}^{n} \lambda_i}.$$

When the modulation spectrum is fully separable, α_{sep} will be close to 1.

2. Asymmetry

The modulation spectrum will be asymmetric if there are more down-sweeps than up-sweeps in the sound ensemble. We quantified the asymmetry by calculating the relative power in the first and second quadrants:

$$\alpha_{\text{asym}} = \frac{P_{\text{down}} - P_{\text{up}}}{P_{\text{down}} + P_{\text{up}}},$$

where P_{down} is the total power in upper right quadrant (of positive ω_t and ω_f) and P_{up} is the total power in the upper left quadrant (of negative ω_t and positive ω_f). If α_{asym} is close to 0, the modulation spectrum is symmetric. If α_{asym} is positive, then there are more down-sweeps than up-sweeps in the sound ensemble and vice-versa.

3. Low-pass coefficient and starriness

We observed that in natural sounds most of the energy is concentrated in the low temporal and spectral frequencies and that the energies in the higher temporal and spectral modulations are not distributed uniformly but instead are

concentrated along the axes. In particular, for animal vocalizations and human speech, most of the spectral modulations are found only for low temporal modulations. To quantify these effects, we calculated two parameters. The first parameter measures the energy in the low frequencies relative to the total energy:

$$\alpha_{\text{low}} = \frac{P_{\text{low}}}{P_{\text{total}}},$$

where

$$P_{\text{low}} = \int_{-\Delta\omega_t}^{+\Delta\omega_t} \int_{-\Delta\omega_f}^{+\Delta\omega_f} P(\omega_t, \omega_f) d\omega_t d\omega_f .$$

We chose $\Delta\omega_t = 10$ Hz and $\Delta\omega_f = 0.195$ kHz^{-1}.

The second parameter measures the relative energy of the modulation spectrum that excludes the regions of joint high temporal and spectral frequency as well as the region of very low joint temporal and spectral frequencies calculated with P_{low}. This area is found next to the x axis and y axis and includes the high temporal modulations but only at low spectral modulations and vice-versa. It is calculated with

$$\alpha_{\text{star}} = \frac{P_{\Delta\omega_t} + P_{\Delta\omega_f} - 2P_{\text{low}}}{(P_{\text{total}} - P_{\text{low}})},$$

where

$$P_{\Delta\omega_t} = \int_{-\infty}^{+\infty} \int_{-\Delta\omega_f}^{+\Delta\omega_f} P(\omega_t, \omega_f) d\omega_t d\omega_f$$

is the total modulation power in a band of spectral frequencies limited by $\delta\omega_f$ and similarly for $P_{\Delta\omega_f}$. P_{total} is the total power in the modulation spectrum.

4. Shape separability

The measure of separability defined above is critically dependent on the power distribution. Since natural sounds have a high concentration of power in the low frequencies, we found that α_{sep} was relatively high for all the natural sound ensembles. However, we could also observe and further quantify with the starriness parameter that the energy outside the low spectral and temporal frequencies was not uniformly distributed. To examine the shape of the distributions, we calculated the separability for an occupancy matrix: given a contour line defined by the percent of the total power within the contour, we set all the values within the contour to have a value of 1 and all values outside to have a value of 0. We then calculated a separability index for this occupancy matrix.

5. Modulation depth

A measure of modulation depth can be estimated by looking at the ratio between the dc power and the power in the rest of the frequencies:

$$\alpha_{\text{mod}} = \sqrt{\frac{P_{\text{total}} - P_{\text{dc}}}{P_{\text{dc}}}}$$

where P_{total} is the total power and P_{dc} is the power at dc, i.e., power at $\omega_t = 0$ and $\omega_f = 0$. We used the square root because modulation depth is usually defined from the amplitude of the envelopes. This measure was applied to the modulation spectrum obtained without the logarithmic transformation.

D. Generating synthetic sounds from a modulation spectrum

We describe a straightforward methodology to generate complex sounds that can match both the frequency and modulation spectrum of a sound ensemble. Our interest is in designing synthetic sounds that match the first and second order modulation statistics of natural sounds. In particular, these sounds can be used to estimate the spectro-temporal tuning of auditory neurons as well as their potential sensitivity to the phase of the modulations that are present in natural vocalizations. Similarly, one could use these synthetic sounds to study perceptual sensitivity to specific spectral-temporal modulations or phase in human or animals.

The method is similar to the method used by Klein *et al.* (2000) and Escabi and Schreiner (Escabi and Schreiner, 2002) to generate synthetic sounds with a band-limited flat modulation spectrum, which has been called noise ripple. For noise ripple the space of ω_t and ω_f is sampled uniformly within some frequency bounds. In our case, we wanted to match our sampling to the modulation spectrum obtained from a particular sound ensemble. For this purpose, we sampled the desired modulation spectrum by normalizing the power spectral density to obtain a probability density function and randomly choosing N distinct pair of values for ω_t and ω_f from that distribution.

To generate the function that describes the amplitude envelope for our synthetic sound, we then obtained the envelope function for a sum of ripple sounds (see Sec. II.A):

$$S(t,f) = \sum_{i=1}^{N} \cos(2\pi\omega_{t,i}t + 2\pi\omega_{f,i}f + \varphi_i),$$

where φ_i is a random phase for each ripple component. In our implementation we generated synthetic sound ensembles made of 20 synthetic sounds, each 2 s in duration. To synthesize the envelope of each noise ripple sound, we used $N = 100$ ripple components.

An ensemble of sounds with an amplitude envelope given by $S(t,f)$ will have the same modulation spectrum as the original sound ensemble but will also have, on average, a flat frequency spectrum and, on average, a flat temporal envelope. The flat average temporal envelope will also be found in the original sound (if the sound ensemble is stationary in time) but the average flat frequency spectrum is unlikely to be found in natural sounds. To match the overall frequency spectrum and the dc value of the modulation spectra (the modulation depth), we normalized $S(t,f)$ by the average standard deviation of the amplitude modulation in each frequency band f and added the mean amplitude envelope. Calling $A(f)$ the average amplitude in each frequency band measured in the original ensemble, $\sigma(f)$ the standard deviation in each frequency band measure in the original ensemble and $\sigma_S(f)$ the standard deviation obtained from

the ensemble of $S(t,f)$ functions for the synthetic ensemble, we generate a new function for the amplitude envelopes given by

$$S_{\text{Norm}}(t,f)=A(f)+\frac{\sigma(f)}{\sigma_S(f)}S(t,f).$$

Finally, to synthesize the sound a direct method or a more precise iterative method can be used. For the direct method, one simply creates an ensemble of carrier frequencies that will be modulated by S_{Norm}. To prevent any artifacts in periodicity, the frequency of the carrier frequencies should be chosen randomly with a uniform distribution between the lower and upper bounds of the desired frequency range. In our case, we set the lower frequency bound $f_{\min}=250$ Hz and the upper bound $f_{\max}=8$ kHz. Each carrier sound has the form

$$s_i(t)=\cos(2\pi f_i t+\theta_i),$$

where f_i is a random frequency between $f_{\min}$ and $f_{\max}$ and θ_i is a random phase and $i=1$ to Nc. We found that $Nc=1000$ carrier frequencies were more than sufficient to sample our range of frequencies. The synthetic sound is finally given by

$$s_{\text{syn}}(t)=\sum_{i=1}^{Nc} S_{\text{Norm}}(t,f_i)s_i(t).$$

The more precise iterative method is called spectrographic inversion and effectively involves iteratively adjusting the phase of the carrier sounds, θ_i, in order to minimize the difference between the desired spectrogram S_{Norm} and the spectrogram obtained from the synthesized sound (Griffin and Lim, 1984). We used the implementation of the Griffin and Lim algorithm provided by Malcolm Slaney as a Matlab program (1994).

When we used the simple direct method, we found that the ensemble of synthesized sounds had very similar frequency spectrum, modulation depth and modulation spectrum as the original sounds. However, the spectrogram obtained from specific sample sounds from the synthetic ensemble could be quite different from the desired spectrogram due to random phase interferences. The iterative method yielded a much better one-to-one match of the spectrogram and a slight improvement on the match between the ensemble modulation spectra of the natural and synthesized sounds.

E. Natural sound ensembles

We analyzed the statistics of three natural sound ensembles: two types of animal vocalizations (speech and zebra finch song) and an ensemble of environmental sounds.

The speech ensemble was made of 20 sentences chosen randomly from the audio-visual speech test library recorded by the Otolaryngology Department at the University of Iowa (Tyler *et al.*, 1990). The sentences *corpus* consists of 100 short complete sentences read by six different adult male and female speakers. Examples of these sentences are ''It rained all day yesterday,'' ''The book tells a story,'' ''The mother reads a paper'' and ''They have only one son.'' The sentences are read out of context, in an acoustically controlled environment. The total length of sound sampled was approximately 40 s. These speech signals have been used previously in speech perception research (Shannon *et al.*, 1995; Dorman *et al.*, 1997).

The zebra finch song ensemble consisted of the songs of 20 different adult males (age$>$100 days) that were raised by their parents in a large zebra finch colony in our laboratory. The recordings were obtained in a noise-free environment by isolating individual male birds in a sound proof recording chamber. Multiple samples of each song were obtained and a particularly clean exemplar was chosen. Each song lasted approximately 2 s and the 20-song ensemble was approximately 40 s in duration.

The ensemble of environmental sounds was 45 s in duration and consisted of a rustling brush, crunching leaves and twigs, rain, fire and forest and stream sounds. These were recorded and provided to us by Michael Lewicki. Lewicki used these sounds to study the higher order statistics of the sound pressure waveform (Lewicki, 2002).

III. RESULTS

We examined the statistics of the temporal-spectral envelope obtained from spectrographic representations of speech, zebra finch song and environmental sounds.

A. Probability distributions of the modulation amplitude

We first examined the probability distribution of the amplitude of the modulation envelopes (Fig. 4). In this analysis, we used a spectrographic representation based on our intermediate value for the time-frequency scale [$\text{BW}(\omega_t)=392.7$ Hz]. The modulation envelopes were obtained as described in Sec. II.A. The middle row in Fig. 4 shows the distribution obtained for $p(A)$ in each frequency band. The distribution of amplitudes for all three natural sounds is strikingly different from that of white noise. The distribution of amplitudes for Gaussian white noise is given by the Rayleigh distribution: $p(A)\alpha Ae^{-A^2}$. The fit of our data with the theoretical distribution is shown in the bottom panel and the two distributions are indistinguishable (K-S test). On the other hand, the distributions for the natural sounds examined here have a strong exponential component and are best fitted with an exponential distribution or a gamma distribution. The exponential distribution gave good fits for song and speech and for the higher frequencies of environmental sounds. The gamma distribution gave good fits for the lower frequencies of the environmental sounds. The exponential shape of these distributions reflects the fact that, for vocalizations and for the higher frequencies of environmental sounds, there is a finite probability of finding sounds that are arbitrarily soft as in, for example, the silent pauses between speech syllables. We also found that there were systematic trends as functions of the center frequency of the band. Some of the differences can be explained simply by changes in amplitude and not by changes the shape of the probability distribution and these differences would not appear if we had normalized our probability distributions by their variance. For example, the coefficient of the exponential fit decreased from low to high

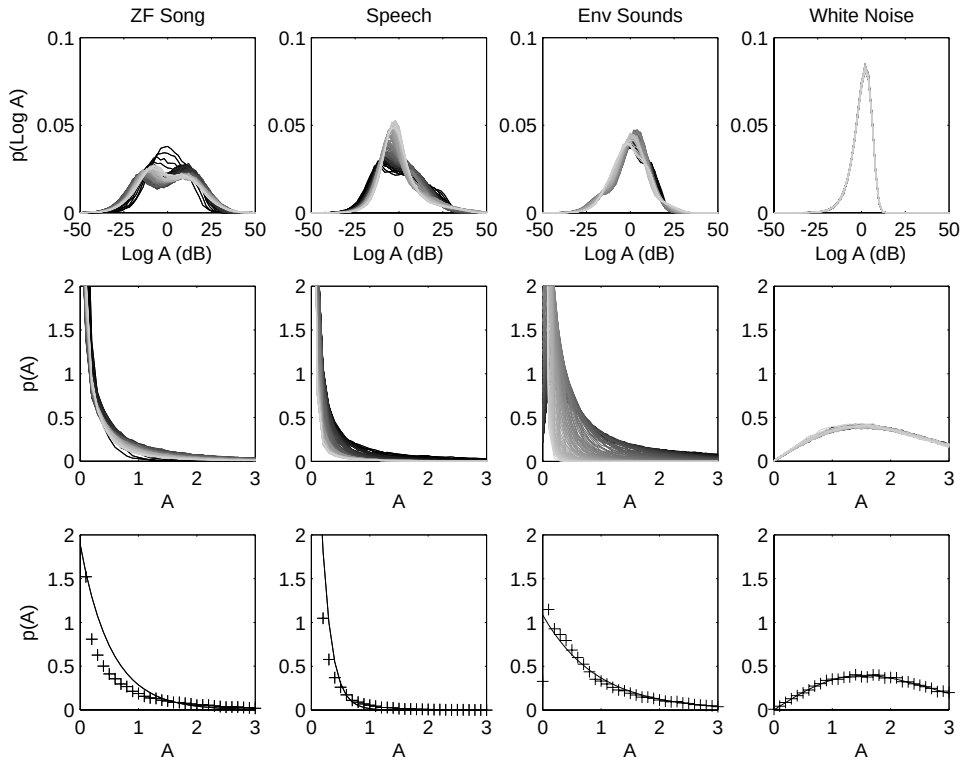


FIG. 4. *Amplitude distributions in different sound ensembles*. Probability density functions were estimated for the amplitude (bottom two rows) and the log amplitude (top row) of the envelope of the four sound ensembles examined in this study. The amplitudes of the envelopes were obtained by calculating the Hilbert transform of the decomposition of the sound into narrow bands as shown in Fig. 2 and explained in Sec. II.A. The filter bandwidth of $\sigma_{\text{filt}} = 125$ Hz was used [$\text{BW}(\omega_t) = 392.7$ Hz]. The gray scale lines show the results for the different frequency bands with the dark black line corresponding to $f = 875$ Hz and the lightest gray line corresponding to $f = 7312.5$ Hz. In the top row the probability distribution of the log amplitude is shown with the x-axis in dB units and with 0 dB corresponding to the mean of the distribution in each band. The middle row shows the probability distribution of the amplitude. The bottom row shows the probability distribution for the frequency band centered at 2500 Hz (crosses) and the best fit (solid) given by an exponential distribution for the natural sounds and given by the Rayleigh distribution for white noise.

frequencies, reflecting the higher probability of soft sounds in the upper frequency range. However, for speech and birdsong, we also observed systematic changes in the shape of the probability distribution as described in more detail in the next paragraph. Finally, although the fits were good in a mean square sense, the data showed systematic deviations from either the exponential form or the gamma form and these two theoretical models were rejected by the K-S test. The theoretical distribution of these natural sounds is therefore complex, potentially composed of multiple components.

The probability distributions were also examined in the logarithmic scale as shown in the top row of Fig. 4 where we plotted $p(\log(A))$ as a function of $\log(A)$. Here, we can also distinguish features that distinguish the distributions of the natural sounds from those of Gaussian white-noise. First, for zebra finch song and the lower frequency bands in speech, the distributions are approximately rectangular (low kurtosis) with similar probability distributions in a 40-dB range (-20 to 20). The distribution for zebra finch song has two peaks (bi-modal) corresponding to syllables and intersyllable silences. The relative peak probability of each peak changes as a function of frequency since there are more song syllables with energy only in the lower frequency range. The probability for speech sounds in the lower frequency range also has low kurtosis but it is not bimodal. Instead the distribution is asymmetric with an almost linear decrease in probability as a function of sound intensity. As frequency increases, the kurtosis also increases reflecting the much steeper slope in the linear region: the kurtosis is below 3 for frequencies below 3500 Hz and above 3 from all frequencies above. The mean kurtosis above 6000 Hz is 4.6. In comparison, the distribution of the log amplitude for the environmental sounds is more symmetric and has a kurtosis close to the normal distribution (mean kurtosis across all frequency bands is 3.26 relative to 3 for the normal distribution).

B. Modulation spectra of natural sounds and time-frequency scale

We calculated the modulation spectrum of the three natural sound ensembles using the methodology described in Sec. II.A and Sec. II.B. Figure 5 shows the modulation spectra for zebra finch song, speech and environmental sounds calculated for the three values of time-frequency scale that we investigated [$\text{BW}(\omega_t) = 196.35$ Hz, $\text{BW}(\omega_t) = 392.7$ Hz and $\text{BW}(\omega_t) = 785.4$ Hz]. The three time-frequency scales capture the principal features in the spectra since most of the energy is found at low spectral and temporal modulations. We can also visually verify the validity of the chosen range of time-frequency scale by noting that, for all three ensembles, the energy for the fastest temporal modulations decay to zero for the wide temporal bandwidth filter

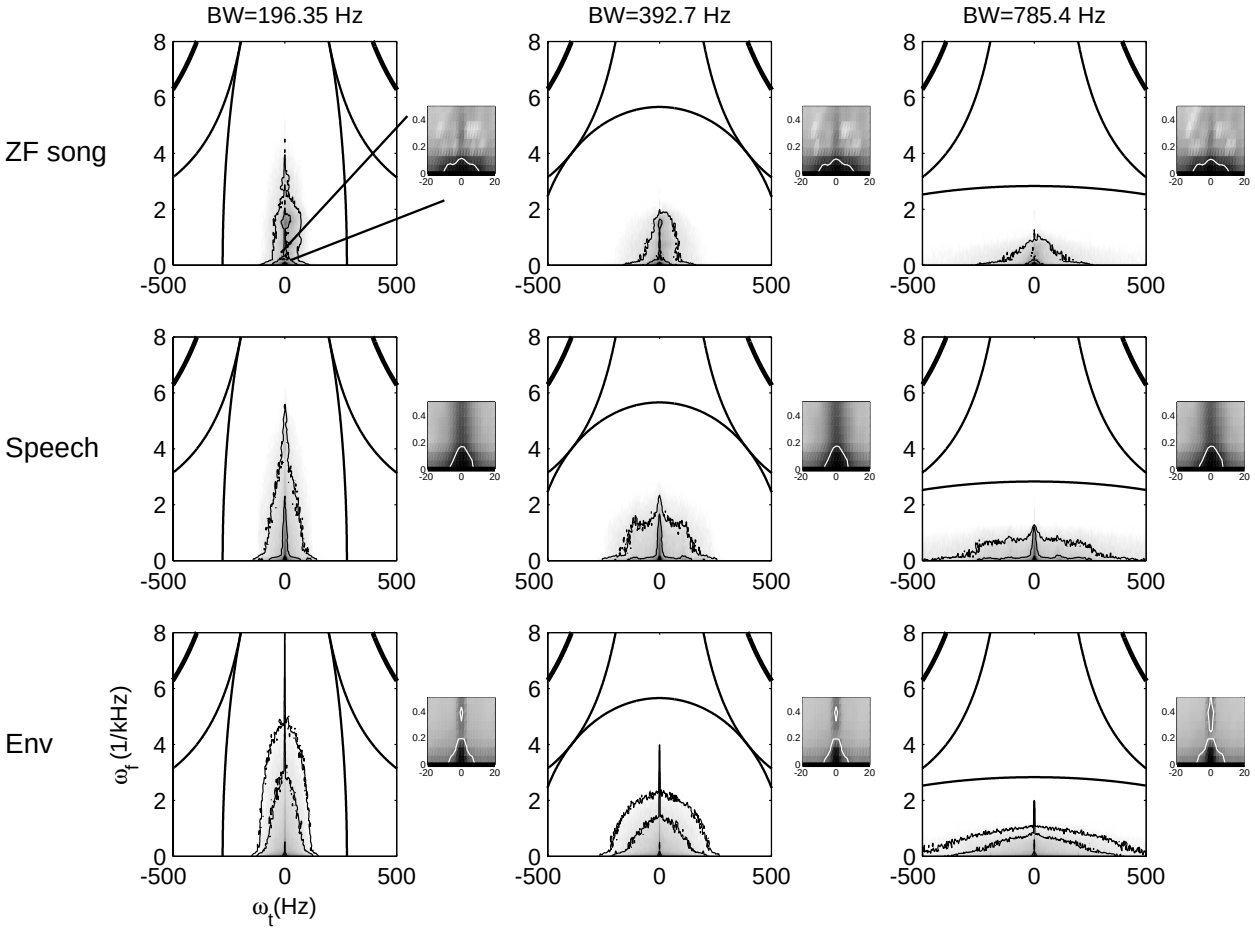


FIG. 5. *Modulation spectra for three different sound ensembles—zebra finch song, speech and environmental sounds obtained at three different time-frequency scales*. The modulation spectra of these three sound ensembles are shown for an effective temporal bandwidth [BW(ω_t)] of 196.35, 321 and 642 Hz. The contours are drawn to circle 50%, 80% and 90% of the total power. The small inset zooms in on the modulation spectra for the lower frequency range. The 50% contour is shown in white on these insets. Since a large fraction of the energy in modulation spectra of these natural sounds is concentrated in the low frequencies, the 50% contour is identical at all three time-frequency scales. Small differences are observed for the 80% contour and large differences for the 90% contour illustrating the fact that the time-frequency compromise affects approximately 20% of the modulation power found at the higher spectral and temporal modulations.

[BW(ω_t) = 785.4 Hz]. Similarly, the energy for the fastest spectral modulation decays to zero for the narrow temporal bandwidth filter [BW(ω_t) = 196.35 Hz]. Nonetheless, the smallest temporal bandwidth misses some of the fast temporal modulations and the widest frequency bandwidth misses some of the fast spectral modulations observed in these natural sounds.

We wanted to find a single time-frequency scale that yielded the best compromise for the representation of the fastest spectral-temporal modulations so that we could quantitatively describe and compare the spectra of these three natural sound ensembles. For this purpose, we calculated an information theoretic entropy measure as described in Sec. II.B. The results of that analysis are shown in Fig. 6. The entropy measures were similar for all three time-frequency scales, reflecting the fact that the probability distributions are well captured at any one of the three scales or that similar compromises are achieved. The highest entropy for the speech ensemble was obtained at the widest temporal bandwidth BW(ω_t) = 785.4 Hz whereas for zebra finch song ensemble the entropy is the highest at the narrowest bandwidth BW(ω_t) = 196.35. The differences, however, were not statistically significant. In the remainder of the paper, we show the results of our analysis of the modulation spectra at the intermediate time frequency scale [BW(ω_t) = 392.7 Hz] but very similar results were obtained at all three scales. To further estimate the effect of the potential compromise, we also calculated the power in the modulation spectra found outside the sampled ellipse. The power outside the area sampled by the intermediate time-frequency scale and found in the area sampled by the BW(ω_t) = 196.35 Hz scale was 1.6% of the total power for zebra finch song, 2.3% for human speech, and 3.7% for environmental sounds.

C. Modulation spectra of natural sounds

Within the allowed space imposed by the uncertainty principle and the time-frequency scale used in the measurements, we observed that the modulations in natural sounds have a characteristic distribution. As mentioned above, for all three sound ensembles, most of the energy is found for low spectral and temporal modulations. These natural sounds and in particular the vocalizations (Zebra finch song and human speech) are further characterized by nonoval distribu-

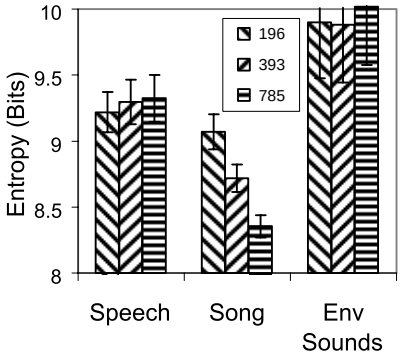


FIG. 6. *Entropy of the modulation spectrum*. The entropy of the modulation spectrum was calculated by treating it as a discrete probability distribution. The entropy was calculated to investigate the optimal time-frequency scale, which would be defined as the scale with the highest entropy value. The error bars show the standard error of the measure obtained by the jack-knife resampling method on the sound ensemble. The legend shows the temporal bandwidth BW(σ_t) in Hz.

tions, reflecting the fact that most of the high frequency spectral modulation power is found at the very lowest temporal modulation and vice versa. In other words, there is a scarcity of sounds with both high spectral and high temporal modulations. This property is best seen in Fig. 7 where we display a contour plot of the modulation spectra of the three natural sound ensembles and white noise for comparison. In this figure, we show all four quadrants of the modulation spectrum to visually emphasize the shape difference. The white noise contours are oval, reflecting the Gaussian-shaped filters, which are symmetric in time-frequency and the relative length of the temporal and spectral axes is determined by the time-frequency scale set by the bandwidth of the filters. For the natural sound ensembles, the contours that bound 50% of the energy are very close to the origin reflecting the low-passed property. The contours that bound 70% and 80% of the total energy draw a star-shape pattern reflecting the low probabilities of finding sound components with jointly high spectral and high temporal modulations. We quantified these observations by calculating various parameters describing the shape of these spectra as described in Sec. II.C. The results of these analyses are shown in Figs. 8 and 9.

First the separability index shows that the three natural sounds ensembles are quite separable. Only the speech ensemble, with an index of 0.84, is significantly different from the index found for the white noise ensemble, which is completely separable [Fig. 8(a)]. On the other hand, our starriness index, which calculates the relative energy in a the low temporal modulation band (the band of ± 10 Hz along the y axis) added to the energy in a low spectral modulation band ($\pm 0.195\ \mathrm{kHz}^{-1}$ along the x axis), is much larger in natural sounds than it is in white noise, reflecting the star-shaped pattern observed in Fig. 7 [Fig. 8(d)]. Although these two results seem contradictory, they have a simple explanation. As quantified by the low pass coefficient, a large fraction of the modulation energy spectrum (64% for zebra finch song, 61% for speech and 51% for environmental sounds) is found at the very low spectral and temporal modulations and the

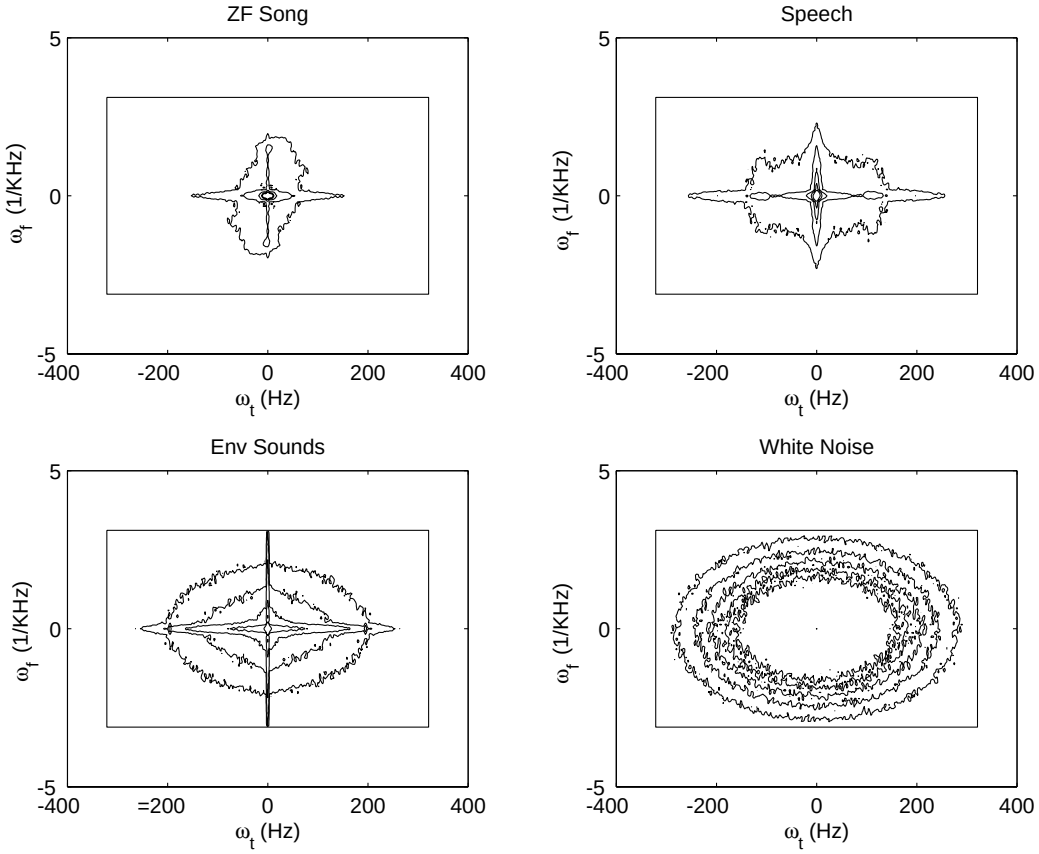


FIG. 7. *Modulation spectra displayed as contour plots*. The modulation spectrum for zebra finch song, speech, environmental sounds and white-noise are displayed with contour plots. The contours plots are drawn at the fixed power values that surround 50%, 60%, 70%, 80% and 90% of the total power.

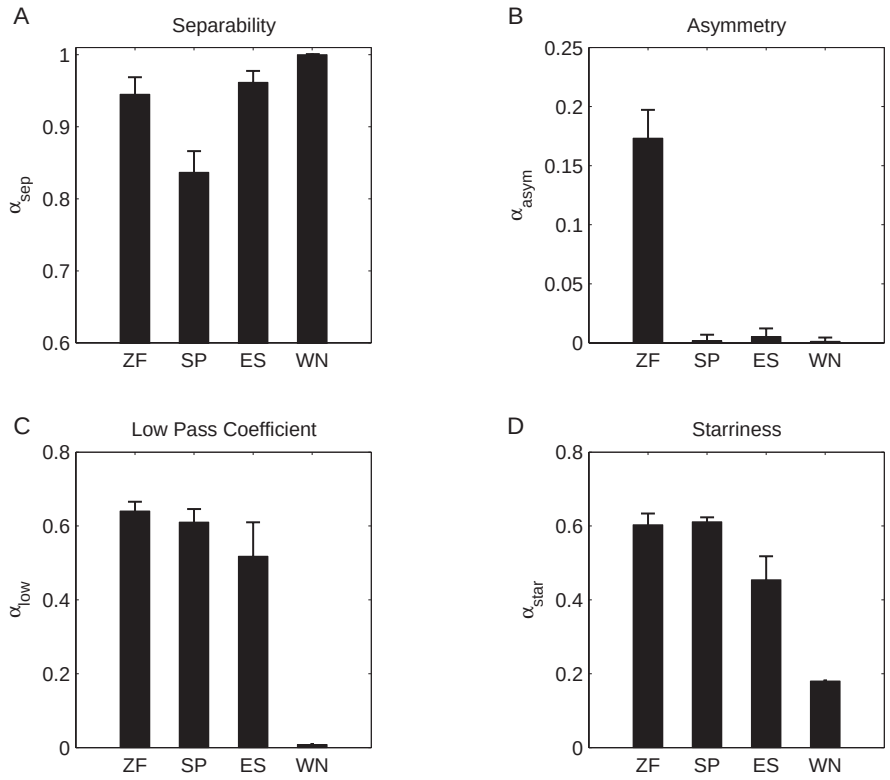


FIG. 8. *Separability, asymmetry, low-pass coefficient and starriness coefficients*. Four quantifiers that measure different aspects of the shape of the modulation spectrum were calculated for each ensemble (ZF: zebra finch song, SP: speech, ES: environment sounds and WN: white noise). The error bars show one standard error obtained with the jack-knife resampling technique. See Sec. II.C for the exact definition of the coefficients.

modulation distribution in this area is highly separable as illustrated by the circular 50% contour in Fig. 7 [Fig. 8(c)]. For those reasons, the separability index remains high. However, when one looks at the tail of the distributions found at the high spectral and high temporal modulations, the natural sounds lack the joint high spectral and temporal power found in the random signal. To measure this effect, we calculated the separability of an occupancy matrix defined by the contours that bound 50% to 90% of the total energy. This analysis showed that speech and environmental sounds are inseparable relative to noise for energy found between the 60% and 80% contours and zebra finch song for energies found between the 80% and 90% contours: the power in the higher modulation frequencies of the spectrum making approximately 30% of the energy for speech and environmental sounds and 20% of the energy for song are particularly concentrated along the spectral and temporal modulation axes.

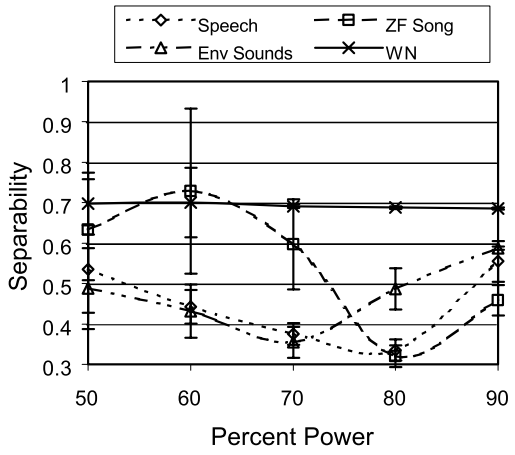


FIG. 9. *Occupancy separability at different thresholds*. An occupancy separability index was defined by calculating the separability coefficient for the space covered by the modulation spectrum. The space covered was defined by the contour line that bounded a given percent of the total power (shown on the *X* axis). The error bars show one standard error obtained with the jack-knife resampling technique.

Besides the low-pass filter characteristics and the lack of jointly high spectral and temporal modulations, the modulation spectra of speech and environmental sounds are remarkably symmetric. These sounds have equal representations of up-sweep and down-sweep ripple sound components (see Figs. 5 and 7). Zebra finch song, on the other hand, exhibits some asymmetry, with slightly more energy in down-sweeps (see Fig. 5). These observations are reflected in the asymmetry index, which measures the relative difference in the two quadrants [Fig. 8(b)]: only the zebra finch song ensemble shows a value of asymmetry that is different from zero.

Finally, we measured the modulation depth of the amplitude envelopes for the four sound ensembles. The modulation depth is traditionally defined as one minus the relative value of the amplitude minimum relative to the maximum: a signal with a modulation depth of 1 is intermittently silent. We used an alternative measure in which we quantified the modulation depth of our signals from their modulation spectra. To estimate the ''size'' of the joint temporal and spectral modulations, we calculated the square root of the ratio of the non-dc power relative to the dc power (see Sec. II.C). Although the actual amplitude modulation observed in the time domain will also depend on the phase of the ripple components of the sound, our measure will be large for signals that are dominated by non-dc ripple components. For example, one would expect isolated animal vocalizations to show larger amplitude modulations than white noise or environmental sounds. Indeed we found that speech had the largest

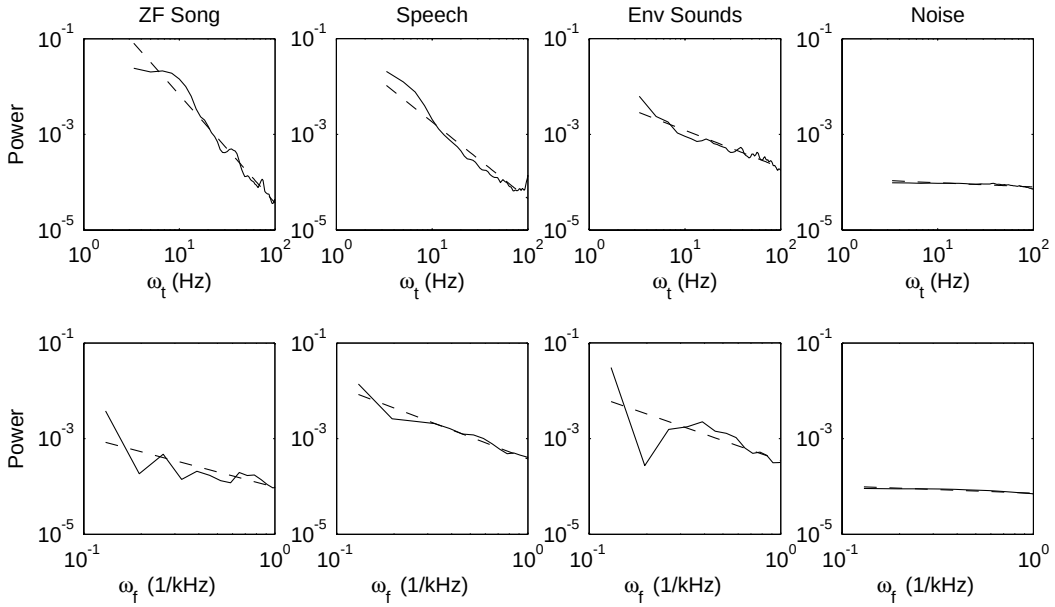


FIG. 10. *Average temporal modulation power and spectral modulation power*. The average temporal modulation power (upper row) and spectral modulation power (bottom row) is plotted as a function of modulation frequency on a log-log plot. The average spectral and average temporal modulations were obtained with a singular value decomposition of the joint modulation spectrum. The data is plotted with a solid line and the power function fit is plotted with a dashed line.

modulation depth (13.8 ± 2.0), followed by Zebra finch song (5.9 ± 0.2), environmental sounds (4.5 ± 0.2) and white-noise (3.92 ± 0.006).

Ultimately, one might want to fit the modulation spectra with a function or a theoretical model. We began this process by fitting the average temporal $[P(\omega_t)]$ and average spectral $[P(\omega_f)]$ components of the modulation spectrum with a power-law function: $P(\omega) \alpha \omega^{-\alpha}$. The average spectral and temporal components were obtained from the first $g(\omega_t)$ and $h(\omega_f)$ functions calculated in the singular value decomposition of the modulation spectrum that was used for the separability analysis (see Sec. II.C). Figure 10 shows these functions (solid line) and the corresponding fits (dashed lines) for all four sound ensembles. The fits were performed for temporal modulations between 3 and 100 Hz and for spectral modulations between 0.1 and 1 kHz^{-1}. This range corresponded to the area of the modulation spectrum where white-noise had a flat distribution, as shown in Fig. 10. Note that, although our effective bandwidth set by the standard deviation parameter of the Gaussian-shaped filters is given by $\mathrm{BW}(\omega_t) = 397.2$ Hz and $\mathrm{BW}(\omega_f) = 4\ \mathrm{kHz}^{-1}$, the boundaries for the areas that showed flat modulation power for white noise are approximately $\frac{1}{4}$ of the total effective bandwidth. Our estimation of the shape of the power distribution must be restricted to that central area or it will be affected by the shape and bandwidth of the filters.

The average temporal components of the modulation spectrum for all three natural sounds were well fitted by the power law (with $R^2 > 0.9$ and $P < 10^{-4}$ in all cases). The vocalizations had steeper slope coefficients with a value for α close to 2 (ZF song: $\alpha = 2.26$, lower 95% = 2.15, upper 95% = 2.36; speech $\alpha = 1.6$, lower 95% = 1.48, upper 95% = 1.72). The slope for environmental sounds was between that of vocalizations and white noise (env sounds: $\alpha = 0.78$, lower 95% = 0.72, upper 95% = 0.84). The approximate $1/\omega_t^2$ relationship is reminiscent of the $1/f^2$ relationship found for spatial frequencies in natural images and its significance is discussed below.

The average spectral components of the modulation spectrum could also be fitted reasonably well with the power law function although additional structure can clearly be observed in the zebra finch song (ZF) and the environmental sounds ($R^2 = 0.95$ $P < 10^{-4}$ for speech; $R^2 = 0.57$ $P < 10^{-3}$ for ZF song, $R^2 = 0.54$ $P = 0.001$ for env sounds). The slope of the power law was shallower than for the temporal component closer to 1 for zebra finch song (ZF song $\alpha = 1$, lower 95% = 0.5 upper 95% = 1.6) and close to 1.5 for speech and environmental sounds ($\alpha = 1.52$, lower 95% = 1.32 upper 95% = 1.72; env sounds $\alpha = 1.4$, lower 95% = 0.6 upper 95% = 2.1).

D. Synthetic sounds with matched modulation spectrum

A final goal of our analysis was to demonstrate how one could synthesize sounds that had similar modulation spectra as arbitrary sound ensembles but different phases in their ripple components. Using the methodology described in Sec. II.D, we generated synthetic zebra finch song, which we called song ripples, and synthetic speech, which we called speech ripples. The top row of Fig. 11 shows the modulation spectrum of zebra finch song and speech and the bottom row shows the modulation spectrum obtained from 400 s of synthetic song ripples and speech ripples. The contour lines surround the area that encloses 50%, 80% and 90% power and the gray scale showing the power is logarithmic. By visual inspection, one can see that the match is relatively good. In particular, the areas of high power are practically identical. The contours that enclose 80% and 90% of the total power are different in the synthetic and natural ensembles but one

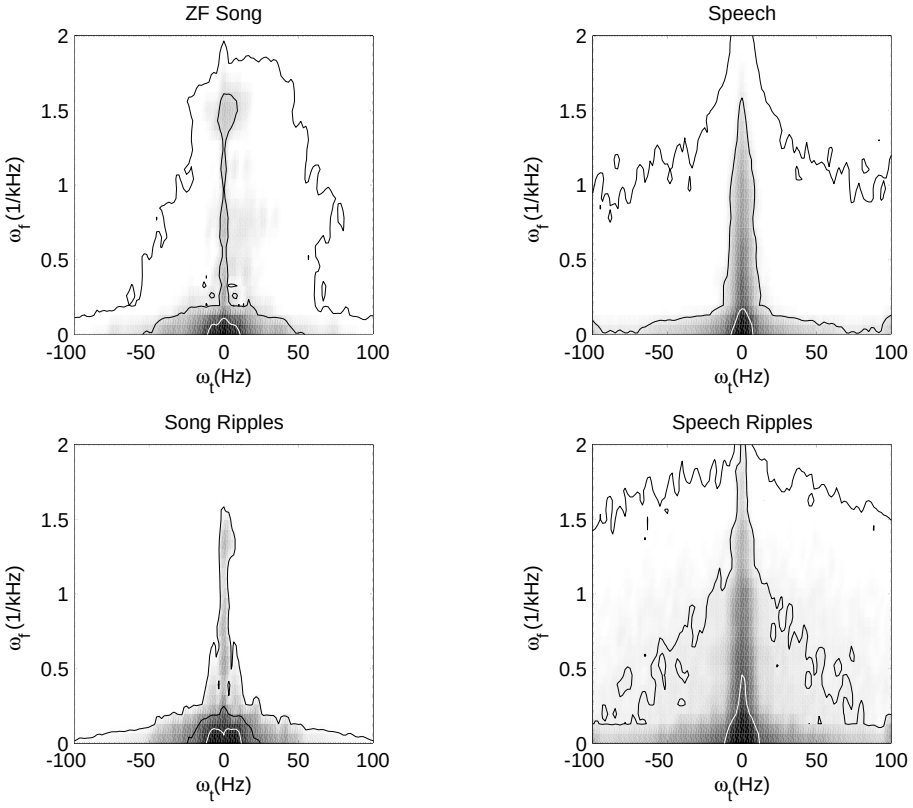


FIG. 11. *Modulation spectrum of natural sounds and their synthetic models*. The top panels show the modulation spectra for zebra finch song and speech. The bottom panels show the corresponding modulation spectra of synthetic sounds that we generated by sampling the spectra of the original sound. The contour lines surround the areas that enclose 50%, 80% and 90% of the total power. The 50% contour lines are drawn in white.

should realize that they correspond to contours drawn at 0.2% and 0.03% of max power for speech and at 0.3% and 0.04% of max power for zebra finch song. These areas of modulation space are therefore infrequently sampled in our synthesis. To better evaluate the quality of the fit, we calculated the cross-correlation coefficient between the modulation spectra of natural sound and that of the synthetic sound: for zebra finch song and song ripples it is 0.9649 and for human speech and speech ripple it is 0.9545.

Spectrograms of exemplars of song ripples and speech ripples are shown in Fig. 12. These sounds have a distinct quality that can be described as zebra-finch-song like and

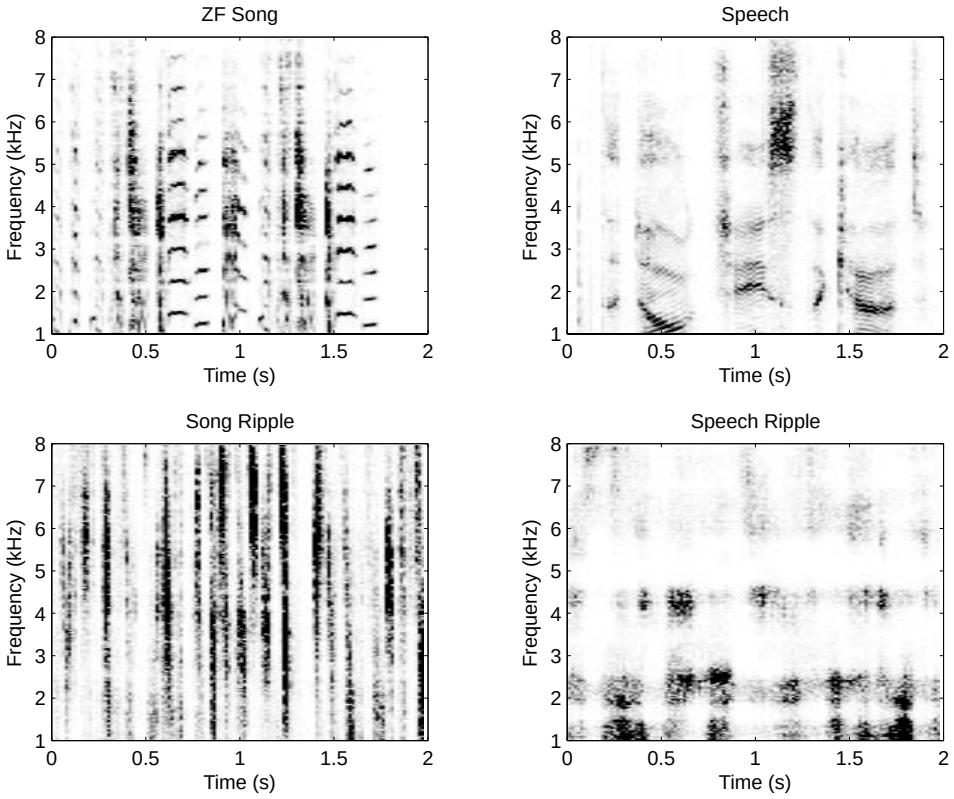


FIG. 12. *Representative spectrograms of a natural zebra finch song and a speech sample (top row) and a synthetic song ripple and speech ripple (bottom row).*

speechlike. Similarly, one can observe, in the spectrograms, similar temporal and spectral modulations in the natural and synthetic ensemble. On the other hand, it is clear that the phase of the ripple components plays an important role in the natural sounds. For example, in the temporal domain, the presence of complete silence is more common in the natural vocalizations than in the ripple sounds. The phase of the ripple sound components also plays a crucial role in the spectral domain, both in generating natural harmonic stacks (all cosine components) and in determining the exact frequency location of the formants in speech. For these reasons and also because the correlations are calculated in 300-ms windows (and 300-ms bits of speech sound parsed together randomly would be unintelligible), the speech ripple sounds are completely unintelligible.

IV. DISCUSSION AND THEORETICAL PREDICTIONS

We have shown that the statistical redundancies that are observed in natural sounds can be, in part, described by the lower-order joint temporal and spectral statistics of the envelopes of the sounds obtained from a time-frequency decomposition of the sound. The distribution of the amplitude of the envelopes of natural sounds has a strong exponential component, which distinguishes it from that of white noise. The modulation spectrum, given by the 2-D Fourier transform of the auto-correlation matrix of the amplitude envelopes, shows that the spectro-temporal envelope modulations in natural sounds are concentrated in the low frequencies, that the average temporal and spectral modulation power can be fitted with a power law and that vocalizations have most of their power in the higher spectral modulation frequencies concentrated at low temporal modulation frequencies.

A. Probability distribution and second order statistics of amplitude envelopes

Our results complement and support previous work that analyzed the statistics of natural sounds. The exponential form of the distribution for amplitude envelopes in speech has been known and exploited in speech processing applications for a long time (e.g., Paez and Glisson, 1972). More recently, Attias and Schreiner (1997) analyzed the amplitude distributions and the temporal modulation power in a larger ensemble of natural sounds which included symphonic music, speech, cat vocalizations and environmental sounds. They also noted the exponential form of the distribution of amplitudes in the natural sounds, which reflects the high probability of finding arbitrarily soft sounds. In their analysis, they found that the distribution in different frequency bands were very similar whereas we found systematic differences in the shape of the probability distribution as a function of the center frequency. However, simple methodological differences between our analyses and theirs could explain the discrepancy: first, we used smaller and more homogenous ensemble of sounds that they did and, second, we used a filter-bank with filters of linearly spaced center frequencies and of fixed bandwidth, whereas they used logarithmically space filters of $\frac{1}{8}$-oct bandwidth.

Attias and Schreiner also calculated the power spectrum of the temporal modulations and fitted their data with a modified power law. There results are similar to those we found by fitting the temporal component of the joint modulation spectrum. In their analysis, they found, as we did, that the power coefficient, α, was between 1 and 2.5. Similar results were also found in the analysis of the distribution of the overall amplitude envelope of speech and music (Voss and Clarke, 1975; Brillinger and Irizarry, 1998). In addition, we found that environmental sounds and animal vocalizations can be segregated into different groups based on these statistics as well as those obtained from the joint time-frequency modulation spectrum (see below). First, the distributions of the log amplitude of the envelopes in those animal vocalizations have low kurtosis, exhibiting a relatively uniform distribution for a 40-dB range of sound intensity, whereas environmental sounds have a log distribution of amplitudes that is approximately normal. Second, animal vocalizations exhibit temporal modulation power relationships that are approximately $1/\omega^2$ whereas environmental sounds are characterized by a significantly flatter power curve. A similar separation of natural sounds into these two broad classes was also suggested by Lewicki (2002) who analyzed the statistically independent components of the sound pressure waveform in these two classes of sounds. He found that vocalizations were best decomposed by Fourier filters and environmental sounds by wavelet filters (Lewicki, 2002).

B. Joint spectro-temporal statistics

In addition, we calculated the joint spectro-temporal modulation spectrum of the natural sounds. It is the natural extension of the purely temporal power-spectra analysis performed on amplitude envelopes or on the overall loudness as described above and the purely spectral power-spectra analysis of the cepstrum performed for speech analysis. We showed how to calculate the joint modulation spectrum and how its calculation was dependent on the choice of the time-frequency representation and on the time-frequency scale of the chosen filters. We chose to perform our analysis using linearly spaced filters with Gaussian shape and fixed bandwidth. The advantages to that representation are that (1) it is symmetric in time and frequency, (2) the limits of the sampled space are well defined and (3) the harmonic sounds, which are very common in animal vocalizations, have well localized occupancy on the spectral axis of modulation spectrum at the value corresponding to the inverse of their fundamental frequency. Since harmonic sounds are critical in acoustical behaviors (e.g., Lohr and Dooling, 1998) and contribute significantly to the statistical redundancy of the sound, the Fourier filter decomposition is in our opinion superior to the wavelet transformation for the characterization of the statistical properties of natural vocalizations. On the other hand, the wavelet transform has the advantage of spanning multiple time-frequency scales. The wavelet transform could therefore be more efficient to estimate the modulation spectrum of sounds that have a broad range of modulation components, such as environmental sounds. These facts and hypotheses are well supported by the analysis performed on the higher order statistics of the sound pressure waveform of vocalizations versus environmental sounds mentioned above (Lewicki, 2002). Moreover, a form of wavelet decomposition

is performed by the mammalian cochlea and the frequency sampling throughout most of the auditory system is approximately uniform on a logarithmic scale. For this reason, the spectro-temporal receptive fields of mammalian auditory neurons have also, until now, been exclusively estimated with a wavelet decomposition (e.g., Depireux *et al.*, 2001; Escabi and Schreiner, 2002). It would therefore be of great value to repeat the statistical analysis performed here using wavelet transforms or other biologically inspired time-frequency representations as those obtained from models of the auditory system (Chi *et al.*, 1999). Note, however, that, because of the physical limits on the frequency range of sounds (and of hearing), a wavelet transform will still result in a modulation spectrum that is bounded to a particular region of spectral and temporal modulations. Also, within the sampled region different frequencies would be analyzed at different scales. Therefore, the results obtained in such analyses would have to be carefully compared to those obtained for white noise and for colored noise, with identical overall frequency power as the ensemble of interest.

Our data suggest that the modulation spectrum for vocalizations can also be distinguished from that of environmental sounds. The vocalizations studied here show significant power in spectral modulations in a narrow band of temporal modulation frequencies (<5 Hz). This power corresponds to the voiced sections of speech and to the harmonic sounds found in zebra finch song. In both vocalizations, there is a scarcity of sound components with both high spectral and high temporal modulations, giving the modulation spectra for vocalizations a characteristic star shape. Environmental sounds are principally characterized by their low-passed quality both spectrally and temporally with a power law function describing both the average temporal and spectral modulation power. The power coefficient describing temporal modulations is smaller (flatter curve) and the power coefficient describing spectral modulations is greater (steeper curve) for environmental sounds than that for vocalizations. We also found that the zebra finch vocalizations are asymmetric with more energy in the down-sweeps than the up-sweeps. We found similar results for other animal vocalizations (bengalese finch song and bat calls; data not shown) and expect this asymmetry to be a common feature of the vocalizations of some animal species. It was strikingly absent from human speech, however.

C. Implications for audio processing

The statistical structure of the spectro-temporal envelopes of natural sounds could have direct implications for various forms of sound processing such as sound compression algorithms for storing music on digital media or speech pre-processing for auditory prosthetics. For example, most current sound compression algorithms (such as MP-3) use standard entropy compression methods on a time-frequency representation of the sound obtained with a filter bank (Painter and Spanias, 2000). The fact that these methods can obtain relatively high compression factors demonstrates that the redundancy in sounds like human music is well captured in the amplitude envelopes of the sound. Since the entropy compression is performed for short segments of time (and therefore optimized for the statistical redundancy at that point in time), it is not clear how the compression could be improved by prior knowledge of the average modulation spectra. On the other hand, information from the modulation spectra could be used to choose appropriate time-frequency decompositions for different classes of sound. Similarly, the knowledge of the differences in modulation spectra for speech versus environmental sounds could be of use for designing preprocessing modules in auditory prosthetics or hearing aids that maximize signal-to-noise ratios for particular signals and noises (see also Chi *et al.*, 1999).

D. Implications for neural coding

Our principal interest is to generate a theoretical framework with which to study the processing of complex sounds in the auditory system of animals. Following the school of thought started by Barlow (1961), we argue that the auditory system has evolved to process behaviorally relevant sounds. For that reason, we expect that the neural representations and computations in the auditory system will be affected by the statistics of behaviorally relevant sounds. In particular, we postulate that the statistics of the spectro-temporal amplitude envelopes of the sound are important for auditory brain areas that are responsible for sound identity. Our results generate various testable hypotheses, which are in some cases, at least qualitatively, supported by psychoacoustical or physiological data.

1. Amplitude coding

The nature of the distributions of the amplitude envelopes leads to a first set of hypotheses. The relatively flat distribution obtained for the log of the amplitude of the envelopes for vocalizations suggests that, in order to discriminate among natural sounds by their amplitude level, an approximate logarithmic amplitude-response curve should be used, as described by Weber's law. Although the psychoacoustical literature on the subject of sound loudness is complex, it is generally accepted that a power law with a coefficient around 0.6 relates loudness and sound pressure amplitude (Stevens, 1956). This power law is not as compressive as the log function but it will ''flatten out'' the distribution of amplitude. Similarly a compressive nonlinearity is a common property of the auditory system found both at the level of the basilar membrane (Schlauch *et al.*, 1998; Ruggero and Rich, 1991) and in many auditory neurons (Sachs and Abbas, 1974; Palmer and Evans, 1982; Phillips, 1990). In addition, synthetic stimuli that have the natural amplitude distribution were shown to increase the coding efficiency of auditory neurons in the cat inferior colliculus (Attias and Schreiner, 1998) and in auditory neurons of the grasshopper (Machens *et al.*, 2001).

2. Spectro-temporal modulation coding

The characteristic modulation spectrum of natural sounds leads to a second set of hypotheses on neural coding in which the spectro-temporal receptive fields (STRFs) of auditory neurons would be matched to the modulation spectrum of behaviorally relevant sounds. This ''matching'' could take various forms. In the simplest form, a matched-filter

hypothesis, we would expect an approximate one-to-one match between the modulation spectrum of behaviorally relevant sounds and the ensemble modulation transfer function of an auditory processing stage. The modulation transfer function (MTF; also called ripple transfer function or RTF) of a neuron is given by the amplitude of the 2-D Fourier transform of its STRF. The MTF is the equivalent of the gain response (or Bode plot) of one-dimensional filters and shows the spectro-temporal modulations in the sound that will drive the cells (see also Chi *et al.*, 1999; Miller *et al.*, 2002). Given that the modulation spectra of all natural sounds is concentrated in the low frequencies, we would expect to find a concentration of best temporal modulation and best spectral modulation tuning in the low frequencies. This concentration has been observed experimentally in the mammalian thalamus and cortex (Miller *et al.*, 2002) although the quantification of the match has not yet been performed. A second line of evidence for neural tuning to the modulation spectrum of natural sounds was described in the songbirds. In the auditory forebrain, neurons that are selective for conspecific song show the greatest responses to synthetic sounds that have similar modulation spectra as the natural vocalizations. Moreover, to obtain responses similar in strength to those of the natural sounds matching the modulation spectrum was more critical than matching the frequency spectrum (Grace *et al.*, 2003).

An additional prediction can be made from the power law relationship between power and temporal modulation frequency that was observed in this analysis (Fig. 10): if one desires equal driving power for each neuron, then the approximate $1/\omega_t^2$ relationship found for zebra finch song and speech requires the bandwidth of the temporal modulation tuning to be fixed in octave units. Recent data on the temporal MTF is also consistent with this hypothesis (Miller *et al.*, 2002). The $1/\omega_t^2$ relationship is reminiscent of the $1/f^2$ power relationship found in natural images as a function of the spatial frequency, f (Field, 1987). For natural images, the $1/f^2$ relationship implies that the second order statistics of natural images are scale invariant. The equivalent statement for natural vocalizations is that the second order temporal statistics of the amplitude envelopes are invariant to time compression or dilation. This mathematical relationship might explain the perceptual effect whereby sped up vocalizations of a particular animal species often sound like the vocalizations of a different animal species.

A more complex form of matching between the modulation spectrum and the MTF of neurons is predicted if one theorizes a spectral-whitening of the input space. In this framework, neurons effectively amplify stimulus regions of low stimulus power to generate a white response output for each neuron or for an ensemble of neurons. The result is that entropy of the output is maximized and the transmission capacity of the neuronal channel is optimized from an information theoretic perspective (Atick, 1992; van Hateren, 1992b). Experimental data in support for this theoretical framework has been found in the visual system of insects (van Hateren, 1992a) and mammals (Dan *et al.*, 1996). With this perspective, one would then expect asymmetric MTF of neurons, with an amplification of the higher temporal modulation frequencies relative to the lower temporal modulations. In these models, the degree of amplification of the higher frequencies depends critically on the signal-to-noise ratio of the input to the neuron. For flat noise power, the theory predicts that at low signal-to-noise the amplification of higher frequencies is reduced or non existent and the neurons are low-pass filters whereas, at higher signal-to-noise ratio, the amplification comes into play and the neurons become band-pass filters (van Hateren, 1992b). Our data on the modulation spectra of natural sounds could be used to extend such noise analysis. If, for example, the role of an auditory area would be to encode speech in a noisy background of environmental sounds, we would expect to find MTF that amplify the areas of high signal-to-noise ratio and filter out the areas of low signal-to-noise: more explicitly, MTF would be tuned to the intermediate modulation frequencies that are present in speech and not so dominant in the environmental sounds. Interestingly, both psychophysically determined thresholds for ripple stimuli (Chi *et al.*, 1999) and ensemble MTF of neurons in the auditory thalamus and cortex of the cat (Miller *et al.*, 2002) exhibit this type of band-pass filtering.

E. Concluding remarks

In summary, the statistics of the envelopes of natural sounds are characteristically different from those of white noise stimuli. This statistical structure allows us to make predictions on theories of auditory processing based on natural sound statistics. Current psychological and physiological data is, in a qualitative fashion, consistent with these predictions. But these theories remain hypothetical until experimental data are generated to prove or disprove them directly. Also, it is almost certain that different coding hypotheses will apply to different stages of the auditory system and that other biological and physical constraints will be of importance. Given the recent work in the analysis of natural sounds (Attias and Schreiner, 1997; Brillinger and Irizarry, 1998; Lewicki, 2002) and in the development of analytical tools to generate complex synthetic sounds and to extract auditory receptive fields from responses to such sounds or to natural sounds (Theunissen and Doupe, 1998; Klein *et al.*, 2000; Theunissen *et al.*, 2001; Escabi and Schreiner, 2002), we are now in a position where we can directly test this theoretical framework and further advance our understanding of the computations occurring in the auditory system for sound identification.

ACKNOWLEDGMENTS

The authors would like to thank Sarah Woolley, Noopur Amin, and Lee Miller for critical comments on a previous version of the manuscript. The paper was greatly improved by following the suggestions and addressing the criticisms of two anonymous reviewers. The work was funded by NIMH Grant Nos. MH-58189 and MH-66990 to FET.

Atick, J. (**1992**). ''Could information theory provide an ecological theory of sensory processing?'' Network **3**, 213–251.

Attias, H., and Schreiner, C. E. (**1997**). ''Temporal low-order statistics of natural sounds,'' Adv. Neural Info. Process. Syst. **9**, 27–33.

Attias, H., and Schreiner, C. E. (**1998**). ''Coding of naturalistic stimuli by auditory midbrain neurons,'' in *Advances in Neural Information Processing Systems* (MIT, Cambridge, MA).
Attneave, F. (**1954**). ''Some informational aspects of visual perception,'' Psychol. Rev. **61**, 183–193.
Barlow, H. B. (**1961**). ''Possible principles underlying the transformation of sensory messages,'' in *Sensory Communication*, edited by W. A. Rosenbluth (MIT, Cambridge, MA), pp. 217–234.
Brillinger, D. R., and Irizarry, R. A. (**1998**). ''An investigation of the second- and higher-order spectra of music,'' Signal Process. **65**, 161–179.
Calhoun, B., and Schreiner, C. (**1998**). ''Spectral envelope coding in cat primary auditory cortex: linear and non-linear effects of stimulus characteristics,'' Eur. J. Neurosci. **10**, 926–940.
Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. (**1999**). ''Spectro-temporal modulation transfer functions and speech intelligibility,'' J. Acoust. Soc. Am. **106**, 2719–2732.
Cohen, L. (**1995**). *Time-Frequency Analysis* (Prentice Hall, Englewood Cliffs, NJ).
Dan, Y., Atick, J. J., and Reid, R. C. (**1996**). ''Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory,'' J. Neurosci. **16**, 3351–3362.
deCharms, R. C., Blake, D. T., and Merzenich, M. M. (**1998**). ''Optimizing sound features for cortical neurons,'' Science **280**, 1439–1443.
Depireux, D. A., Simon, J. Z., Klein, D. J., and Shamma, S. A. (**2001**). ''Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex,'' J. Neurophysiol. **85**, 1220–1234.
Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). ''Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs,'' J. Acoust. Soc. Am. **102**, 2403–2411.
Drullman, R. (**1995**). ''Temporal envelope and fine structure cues for speech intelligibility,'' J. Acoust. Soc. Am. **97**, 585–592.
Drullman, R., Festen, J. M., and Plomp, R. (**1994**). ''Effect of temporal envelope smearing on speech reception,'' J. Acoust. Soc. Am. **95**, 1053–1064.
Eggermont, J. J. (**2002**). ''Temporal modulation transfer functions in cat primary auditory cortex: separating stimulus effects from neural mechanisms,'' J. Neurophysiol. **87**, 305–321.
Eggermont, J. J., Aertsen, A. M., and Johannesma, P. I. (**1983**). ''Prediction of the responses of auditory neurons in the midbrain of the grass frog based on the spectro-temporal receptive field,'' Hear. Res. **10**, 191–202.
Escabi, M. A., and Schreiner, C. E. (**2002**). ''Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain,'' J. Neurosci. **22**, 4114–4131.
Field, D. J. (**1987**). ''Relations between the statistics of natural images and the response properties of cortical cells,'' J. Opt. Soc. Am. A **4**, 2379–2394.
Flanagan, J. L. (**1980**). ''Parametric coding of speech spectra,'' J. Acoust. Soc. Am. **68**, 412–419.
Grace, J. A., Amin, N., Singh, N. C., and Theunissen, F. E. (**2003**). ''Selectivity for conspecific song in the zebra finch auditory forebrain,'' J. Neurophysiol. **89**, 472–487.
Green, D. (**1986**). ''Frequency and the detection of spectral shape change,'' in *Auditory Frequency Selectivity*, edited by B. C. Moore and R. Patterson (Plenum, Cambridge), pp. 351–359.
Griffin, D., and Lim, J. (**1984**). ''Signal estimation from modified short-time Fourier transform,'' IEEE Trans. Acoust., Speech, Signal Process. **32**, 236–242.
Klein, D. J., Depireux, D. A., Simon, J. Z., and Shamma, S. A. (**2000**). ''Robust spectro-temporal reverse correlation for the auditory system: Optimizing stimulus design,'' J. Comput. Neurosci. **9**, 85–111.
Lewicki, M. S. (**2002**). ''Efficient coding of natural sounds,'' Nat. Neurosci. **5**, 356–363.
Lohr, B., and Dooling, R. J. (**1998**). ''Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*),'' J. Comp. Psychol. **112**, 36–47.
Machens, C. K., Stemmler, M. B., Prinz, P., Krahe, R., Ronacher, B., and Herz, A. V. (**2001**). ''Representation of acoustic communication signals by insect auditory receptor neurons,'' J. Neurosci. **21**(9), 3215–3227.
Margoliash, D. (**1983**). ''Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow,'' J. Neurosci. **3**, 1039–1057.
Miller, L. M., Escabi, M. A., Read, H. L., and Schreiner, C. E. (**2002**). ''Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex,'' J. Neurophysiol. **87**, 516–527.
Newman, J., and Wollberg, Z. (**1978**). ''Multiple coding of species-specific vocalizations in the auditory cortex of squirrel monkeys,'' Brain Res. **54**, 287–304.
Paez, M. D., and Glisson, T. H. (**1972**). ''Minimum Mean-Squared-Error Quantization in speech PCM and DPCM Systems,'' IEEE Trans. Commun. **COM-20**(2), 225–230.
Painter, T., and Spanias, A. (**2000**). ''Perceptual Coding of Digital Audio,'' Proc. IEEE **88**, 451–513.
Palmer, A. R., and Evans, E. F. (**1982**). ''Intensity coding in the auditory periphery of the cat: responses of cochlear nerve and cochlear nucleus neurons to signals in the presence of bandstop masking noise,'' Hear. Res. **7**, 305–323.
Phillips, D. P. (**1990**). ''Neural representation of sound amplitude in the auditory cortex: effects of noise masking,'' Behav. Brain Res. **37**, 197–214.
Phillips, D. P., and Hall, S. E. (**1987**). ''Responses of single neurons in cat auditory cortex to time-varying stimuli: linear amplitude modulations,'' Exp. Brain Res. **67**, 479–492.
Popper, A. N., and Fay, R. R. (**1992**). *The Mammalian Auditory Pathway: Neurophysiology*. (Springer-Verlag, New York).
Rieke, F., Bodnar, D. A., and Bialek, W. (**1995**). ''Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents,'' Proc. R. Soc. London, Ser. B **262**, 259–265.
Ruggero, M. A., and Rich, N. C. (**1991**). ''Application of a commercially-manufactured Doppler-shift laser velocimeter to the measurement of basilar-membrane vibration,'' Hear. Res. **51**, 215–230.
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.
Schlauch, R. S., DiGiovanni, J. J., and Ries, D. T. (**1998**). ''Basilar membrane nonlinearity and loudness,'' J. Acoust. Soc. Am. **103**, 2010–2020.
Schreiner, C. E., and Calhoun, B. M. (**1994**). ''Spectral envelope coding in cat primary auditory cortex: properties of ripple transfer functions,'' Aud. Neurosci. **1**, 39–61.
Sen, K., Theunissen, F. E., and Doupe, A. J. (**2001**). ''Feature analysis of natural sounds in the songbird auditory forebrain,'' J. Neurophysiol. **86**, 1445–1458.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.
Simoncelli, E. P., and Olshausen, B. A. (**2001**). ''Natural image statistics and neural representation,'' Annu. Rev. Neurosci. **24**, 1193–1216.
Slaney M. (**1994**). ''An introduction to auditory model inversion,'' Interval Technical Report IRC1994-014.
Stevens, S. S. (**1956**). ''The direct estimation of sensory magnitudes: loudness,'' Am. J. Psychol. **69**, 1–25.
Suga, N., O'Neill, W. E., and Manabe, T. (**1978**). ''Cortical neurons sensitive to combinations of information-bearing elements of biosonar signals in the moustache bat,'' Science **200**, 778–781.
Theunissen, F. E., and Doupe, A. J. (**1998**). ''Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVc of male zebra finches,'' J. Neurosci. **18**, 3786–3802.
Theunissen, F. E., Sen, K., and Doupe, A. J. (**2000**). ''Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds,'' J. Neurosci. **20**, 2315–2331.
Theunissen, F. E., David, S. V., Singh, N. C., Hsu, A., Vinje, W., and Gallant, J. L. (**2001**). ''Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli,'' Network Comput. Neural Syst. **12**, 1–28.
Tyler, R. S., Preece, J. P. and Tye-Murray, K. (**1990**). ''Iowa Audiovisual Speech Perception Tests,'' Department of Otolaryngology. The University of Iowa, Iowa City, IA 52242.
van Hateren, J. H. (**1992a**). ''Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation,'' J. Comp. Physiol. [A] **171**, 157–170.
van Hateren, J. H. (**1992b**). ''A theory of maximizing sensory information,'' Biol. Cybern. **68**, 23–29.
Viemeister, N. F. (**1979**). ''Temporal modulation transfer functions based upon modulation thresholds,'' J. Acoust. Soc. Am. **66**, 1364–1380.
Voss, R. F., and Clarke, J. (**1975**). ''1/*f* noise in music and speech,'' Nature (London) **258**, 317–318.

Acoustic variability and distinguishability among mouse ultrasound vocalizations[a)]

Robert C. Liu[b)] and Kenneth D. Miller
Sloan-Swartz Center for Theoretical Neuroscience, Department of Physiology, Box 0444, University of California at San Francisco, 513 Parnassus Avenue, San Francisco, California 94143

Michael M. Merzenich and Christoph E. Schreiner
Department of Otolaryngology, Box 0732, University of California at San Francisco, 513 Parnassus Avenue, San Francisco, California 94143

(Received 21 February 2003; revised 28 August 2003; accepted 15 September 2003)

Auditory neurobiology has benefited significantly from ethological approaches using acoustic communication signals. Developing an ethological model in a genetically manipulable system such as the mouse would enhance the ability to investigate the processing, learning, and recognition of sounds. Characterizing the basic acoustic structure of mouse vocalizations would help lay a foundation for such a future study. Towards this goal, ultrasound vocalizations emitted by isolated mouse pups and pairs of adult males and females have been digitally recorded and examined. Previous work suggests that these calls may have communicative significance. An analysis of the natural variability in their spectral content, median frequency, duration, and repetition period reveals acoustic structure that could be used for recognizing the calls. Other parameters, like the rate of frequency modulation, may also be informative, but have not been examined. Pup isolation calls develop systematically between postnatal day 5 and 12 towards a more stereotyped vocalization—contracting from a wide range of values into narrower clusters of frequency and duration, and shifting from longer to shorter repetition periods. Most significantly, pup isolation and adult encounter calls fall into two distinct spectral and temporal categories, making it possible for a receiver to acoustically distinguish between them, and to potentially categorically perceive them along those dimensions. [DOI: 10.1121/1.1623787]

PACS numbers: 43.80.Ka, 43.66.Gf, 43.64.Tk [WA] Pages: 3412–3422

I. INTRODUCTION

Acoustic communication has played an important role in expanding our understanding of audition. Studies using primate, bat, frog, and bird vocalizations have all helped clarify mechanisms of auditory processing, learning, and sound recognition (Hauser, 1996). The extension of this neuroethological strategy to the mouse would enable new research opportunities. Mutant mice and transgenic techniques could be applied to elucidate neural circuitry in the auditory system (Willott, 2001). Comparisons of vocalizations between different mouse strains could reveal genetic contributions to behavior (Bell *et al.*, 1972; Hahn *et al.*, 1987; Roubertoux *et al.*, 1996).

Implementing this approach requires characterizing the acoustic variability of specific communication calls. Our work focuses on certain vocalizations heard by an adult female mouse—adult encounter and pup isolation calls in the ultrasound frequency range (>25 kHz). Adult encounter calls are detected when females meet adult males (Nyby, 2001). Experiments that anesthetize or devocalize one of the pair suggest that they are mainly produced by the male (Whitney *et al.*, 1973; Nunez *et al.*, 1985; Warburton *et al.*, 1989). It has been hypothesized that these vocalizations attract the female (Whitney *et al.*, 1973; Pomerantz *et al.*, 1983), although definitive evidence for this communicative role has been lacking. Pup calls are emitted by isolated mouse pups between postnatal day 3 (P3) and ~P13 (Smith and Sales, 1980; Haack *et al.*, 1983). The vocalizations clearly elicit a search and retrieval of pups by the mother (Sewell, 1970; Smith, 1976; Haack *et al.*, 1983). Lactating females can categorically perceive synthetic ultrasounds as pup calls along the bandwidth and duration dimensions (Ehret and Haack, 1981, 1982; Ehret, 1992). These facts suggest that the isolation calls serve an important communicative function.

Describing the basic acoustic properties of these ultrasounds, with attention to dimensions that distinguish them as adult- or puplike, is the main goal of this paper. The distributions of four of their acoustic parameters are explored: spectral content, typical frequency, duration, and repetition period. High-bandwidth recording equipment, digital streaming technology, spectral-subtraction noise reduction (Boll, 1979), and automated detection are combined to build a large library of calls produced by the CBA/CaJ[1] inbred mouse. Our statistical analysis of call parameters extends prior research into vocalizations by adults [see recent review in Nyby (2001)] and pups [see early review in Sales and Smith (1978), and some more recent reports (Elwood and Keeling, 1982; Roubertoux *et al.*, 1996; Hahn *et al.*, 1998; Branchi *et al.*, 1998; Motomura *et al.*, 2002)]. This previous work was often based on analog or narrow-bandwidth recordings

[a)]Portions of this work were presented in "The variability and distinguishability of mouse ultrasound vocalizations," poster 588.16, Society for Neuroscience 32nd Annual Meeting, Orlando, FL, November 2002.
[b)]Electronic mail: liu@phy.ucsf.edu

(e.g., bat detector), manual detection and counting, and/or much smaller sample sizes.

The results presented here provide new details relevant to understanding the production and perception of ultrasound vocalizations in the mouse. Specifically, pup isolation vocalizations develop in their spectral content and rates of repetition, becoming more stereotyped with age. These calls divide into one primary and one satellite cluster in the dimensions of typical frequency and duration. Adult encounter calls fall into their own cluster, acoustically distinct from pup isolation calls. Repetition periods for pup and adult calls also separate. Therefore, these acoustic features are potentially informative cues that the receiver, the adult female mouse, can use to recognize and possibly categorize ultrasounds.

II. METHODS

A. Recordings

Vocalization recordings are carried out in an anechoic box ($W30\times D20\times H18$ in.^{3}), where the ambient temperature is approximately 21.5 ± 1 °C. A $\frac{1}{4}$-in. B&K (Bruel and Kjaer, Naerum, Denmark) microphone (either model 4135 or 4939), connected to a (B&K 2669) preamplifier and (B&K 2690) conditioning amplifier (gain 3.16 or 10 V/Pa; filter 20 Hz to 100 kHz), picks up and transmits the vocalizations to a TDT (Tucker Davis Technologies, Gainesville, FL) AD2 analog-to-digital converter, after first filtering through a TDT FT6 100 kHz antialiasing filter. Custom-written MATLAB (Mathworks, Natick, MA) code interfaces a PC computer with the TDT equipment to stream the sounds to a SCSI harddisk at 250 kS/s. The background noise sampled in the sound box at the amplifier gains used for recordings is consistent between days and across different gain levels.

A low-light-sensitive video camera monitors the animals throughout the duration of each recording session, which occurs during the dark phase of the animals' light cycle. Ultrasounds can be heard over headphones with an Ultrasound Advice (London, United Kingdom) Mini-3 bat detector tuned to 65 kHz (for pups) or 80 kHz (for adults).

For pup isolation call recordings, home cages are first placed next to the sound box. An experimenter wearing latex gloves transfers individual pups from the nest to a raised platform (7.5×10.5-in.^{2}) inside the sound box. Acquisition through a microphone aimed towards the platform's center starts 20–25 s after releasing the pup, and lasts 5 min. The same pup is not recorded more than once in a day. Three to four litters of pup have been studied from P5 to P12 (specifically, $N=18$, 16, 17, 17, 17, 20, 16, 20 pups, respectively, with age). Over this age range, pups undergo significant growth of hair and body mass. Such factors can affect a pup's ability to adjust to the ambient temperature change between the nest and platform, and since temperature can affect vocalization rates in mice (Okon, 1970), this may be a mechanism for driving developmental changes in the acoustic parameters of mouse calls as well. Nevertheless, these developing vocalizations reflect what a receiver of these calls would hear from a pup exposed to this consistent natural environment. Figures 1(a)–(c) show examples of these recorded pup calls.

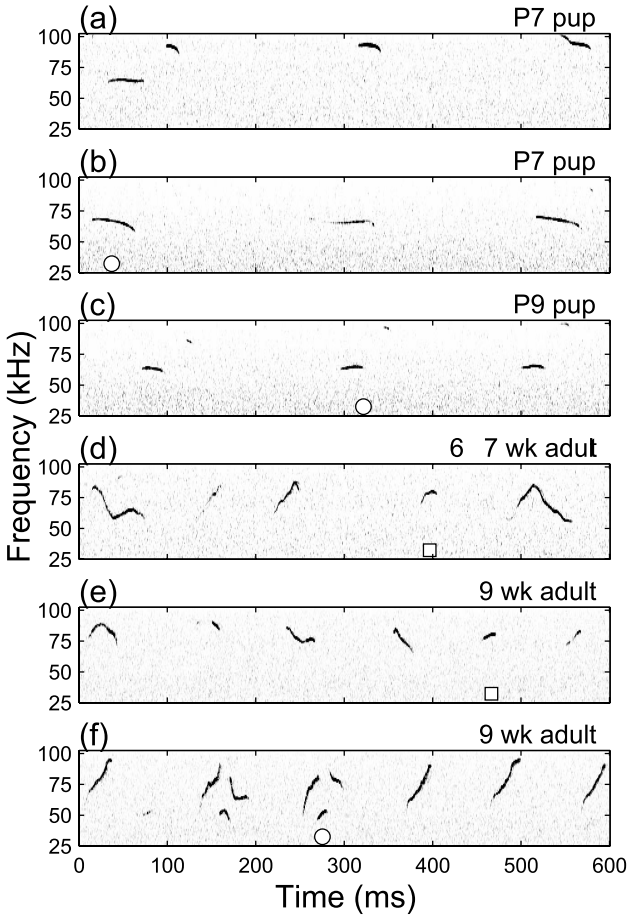


FIG. 1. Vocalization spectrograms. (a)–(c) Pup isolation call bouts. (d)–(e) Adult encounter call bouts. The bouts in (a) and (b) are from pups in different litters; (b) and (c), the same litter. The bouts in (d) and (e) are from different adult pairs; (e) and (f), the same pair. Calls tagged with circles or squares are drawn from the highlighted areas in the frequency-duration contour plots shown in Fig. 9. Consistent differences in the spectral and temporal patterns of pup isolation and adult encounter calls are apparent.

For adult encounter call recordings, the home cage of a single male is placed into the sound box, and a microphone is pointed towards its center. An experimenter wearing latex gloves then adds a female littermate (one that has been separated from the male since their weaning) to the cage, and acquisition starts approximately 20–25 s after release. The duration of a recording varies from 20 to 60 min. Six pairs of six- to seven-week-old mice, and one pair of nine-week-old mice have been used for this study. Samples of recorded adult calls are displayed in Figs. 1(d)–(f).

B. Vocalization processing

Building a library of calls requires first detecting them in a sound file. Given a sufficiently high signal-to-noise ratio, this can be done by setting a threshold for the acoustic signal in the time domain, and marking the locations of threshold crossings. The background in our system is approximately −11 dB SPL/Hz, which is close to the thermal noise of the $\frac{1}{4}$ -in. microphone cartridge. With an original signal bandwidth of nearly 100 kHz, the integrated noise level is around 39 dB SPL, making time domain detection of vocalizations with intensities on that order difficult. As an example, Fig.

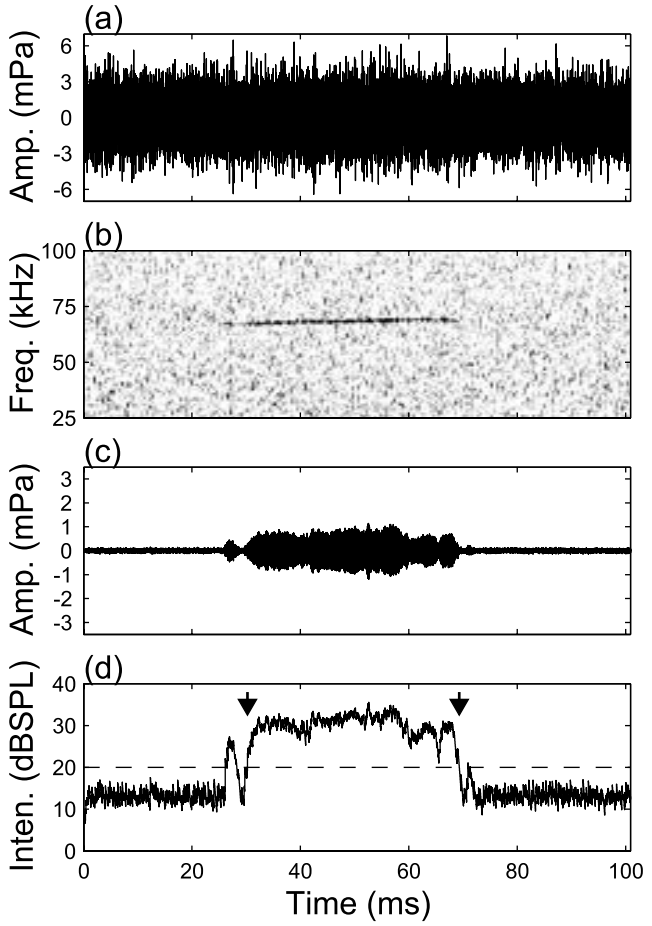


FIG. 2. Spectral subtraction example from a P6 pup. (a) A segment of a sound file. Noise dominates the signal in the time domain. (b) Spectrogram of the segment in (a). A clear ultrasound signal appears around 68 kHz. (c) Time domain signal after denoising. The spectral subtraction algorithm allows the ultrasound call to be revealed in the time domain. (d) Envelope of the denoised signal in (c). The start and end points of the call are found by finding where the envelope of the sound crosses the threshold, which is set at 6× the standard deviation of the residual noise.

2(a) shows the waveform of a portion of a sound file after digital high-pass filtering (25-kHz corner frequency, eighth-order Butterworth filter); no ultrasound calls are distinguishable from the fluctuations. However, a spectrogram representation of the same portion clearly shows a coherent signal near 68 kHz lasting about 43 ms [Fig. 2(b)]. Automatically detecting this or other similar calls in the spectrogram domain is an algorithmically difficult task. Instead, MATLAB code (utilizing the Signal Processing and Osprey spectrogram toolboxes, both obtainable from the Mathworks website, http://www.mathworks.com) implementing a denoising technique called ''spectral subtraction'' reduces the noise so that quiet calls can be detected in the time domain (Boll, 1979).

Spectral subtraction denoises in the frequency domain by ''subtracting out'' an estimate of the background noise. The basic algorithm is as follows: (1) the fast Fourier transforms (FFT) of a chunk of the signal file and the noise file are obtained, yielding amplitudes and phases at each frequency; (2) the noise spectral amplitudes are subtracted from the signal spectral amplitudes; (3) the resulting spectrum is half-wave rectified to eliminate negative spectral amplitudes;

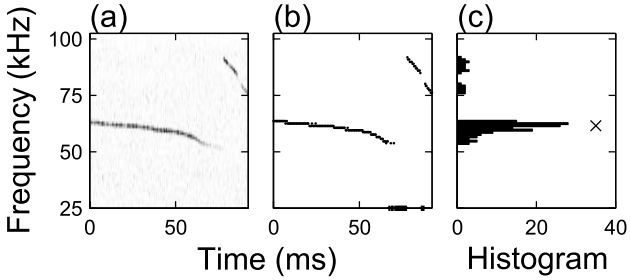


FIG. 3. Extraction of spectral information from a P8 pup's isolation call. (a) Spectrogram of call. (b) Local frequency trajectory of the call in (a). The peak frequency in the spectrum of each 1-ms bin is found as a function of time. Hash marks along the time axis indicate bins where the call amplitude drops into the noise. (c) Histogram distribution of local frequencies in the call. This permits the characterization of the various parameters, including the median frequency (cross) in the call.

(4) a small amplitude, random spectrum is added back to generate a minimum noise floor, reducing ''musical note'' artifacts (isolated, random positive fluctuations in the spectrum); and (5) an inverse FFT is then applied, using the signal's original phases, to obtain the final denoised signal. Figure 2(c) depicts the output of this process for the chunk shown in Fig. 2(a); the call is now clearly revealed in the time domain, simplifying its automatic detection in the time domain [Fig. 2(d)]. Spectral subtraction works well only when the noise spectrum is stationary, as it is in our case. An alternative method would simply be to narrowly filter the signal over a 5–10-kHz bandwidth. However, that would require *a priori* knowledge of the frequency range of interest, and would fail if a vocalization wanders out of that range. By contrast, spectral subtraction succeeds with few assumptions about the actual signals of interest, and is based solely on the properties of the noise.

After denoising, a peak tracking (with decay) algorithm (Beeman, 1998) extracts the sound file's amplitude envelope. Positive crossings of a threshold (set at 6× the standard deviation of the residual noise) mark the beginnings of potential calls, provided the signal stays above threshold for more than 2 ms. At the end of a call, the envelope must drop below threshold for longer than 40 ms. This allows call sequences that feature a high-frequency component above 81 kHz following a lower frequency component with a variable time delay of less than 40 ms [Fig. 3(a)] to be classified as a single call in the library.

Although mice produce many types of calls (Haack *et al.*, 1983), this paper is restricted to the analysis of ultrasound whistles above 25 kHz. Clicklike sounds (transient, broadband), which are often indistinguishable from artifacts, are rejected by eliminating sounds with less than 4-ms durations (although inclusion of these sounds does not significantly affect our conclusions). Following the above procedure, 51 672 pup isolation calls and 11 210 adult encounter calls have been studied—orders of magnitude larger than previous attempts to characterize mouse vocalizations.

C. Spectral and temporal analysis

The spectral content of a typically single-frequency ultrasound whistle is determined by extracting its time-

dependent local frequency with the following procedure. Spectrograms of (not denoised) vocalizations are computed using overlapping 256-point (1.024 ms) Hanning windows (977-Hz resolution). The local frequency is defined as the peak frequency, between 25 and 100 kHz, at each time point (0.512-ms spacing), provided its peak amplitude is at least 3 dB above the background noise spectral density (estimated between 28 and 32 kHz from the same spectrum). If these criteria are not met, no local frequency is recorded for that time. In this way, a local frequency contour line, which is independent of the call's amplitude, can be defined as a function of time [Fig. 3(b)]. Alternatively, local (or instantaneous) frequency is often defined in terms of the derivative of the instantaneous phase derived from the Hilbert transform of the acoustic signal. We expect that such a procedure would not produce significantly different results.

Since the spectral content is a full distribution (like a power spectrum), it cannot be conveniently used as a ''label'' for the typical frequency of a call, a role assigned instead to the median frequency in the call's spectral content [cross symbol in Fig. 3(c)]. Two temporal features are also considered in this paper: the call duration, and the time between sequential calls (repetition period). These parameters have been chosen for their analytical simplicity and their relevance to behavior,[2] but other acoustic variables may also be informative. For example, Fig. 1 suggests that large frequency sweeps occur more often in adult compared to pup calls. However, the characterization of this modulation is complicated by temporal gaps and frequency jumps within some calls. Features like this would be better investigated in the context of the entire trajectory of a call in frequency and time, a subject beyond the scope of this paper.

D. Distinguishing vocalizations

Behavioral testing using adult encounter and pup isolation calls must ultimately be conducted to determine whether a female mouse perceptually distinguishes between them. In this paper, signal detection theory (Green and Swets, 1988) is employed to compute the probability that a receiver correctly classifies calls using various acoustic variables, x. According to this theory, a rational observer uses the conditional probability distributions, $p(x|\text{pup})$ and $p(x|\text{adult})$, to compute the likelihood ratio, $p(x|\text{pup})/p(x|\text{adult})$. The decision of whether a particular call originates from a pup or adult depends on the threshold level for acceptable likelihoods. Once this decision criterion is set, the performance can be quantified by the probability of correctly classifying a pup call (hit pup), and the probability of incorrectly classifying an adult call as a pup call (miss adult). These are computed by integrating $p(x|\text{pup})$ and $p(x|\text{adult})$, respectively, over the x values that give likelihoods exceeding the threshold. As decision criteria change, performance changes, mapping out a receiver-operating-characteristic (ROC) curve in the hit versus miss plane. The probability, P_c, of correctly classifying a call regardless of criteria is the area under the ROC curve, and varies from 50% (chance) to 100%.

The parameters subjected to this analysis include the one-, two- and three-dimensional combinations of call median frequency, duration, and repetition period. The distribution for median frequency is limited to 40 to 100 kHz in 977-Hz bins; 0 to 150 ms in 5-ms bins for duration; and 0 to 500 ms in 10-ms bins for period. A bootstrap resampling technique is used to obtain a debiased estimate and 90% confidence intervals on P_c (Efron and Tibshirani, 1993). Pups are compared to adults both at an individual-to-individual level (987 total pairings, 50 bootstrap resamples per pairing), and at the collective level (100 bootstrap resamples), where all pup and adult calls are separately combined.

III. RESULTS

Figures 1(a)–(f) display sample spectrograms showing the frequency and amplitude trajectories of mouse calls as a function of time. Obvious spectral and temporal differences exist between pup isolation and adult encounter vocalizations. This paper analyzes the variability in their spectral content, median frequencies, durations, and repetition periods, focusing on both the age dependence of pup calls, and the comparison between pup and adult calls.

Most of the recorded ultrasound vocalizations contain only a single frequency peak (the local frequency) at any time point. The collection of these local frequencies constitute a call's spectral content. The normalized spectral content distributions for all calls of individual pups at ages P5 and P12 are shown in Figs. 4(a) and (b), with each line corresponding to a different pup. These distributions generally exhibit two peaks—one below and one above 81 kHz—that usually arise from frequency jumps in a call [but not exclusively, as seen by contrasting the first two calls in Fig. 1(a)].

The spectral content below 81 kHz develops systematically with age. At P5, the distributions from individual pups are centered on many different frequencies. This contributes to a wide P5 collective distribution of local frequencies, derived by combining the spectral content from all calls of all pups at that age.[3] Figure 4(c) shows the progression from P5 to P12 of this collective distribution. Beginning around P9, it narrows substantially, as the centers of the individual distributions converge to ~65 kHz (data not shown). The broadness of a spectral content distribution can be assessed by extracting the spectral content width—the frequency difference between the 75th and 25th percentile local frequencies in the distribution. Figure 4(d) plots, as a function of age, the average of the individuals' widths at each age (white bars), as well as the width measured from the collective distribution at the corresponding ages (black bars). In contrast to the period from P5 to P8, the two widths from P9 to P12 are not significantly different from each other (two-tailed, single-sample t-test, $p<0.01$), implying considerable overlap between individual spectral content distributions. In other words, the spectral content from different individuals converge to a more stereotyped distribution.

In addition to the main long whistle below 81 kHz, a second shorter whistle above 81 kHz is often evident for each pup, centered at ~94 kHz. However, because our recording apparatus rolls-off steeply around 100 kHz, the estimate of the spectral content width in this range (see Fig. 4 legend) represents only a lower bound.

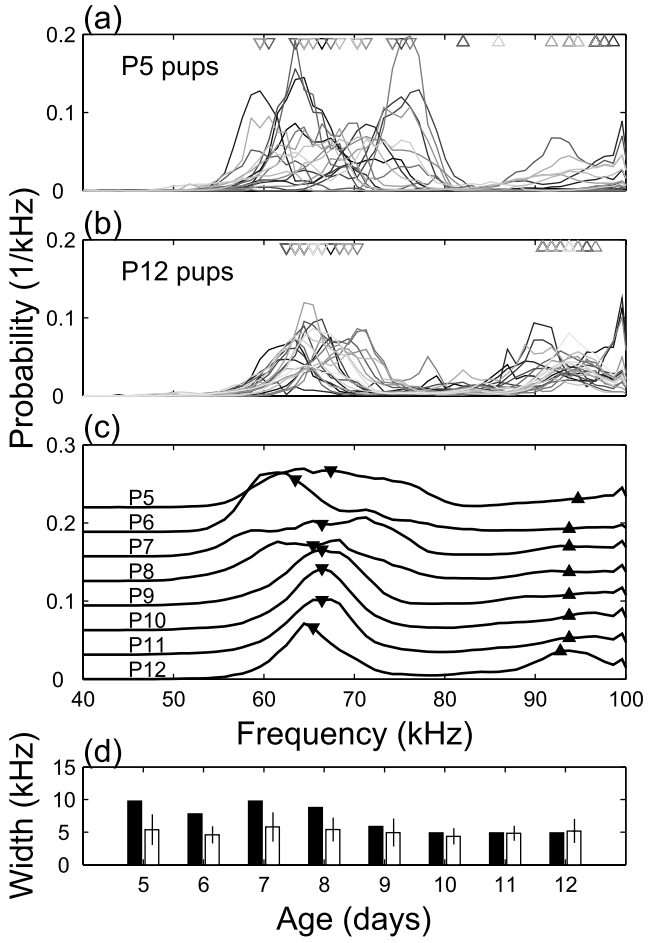


FIG. 4. Local frequency distributions for pup isolation calls. (a) Individual P5 pups. (b) Individual P12 pups. Distributions are calculated with 977-Hz bins. Each downward (upward) triangle indicates an individual pup's median frequency for local frequencies ⩽81 kHz (>81 kHz). At P5, the downward triangles span 17 kHz, but only 8 kHz at P12. Pooling across all ages, the average (mean± standard deviation) of these medians is 66±4 kHz (⩽81 kHz) and 94±3 kHz (>81 kHz). (c) Progression of the collective local frequency distribution (successively offset by 0.031/kHz). Filled triangles mark the medians for the collective distributions, as in (a) and (b). (d) Spectral content widths (see Sec. III) of the collective distributions (black bars), and averages of the widths of individual distributions (white bars) at each age. Error bars are standard deviations. The width (⩽81 kHz) of the collective distribution at P5 is 10 kHz, but contracts to 5 kHz by P12. The average of the individual widths across all ages is 5±2 kHz (⩽81 kHz) and 7±3 kHz (>81 kHz). On average, a pup produces 368±203 calls, or 26319±18809 valid local frequency points per 5 min of recording.

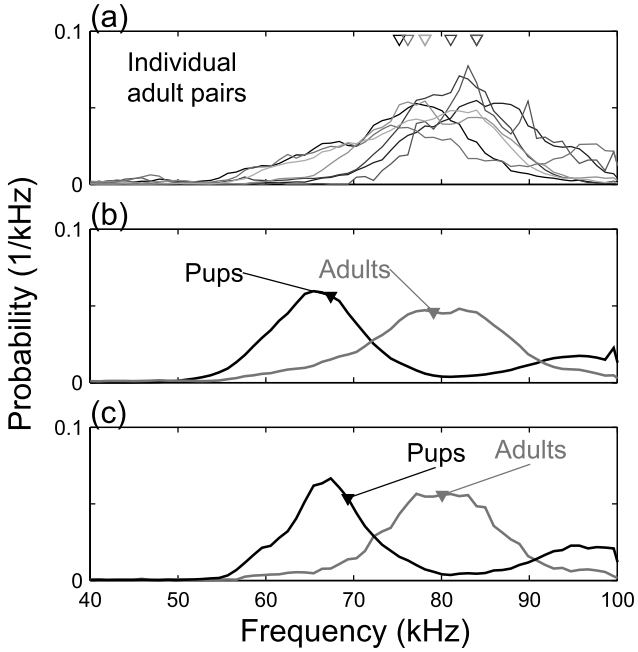


FIG. 5. (a) and (b) Local frequency distributions. (a) Adult encounter calls for seven male/female pairs. Downward triangles indicate each pair's median frequency, and averages 80±4 kHz. The mean of the spectral content widths derived from individual distributions is 11±3 kHz. On average, a pair produces 210±126 calls, or 8679±5862 valid local frequency points per 5 min of recording. (b) Collective local frequency distributions for all pup (black, 3 710 979 frequencies) and all adult (gray, 426 818 frequencies) calls. Downward triangles mark the overall median frequency for pups (67 kHz) and adults (79 kHz). Pup isolation and adult encounter calls clearly separate in spectral content. (c) Collective distributions of call median frequency for all pup isolation (black) and all adult encounter (gray) vocalizations. Downward triangles mark the medians of the call median frequency distributions for pups (69 kHz) and adults (80 kHz). The median frequency distributions mirror the corresponding spectral content distributions.

Contrary to the stereotyped distribution of spectral content from older pups, the vocalizations of adults in the encounter setting show considerable spectral variability. Figure 5(a) displays their normalized distributions. The spectral content widths from individual adult pairs are noticeably larger than those derived from individual pups, probably due to the large proportion of frequency modulated calls in the adult repertoire [Figs. 1(d)–(f)]. No differences are apparent between the nine-week-old pair and the six- to seven-week-old pairs. Combining all adult calls produces a spectral content distribution distinct from that of pup calls, as Fig. 5(b) shows. The local frequencies from all pup isolation calls cluster around the medians of 65 kHz (⩽81 kHz) and 94 kHz (>81 kHz), with both clusters having widths of 7 kHz. A conspicuous dip at ~81 kHz lies between the two peaks. This is precisely the area dominated by adult call frequencies, which exhibit a median of 79 kHz (width of 12 kHz).

In principle, an adult female could use this acoustic distinctness to classify vocalizations by comparing spectral features of the call to the ''template'' distributions for that feature among pup and adult calls.[4] Rather than use a call's entire spectral content, this feature could be simply a typical frequency, such as the call's median local frequency. The distributions of median frequencies among all pup isolation or adult encounter vocalizations are shown in Fig. 5(c). Its similarity to Fig. 5(b) justifies the usefulness of the median frequency as a label for a call's spectral content. The pup distribution clusters around medians of 67 kHz (⩽81 kHz) and 94 kHz (>81 kHz), with frequency widths (difference between 75th and 25th percentile call median frequency) of 7 and 6 kHz, respectively. The distribution for adults is concentrated around a median of 80 kHz, with a width of 9 kHz.

The minimal overlap between the pup and adult distributions suggests that a receiver using only median frequency could do quite well in distinguishing pup calls from adult calls. This can be quantified formally by signal detection theory (Green and Swets, 1988). Figure 6(a) (solid line) plots the receiver-operating-characteristics (ROC) curve for the median frequency distributions. This describes the probability of correctly labeling a pup call as a pup call based on its

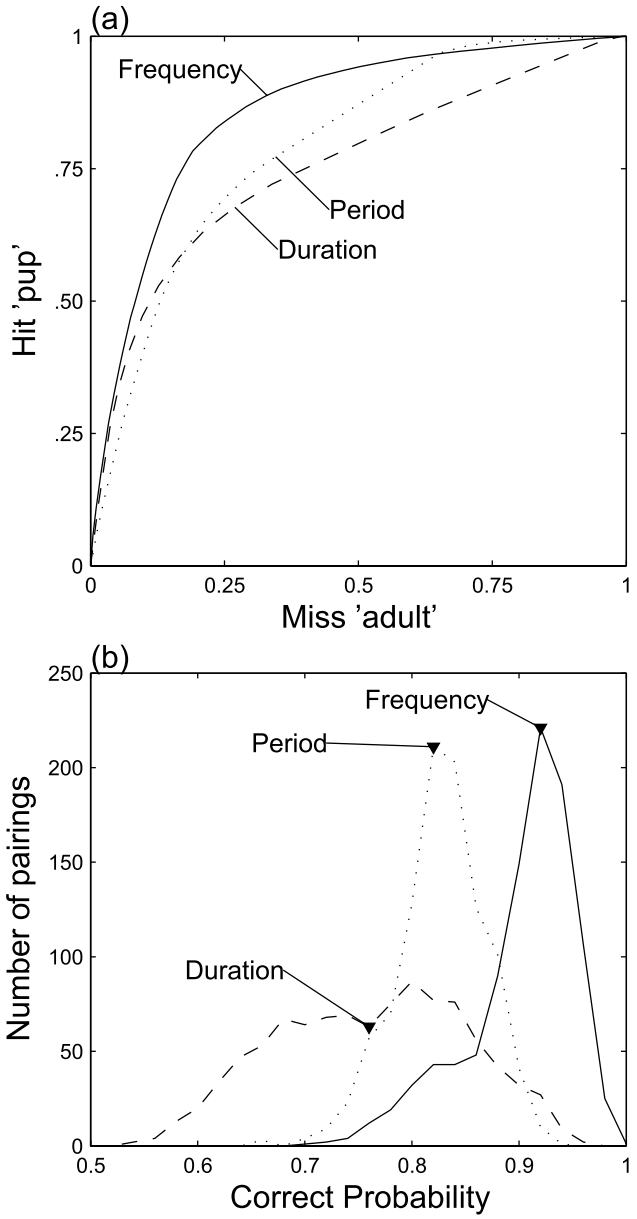


FIG. 6. Classification of calls from pups and adults. (a) ROC curves for call median frequency (solid), duration (dashed), and repetition period (dotted). A point on the curve represents the probability of correctly classifying a pup call (hit pup) and incorrectly classifying an adult call as a pup call (miss adult) under a specific classification criterion used by the receiver (see Sec. II). The area under an ROC curve yields the probability of correctly classifying a pup call, irrespective of criterion. (b) Distributions of the probability of correct classification of an individual pup against an individual adult (987 pairings).

median frequency, as a function of the probability for mislabeling an adult call as a pup call. The top right corner of the plot corresponds to when the receiver is willing to always miss adult calls by assigning all calls to pups; in this case, the receiver's probability for labeling an actual pup call as a pup call is perfect. On the other hand, if the receiver does not allow any adult calls to be mislabeled, then all pup calls will be misassigned (bottom left corner). The curve between these two points is determined by the overlap of the two distributions being distinguished, and represents the tradeoff between correctly labeled and incorrectly labeled calls as the receiver's criterion for comparing the two distributions changes. The better separated the two distributions are, the steeper the initial rise and the flatter the plateau of the ROC curve as the probability for missing an adult call increases.

The area under the ROC curve yields the total probability P_c of correctly labeling a pup call, irrespective of the receiver's criterion. It represents the overall discriminability between the two distributions. For median frequency, this area is 85.6% (the 90% confidence intervals for probabilities always span $<0.7\%$ around the estimate; chance is 50%). Classifying individual pups against adults produces a collection of probabilities for the various pairwise combinations, shown as a histogram in Fig. 6(b). The median probability of correctly classifying an individual pup against an adult [Fig. 6(b), downward triangles] using only this spectral cue is 91.6%, with 50% of the pairwise combinations falling between 88.0% and 93.5%. This is larger than for distinguishing collective distributions of all pups and adults because the distributions of individual animals are usually narrower and overlap less frequently. By either measure though, this simple spectral feature is reasonably informative as an acoustic cue for distinguishing isolated pups from social adults.

We next consider the duration of calls. The duration distributions for individual pups at P5 and P12 are shown in Figs. 7(a) and (b), using different shades of gray. Because of the considerable overlap between individuals of the same age, each pup's distribution resembles the collective distribution for all pups at the same age [Fig. 7(c)]. In particular, as Fig. 7(d) shows, the width (75th-25th percentile duration) of the collective distribution (black bars) is not different from the average of the individuals' widths (white bars) at the $p<0.01$ significance level (two-tailed, single-sample, *t*-test; exception for P7, where the equality hypothesis can be rejected up to the $p<0.001$ level). Moreover, these widths do not vary much between P5 and P12. Nevertheless, the collective distribution gradually shifts its weight towards longer durations. In fact, the distribution peak jumps to ~60–75 ms starting at P8 [Fig. 7(c), upward triangles]. Such a development would be difficult to observe without the full duration distributions since simple statistical measures like the mean or median duration [Fig. 7(c), downward triangles] feature only modest changes with age.

The adult calls are generally much shorter in duration than the pup calls, with similar distributions between individual pairs of adults, as shown in Fig. 8(a). Collected together in Fig. 8(b), the adult call distribution peaks at 10 ms (upward, gray triangle), in contrast to the 65-ms peak for pup calls (upward, black triangle). Although the widths of these two distributions are quite different (26 ms for all adult calls and 42 ms for all pup calls), they overlap significantly at short durations. An ideal receiver basing its call classification on duration would have a 77.1% probability of correctly distinguishing the calls. This is well below the confidence interval for using only a call's median frequency. Moreover, the histogram of individual pairwise comparisons [Fig. 6(b)] is shifted towards chance discrimination (median $P_c=76.7\%$; 50% range from 69.3% to 82.9%).

The combination of spectral and temporal features enhances the contrast between the two call types. In particular,

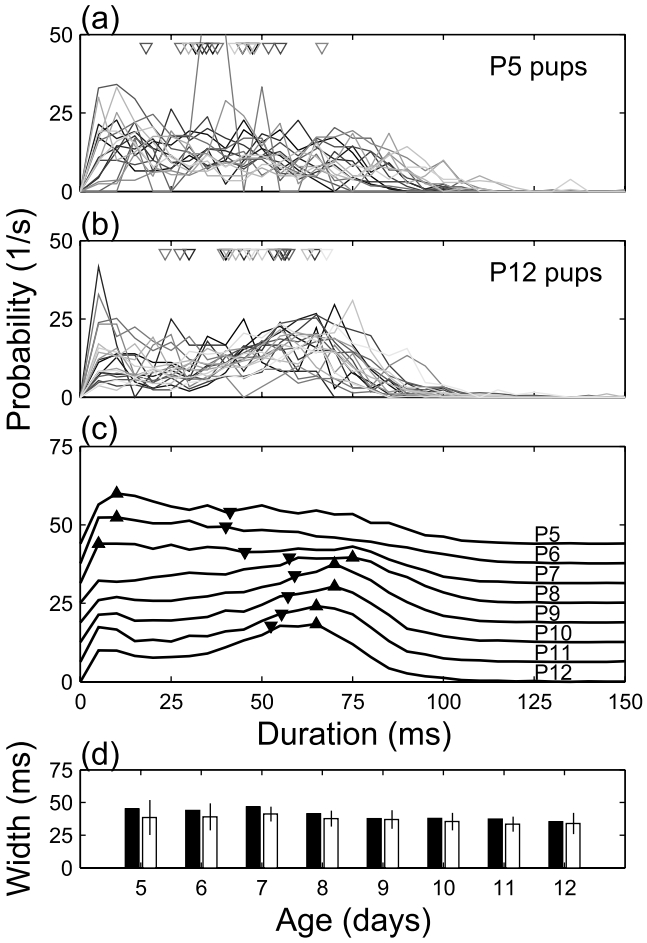


FIG. 7. Duration distributions for pup isolation calls. (a) Individual P5 pups. (b) Individual P12 pups. Distributions are calculated with 5-ms bins. Each downward triangle indicates an individual pup's median duration. At P5, these average 41 ± 11 ms (mean $\pm$ std. dev.); at P12, it is 48 ± 12 ms. Pooling across all ages, the average of the median durations is 48 ± 14 ms. (c) Progression of the collective duration distribution (successively offset by 6.3/s). Downward (upward) triangles mark the medians (peaks) of the collective distributions. (d) Widths (see Sec. III) of the collective distributions (black bars), and averages of the widths of individual distributions (white bars) at each age. Error bars are standard deviations. Pooling across all ages, the width from an individual pup's distribution averages 37 ± 8 ms.

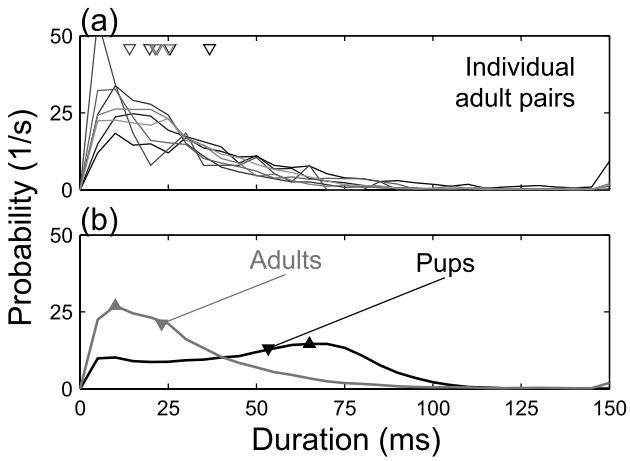


FIG. 8. Duration distributions. (a) Adult encounter calls for seven male/female pairs. Downward triangles indicate each pair's median duration, and averages 23 ± 7 ms. The mean of the duration distribution widths is 28 ± 7 ms. (b) Collective duration distributions for all pup (black) and all adult (gray) calls. Downward triangles mark the median duration for pups (53 ms) and adults (23 ms); upward triangles mark the peak duration for pups (65 ms) and adults (10 ms). Although there is significant overlap between the distributions, their peak durations clearly separate.

the duration of a pup isolation call is influenced by its median frequency. This is apparent in the plot of their joint distribution in Figs. 9(a) and (b) for P5 and P12 pups, respectively. The relatively diffuse cluster of calls near 65 kHz at P5 progresses with age into a tight cluster by P12. At the same time, the short duration, higher frequency calls seem to extend with age towards longer durations. When collected over all ages, pup isolation calls center around 67 kHz/59 ms and 94 kHz/30 ms, with widths of 7 kHz/35 ms and 6 kHz/41 ms, respectively. These vocalizations remain well separated from adult encounter calls, as shown in the contour plots of the joint distributions in Fig. 9(c). By using both a call's median frequency and duration to determine whether an isolated pup or an adult male produced the call, an observer would be correct 90.5% of the time, significantly better than using either feature alone.

Finally, we consider the period between the call onsets. As shown in the examples of Figs. 1(a)–(f), calls from both pups and adults often occur in bouts of at least two individual calls. Figures 10(a) and (b) show the normalized distribution of periods between the onsets of adjacent calls for individual pups at P5 and P12. There is a systematic shift in the collective distributions with age towards shorter peak periods, as reflected in Fig. 10(c). By P12, this peak repetition period is approximately 160 ms, down from 250 ms at P5. Figure 10(d) plots the width (full width at half the peak) of the collective distributions (black bars) along with the average of the widths of the individual distributions (white bars), as a function of age. It indicates a gradual narrowing of the collective repetition period distribution width (140 ms at P6; 60 ms by P12) towards the average individual's width (57 ± 17 ms).

Adults call at a much faster rate than pups, further distinguishing their vocalizations from pup isolation calls. Individual distributions of repetition periods for the different adult pairs are plotted in Fig. 11(a). Combining all the adults into one distribution [Fig. 11(b)], the peak repetition period is 100 ms, with a width of 60 ms. This is clearly separable from the collective pup distribution, which has a peak at 180 ms, and a width of 100 ms. A receiver classifying a call based on the period between adjacent calls would be correct 82.6% of the time. Individual to individual comparisons would give a median probability of being correct of 82.9%, with a 50% range from 80.5% to 85.5% [Fig. 6(b)]. If combined with the collective median frequency distribution, the probability would improve to 91.7%, slightly better than using median frequency and duration. Finally, the combination of all three parameters allows 97.4% correct discrimination. Thus, a receiver could accurately distinguish whether a vocalization is produced by an isolated pup or an adult in an encounter setting, despite the substantial variability within each of the call types' parameters.

IV. DISCUSSION

Our analysis of mouse pup isolation and adult encounter vocalizations produced two main results. First, the ontogeny

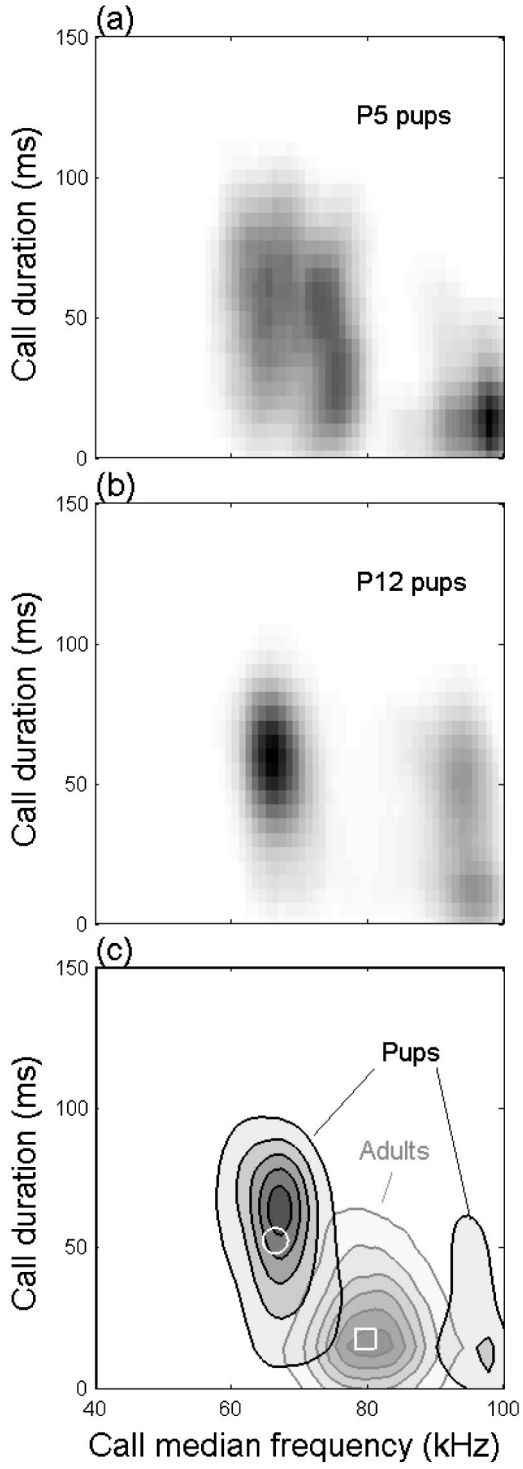


FIG. 9. Joint distributions of call median frequency and duration. (a) P5 pup isolation vocalizations (3423 calls). Calls with median frequencies ≤81 kHz clustered near 69 kHz and 51 ms (medians), with widths of 10 kHz and 39 ms, respectively. Calls with median frequencies >81 kHz clustered near 96 kHz and 17 ms, with widths of 7 kHz and 20 ms, respectively. (b) P12 pup isolation vocalizations (6933 calls). The clusters of calls have evolved since P5. Calls with median frequencies ≤81 kHz converge around 66 kHz and 58 ms, with widths of 5 kHz and 26 ms, respectively. On the other hand, calls with median frequencies >81 kHz shift and broaden into a cluster near 94 kHz and 40 ms, with widths of 7 kHz and 42 ms, respectively. (c) Contour plots of the joint distributions for all pup isolation (black lines, 51 954 calls) and adult encounter (gray lines, 11 248 calls) vocalizations. The majority of adult calls fall between the two pup call clusters. The white circles and squares denote the regions from which the marked calls in Figs. 1(b)–(f) have been drawn.

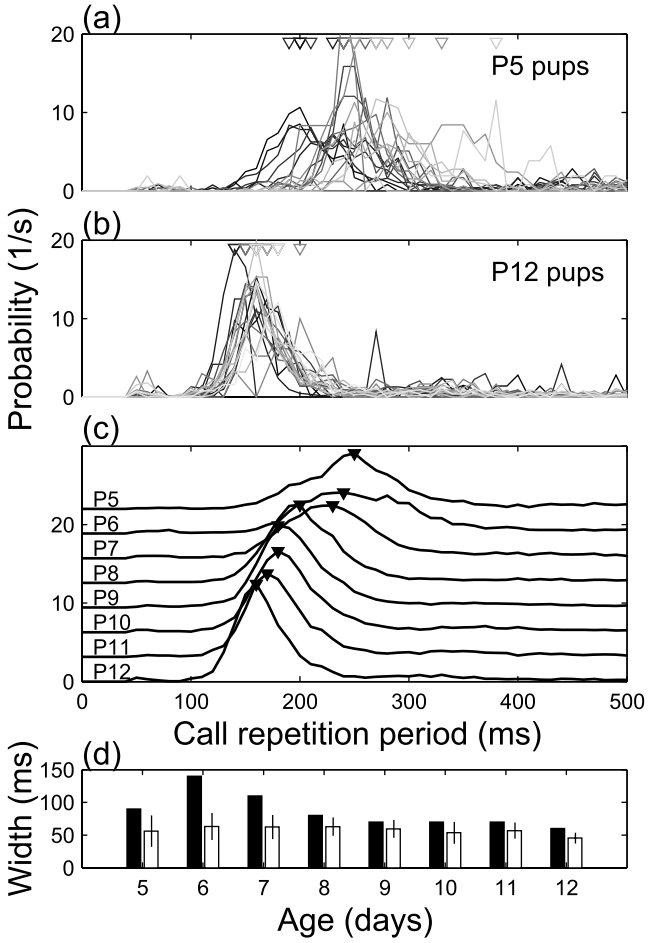


FIG. 10. Repetition period distributions for pup isolation calls. (a) Individual P5 pups. (b) Individual P12 pups. Distributions are calculated with 10-ms bins. A downward triangle indicates an individual pup's peak period. The average of the individual peak periods shifts from 254±47 ms at P5 to 161±14 ms at P12. Pooling across all ages, these peaks average 201 ±43 ms. (c) Progression of the collective repetition period distribution (successively offset by 3.1/s). A gradual shortening of the peak period begins around P8. (d) Widths (see Sec. III) of the collective distributions (black bars), and averages of the widths of individual distributions (white bars) at each age. Error bars are standard deviations. Between P6 and P12, there is a gradual decrease in the collective width from 140 to 60 ms. Pooling across all ages, the width from an individual pup's distribution averages 57 ±17 ms.

of spectral and temporal features in pup isolation calls shows a progression towards a more stereotyped call near 67-kHz frequency and 59-ms duration (main cluster of calls), occurring in bouts with repetition periods of around 180 ms. Second, adult encounter calls concentrate in regions mostly devoid of pup isolation calls—near 80 kHz and 23 ms, with approximately 100-ms repetition periods. Consequently, despite variability within each call type, adult encounter vocalizations remain acoustically distinct from pup isolation vocalizations along combined spectral and temporal dimensions. These findings suggest intriguing possibilities for how the calls might be perceived, processed, and produced.

A. Categorical perception

Categorical perception refers to the perceptual division of a continuous stimulus dimension into discrete sections,

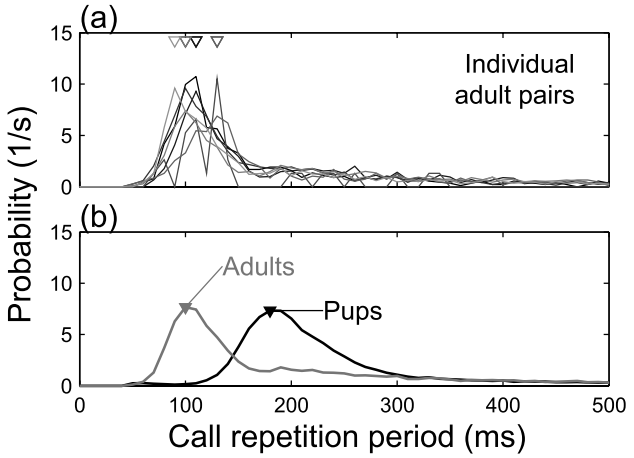


FIG. 11. Repetition period distributions. (a) Adult encounter calls for seven male/pairs of individuals are shown in different shades of gray. Downward triangles indicate each pair's peak period (several distributions have the same peak), and averages 110 ± 15 ms. The mean of the repetition period distribution widths is 51 ± 15 ms. (b) Collective repetition period distributions for all pup (black, 51 672 periods) and all adult (gray, 11 210 periods) calls. Downward triangles mark the peak period among pups (180 ms) and adults (100 ms). The long time (>300 ms), exponentially decaying trend is similar for both pups and adults, and reflects the Poisson nature of bouts. Periods from the two contexts are clearly distinguishable from each other.

where signals within (across) categories are poorly (well) discriminated from each other (Harnad, 1987). Lactating female NMRI mice categorically perceive ultrasounds along the bandwidth and duration dimensions (Ehret and Haack, 1981, 1982; Ehret, 1992). Narrow-band noise in the 30- to 60-kHz frequency range appears to elicit female preferences in two-alternative choice tests similar to natural pup isolation calls in this strain of mice (for which pup isolation vocalizations probably cluster at ~50 kHz) (Ehret and Haack, 1982). Assuming that CBA/CaJ mice also categorize pup isolation calls, the relevant frequency range probably centers at ~65 kHz, the middle of CBA/CaJ pups' spectral content clusters. Indeed, if the natural call distributions help to define perceptual categories, then our results on adult encounter ultrasounds suggest that they too could be categorically perceived along a bandwidth and/or frequency dimension. The existence of an adult encounter call category should be tested behaviorally, along with the ability of an adult female mouse to distinguish those calls from pup isolation calls. If they are capable of this, then adult encounter calls probably do not mimic pup isolation calls, as once hypothesized (Whitney *et al.*, 1973).

Categorical perception implies that acoustic differences between calls of the same category are ignored since they are poorly discriminated. However, testing this perception depends on choosing the right task. For example, we speculate that pup retrieval may not require individual pup recognition, so that simple categorization of sounds as puplike is sufficient to approach the sound. Importantly though, adult mice can distinguish their own versus alien pups (Ostermeyer and Elwood, 1983), as well as individual adults (D'Amato and Moles, 2001; Hurst *et al.*, 2001), using olfactory cues. From our data, ultrasound spectral content could contribute to individual recognition between P5–P8 pups; and bout repetition periods could help distinguish younger from older pups. Whether these acoustic parameters facilitate those finer perceptual distinctions should be investigated in more challenging tasks.

B. Neural processing

The bandwidth boundary found for the categorical perception of pup isolation calls by NMRI mice (Ehret and Haack, 1982) is similar to the critical bandwidth for 50 kHz, suggesting that critical bands are an essential neural mechanism for processing these vocalizations (Ehret, 1983). A critical band represents the hypothetical frequency filter over which energy is summed in the auditory system. Assuming a standard ''power spectrum model'' (Moore, 1997), where the relative phases of frequency components within a critical band are ignored, pup call detection would depend on the spectrum of each call. The spectral content reported here could be considered an averaged estimate of this power spectrum. Since the pup isolation call cluster near 67 kHz has a spectral width less than the estimated 27-kHz critical band around this frequency (Ehret, 1975), a critical band mechanism could be effective for processing these calls. However, some of the pup call spectral content overlaps the 34-kHz critical band centered on the adult calls at 79 kHz. If discrimination of the two calls is behaviorally important, spectral analysis alone may not provide a sufficient mechanism, and other call features such as duration and repetition period must be invoked. For example, neurons involved in processing calls should be sensitive to the appropriate critical band frequency range and the natural distribution of durations. Indeed, some neurons in the inferior colliculus of the mouse exhibit critical band tuning (Vartanyan *et al.*, 2000; Egorova *et al.*, 2002), while others seem to be tuned to durations (Brand *et al.*, 2000). Ultimately, higher order neurons may be sensitive to the appropriate combination of features that define each sound category.

Categorically responding neurons would provide a reasonable code for drawing the acoustic distinction between any pup and adult. However, if individual discriminations are also behaviorally relevant, as mentioned earlier, its neural substrate would probably be distributed across a population of neurons at early stages of processing. Each neuron would then contribute to the representation of the whole range of ultrasound vocalizations. Recent ecological theories of neural coding suggest that this population might nevertheless be adapted to the statistical structure of the natural ultrasound ensemble (Lewicki, 2002). Our extensive library of calls would enable us to predict the receptive fields of these optimized neurons.

C. Vocal development mechanisms

Although we focus on the ethological questions of call variability and discrimination by a hypothetical recipient, our developmental results raise an interesting mechanistic issue. Across all parameters, we usually find more variability between individuals from P5 to P7 compared to later ages. Most strikingly, we observe a contraction of the spectral content around P9, resulting from the convergence of the fre-

quencies in different pups' calls towards the midrange value of ~67 kHz, rather than from a reduction in the spectral content widths of distributions from individual pups. Previous developmental studies have measured only specific spectral parameters rather than the full spectral content distributions, and have not reported such a contraction (Noirot and Pye, 1969; Sales and Smith, 1978; Roubertoux *et al.*, 1996; Hahn *et al.*, 1998).

This convergence may signify one of several proximate causes. First, it could be due to the pup's natural physical development. Ultrasound whistles are produced when air is forced through the trachea and past narrow apertures in the larynx created by a constriction of the laryngeal folds (Roberts, 1975). As the pup ages, its trachea and larynx grow, and its control over the laryngeal muscles presumably develops. The frequencies generated by this mechanism would be expected to change monotonically and gradually, as inferred from studies in rats (Blumberg *et al.*, 2000). However, we observe that similarly sized young pups produce a range of different frequencies, and transition with age towards a common, middle frequency. This results from a shift of both the upper and lower edges of the spectral content distribution towards the center. Moreover, this mechanism would probably not account for the sudden convergence in spectral content around P9.

As another possibility, acoustic feedback may lock a pup's frequencies to ~67 kHz. In the hearing development of the CBA mouse, P9 marks the opening of the ear canal, and the disappearance of most of the gelatinous, embryonic tissue in the middle ear (Mikaelian and Ruben, 1964). In fact, most of the cochlea is anatomically developed by P8 (Mikaelian and Ruben, 1964; Shnerson and Pujol, 1983). Physiologically, cochlear microphonics can be recorded from the mouse as early as P8, indicating electrical activity from the hair cells. Eighth-nerve action potentials begin firing soon afterwards around P9–P10 (Alford and Ruben, 1963; Mikaelian and Ruben, 1964). In C57BL/6J mice, the opening of the ear canal is associated with an immediate decrease in thresholds for eighth-nerve evoked activity, and a gradual lowering of the acoustically driven neural activity threshold in the inferior colliculus (Shnerson and Pujol, 1983). Behaviorally, hand claps can elicit unconditioned responses as early as P9 for some CBA pups (Alford and Ruben, 1963). This time-line indicates that the hardware for detecting acoustic signals is mostly in place by P9. If pups exhibit a dip in their hearing thresholds around 50–70 kHz, as seen in adult mice (Ralls, 1967; Brown, 1973), then the pups may use the acoustic feedback from their own or other pups' vocalizations to lock to the loudest or most often perceived frequency. The initial interanimal variability may then be due to the lack of acoustic calibration for an individual's vocal production. Unfortunately, the frequency range of cochlear potentials at P9 only extends up to 10 kHz for 100 dB SPL tones in CBA mice (Mikaelian and Ruben, 1964), and unconditioned behavioral responses to 5–40 kHz (<80 dB SPL) tones in NMRI mice appear only after P11 (Ehret, 1976). Since the (more sensitive) shock-conditioned behavioral responses of P17–P19 NMRI mice do exhibit a threshold dip (~40 dB SPL) at 50 kHz (Ehret, 1976), a sensitivity dip might exist in younger mice, but at a much higher threshold. Further testing is therefore necessary to determine whether our results indicate that mouse pups show vocal learning based on acoustic feedback. Alternatively, pups may be learning to lock their vocalization frequency to the *mother's* hearing threshold dip because it might be most effective for eliciting a retrieval, although it is less clear why this would happen at P9. Distinguishing the cause of the spectral content convergence from among these and other hypotheses requires future experimentation.

ACKNOWLEDGMENTS

The authors thank J. F. Linden, M. S. Brainard, A. J. Doupe and G. Ehret for helpful discussions. Funding has been provided by a Sloan and Swartz Foundation fellowship (RCL), University of California President's fellowship (RCL), NIDCD NRSA Fellowship No. F32 DC05279 (RCL), and grants from University of California, San Francisco Research Evaluation and Allocation Committee, and NIDCD 002260.

[1]The CBA/CaJ mouse is an excellent hearing strain that is often used as a normal-hearing control in auditory studies (Zheng *et al.*, 1999). Full genealogies of mouse inbred strains are available in Beck *et al.* (2000).

[2]The behavioral relevance of the ultrasound frequency spectrum and call duration is evidenced by their categorical perception. Additionally, repetition period may also serve a behavioral function. One pup call is often insufficient in eliciting a retrieval behavior, but two or more seem equally effective (G. Ehret, private communication).

[3]An "individual distribution" is formed from the calls of a single pup, whereas a "collective distribution" is formed from the calls of all pups matching the desired criteria, e.g., P5 pups. The average of a parameter, e.g., width, across the individual distributions is computed by estimating the parameter for each individual first, and then averaging together those results. This can differ from the value of this parameter for the collective distribution containing the same individuals.

[4]Because the physical contexts are clearly different as well, this does not address whether nonauditory cues may also play a role in discriminating calls. However, the focus here is on the potential information provided by the sounds themselves.

Alford, B. R., and Ruben, R. J. (**1963**). "Physiological, behavioral and anatomical correlates of the development of hearing in the mouse," Ann. Otol. Rhinol. Laryngol. **72**, 237–247.

Beck, J. A., Lloyd, S., Hafezparast, M., Lennon-Pierce, M., Eppig, J. T., Festing, M. F., and Fisher, E. M. (**2000**). "Genealogies of mouse inbred strains," Nat. Genet. **24**(1), 5, 23–25.

Beeman, K. (**1998**). "Digital signal analysis, editing and synthesis," in *Animal Acoustic Communication*, edited by S. L. Hopp, M. J. Owren, and C. S. Evans (Springer-Verlag, New York), pp. 59–103.

Bell, R. W., Nitschke, W., and Zachman, T. A. (**1972**). "Ultra-sounds in three inbred strains of young mice," Behav. Biol. **7**(6), 805–814.

Blumberg, M. S., Sokoloff, G., and Kent, K. J. (**2000**). "A developmental analysis of clonidine's effects on cardiac rate and ultrasound production in infant rats," Dev. Psychobiol. **36**(3), 186–193.

Boll, S. F. (**1979**). "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-27**(2), 113–120.

Branchi, I., Santucci, D., Vitale, A., and Alleva, E. (**1998**). "Ultrasonic vocalizations by infant laboratory mice: a preliminary spectrographic characterization under different conditions," Dev. Psychobiol. **33**(3), 249–256.

Brand, A., Urban, R., and Grothe, B. (**2000**). "Duration tuning in the mouse auditory midbrain," J. Neurophysiol. **84**(4), 1790–1799.

Brown, A. M. (**1973**). "High frequency peaks in the cochlear microphonic response of rodents," J. Comp. Physiol. **83**, 377–392.

D'Amato, F. R., and Moles, A. (**2001**). "Ultrasonic vocalizations as an index of social memory in female mice," Behav. Neurosci. **115**(4), 834–840.

Efron, B., and Tibshirani, R. J. (**1993**). *An Introduction to the Bootstrap* (Chapman and Hall, New York).
Egorova, M. A., Vartanyan, I. A., and Ehret, G. (**2002**). ''Neural critical bands and inhibition in the auditory midbrain of the house mouse (*mus domesticus*),'' Dokl. Biol. Sci. **382**, 5–7.
Ehret, G. (**1975**). ''Masked auditory thresholds, critical ratios, and scales of the basilar membrane of the housemouse (*mus musculus*),'' J. Comp. Physiol. **103**, 329–341.
Ehret, G. (**1976**). ''Development of absolute auditory thresholds in the house mouse (*mus musculus*),'' J. Am. Audiol Soc. **1**(5), 179–184.
Ehret, G. (**1983**). ''Auditory processing and perception of ultrasounds in house mice,'' in *Advances in Vertebrate Neuroethology*, edited by J-P. Ewert, R. R. Capranica, and D. J. Ingle, NATO advanced science institutes series. Series A, Life Sciences (Plenum, New York), 56th ed.
Ehret, G. (**1992**). ''Categorical perception of mouse-pup ultrasounds in the temporal domain,'' Anim. Behav. **43**(3), 409–416.
Ehret, G., and Haack, B. (**1981**). ''Categorical perception of mouse pup ultrasound by lactating females,'' Naturwissenschaften **68**(4), 208–209.
Ehret, G., and Haack, B. (**1982**). ''Ultrasound recognition in house mice: key-stimulus configuration and recognition mechanism,'' J. Comp. Physiol., A **148**(2), 245–251.
Elwood, R. W., and Keeling, F. (**1982**). ''Temporal organization of ultrasonic vocalizations in infant mice,'' Dev. Psychobiol. **15**(3), 221–227.
Green, D. M., and Swets, J. A. (**1988**). *Signal Detection Theory and Psychophysics* (Penninsula, Los Altos, CA).
Haack, B., Markl, H., and Ehret, G. (**1983**). ''Sound communication between parents and offspring,'' in *The Auditory Psychobiology of the Mouse*, edited by J. F. Willott (Thomas, Springfield, IL), pp. 57–97.
Hahn, M. E., Hewitt, J. K., Adams, M., and Tully, T. (**1987**). ''Genetic influences on ultrasonic vocalizations in young mice,'' Behav. Genet. **17**(2), 155–166.
Hahn, M. E., Karkowski, L., Weinreb, L., Henry, A., Schanz, N., and Hahn, E. M. (**1998**). ''Genetic and developmental influences on infant mouse ultrasonic calling. II. Developmental patterns in the calls of mice 2-12 days of age,'' Behav. Genet. **28**(4), 315–325.
Harnad, S. R. (**1987**). *Categorical Perception: The Groundwork of Cognition* (Cambridge U.P., New York).
Hauser, M. D. (**1996**). *The Evolution of Communication* (MIT, Cambridge, MA).
Hurst, J. L., Payne, C. E., Nevison, C. M., Marie, A. D., Humphries, R. E., Robertson, D. H., Cavaggioni, A., and Beynon, R. J. (**2001**). ''Individual recognition in mice mediated by major urinary proteins,'' Nature (London) **414**(6864), 631–634.
Lewicki, M. S. (**2002**). ''Efficient coding of natural sounds,'' Nat. Neurosci. **5**(4), 356–363.
Mikaelian, D., and Ruben, R. J. (**1964**). ''Development of hearing in the normal CBA-J mouse,'' Acta Oto-Laryngol. **59**, 451–461.
Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, San Diego).
Motomura, N., Shimizu, K., Shimizu, M., Aoki-Komori, S., Taniguchi, K., Serizawa, I., and Saito, T. R. (**2002**). ''A comparative study of isolation-induced ultrasonic vocalization in rodent pups,'' Exp. Anim. **51**(2), 187–190.
Noirot, E., and Pye, D. (**1969**). ''Sound analysis of ultrasonic distress calls of mouse pups as a function of their age,'' Anim. Behav. **17**, 340–349.
Nunez, A. A., Pomerantz, S. M., Bean, N. J., and Youngstrom, T. G. (**1985**). ''Effects of laryngeal denervation on ultrasound production and male sexual behavior in rodents,'' Physiol. Behav. **34**(6), 901–905.
Nyby, J. G. (**2001**). ''Auditory communication among adults,'' in *Handbook of Mouse Auditory Research: From Behavior to Molecular Biology*, edited by J. F. Willott (CRC, New York), pp. 3–18.
Okon, E. E. (**1970**). ''The effect of environmental temperature on the production of ultrasound by isolated, non-handled, albino mouse pups,'' J. Zool., London **162**, 71–83.
Ostermeyer, M. C., and Elwood, R. W. (**1983**). ''Pup recognition in *mus musculus*: parental discrimination between their own and alien young,'' Dev. Psychobiol. **16**(2), 75–82.
Pomerantz, S. M., Nunez, A. A., and Bean, N. J. (**1983**). ''Female behavior is affected by male ultrasonic vocalizations in house mice,'' Physiol. Behav. **31**(1), 91–96.
Ralls, K. (**1967**). ''Auditory sensitivity in mice, *peromyscus* and *mus musculus*,'' Anim. Behav. **15**(1), 123–128.
Roberts, L. H. (**1975**). ''The rodent ultrasound production mechanism,'' Ultrasonics **13**(2), 83–88.
Roubertoux, P. L., Martin, B., Le Roy, I., Beau, J., Marchaland, C., Perez-Diaz, F., Cohen-Salmon, C., and Carlier, M. (**1996**). ''Vocalizations in newborn mice: genetic analysis,'' Behav. Genet. **26**(4), 427–437.
Sales, G. D., and Smith, J. C. (**1978**). ''Comparative studies of the ultrasonic calls of infant murid rodents,'' Dev. Psychobiol. **11**(6), 595–619.
Sewell, G. D. (**1970**). ''Ultrasonic communication in rodents,'' Nature (London) **227**(256), 410.
Shnerson, A., and Pujol, R. (**1983**). ''Development: anatomy, electrophysiology and behavior,'' in *The Auditory Psychobiology of the Mouse*, edited by J. F. Willott (Thomas, Springfield, IL), pp. 395–425.
Smith, J. C. (**1976**). ''Responses of adult mice to models of infant calls,'' J. Comp. Physiol. Psychol. **90**(12), 1105–1115.
Smith, J. C., and Sales, G. D. (**1980**). ''Ultrasonic behavior and mother-infant interactions in rodents,'' in *Maternal Influences and Early Behavior*, edited by R. W. Bell and W. P. Smotherman (SP Medical and Scientific Books, New York), pp. 105–133.
Vartanyan, I. A., Egorova, M. A., and Ehret, G. (**2000**). ''Critical bandwidths of different types of neurons in the mouse auditory midbrain,'' Dokl. Biol. Sci. **373**(1-6), 364–366.
Warburton, V. L., Sales, G. D., and Milligan, S. R. (**1989**). ''The emission and elicitation of mouse ultrasonic vocalizations: the effects of age, sex and gonadal status,'' Physiol. Behav. **45**(1), 41–47.
Whitney, G., Coble, J. R., Stockton, M. D., and Tilson, E. F. (**1973**). ''Ultrasonic emissions: do they facilitate courtship of mice,'' J. Comp. Physiol. Psychol. **84**(3), 445–452.
Willott, J. F. (editor) (**2001**). *Handbook of Mouse Auditory Research: From Behavior to Molecular Biology* (CRC, New York).
Zheng, Q. Y., Johnson, K. R., and Erway, L. C. (**1999**). ''Assessment of hearing in 80 inbred strains of mice by ABR threshold analyses,'' Hear. Res. **130**(1-2), 94–107.

Erratum: "Estimation of broadband power levels radiated from turbulent boundary layer-driven ribbed plates having dissimilar sections" [J. Acoust. Soc. Am 114, 737–744 (2003)]

M. L. Rumerman
Naval Surface Warfare Center, Carderock Division, Signatures Directorate (Code 7204), 9500 MacArthur Boulevard, West Bethesda, Maryland 20817-5700

(Received 15 September 2003; accepted for publication 23 September 2003)

[DOI: 10.1121/1.1626311]

PACS numbers: 43.30.Jx, 43.40.Rj, 43.20.Tb, 43.10.Vx [ANN]

The right hand sides of Eqs. (3b)–(3f) and Eq. (4) should be multiplied by the panel width, *W*. The final right hand side of Eq. (16) and the left hand side of Eq. (17) should also be multiplied by *W*. (The final right hand side of Eq. (16) will then contain the square of *W*.) Results for calculated power presented in Figs. 4 and 6 are unaffected.

The quantity v_n^2, first appearing in Eq. (3b), was undefined and is intended to be the mean squared velocity of a mode (in the *y* direction) of the *n*th section. There is an error in the notations on Fig. 2, in which the reflected wave should have been represented as $V_{\mathrm{ref}} \exp(-i\alpha_L x)$. There is a typographical error in the square root expression of Eq. (A2); the quantity β should be raised to the second power. The symbol "*w*" in the second term on the right hand side of Eq. (16) should be upper case, and represents the plate width.

Erratum: "Correction of ultrasonic wave aberration with a time delay and amplitude filter" [J. Acoust. Soc. Am. 113, 2009–2020 (2003)]

Svein-Erik Måsøy, Tonni F. Johansen, and Bjørn Angelsen
Department of Circulation and Imaging, Norwegian University of Science and Technology, Trondheim, Norway

(Received 9 September 2003; accepted for publication 12 September 2003)

[DOI: 10.1121/1.1623451]

PACS numbers: 43.58.Ry, 43.20.Fn, 43.10.Vx [AJZ]

TABLE I. Body wall models statistics and waveform similarity factors (WSF).

	Arrival time fluctuations		Energy level fluctuations		WSF
Wall model	rms value (ns)	Corr. len. (mm)	rms value (dB)	Corr. len. (mm)	(*1e-3)
w2	36.0	2.8	2.9	1.6	971
w4	37.5	2.8	2.4	1.6	984
w6	**49.4**	**3.0**	**3.1**	**1.8**	**975**
w8	47.7	3.0	2.5	2.2	992
s2	39.6	2.8	3.6	2.0	957
s4	54.5	3.4	4.0	2.0	924
s6	**53.7**	**2.4**	**4.1**	**1.4**	**914**
s8	69.8	4.0	3.8	1.8	949

Incorrect values were submitted in the article ''Correction of ultrasonic wave aberration with a time delay and amplitude filter'' due to a programming error. A corrected version of Table I is included here and the effect of the errors are discussed below.

The correlation lengths for the arrival time fluctuations displayed in Table I were incorrect due to a programming error. The correct values are given in Table I. The errors in the correlation lengths have no impact on any other result presented in the article, and therefore does not challenge any of the conclusions made.