CODEN: JASI

0001-4966

The Journal of the Acoustical Society of America

Vol. 108, No. 2 August 2000

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 108, NO. 2, AUGUST 2000

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 108, NO. 2, AUGUST 2000

CONTENTS—*Continued from preceding page*

ACOUSTICAL NEWS—USA

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

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

New Fellow of the Acoustical Society of America

Peter Tyack—For contributions to the understanding of acoustic communication by marine mammals.

Christy Holland and J. Brian Fowlkes receive AIUM award

ASA members Christy K. Holland and J. Brian Fowlkes were presented the American Institute for Ultrasound in Medicine (AIUM) presidential recognition award in 2000 for their efforts in chairing a conference on ''Mechanical Bioeffects from Diagnostic Ultrasound.'' The conference was held in April 1998 to debate ultrasound bioeffects. Consensus statements and supporting materials were published in the February 2000 issue of the *Journal of Ultrasound in Medicine*.

Christy Holland is Research Associate Professor at the University of Cincinnati, Department of Radiology, in Cincinnati. She is a Fellow of the Acoustical Society of America and currently serves as a Member of the ASA Executive Council. J. Brian Fowlkes is Senior Associate Research Scientist at the University of Michigan Medical Center, Department of Radiology, in Ann Arbor, and serves as a member of the ASA Technical Committee on Biomedical Ultrasound/Bioresponse to Vibration.

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.

2000

2–4 August	2000 Ultrasonic Transducer Engineering Conference, University Park, PA [K. Kirk Shung, NIH Resource on Medical Ultrasonic Transducer Technology, Pennsylvania State Univ., 231 Hallowell Bldg., University Park, PA 16802; Tel.: 814-865-1407; Fax: 814-863-0490; Email: kksbio@engr.psu.edu].
21–23 September	Eighth Annual Conference on the Management of the Tinnitus Patient, Iowa City, IA [Richard Tyler, Tel.: 391-356-2471; Email: tyler@uiowa.edu; WWW: www.medicine.uiowa.edu/otolaryngology/news/news].
22–25 October	IEEE Ultrasonics Symposium, San Juan, Puerto Rico [R. Almar, 896 Buttonwood Ln., Altamonte Springs, FL 32714; Fax: 407-290-5181; WWW; www.uffcsymp2000.org].
4–8 December	Joint Meeting: 140th Meeting of the Acoustical Society of America/NoiseCon 2000, Newport Beach, CA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: asa.aip.org].

2001

22–25 March	''New Frontiers in the Amelioration of Hearing Loss,'' St. Louis, MO [Sarah Uffman, CID Department of Research, 4560 Clayton Ave., St. Louis, MO 63110; Tel.: 314-977-0278; Fax: 314-977-0030; Email: suffman@cid.wustl.edu].
30 April–3 May	2001 SAE Noise & Vibration Conference & Exposition, Traverse City, MI [Patti Kreh, SAE Int'l., 755 W. Big Beaver Rd., Suite 1600, Troy, MI 48084; Tel.: 248-273-2474; Fax: 248-273-2494; Email: pkreh@sae.org].
4–8 June	141st Meeting of the Acoustical Society of America, Chicago, IL [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: asa.aip.org].
7–10 October	2001 IEEE International Ultrasonics Symposium Joint with World Congress on Ultrasonics, Atlanta, GA [W. O'Brien, Electrical and Computer Engineering, Univ. of Illinois, 405 N. Mathews, Urbana, IL 61801; Fax: 217-244-0105; WWW: www.ieee-uffc.org/2001].

 0001-4966/2000/108(2)/463/1/$17.00

ACOUSTICAL NEWS—INTERNATIONAL

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

Papers published in JASJ(E)

A listing of Invited Papers and Regular Papers appearing in the latest issue of the English language version of the *Journal of the Acoustical Society of Japan*, JASJ(E), was published for the first time in the January 1995 issue of the *Journal*. This listing is continued below.

The May 2000 issue of JAS(E), Vol. 21, No. 3, contains the following contributions:

A review paper by H. Goydke, ''The present circumstances of international harmonisation of building acoustic standards in Europe'' and the following regular papers:

D. Takahashi, Y. Kato, and K. Sakamoto, ''Sound fields caused by diffused-type reflectors with periodic profile''
T. Hasegawa, T. Kido, T. Iuzuka, and C. Matsuoka, ''A general theory of Rayleigh and Langevin radiation pressures''
N. Takahashi, K. Futa, T. Tsuchiya, and T. Kikuchi, ''Calculation of eigenray with equi-sound-speed division of sound speed profile''

International Meetings Calendar

Below are announcements of meetings to be held abroad. Entries preceded by an * are new or updated listings with full contact addresses given in parentheses. *Month/year* listings following other entries refer to meeting announcements, with full contact addresses, which were published in previous issues of the *Journal*.

August 2000

28–30 **INTER-NOISE 2000**, Nice. (Fax: +33 1 47 88 90 60; Web: internoise2000.loa.espci.fr) *6/99*

31–2 **International Conference on Noise & Vibration Pre-Design and Characterization Using Energy Methods (NOVEM)**, Lyon. (Fax: +33 4 72 43 87 12; Web: www.insa-lyon.fr/laboratoires/lva.html) *6/99*

September 2000

3–6 **5th French Congress on Acoustics—Joint Meeting of the Swiss and French Acoustical Societies**, Lausanne. (Fax: +41 216 93 26 73) *4/99*

13–15 **International Conference on Noise and Vibration Engineering (ISMA 25)**, Leuven. (Fax: +32 16 32 24 82; e-mail: lieve.notre@mech.kuleuven.ac.be) *12/99*

17–21 **Acoustical Society of Lithuania 1st International Conference**, Vilnius. (Fax: +370 2 223 451; e-mail: daumantas.ciblys@ff.vu.lt) *8/99*

18–22 **47th Seminar on Acoustics (OSA2000)**, Zalew Solinski, Poland. (e-mail: osa@atena.univ.rzeszow.pl) *4/00*

29–1 **Tone Wood Forum**, Quesnel, BC, Canada. (J. Griffin, Island Mountain Arts, P.O. Box 65, Wells, BC V0K 2R0, Canada; Fax: +1 250 994 3433; Web: www.imarts.com) *6/00*

October 2000

3–5 **WESTPRAC VII**, Kumamoto. (Web: cogni.eecs.kumamoto-u.ac.jp/others/westprac7) *6/98*

3–6 **EUROMECH Colloquium on Elastic Waves in NDT**, Prague. (Fax: +420 2 858 4695; e-mail: ok@bivoj.it.cas.cz) *10/99*

12–14 **International Conference on Newborn Hearing Screening**, Milan. (Fax: +39 2 23993367/.60; Web: www.biomed.polimi.it/nh2000) *2/00*

16–18 **2nd Iberoamerican Congress on Acoustics, 31st National Meeting of the Spanish Acoustical Society, and EAA Symposium**, Madrid. (Fax: +34 91 411 7651; e-mail: ssantiago@fresno.csic.es) *12/98*

16–20 **6th International Conference on Spoken Language Processing**, Beijing. (Fax: +86 10 6256 9079; Web: www.icslp2000.org) *10/98*

November 2000

10–12 ***Institute of Acoustics Autumn Conference (Industrial Noise)**, Stratford-upon-Avon, UK. (IOA, 77A St. Peter's Street, St. Albans, Herts AL1 3BN, UK; Fax: +44 1727 850553; Web: www.ioa.org.uk)

17–19 ***Reproduced Sound 16**, Stratford-upon-Avon, UK. (IOA, 77A St. Peter's Street, St. Albans, Herts AL1 3BN, UK; Fax: +44 1727 850553; Web: www.ioa.org.uk)

24–27 **21st Tonmeistertagung (VDT International Audio Convention)**, Hanover, Germany. (Convention Office VDT, Am Zaarshäuschen 9, 51427 Bergisch-Gladbach, Germany; Fax: +49 2204 21584; Web: www.tonmeister.de) *6/00*

April 2001

9–11 ***Acoustical Oceanography**, Southampton, UK. (IOA, 77A St. Peter's Street, St. Albans, Herts AL1 3BN, UK; Fax: +44 1727 850553; Web: www.ioa.org.uk)

July 2001

2–6 ***8th International Congress on Sound and Vibration**, Kowloon, Hong Kong. (K M Li, Mechanical Engineering, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong; Fax: +852 2365 4703; Web: www.iiav.org)

August 2001

28–30 **INTER-NOISE 2001**, The Hague. (Web: internoise2001.tudelft.nl) *6/99*

September 2001

2–7 **17th International Congress on Acoustics (ICA)**, Rome. (Fax: +39 6 4424 0183; Web: www.uniroma1.it/energ/ica.html) *10/98*

10–13 **International Symposium on Musical Acoustics (ISMA 2001)**, Perugia. (Fax: +39 75 577 2255; e-mail: perusia@classico.it) *10/99*

October 2001

17–19 **32nd Meeting of the Spanish Acoustical Society**, La Rioja. (Fax: +34 91 411 76 51; Web: www.ia.csic.es/sea/index.html) *10/99*

September 2002

16–21 **Forum Acusticum 2002 (Joint EAA–SEA–ASJ Meeting)**, Sevilla. (Fax: +34 91 411 7651; Web: www.cica.es/aliens/forum2002) *2/00*

 0001-4966/2000/108(2)/465/1/$17.00

REPORTS OF RELATED MEETINGS

This Journal department provides concise reports of meetings that have been held by other organizations concerned with acoustical subjects; and of meetings co-sponsored by the Acoustical Society but planned primarily by other co-sponsors.

Concert Hall Acoustics Summer Institute Held at Tanglewood Music Center, Lenox, MA

''Tanglewood 99,'' the first Concert Hall Acoustics Summer Institute, was held August 29 to September 1 1999 at the Tanglewood Music Center, Lenox, Massachusetts. The Institute was sponsored by the Concert Hall Research Group (CHRG), with the cooperation of the Acoustical Society of America, Technical Committee on Architectural Acoustics (ASA/TCAA), The National Council of Acoustical Consultants (NCAC), and the Robert Newman Student Award Fund. Thirty-one attendees included students, acoustical researchers, consultants, and architects interested in the design of music facilities. Five students received CHRG-funded full fellowships to cover tuition, room and board, and a grant for travel expenses. Six additional students received partial tuition grants. The Institute was conceived and organized by Bill Cavanaugh and Tim Foulkes of the CHRG who served as associate directors of Tanglewood 99.

Thirty-one students and fourteen faculty and staff participated (see Fig. 1). The presenters for six scheduled 3-hour workshop sessions included: John Bradley on ''Concert Hall Acoustics Measurements;'' Russell Johnson, Robert Wolff, and Andrew Monro on ''Design of Concert Chambers;'' Leo Beranek on ''Critical Factors in Concert Hall and Opera House Design;'' Larry Kirkegaard on ''Concert Halls in Search of an Acoustic Aesthetic;'' Chris Jaffe and Paul Scarborough on ''Concert Halls for a New Century;'' and Dick Campbell and Mendel Kleiner on ''Auralization and Modeling.''

The Keynote Lecture was given by Gary Siebein on Sunday evening, 29 August, after faculty, staff, and participants attended an afternoon Boston Symphony Orchestra (BSO) performance of Beethoven's 9th Symphony. Renowned maestro James DePreist conducted the BSO in a stirring performance with John Oliver leading the Tanglewood Festival Chorus in the 5000-seat Koussevitzky Music Shed. In 1959, the consulting firm Bolt Beranek and Newman, with architects Saarinen Associates, designed the highly successful and architecturally pleasing orchestra canopy and other acoustical improvements for the Shed. On Monday morning, 30 August, David Sturma, Director of Tanglewood Facilities, led the participants on an initial ''behind-the-scenes'' technical tour of the Tanglewood Campus facilities. After the morning tour, the participants reconvened at Highwood Manor House on the Tanglewood Campus for a round-table discussion of music performance and listening issues with Tanglewood staff, performers, and Leo Beranek, a Life Trustee of the BSO. After lunch, the participants embarked on the first of the six scheduled workshops, two of which were held at the widely acclaimed Seiji Ozawa Concert Hall. Larry Kirkegaard led the Tuesday afternoon session at Ozawa Hall (see Fig. 2).

The course participants received 262 pages of printed notes, prepared to complement the lectures and workshops, and two books: L. L. Beranek, *Concert and Opera Halls: How They Sound* (Acoustical Society of America, 1996), and W. J. Cavanaugh and J. A. Wilkes (eds.), *Architectural Acoustics: Principles and Practice* (Wiley, 1999). The books were a special con-

FIG. 1. Group photo taken on the lawn overlooking the Berkshire Hills near Highwood Mansion on Monday, August 30, 1999 after touring the Tanglewood campus. Note there were a few missing participants, presumably those preparing for later sessions or just enjoying the ambience and serenity.

 0001-4966/2000/108(2)/466/2/$17.00

FIG. 2. Larry Kirkegaard, acoustical designer of the renowned Seiji Ozawa Hall, leads the Tuesday afternoon workshop. John Bradley had conducted the session at Ozawa the previous afternoon on basic acoustical measurements in halls to set the stage for Larry's session.

tribution by the Robert Newman Student Award Fund to the Tanglewood '99 students. Together the participants' resources totaled 1237 pages for study and reference during the four day Institute! Following the Tuesday, 31 August, afternoon workshop at Ozawa Hall, participants had the opportunity to listen to music in Ozawa Hall and, later, to portions of a Boston Pops Orchestra rehearsal being conducted by John Williams in the Shed. The final session on Wednesday afternoon, 1 September, included Yoichi Ando's invited lecture on ''Concert Hall Acoustics, Blending Music Performances, Sound Fields, and Listener,'' Jerry Hyde's ''Overview and Endnote Remarks,'' and a lively open forum which concluded this unique, intensive acoustics educational experience on the art, science, and technology of concert hall design.

Students and faculty members were affiliated with seventeen universities worldwide. Attendees were from ten nations: Australia, Canada, Denmark, England, France, Greece, Japan, New Zealand, Sweden, and the United States. In the workshops, the design, construction, and evaluation of many new successful halls for music were described along with data for their objective acoustical characteristics measured by reverberation time (RT), early decay time (EDT), bass ratio (BR), relative strength (G), clarity index (C80), lateral energy fraction (LEF), and interaural cross-correlation coefficient (IACC). It was evident from the lectures and workshops that significant progress on concert hall acoustics has been made during the concluding half of the 20th Century, but as Dr. Bradley noted in his workshop: ''There is clearly more to do.'' It is anticipated that future institutes will be organized in the years ahead to build on the enthusiasm and student-researcher-consultant-architect links established during Tanglewood 99. For information on Concert Hall Acoustics Summer Institutes, contact CHRG at 978/443-7871 or E-mail: tfoulkes@cavtocci.com.

M. David Egan
Professor Emeritus, Clemson University
P. O. Box 365
Anderson, SC 29622-0365

OBITUARIES

This section of the Journal publishes obituaries concerning the death of Fellows of the Society and other acousticians eminent in the world of acoustics. When notified, the Editor-in-Chief solicits a summary of the person's life and contributions from an ASA member thoroughly familiar with the details, if possible. If a promised obituary is never received, a brief obituary notice may be published later.

Jean-Noël Decarpigny • 1948–1999

Jean-Noël Decarpigny, a Fellow of our Society, passed away unexpectedly after a brief battle with liver cancer on 29 April 1999 at age 50. At the time of his death, he was serving as the Director of the Institute Supérieur d'Electronique du Nord (ISEN) in Lille, France. He was also a Professor of Acoustics at the University of Valencienne.

He received his Diplome Ingenieur from ISEN in 1971. He was awarded a Ph.D. in Solid State Physics (doctorat de troisièm cycle) in 1973 from the Université des Sciences et Techniques de Lille, France, and a doctorat d'Etat in 1984 from the same institution for work in acoustics.

He is best known in acoustics for his contributions to numerical methods for the design of electroacoustic transducers. Together with Jean Claude Debus and other colleagues at ISEN, he developed the finite-element code, ATILA, with support from the French Navy, which has been and continues to be very useful in many parts of the world for the design of high-power piezoelectric and piezomagnetic sonar transducers.

He was associated with ISEN for his entire professional career, first as a student, then a teacher and researcher in physics, acoustics, and electronics, then Director of Studies, and finally, in 1992, until his death, Director. He was the leader of efforts to encourage research activities of the teaching staff, the development of high academic standards, including the requirement that students develop skills with two foreign languages. The result has been a high demand for ISEN's Diplome Ingenieur graduates in French industry, especially in the fields of microelectronics and acoustics. He was highly respected for his enthusiastic teaching and inspiring leadership in research and in academic administration. He always had a clear idea and plan for the future of his research and of engineering education. He was the leader in organizing the establishment of two private electrical engineering schools, sisters to ISEN, the ISEM in Toulon and the ISEB in Brest, during the last 10 years. For this, he was given in 1998 the rank of *Chevalier de l'Ordre National du Mérite*. Among his awards were the *Prix Wicar and Hagelstein* from the Société Scientifique de Lille, France in 1985, and the *Prix Chavasse* from the Société Française d'Acoustique in 1987.

He was the author or coauthor of 28 scientific papers, many published in this journal, and the editor of at least two proceedings of workshops on transducer design, which he organized. He supervised the thesis work of 16 doctoral students.

During several summers, he was a Visiting Scientist at the Naval Postgraduate School in Monterey to work with me and my colleagues on transducer analysis and design. I had the privilege of spending a sabbatical year in his lab at ISEN in 1989–1990. He will be sorely missed by his family, his many friends, colleagues, and students.

OSCAR B. WILSON

Jack B. C. Purcell • 1920–2000

Jack Purcell, Fellow of the Acoustical Society of America and past president of the acoustical consulting firm, Purcell+Noppe+Associates, Inc. (PNA) of Chatsworth, California, died on 18 January 2000 of complications from a stroke suffered two weeks earlier. For 32 years, Jack served as president of the consulting firm that he founded with Roger Noppe in the mid 1960's after heading the architectural acoustics group at Bolt Beranek and Newman's Los Angeles office which he helped to found a decade earlier. Up until his recent illness, he was active as a principal consultant to PNA and served as senior supervisory consultant in architectural acoustics, noise control, and in audio systems. He was actively engaged in acoustical consulting for nearly 50 years, entering the field in 1950 while taking special courses in physics and acoustics during his attendance at the Massachusetts Institute of Technology (MIT) in pursuit of his Master's Degree in Architecture, which he earned in 1952. While at MIT, he worked with the newly founded Bolt Beranek and Newman (BBN) and became teaching assistant to the late Professor Robert B. Newman in a new acoustics course being taught in the Department of Architecture. Later, he was a visiting lecturer at the University of Southern California, where he established the Architectural Acoustics program in the Department of Architecture, teaching that subject for over 8 years.

He was active in the Acoustical Society of America, was awarded Fellowship status in 1959 in recognition of his contributions to the field of architectural acoustics, and served as the first chairman of the Technical Committee on Architectural Acoustics from 1960 to 1962. Jack was a board-certified member of the Institute of Noise Control Engineering, a member of the American Association for the Advancement of Science, the Audio Engineering Society, the Society of Motion Picture and Television Engineers, the United States Institute of Theater Technology, and other civic and professional organizations.

In addition to his teaching activities, he lectured often to building industry-related organizations such as American Society for Heating, Refrigerating and Air Conditioning Engineers, and the American Institute of Architects, as well as to music education groups, and others on acoustical concerns in their respective fields. He authored numerous articles for technical journals and for architectural and building industry publications. He was responsible for the acoustical design of so many technical projects that an accurate count is difficult without a lengthy search of the consulting archives of both BBN and PNA. His background in the profession of architecture and acoustics provided him with the unique talent of being equally at home in the technical and creative aspects of the field of architectural acoustics. Mr. Purcell acquired his architectural registration in 1957 in Massachusetts. However, he chose not to pursue a professional practice in architecture, but rather to concentrate on the application of acoustical technologies as they might visually enhance and acoustically contribute to the ''products'' of architecture. He served as a peer review consultant on the Walt Disney Concert Hall now in construction and the recent reconstruction of the renowned Santa Fe Opera, where he provided ongoing consulting in connection with the growth and development of that venue since 1956. His long-term consulting relationship with the Santa Fe Opera earned him a coveted 1999 Merit Award from the US Institute of Theater Technology. Under Jack's direction, the firm of Purcell+Noppe+Associates, Inc. provided comprehensive acoustical services on countless assignments including the Grand Ole Opry, the Universal Amphitheater, Universal Studios in Florida and California, more than 30 large-format OMNIMAX and IMAX theaters, performing arts centers at Pierce College, Whittier College, Loyola Marymount University, Occidental College, Pepperdine University, and also for the Folk Arts Theatre in Manila, the Philippines International Conference Center in Manila, among numerous others in the US and abroad. During his 18 years with Bolt Beranek and Newman, he consulted on the Chapel and Kresge Auditorium at MIT, both designed by the internationally prestigious architectural firm, Eero Saarinen Associates, the Air Force Academy Chapel in Colorado Springs, the Bio-Acoustics Research Facility at Wright-Patterson Air Force Base in Dayton, Ohio, and on countless jet engine test cells that were being built throughout the US as the ''jet age'' dawned in the late forties and early fifties. Jack's great talent and strength, like that of Bob Newman and his other early colleagues at BBN, was his ability to easily communicate acoustical concepts to architects and engineers and persuade them to design buildings with good acoustical environments.

Bernard Charles Purcell was born on 13 November 1920 in Washington, DC. From his earliest years, he was known as ''Jack,'' a name that he legally adopted during his early professional years, so he could sign his

 0001-4966/2000/108(2)/468/2/$17.00

reports ''officially'' as Jack B. C. Purcell, which he did with great relish and flourish. Jack attended Columbus University, studying business law and accounting from 1940 to 1942, prior to joining the US Navy. While in the Navy, he met and married the love of his life, Jackie, on 26 April 1943. Jack served in the Pacific Theater of Operations during World War II as an aircraft control tower operator and aircraft carrier fuel systems specialist, and was honorably discharged at the war's end in 1945. In 1946, he attended Emerson Institute majoring in physics and mathematics, transferring to the Catholic University of America in Washington, DC where he was awarded the Bachelor of Architecture degree in 1950. He entered graduate studies at the Department of Architecture and Planning at MIT in Cambridge in the fall of 1950, where his professional life-long love affair with acoustics began.

Jack is survived by his wife of 57 years, Jackie; they always took great pleasure in signing their Christmas cards and personal notes as ''Jack and Jackie.'' They had five children: Marianne Purcell, Cathy Celano, Judy Purcell, Bob Purcell, and Kelly Mumford, all living in the southern California area, and were blessed with three grandchildren. When Jack was not ''doing acoustics,'' he was tinkering with his classic automobile collection, one of which was a favorite Lincoln limousine formerly owned by Gene Autry, or with one of his many radio-controlled model airplanes.

You might say Jack Purcell, who truly loved life as well as his life-long work consulting in architectural acoustics, passed away still very much ''in the saddle'' of the profession. He will be greatly missed by his colleagues, friends, and family. Donations in his memory may be made to the Acoustical Society Foundation.

ROGER C. NOPPE AND WILLIAM J. CAVANAUGH

BOOK REVIEWS

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

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

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

Acoustics: Basic Physics, Theory and Methods

P. Filippi, D. Habault, J P. Lefebvre, and A. Bergassoli

Academic Press, 1999, 24-28 Oval Road, London NW1 7DX, UK. xiii+316 pp. Price: $69.99 ISBN: 0122561902.

This text is a compilation of lecture notes for a graduate course in Acoustics presented at the University of Aix-Marseille by four scientists from CNRS (Center National de la Recherche Scientific). Each chapter is attributed to a single author. As such, the chapters strongly reflect the research interests of each author. The principal subject is the calculation of sound propagation outdoors and in enclosures. The structure of the book follows that of lecture notes in that each chapter is in the form of a lecture on a single subject. Material is often repeated between chapters by different authors. One also feels that some explanatory material has been omitted in the transition from lecture to book.

Chapter 1 is a concise summary of the basic equations of sound propagation in fluids and elastic media. The text assumes the reader has a very good background in mechanics and fluid mechanics and presents starting equations with little accompanying discussion.

Chapter 2 presents commonly used methods of calculation for enclosures: eigenmode and eigenfrequency analysis, Green's function solutions, and separation of variables. In addition, transient solutions are treated and Sabine's formula is introduced.

The diffraction of acoustic waves and the use of the boundary integral equations are introduced in Chap. 3. The use of single- and double-layer potentials in calculations is given more weight than in similar treatments I have seen. The simple example calculations, which conclude this chapter, explicate the mathematics.

The chapter on outdoor sound propagation, Chap. 4, includes calculation methods for predicting the effect of homogeneous and inhomogeneous ground surfaces, the diffraction of sound by screens and objects, and the refraction of sound by speed of sound gradients. The emphasis is on the calculations, not on the characterization of the environment for modeling. Research results are used as examples, but the figures are only briefly described and it is sometimes difficult to interpret them without referring to the original research paper.

Chapter 5 is a compilation of approximations and analytic expansions used in acoustic calculations. Although the chapter contains much interesting material, the mathematics might have been easier to follow in context. I found myself constantly flipping backward or forward in the text. The repetition of the material mentioned above is particularly evident in this chapter.

Chapter 6 briefly describes different numerical methods used to solve boundary integral equation methods. Guidance from the author's experience is provided on which methods work best and selected research results are presented.

The physics of guided waves in atmospheric ducts, underwater channels, and rectangular ducts is presented in Chap. 7. The emphasis in this chapter is on developing an understanding of the behavior of sound fields in guided propagation. Many of the approximation methods described in the earlier chapters are revisited with a more applied viewpoint.

Chapter 8 studies the transmission and reflection of sound by thin plates. The chapter presents an in-depth treatment of waves in plates and the interaction of these waves with waves in the fluid. The chapter concludes with a presentation of research on the coupling of turbulent flow to thin plates using the boundary integral equation method.

Chapter 9 contains homework problems. The problems presented are mostly theoretical in nature and some are quite daunting.

Acoustics: Basic Physics, Theory and Methods should be retitled to accurately reflect its contents. The subject matter is strictly calculational methods for linear sound propagation. Since my research interests overlap somewhat with those of the authors, I found the text to be quite interesting and to contain much useful information. This is a necessary book for acousticians involved in the calculation of acoustic propagation in fluids. I would find it difficult to use it as a text for any class, though. The rigorous but brief mathematical treatment of acoustics in this book required further reading in the references to achieve full understanding. The book does serve well as a quick reference for possible calculation methods and as a check on the mathematical limits of various methods. In addition, the book is a good introduction to research previously published in French.

RICHARD RASPET
Department of Physics and Astronomy
University of Mississippi
University, Mississippi 38677

Fundamentals of Acoustic Signal Processing

Mikio Tohyama and Tsunehiko Koike

Academic Press, San Diego, 1998.
xiii+321 pp. Price: $68.00 hardcover. ISBN: 0126926603.

The specialty of this reviewer is acoustic signal processing. It therefore is only natural that *Fundamentals of Acoustic Signal Processing* should have attracted his attention. Upon reading the book, one finds that the authors intend to serve two constituencies simultaneously. In the Preface, they explain their rationale: According to their observations, rapid changes in digital signal processing technology have left acoustic engineers behind, and at the same time have preoccupied digital processing engineers to the extent that they have not had time to learn classical acoustic wave physics. Therefore, the authors have set out to bridge this gap.

The resulting book is rather peculiarly constructed. To quote the authors in Chapter 1, Introduction, ''More or less all the chapters are independent of each other.'' The sequence of material covered in the nine chapters is interesting. In their words, ''Chapters 2–5 and Chapter 9 deal with discrete linear system theory, and Chapter 8 describes the properties of continuous models for linear acoustic systems. Chapter 6 provides the fundamental basis of acoustic wave physics and Chapter 7 discusses the statistical model for acoustic transfer functions.'' Thus, it appears that in the first set of chapters the authors are teaching digital signal processing to acousticians, while in the second set they are teaching linear acoustics to signal processing engineers.

Let us look more closely at the selected subject matter. First, the ques

 0001-4966/2000/108(2)/470/3/$17.00

tion arises: why should the signal processing be digital? The authors answer this question in Chapter 2, Discrete Representation of Signals: ''Recent advances in high-performance computer technology have made possible more complex signal processing by computers or digital signal processors than was possible using analog circuits. In addition, digital processing can perform tasks which analog circuits could not.'' The material in Chapters 2, 3, and 5 is certainly what one might expect in a course in digital signal processing: discrete signal expressions, the sampling theorem, the *z*-transform, and the discrete Fourier transform.

Chapter 6 presents results in linear acoustics, written in continuous form. As pointed out by the authors, wave equations, eigenfrequencies and eigenfunctions, Green's functions, and Kirchhoff's integral formula are treated in this chapter.

Transfer functions are treated in Chapter 4 in discrete form, in Chapter 6 in continuous form. The inverse filter is covered in Chapter 4 in discrete form, in Chapter 8 in continuous form.

Chapter 7 is philosophically interesting. This reviewer, who ''cut his teeth'' in the field of underwater acoustics, specifically sonar, sees no mention in the book under review of additive noise, nor a related quantity that always seemed so important to the performance of a processing system: signal-to-noise ratio. A consideration of this factor naturally leads to the *matched* filter, which does not appear in the book either.

Randomness does enter, however, in terms of a random sound field model, accounting for the superposition of random reflections. This gives rise to two observations: First, despite the general title, the topics taken up in this book are those required for a particular application: room acoustics. That is all right; this reviewer has long felt that the field of acoustic signal processing is a broad one, and should be available across the broad spectrum of acoustic specialties. Second, there is, in fact, an analog in the sonar signal-processing field. When the signal is much stronger than the ambient noise, that is, the signal-to-noise ratio is very high, it often happens in practice that the dominant interference is caused by reverberation. This situation naturally leads to thoughts of *inverse* filtering, with all its problems. This subject is very well treated in Chapters 4, 8, and 9.

Looking closely at Chapter 8, we finally are treated to Figure 8.3.11, Two-channel sound image projection system. This is a block diagram showing the principal application the authors had in mind all along. It shows sound image projection filters: signal processing in action. Also appearing in this chapter is a brief treatment of source waveform recovery or dereverberation of speech.

Finally, Chapter 9, though it certainly covers important material on matrix analysis, appears to have been added at the end of the book as an afterthought.

A few comments are in order about the pedagogic decisions made by the authors. In the Preface, they have this to say: ''Mathematical equations are not numbered ... many deductive steps are described without omission, as lecturers write equations on a whiteboard in a lecture room.'' This conjures up in this reviewer's mind a terrifying vision of a lecturer standing at a white-board writing equation after equation without a word of explanation, with no sound but the seemingly endless tap-tapping of his marker on the white-board!

In the same Preface, the authors say, ''Engineers who are interested in a variety of application areas of acoustics are recommended to read *The Nature and Technologies of Acoustic Space*.''[1] This book has among its authors Professor Yoichi Ando, who had another book, entitled *Architectural Acoustics: Blending Sound Sources, Sound Fields, and Listeners*,[2] reviewed in the December 1998 issue of this Journal. From a reading of that review, written by Daniel R. Raichel, it can be seen that these two referenced books are quite similar to each other.

A quick reading of Ref. 1 leads to an interesting conclusion: It is better written than *Fundamentals of Acoustic Signal Processing*. The equations are numbered and there is a much more substantial bibliography. Curiously, the former, which is not a signal-processing book, covers the subject of time-frequency domain analysis, while the latter, which is, does not!

Fundamentals of Acoustic Signal Processing is technically correct, but, as mentioned, a bit idiosyncratic. Its title is misleading, leading one to believe that it provides a foundation for the application of signal processing to acoustics across the board. A better title might be *Fundamentals of Signal Processing for Room Acoustics*. Practitioners in that field will be well advised to purchase a copy. On the other hand, practitioners in the underwater acoustics field, for example, would be poorly served by it.

[1] M. Tohyama, H. Suzuki, and Y. Ando, *The Nature and Technology of Acoustic Space* (Academic, London, 1995).
[2] Y. Ando, *Architectural Acoustics: Blending Sound Sources, Sound Fields, and Listeners* (Springer-Verlag, New York, 1998).

JAMES F. BARTRAM
JFB Consultants
94 Kane Avenue
Middletown, Rhode Island 02842

Wave Propagation in Layered Anisotropic Media with Applications to Composites

Adnan H. Nayfeh

Elsevier Science, Amsterdam, 1995.
xiii+332 pp. Price: $163.00 (hardcover).

We have come to the time when many important applications of acoustics exploit the propagation of waves through anisotropic elastic media. In addition, it is often the case that the application requires electromechanical energy conversion, so that the waves propagate in a material that is piezoelectric as well. Characterizing composite materials with ultrasound, propagation of surface waves on lithium niobate or lithium tantalate for SAW filters, and use of electromechanical energy conversion in medical imaging transducers or ultrasonic motors are important examples of useful applications. One might extrapolate and predict that as materials and devices become increasingly diverse and inhomogeneous at smaller and smaller dimensions, the need to understand the propagation of acoustic waves in anisotropic and/or piezoelectric elastic solids will become even greater.

Fortunately for those who would like to understand and exploit acoustic wave propagation in anisotropic elastic solids, a new important reference has been published. The book, authored by Adnan H. Nayfeh, is entitled *Wave Propagation in Layered Anisotropic Media with Applications to Composites*, and was published by Elsevier in 1995.

Wave Propagation in Layered Anisotropic Media is based upon a decade-long series of articles authored by Dr. Nayfeh and his colleagues. These articles have appeared in several technical journals, including the *Journal of the Acoustical Society of America*. The book is a very convenient and concise summary of the work contained in these articles, and similar work reported by other authors other than Dr. Nayfeh.

The contributions of *Wave Propagation in Layered Anisotropic Media* are many and valuable. In my opinion, the single most valuable contribution provided by this book is the wide variety of problems that are solved in a consistent set of notation that is convenient for computation. The complexity of wave propagation in anisotropic elastic media is enormous. In monoclinic materials, there are 13 stiffness coefficients, and if the material is piezoelectric, there are even more coefficients. For materials with little symmetry, and for wave propagation in directions that do not correspond with axes of symmetry that do exist, wave modes are frustratingly coupled, and require numerical separation. In addition to the contents described below, *Wave Propagation in Layered Anisotropic Media* is well referenced; 258 are cited within the text, and 115 additional references are contained in a list at the end of the book.

The majority of the book is concerned with detailed discussion of harmonic wave propagation in planar geometry anisotropic media, culminating with the case of general layered media coupled to a surrounding fluid medium. The remainder considers piezoelectric materials, transient waves, and cylindrical geometry.

After an introduction in Chapter 1 placing the subject material in a context of previous works concerning isotropic layered media such as Brekhovskikh,[1] Viktorov,[2] and Ewing,[3] and other works on elastic waves by Achenbach[4] and Auld,[5] Chap. 2 contains a systematic discussion of material symmetries and geometrical transformations. Symmetries ranging in complexity from monoclinic to isotropic are outlined. Geometric transformations are described so that wave propagation on and off planes of symmetry can

later be analyzed. Chapter 2 provides a useful notational function for the topics discussed in the remainder of the book.

Chapter 3 discusses the family of quasi-transverse and -longitudinal bulk waves that can propagate in anisotropic elastic materials as predicted by the Christoffel equation. The definition and properties of quasi-longitudinal and -transverse waves are discussed with reference to slowness diagrams for several examples. As material symmetries reduce in complexity and wave propagation directions coincide with planes of symmetry, resulting decoupling of transverse and longitudinal waves are systematically noted and discussed. This has important consequence for computational procedures, and give opportunity for physical insight. Where decoupling is not possible, advice is given for the computational strategies that can be used to sort wave modes and generate slowness diagrams for arbitrary materials.

Interaction of harmonic waves with planar boundaries, surface wave propagation, harmonic free waves in plates, and general layered media are the subjects of Chaps. 4–11. The layer wave interactions are analyzed with the propagator method. The formulas contained in these chapters are very useful to practitioners. Quick, what is the formula that can be used to determine the Rayleigh wave speed in an orthotropic material? What is the reflection coefficient for a composite material composed of nine layers of fiber reinforced composite immersed in water? Answers to these types of questions are readily found and are presented in a manner that is easy to implement in a computational procedure.

In addition to the wide range of completed analyses, Chaps. 4–11 contain qualitative discussions of the physical behavior of acoustic waves propagating in planar geometry for anisotropic elastic solids. Decoupling due to propagation along planes of symmetry, the significant effect of fluid loading on low-frequency A0 and S0 plate modes for high fluid densities, and the interactions of finite width beams with plates are noted. The discussion of finite beam interactions with fluid-coupled plates is of interest to experimentalists and engineers. However, like most of the literature on the topic, the discussion is limited to two-dimensional beams.

A collection of other topics is the subjects of Chaps. 12–15. Free plate and surface wave propagation, and reflection and transmission from planar geometry piezoelectric materials are addressed in Chap. 12. Transient waves caused by line loads on anisotropic elastic solids are analyzed with the Cagniard–de Hoop transformation in Chap. 13. Chapter 14 discusses harmonic scattering from cylinders, and considers general properties and several examples. A thoughtful component of the book is Chap. 15, which contains models of composite materials which may be used to estimate the overall stiffness coefficients of the material given the properties of the constituents.

Some understanding of *Wave Propagation in Layered Anisotropic Media* is also obtained by considering what is not in the book. Unlike Brekhovskikh,[1] continuously variable anisotropic materials are not discussed. A comprehensive description of applications for piezoelectric materials, like that contained in Auld[5] and Kino,[6] is not included. Unlike Auld[5] and Kino,[6] the Appendix contains little property data on materials. Green's function approaches are not included. By limiting the discussion primarily to planar geometries, and employing the propagator methodology throughout, Dr. Nayfeh has kept the book to a reasonable size of 288 pages while providing useful information not found elsewhere in the bound literature.

MICHAEL J. ANDERSON
Department of Mechanical Engineering
University of Idaho
Moscow, Idaho 83844-0902
anderson@uidaho.edu

[1] L. M. Brekhovskikh, *Waves in Layered Media*, 2nd ed. (Academic, New York, 1980).
[2] I. A. Viktorov, *Rayleigh and Lamb Waves* (Plenum, New York, 1967).
[3] W. M. Ewing, W. S. Jardetsky, and, F. Press, *Elastic Waves in Layered Media* (McGraw–Hill, New York, 1957).
[4] J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, Amsterday, 1973).
[5] B. A. Auld, *Acoustic Fields and Waves in Solids, Volumes I and II*, 2nd ed. (Krieger, Malabar, FL, 1990).
[6] G. S. Kino, *Acoustic Waves: Devices, Imaging, and Analog Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1987).

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.

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
DAVID PREVES, *Songbird Medical, Inc., 5 Cedar Brook Drive, Cranbury, New Jersey 08512*

5,978,450

43.38.Fx PERSONAL DIAL TONE

Alexander I. McAllister *et al.*, assignors to Bell Atlantic Network Services, Incorporated
2 November 1999 (Class 379/88.02); filed 28 March 1997

The essence of the procedure described here is that all legitimate users of the telephone system would be assigned a ''virtual'' phone number. On attempting to place a call from any location, using any type of transmission service, the caller's voice would be analyzed to identify the caller as one who is authorized to use any of the available network services. A caller's profile recalled from a central database would specify which services are authorized for that caller. The voice patterns would be continuously monitored during the call to detect whether a different party came on the line during that time.—DLR

5,929,392

43.38.Ja SPEAKER SYSTEM

Takashi Sabato *et al.*, assignors to Matsushita Electric Industrial Company, Limited
27 July 1999 (Class 381/152); filed 29 April 1998

Japanese audio engineers are currently fascinated by the notion of locating loudspeakers in the rear of a TV set, conducting sound to the front of the cabinet through waveguides, and then adding various tuned chambers and absorbers to counteract resonances in the waveguides. The patent describes an improved ''resonance absorption section'' that is easy to fabricate and easy to tune.—GLA

5,949,893

43.38.Ja LOUDSPEAKER BOX

Heinz-Juergen Augustin, Storkau, Germany
7 September 1999 (Class 381/182); filed in Germany 24 September 1996

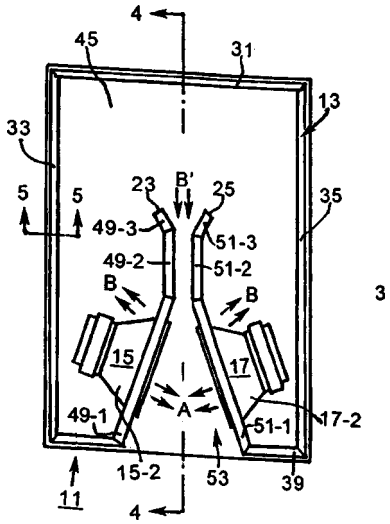
A loudspeaker box is divided into two compartments, each housing a loudspeaker. A conventional forward-facing speaker is mounted on the front panel. Sound from an upward-facing speaker on the top panel is reflected from a diffusor, as is sound from a third, downward facing speaker. The goal is to produce an enhanced stereo sound field by controlling individual speakers with respect to level or frequency response.—GLA

5,975,236

43.38.Ja SPEAKER ASSEMBLY

Shuji Yamamoto, Framingham, MA and Richard G. Plourde, Nashua, New Hampshire
2 November 1999 (Class 181/152); filed 8 January 1998

Every five years or so someone reinvents the R-J loudspeaker enclosure. In this dual-speaker version, back-pressure waves B′ combine with front pressure waves A in the flared vent opening. Although not mentioned



in the patent document, this geometry might well be used to make a powerful, inertia-canceling, subwoofer system.—GLA

5,982,907

43.38.Lc AUDIO SIGNAL WAVEFORM EMPHASIS PROCESSING DEVICE AND METHOD

Norio Akamatsu and Jun-Ichi Kakumoto, assignors to Jun-Ichi Kakumoto
9 November 1999 (Class 381/98); filed in Japan 22 October 1996

This patent makes the assertion that typical prior-art audio reproduction systems introduce significant amounts of phase distortion in the output, particularly as a result of gain alterations in the form of high- and low-frequency boost circuitry. That effect would be reduced in this device by the use of even-order integrators and differentiators to perform the desired gain modifications.—DLR

 0001-4966/2000/108(2)/473/7/$17.00

5,907,623

43.38.Md AUDIO NOISE REDUCTION SYSTEM IMPLEMENTED THROUGH DIGITAL SIGNAL PROCESSING

Laura Mercs and Paul M. Embree, assignors to Sony Corporation of Japan and Sony Pictures Entertainment, Incorporated
25 May 1999 (Class 381/94.1); filed 22 November 1995

In today's all-digital audio world it should be easy to mimic almost any kind of analog processing by using a DSP with suitable programming. Such is not the case with Dolby A noise reduction playback. The patent describes a practical method of emulating an analog Dolby A decoder.—GLA

5,946,400

43.38.Vk THREE-DIMENSIONAL SOUND PROCESSING SYSTEM

Naoshi Matsuo, assignor to Fujitsu Limited
31 August 1999 (Class 381/17); filed in Japan 29 August 1996

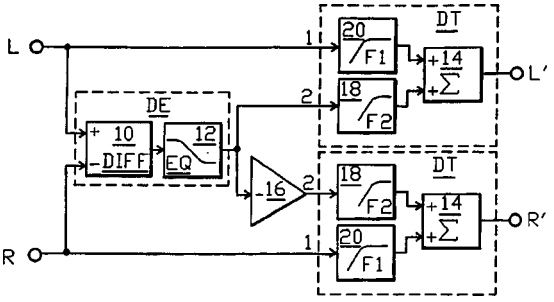
The invention is mainly concerned with accurate localization of sound images in a virtual reality environment. It is a sophisticated combination of memorized filter coefficients, characteristics of the original sound field, and inverse characteristics of the reproduced sound field, plus algorithms for calculating distance and motion.—GLA

5,970,153

43.38.Vk STEREO SPATIAL ENHANCEMENT SYSTEM

Michael L. Petroff, assignor to Harman Motive, Incorporated
19 October 1999 (Class 381/17); filed 16 May 1997

The primary object of the invention is ''...a stereophonic signal processing system, to be interposed in a dual channel stereo signal path, that derives a pair of oppositely polarized modified difference signals and combines them with the stereo signals in a manner to provide spatial enhancement in a small stereo sound system having limited physical separation between two small stereo loudspeakers.'' There are literally dozens of prior patents dealing with sound field expansion in automobile stereo systems, yet the inventor has come up with a new, simple, novel approach.—GLA



5,974,153

43.38.Vk METHOD AND SYSTEM FOR SOUND EXPANSION

Terry K. Cashion *et al.*, assignors to QSound Labs, Incorporated
26 October 1999 (Class 381/11); filed 19 May 1997

A combination of delay and filtered cross-feeding is used to create enhanced spatial effects from mono or stereo program material. ''The system operates to preserve full tone range at both speakers and does not split the sound from a particular instrument between both speakers.''—GLA

5,999,630

43.38.Vk SOUND IMAGE AND SOUND FIELD CONTROLLING DEVICE

Masayuki Iwamatsu, assignor to Yamaha Corporation
7 December 1999 (Class 381/17); filed in Japan 15 November 1994

Every three or four months, Yamaha comes up with yet another way to control the sound field produced by a stereo reproducing system. In the case at hand, 34 earlier patents are cited and a good percentage of them were issued to Yamaha. The circuitry includes delays, crosstalk cancellation, phase processing, and reflected sound simulation. The main object of the invention is to enlarge the sound field and create a feeling of spaciousness without adversely affecting sound image localization.—GLA

6,021,207

43.66.Ts WIRELESS OPEN EAR CANAL EARPIECE

Steven Puthuff *et al.*, assignors to ReSound Corporation
1 February 2000 (Class 381/330); filed 3 April 1997

An earpiece is described that can be used as a hearing aid/communication device. It enables both those with normal hearing and hearing loss to communicate via wireless link to a cellular telephone or communications node. The earpiece allows the ear canal to remain partially open and comprises a sound pickup tube that connects to a microphone that may be packaged in a sound processor case. The processed and amplified microphone signals are converted to digital form and transmitted by a transceiver via wireless link to a remote unit. Signals transmitted back from the remote unit are received by the transceiver and are sent with the microphone signals after suitable amplification for the hearing requirements of the wearer to a speaker that may be packaged in the sound processor case. The speaker output goes through a tubing down to the earpiece in the ear canal.—DAP

6,022,311

43.66.Ts APPARATUS AND METHOD FOR A CUSTOM SOFT-SOLID HEARING AID

Roger P. Juneau *et al.*, assignors to General Hearing Instrument, Incorporated
8 February 2000 (Class 600/25); filed 25 May 1998

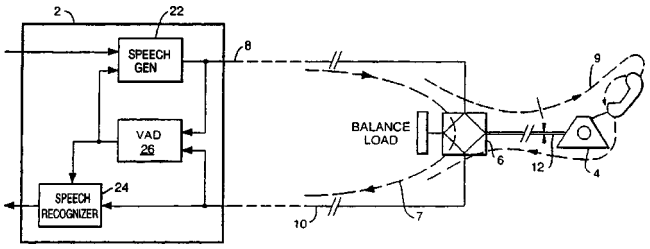
A combination soft/solid earshell is said to eliminate the need for earmold technician modification of ear impressions in order to achieve a good acoustic seal and comfort for a hearing aid wearer. The outer layer, typically 35 durometer shore A hardness, provides a bacteria barrier, resistance to tearing and the appropriate stiffness for insertion into ear canals of various shapes. The soft/solid center, molded to the shape of the ear canal, is typically 20 durometer shore A and accommodates changes in the external ear canal caused by jaw movement while achieving sufficient rigidity to house hearing aid components without compromising their reliability. For custom in-the-ear hearing aids, both the outer layer and the center adhere to a plastic faceplate.—DAP

5,978,763

43.72.Ar VOICE ACTIVITY DETECTION USING ECHO RETURN LOSS TO ADAPT THE DETECTION THRESHOLD

James A. Bridges, assignor to British Telecommunications public limited company
2 November 1999 (Class 704/233); filed in European Patent Office 15 February 1995

Designed for use in an automated phone system, this voice activity detector **26** operates by comparing incoming speech power levels to a threshold. The threshold is adjusted if speech is currently active on the outgoing line. This allows the far-end talker to interrupt an automated



prompt. Error levels are set such that an interrupting speech signal may be missed rather than to erroneously stop the outgoing speech when an interrupt condition was falsely detected.—DLR

5,978,824

43.72.Dv NOISE CANCELER

Shigeji Ikeda, assignor to NEC Corporation
2 November 1999 (Class 708/322); filed in Japan 29 January 1997

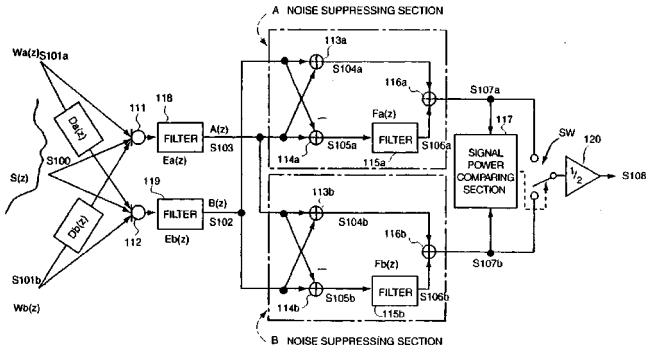
This is a two-microphone noise canceller which adjusts the gain of cross-subtracted mic signals to produce the clean speech signal. Gain adjustments are computed on the basis of speech and noise power to overall signal power ratios in both microphone channels. The tracking rate (gain step size) is also adjusted according to the power ratios.—DLR

5,982,906

43.72.Dv NOISE SUPPRESSING TRANSMITTER AND NOISE SUPPRESSING METHOD

Yoshihiro Ono, assignor to NEC Corporation
9 November 1999 (Class 381/94.2); filed in Japan 22 November 1996

The patent describes a multi-microphone noise reduction system intended for use in a small, hand-held device, such as a cellular phone. The processing is symmetrical for the mics of a two-mic version, in effect, resulting in a type of beamformer system. The description states that a phase



difference may be expected for unwanted noise at the microphones. A three-mic version uses the third mic signal in a different manner, treating it more like a traditional reference signal.—DLR

5,978,759

43.72.Ew APPARATUS FOR EXPANDING NARROWBAND SPEECH TO WIDEBAND SPEECH BY CODEBOOK CORRESPONDENCE OF LINEAR MAPPING FUNCTIONS

Mineo Tsushima *et al.*, assignors to Matsushita Electric Industrial Company, Limited
2 November 1999 (Class 704/223); filed in Japan 13 March 1995

This patent describes a method of expanding the bandwidth of a speech signal originally transmitted over a narrowband channel so as to provide higher quality playback in a wideband system. The spectral and residual components are extracted by linear prediction and separately passed to wideband converters. Several versions of the spectral converter are described. These include a linear mapping, implemented as a conversion matrix, a codebook mapping, in which codes from a narrowband analysis are used to look up spectral envelopes in a corresponding wideband codebook, and a neural network, which selects an output wideband spectrum based on the input spectral coefficients. A nonlinear converter expands the residual signal based on gain or on the multipulse pattern. A wideband speech signal is then synthesized and filtered and the extended frequency portions are added to the original narrowband signal.—DLR

5,978,761

43.72.Gy METHOD AND ARRANGEMENT FOR PRODUCING COMFORT NOISE IN A LINEAR PREDICTIVE SPEECH DECODER

Ingemar Johansson, assignor to Telefonaktiebolaget LM Ericsson
2 November 1999 (Class 704/226); filed in Sweden 13 September 1996

Low bitrate speech coding techniques typically include a voice presence detector and reduce the coding rate to a minimum during nonspeech intervals. This can result in a variety of unpleasant sound effects being generated as artifacts of the reduced coding rate. This patent describes a method by which uniform, appropriate background sounds are substituted at the receiver to replace the transmitted nonspeech frames. The receiver background sound is continually updated to represent a composite of the background sounds at the transmitter, excluding breath sounds, clicks, pops, and the like.—DLR

5,982,817

43.72.Gy TRANSMISSION SYSTEM UTILIZING DIFFERENT CODING PRINCIPLES

Friedhelm Wuppermann, assignor to U.S. Philips Corporation
9 November 1999 (Class 375/244); filed in European Patent Office 6 October 1994

The vocoder system described in this patent adds a small additional overhead to an existing vocoder, such as a G.728 Code Excited LP vocoder, increasing the speech quality with a minimal change in bitrate or computational complexity. In effect, the device operates a G.728, or other, coder and then computes the overall error signal, by comparing the coder output to the original speech. This resulting high-level residual is then itself coded for transmission using a filterbank form of spectrum coding.—DLR

5,983,172

43.72.Gy METHOD FOR CODING/DECODING, CODING/DECODING DEVICE, AND VIDEOCONFERENCING APPARATUS USING SUCH DEVICE

Makoto Takashima and Yoshiaki Asakawa, assignors to Hitachi, Limited
9 November 1999 (Class 704/203); filed in Japan 30 November 1995

This patent describes an improvement to the known methods of transform coding such as may be used to code the speech signals in a video conferencing system. Transform coding typically applies a fixed scaling or gain to normalize all frequency coefficients prior to band quantization. In this patent, the amount of normalization gain is tuned so as to minimize coding errors across the spectrum.—DLR

5,983,173

43.72.Gy ENVELOPE-INVARIANT SPEECH CODING BASED ON SINUSOIDAL ANALYSIS OF LPC RESIDUALS AND WITH PITCH CONVERSION OF VOICED SPEECH

Akira Inoue *et al.*, assignors to Sony Corporation
9 November 1999 (Class 704/219); filed in Japan 19 November 1996

This high-quality speech vocoder uses a combination of sinusoidal analysis and linear prediction coding to perform the vocoding while, at the same time, incorporating a pitch modification capability. The sinusoidal analysis is typically harmonic coding or multiband excitation. The linear prediction coder uses a codebook for compression of the spectral coefficient vectors.—DLR

5,987,405

43.72.Gy SPEECH COMPRESSION BY SPEECH RECOGNITION

David Frederick Bantz and Robert Joseph Zavrel, Jr., assignors to International Business Machines Corporation
16 November 1999 (Class 704/220); filed 24 June 1997

The idea of a phonetic vocoder has been around a long time, but the device has never achieved usable quality levels. This patent represents a new method for improving the results of such a vocoder. The simple idea, widely used in linear predictive vocoders, is to transmit an error residual along with the phonetic code sequence. The error signal is computed after correlating a resynthesized speech signal with the original speech input. The residual is coded using a type of waveform vector coding which allows variable compression. The error signal includes speech rate errors as detected by the correlator. These are corrected at the receiver.—DLR

5,987,406

43.72.Gy INSTABILITY ERADICATION FOR ANALYSIS-BY-SYNTHESIS SPEECH CODECS

Tero Honkanen *et al.*, assignors to Universite de Sherbrooke
16 November 1999 (Class 704/220); filed 7 April 1997

The patent discloses a method of reducing instabilities in certain types of linear prediction vocoders widely used in cellular telephones. The problem is often seen when high-frequency sine waves such as DTMF tones occur mixed with the speech signal. Using the method, a period of instability described as a resonance is detected by its spectrum, duration, and gain characteristics. When these conditions are detected, the encoder pitch gain is reduced.—DLR

5,987,407

43.72.Gy SOFT-CLIPPING POSTPROCESSOR SCALING DECODED AUDIO SIGNAL FRAME SATURATION REGIONS TO APPROXIMATE ORIGINAL WAVEFORM SHAPE AND MAINTAIN CONTINUITY

Shuwu Wu and John Mantegna, assignors to America Online, Incorporated
16 November 1999 (Class 704/224); filed 28 October 1997

The patent describes a method of compressing audio signals which may include music as well as speech. The non-model-based encoder uses a modified DCT spectrum analysis which is similar to the MPEG method, followed by mel-scale perceptual masking and a quantization technique known as adaptive sparse vector quantization. Nonzero bands are collapsed to a populated vector, which is then classified into one of six types and vector quantized.—DLR

5,987,413

43.72.Ja ENVELOPE-INVARIANT ANALYTICAL SPEECH RESYNTHESIS USING PERIODIC SIGNALS DERIVED FROM REHARMONIZED FRAME SPECTRUM

Thierry Dutoit, Sirault, Belgium *et al.*
16 November 1999 (Class 704/267); filed 5 June 1997

This audio synthesis system, based on reconstructions from a table of periodic waveform segments, is said to be usable for music as well as for speech. The phases of the stored waveform segments are adjusted such that one stored segment may be used to reproduce a given spectral envelope with any desired fundamental frequency. The method seems to be quite similar to an approach patented by Forrest Mozer in the 1980's.—DLR

5,987,143

43.72.Kb METHOD AND APPARATUS FOR ERASING ACOUSTIC ECHO

Takatoshi Okuno *et al.*, assignors to Pioneer Electronic Corporation
16 November 1999 (Class 381/66); filed in Japan 25 March 1996

The speakerphone echo canceller described here seems to be based on a typical FIR adaptive filter, except that adaptation proceeds only when the estimated step is small. This is said to prevent the filter from attempting to track a false echo which results when both parties talk at the same time,

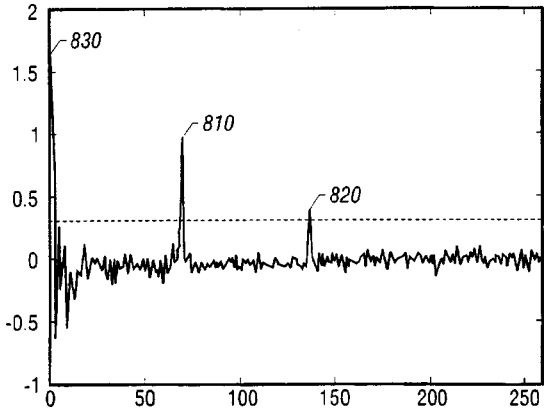
with the far-end speech being picked up by the mic from a near-end loudspeaker.—DLR

5,991,277

43.72.Kb PRIMARY TRANSMISSION SITE SWITCHING IN A MULTIPOINT VIDEOCONFERENCE ENVIRONMENT BASED ON HUMAN VOICE

Joon Maeng *et al.*, assignors to Vtel Corporation
23 November 1999 (Class 370/263); filed 20 October 1995

The patent presents a voice activity detector for use in a distributed videoconferencing system. Voice detection at the individual sites is done by a fairly typical combination of speech and noise power level detections and comparisons with various adapting thresholds. The voice detection status



results for all sites are then sent to a conference control center and processed to determine which site will be designated as the currently talking site.—DLR

5,991,385

43.72.Kb ENHANCED AUDIO TELECONFERENCING WITH SOUND FIELD EFFECT

James M. Dunn *et al.*, assignors to International Business Machines Corporation
23 November 1999 (Class 379/202); filed 16 July 1997

This videoconferencing system uses information other than the voice signals from each site to provide an enhanced conference experience. In one version, background sounds from the currently talking site are reproduced during speech silences to provide a more realistic conference room ambience. In a second version, the seating arrangement at the current talking site is used to generate a virtual seating arrangement at each of the listening sites.—DLR

5,983,177

43.72.Ne METHOD AND APPARATUS FOR OBTAINING TRANSCRIPTIONS FROM MULTIPLE TRAINING UTTERANCES

Jianxiong Wu *et al.*, assignors to Nortel Networks Corporation
9 November 1999 (Class 704/244); filed 18 December 1997

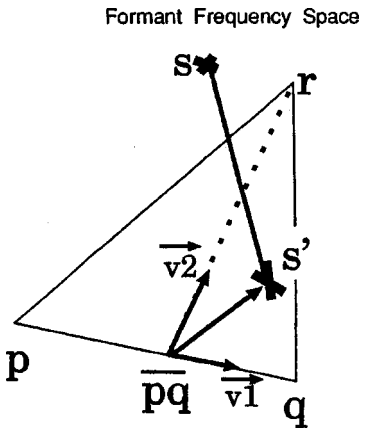
This patent describes a method for adding new words to the training data of a speech recognition system. As is usual with such systems, the user is asked to provide multiple pronunciations of the new word to be added. Instead of simply storing all versions of the new item, this system combines the various phonetic sequence graphs into a single best approximation to represent the phonetic sequence of the new word.—DLR

5,983,178

43.72.Ne SPEAKER CLUSTERING APPARATUS BASED ON FEATURE QUANTITIES OF VOCAL-TRACT CONFIGURATION AND SPEECH RECOGNITION APPARATUS THEREWITH

Masaki Naito *et al.*, assignors to ATR Interpreting Telecommunications Research Laboratories
9 November 1999 (Class 704/245); filed in Japan 10 December 1997

This speaker-independent speech recognition system performs a cluster analysis of speaker characteristics based on a computed vocal tract configuration. The vocal tract features include shape parameters representative



of formant frequencies and a tract length parameter. These parameters are converted to hidden Markov model distance values for the clustering computations.—DLR

5,983,179

43.72.Ne SPEECH RECOGNITION SYSTEM WHICH TURNS ITS VOICE RESPONSE ON FOR CONFIRMATION WHEN IT HAS BEEN TURNED OFF WITHOUT CONFIRMATION

Joel M. Gould, assignor to Dragon Systems, Incorporated
9 November 1999 (Class 704/251); filed 13 November 1992

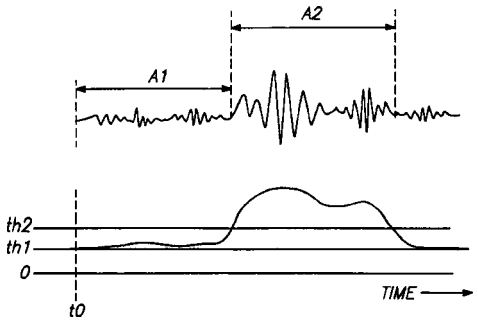
The patent describes a PC-based speech recognition system such as might be used for dictation or data entry and which includes a voice command to turn off the recognizer system. In order to avoid a clumsy sequence of corrective operations, should the ''off'' command be falsely recognized, the system includes a confirmation protocol which may be optionally included in the shutdown command sequence.—DLR

5,983,186

43.72.Ne VOICE-ACTIVATED INTERACTIVE SPEECH RECOGNITION DEVICE AND METHOD

Yasunaga Miyazawa *et al.*, assignors to Seiko Epson Corporation
9 November 1999 (Class 704/275); filed in Japan 21 August 1995

This patent discloses a speaker-independent voice-operated controller for use in a device such as a toy. It is desirable that the speech system in such a device appears to be operating at all times and not have to be ''turned on'' for use. The patent presents a protocol by which a minimal audio input circuit operates at all times. If an input sound occurs greater than a threshold



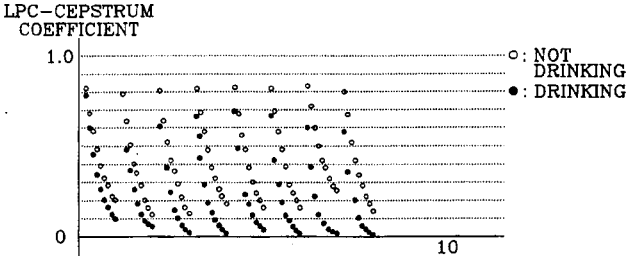
level and persists for a set period of time, then the recognition system is powered up. If no further speech inputs arrive within a set period of time, the recognition system is again powered down in order to extend the battery life.—DLR

5,983,189

43.72.Ne CONTROL DEVICE FOR CONTROLLING THE STARTING OF A VEHICLE IN RESPONSE TO A VOICE COMMAND

See-Woo Lee, assignor to Samsung Electronics Company, Limited
9 November 1999 (Class 704/275); filed in Republic of Korea 27 August 1996

This patent presents a speech analyzer which attempts to determine whether the talker is sufficiently sober as to be capable of operating a motor vehicle. The prospective driver is asked to pronounce five vowel sounds into the microphone. A test parameter, apparently a weighted average, is computed from the linear prediction cepstrum coefficients and compared with reference data for the vowels. A synthesized voice warning announces a decision that the person is judged incapable of driving and the engine starter is disabled.—DLR



5,987,408

43.72.Ne AUTOMATED DIRECTORY ASSISTANCE SYSTEM UTILIZING A HEURISTICS MODEL FOR PREDICTING THE MOST LIKELY REQUESTED NUMBER

Vishwa Gupta, assignor to Nortel Networks Corporation
16 November 1999 (Class 704/231); filed 16 December 1996

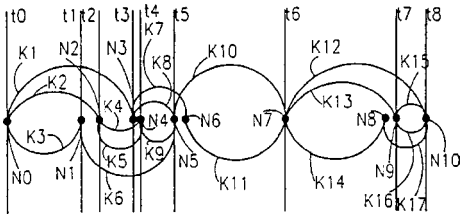
This automated telephone directory assistance system uses speech recognition to obtain an initial estimate of the name of the desired party. Additional information, primarily derived from the caller's location, is then used to establish probabilities that high-scoring recognized phrases correctly represent the desired party. Additional phrases may be requested for the final decision, the idea being to avoid presenting the caller with a long menu of choices.—DLR

5,987,409

43.72.Ne METHOD OF AND APPARATUS FOR DERIVING A PLURALITY OF SEQUENCES OF WORDS FROM A SPEECH SIGNAL

Bach-Hiep Tran *et al.*, assignors to U.S. Philips Corporation
16 November 1999 (Class 704/240); filed in Germany 27 September 1996

The patent describes a method of reevaluating the scores of multiple high-scoring phrases returned by a speech recognition system. The words from the top scoring item are used to build a new word graph as shown in the figure. This word graph is then used to build a new search tree which, because of its relative simplicity, can be exhaustively searched for an improved phrase matching score.—DLR

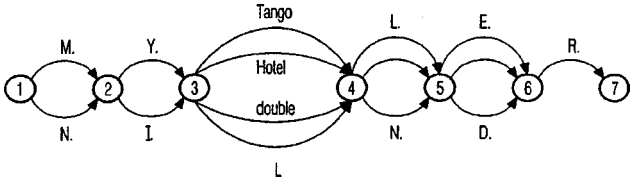


5,987,410

43.72.Ne METHOD AND DEVICE FOR RECOGNIZING SPEECH IN A SPELLING MODE INCLUDING WORD QUALIFIERS

Andreas Kellner and Frank Seide, assignors to U.S. Philips Corporation
16 November 1999 (Class 704/255); filed 10 November 1997

This speech recognition system allows the user to spell input items using a combination of typical spelling aids. For example, the name Miller might be spoken as ''M as in Mike, I, double L, E, R.'' After the recognized input is converted to a simple sequence of individual letters, a large database of proper names is consulted to determine the most probable intended spelling.—DLR



5,987,411

43.72.Ne RECOGNITION SYSTEM FOR DETERMINING WHETHER SPEECH IS CONFUSING OR INCONSISTENT

Marco Petroni and Hung S. Ma, assignors to Northern Telecom Limited
16 November 1999 (Class 704/255); filed 17 December 1997

The patent describes a strategy for adding phrases to a large speech recognition telephone dialing system. New spoken items may be the same as, or confusingly similar to, other items already in the system. In that case, the caller is prompted for additional information to resolve whether the new item is a duplicate or a new entry which sounds like another name already in the system.—DLR

5,991,364

43.72.Ne PHONETIC VOICE ACTIVATED DIALING

Alexander I. McAllister *et al.*, assignors to Bell Atlantic Network Services, Incorporated
23 November 1999 (Class 379/88.01); filed 27 March 1997

This automated recognition-based telephone dialing system stores expected phonetic sequences as used by individual subscribers in the database at an integrated service control point of an advanced intelligent network telephone system. If the spoken dialing phrase matches one of the prestored phonetic sequences, the desired party may be dialed immediately. If no match occurs, then the caller may be prompted for additional information to locate a number for the desired party.—DLR

FORUM

Forum is intended for communications that raise acoustical concerns, express acoustical viewpoints, or stimulate acoustical research and applications without necessarily including new findings. Publication will occur on a selective basis when such communications have particular relevance, importance, or interest to the acoustical community or the Society. Submit such items to an appropriate associate editor or to the Editor-in-Chief, labeled FORUM. Condensation or other editorial changes may be requested of the author.

Opinions expressed are those of the individual authors and are not necessarily endorsed by the Acoustical Society of America.

(Ed. Note: The following submission is an additional contribution to the issues raised by Robert Hickling and Clarence S. Clay in their Forum contributions in the December 1999 issue of the Journal.)

Decibels, SI units, and standards

David M. F. Chapman
Defence Research Establishment Atlantic, P.O. Box 1012, Dartmouth, Nova Scotia B2Y 3Z7, Canada

(Received 23 December 1999; accepted for publication 8 May 2000)
[S0001-4966(00)02908-8]

Provided they are used correctly, decibels provide a convenient centrigrade scale of sound levels in air that nicely matches human auditory experience: 0 dB *re* 20 μPa is about the smallest sound a human can detect and 100 dB *re* 20 μPa is about as loud a sound that we will tolerate for a short duration. The rationale for using decibels in other fields is not so well established, but decibel use has spread to applications in underwater acoustics, radio wave communications, and electronics in general. For acoustical measurements in liquids, the standard reference pressure has been chosen to be 20 times lower, at 1 μPa. Using different reference pressures in gases and in liquids is confusing, especially when considering cross-disciplinary acoustic issues that involve dissimilar acoustic media (mammal hearing in air and in water, for example). However, this dichotomy is entrenched in an ANSI standard[1] and it is unlikely to change soon. Many guidelines, regulations, and laws are written in the language of decibels, and it would take time to expunge the decibel from these, even if all the acoustical experts in the world woke up tomorrow sharing the conviction that the decibel had outlived its purpose.

The general public, most journalists, and even some scientists have difficulty with decibels. Nonexperts use the decibel as if it were a physical unit itself, rather than a logarithmic measure of the ratio of like physical quantities. Perhaps they imagine that sound comes in decibels in the same way that cheese is sold by the pound. ''Unadorned'' decibels make sense only when reporting truly dimensionless ratios such as the gain of an amplifier or the attenuation of a filter. When decibels are used to represent absolute quantities having physical dimensions, such as sound pressure or sound intensity, it is *imperative* that the reference quantities be clearly stated, otherwise decibels are meaningless. This dictum is not always observed, despite the fact that the ANSI standard demands it. No wonder we see inappropriate comparisons between transmitted source levels, received sound pressure levels, and noise spectral density levels: often the reference pressures, distances, and bandwidths are missing! Even respected scientific journals and newsmagazines are guilty of mixing ''apple'' decibels with ''orange'' decibels.[2]

In recent years, the argument over safe levels of underwater sound for marine animals has been a fertile breeding ground for decibel confusion. Correcting misconceptions due to ignorance of decibels distracts us and drains our resources. Abandoning the decibel might allow us to focus on the real issues; however, this alone will not terminate the debate. Underwater acousticians may report radiated sound powers in watts and received sound pressures in Pascals if they like, but we should take care not to extrapolate the auditory experience of humans in air to animals that have adapted to living in entirely different habitats. Claiming that an underwater sound projector is ''quieter'' for a whale than a jet aircraft is for a human on the basis of lower radiated power is a specious argument; so is the claim that the projector is ''louder'' for the whale than the jet is for a human on the basis of higher received sound pressure. Loudness is subjective, particularly so when your sample space includes different species inhabiting different environments.

The attractiveness of a logarithmic system of units for reporting auditory perception cannot be overlooked. Anyone who listens to the ASA's *Auditory Demonstrations on Compact Disk* will soon be convinced of this. Also, it is evident that fluctuations of underwater noise, sonar pings, and reverberation are more sensibly displayed on a logarithmic scale. In cases where logarithmic units seem ''natural,'' reporting quantities in linear units might overemphasize differences that are actually of modest proportion in perceptual terms. The generation, transmission, and transduction of sound can be presented handily in SI units without recourse to decibels. The audition and perception of sound by humans—and possibly other mammals—could be presented in SI units, but our experience strongly suggests a logarithmic system of units that matches our senses. Between these poles lies a gray area in which either or both systems could be used if good judgment is applied.

For scientific purposes, the arithmetic convenience of decibels has largely been superseded—but not entirely. Computer-assisted calculations are now the norm, and the choice between linear and logarithmic plotting of data is only a few mouse clicks away. There is nothing to stop the scientific community from voluntarily abandoning decibels in favor of SI units; would any AIP editor or reviewer turn down a paper on that score? Changing the standards, guidelines, regulations, etc. will take more time. In making any changes, we might introduce additional confusion while everyone adjusts to the ''new'' modes of expression; there is no doubt that nonexperts will find new ways to confuse themselves! Whether or not the decibel falls from favor, we—as scientists, authors, reviewers, and editors—are responsible for ensuring that decibel quantities are reported with appropriate reference values, according to the defined and accepted standards.

To end on a lighter note, I would like to pass on the following working definition of the decibel, collected at a recent acoustics meeting: ''When you complain that the television is too loud, your children reduce the volume by precisely one decibel.''

Many thanks to DREA colleagues Dale Ellis, Harold Merklinger, and Paul Hines for their contributions to this article.

[1] *American National Standard Preferred Reference Quantities for Acoustical Levels*, ANSI S1.8-1969, p. 8.

[2] D. M. F. Chapman and D. D. Ellis, ''The Elusive Decibel: Thoughts on Sonars and Marine Mammals,'' Can. Acoust. **26**(2), 29–31 (1996). Also available on the Internet at www.drea.dnd.ca, in PDF format. (Follow the ''documentation'' link.)

Acoustical wave propagator

J. Pan
Department of Mechanical and Materials Engineering, The University of Western Australia, Nedlands, WA 6907, Australia

J. B. Wang
Department of Physics, The University of Western Australia, Nedlands, WA 6907, Australia

(Received 15 April 1999; accepted for publication 13 April 2000)

In this paper, an explicit acoustical wave propagator (AWP) is introduced to described the time-domain evolution of acoustical waves. To implement its operation on an initial state of wave motion, the acoustical wave propagator is approximated as a Chebyshev polynomial expansion, which converges to machine accuracy. The spatial gradient in each polynomial term is evaluated by a Fourier transformation scheme. Analysis and numerical examples demonstrated that this Chebyshev–Fourier scheme is highly accurate and computational effective in predicting time-domain acoustical wave propagation and scattering. © *2000 Acoustical Society of America.* [S0001-4966(00)04607-5]

PACS numbers: 43.20.Bi, 43.20.Fn, 43.20.Gp [ANN]

INTRODUCTION

Although the analytical D'Alembert's time-domain solution of sound waves in homogeneous and boundary free space was available more than two centuries ago, the time-domain equations of acoustical waves in bounded or inhomogeneous spaces have remained unsolved in closed form even for simple geometric boundaries. Various numerical schemes for solving sets of partial differential equations have been developed in parallel in many fields with little cross referencing. For example, acousticians have been using the finite-difference time-domain (FDTD), finite-element and boundary-element time domain (FETD and BETD) simulations for room acoustic problems and for transient wave propagation and scattering problems.[1–6]

On the other hand, several time-domain schemes were also used by quantum chemists and quantum physicists to solve the time-dependent Schrödinger's equation[7–10]

$$-i\frac{\partial\psi(x,t)}{\partial t}=\hat{H}\psi(x,t), \tag{1}$$

where $\psi(x,t)$ is the probability wave function, x denotes collectively the spatial coordinates, and $\hat{H}$ is the system Hamiltonian. The formal solution to this equation is

$$\psi(x,t)=e^{-i(t-t_0)\hat{H}}\psi(x,t_0), \tag{2}$$

where $\hat{U}=e^{-i(t-t_0)\hat{H}}$ is the quantum wave propagator, which maps the wave function at time t_0 to that at time t. To implement the operation of the exponential propagator, $\hat{U}$ is often approximated to a finite polynomial expansion. A comprehensive discussion on various expansion schemes can be found in the review article written by Balakrishnan *et al.*[11] The most accurate and stable method to date is the Chebyshev polynomial expansion scheme, pioneered by quantum chemists Tal-Ezer and Koslff.[12] This scheme is regarded as a global propagator method, especially applicable to an explicit time-independent Hamiltonian operator of the time-dependent Schrödinger equation. It allows very long time steps and often a single time step completes the calculation. This scheme has been successfully employed in the field of quantum chemistry and atomic physics,[11–13] but is yet to be used in the acoustical community.

The aim of this work is to extend the wave propagator approach to investigate the propagation and scattering properties of acoustic waves. The outline of the paper is as follows. In Sec. I we derive, in analogy to the quantum propagator, an explicit acoustical wave propagator (AWP), which includes sound waves in air and elastic waves in structures. Section II discusses the implementation of the AWP scheme using matrix operator algebra and also studies the convergence property of real Chebyshev polynomials. In Sec. III, the AWP scheme is applied to study (1) flexural wave evolution in a thin beam, and (2) sound wave propagation and scattering in a one-dimensional duct.

I. THEORY OF ACOUSTICAL WAVE PROPAGATOR

Acoustical wave motion in air and in solid is described by the following partial differential equation:

$$\frac{\partial}{\partial t}\Phi(x,t)=-\hat{\mathbf{H}}\Phi(x,t), \tag{3}$$

where x denotes, collectively, the spatial coordinates, Φ is the state vector, and $\hat{\mathbf{H}}$ is a spatial derivative operator. For example, if Φ represents sound pressure $p(x,t)$ and particle velocity $v(x,t)$ in a one-dimensional duct, i.e.,

$$\Phi=\begin{bmatrix}p\\ v\end{bmatrix},$$

then

$$\hat{\mathbf{H}}=\begin{bmatrix} 0 & \rho_0 c_0^2 \dfrac{\partial}{\partial x} \\ \dfrac{1}{\rho_0}\dfrac{\partial}{\partial x} & 0 \end{bmatrix},$$

where c is the speed of sound in air and ρ_0 the air density. If

$$\Phi=\begin{bmatrix} M \\ V \end{bmatrix}$$

represents the bending moment and transverse velocity in a thin flexible beam, then

$$\hat{\mathbf{H}}=\begin{bmatrix} 0 & \rho A \gamma^2 \dfrac{\partial^2}{\partial x^2} \\ -\dfrac{1}{\rho A}\dfrac{\partial^2}{\partial x^2} & 0 \end{bmatrix},$$

where $\gamma^4=EI/\rho A$, E is the Young's modulus, ρ is the density, $I=\frac{1}{12}h^2A$ is the cross-sectional area moment of inertia, h and A are the thickness and the cross section area of the beam, respectively.

Integrating Eq. (3) with respect to time, we have

$$\Phi(x,t)=e^{-(t-t_0)\hat{\mathbf{H}}}\Phi(x,t_0). \tag{4}$$

In this way, the state vector $\Phi(x,t)$ at any time t can be evaluated by the operation of the acoustical wave propagator $e^{-\hat{\mathbf{H}}(t-t_0)}$ acting upon the initial state vector. Boundaries and spatial variation of the acoustical media can be readily included in the system operator $\hat{\mathbf{H}}$. For example, the effect of a solid boundary on one-dimensional sound waves in a duct can be described by assigning sound speed in air to the duct section and speed of longitudinal structural waves to the boundary section. Equation (4) can also be extended to describe multiple wave types in two- and three-dimensional media by adding more elements to the state vector and to the acoustical wave propagator.

II. IMPLEMENTATION OF THE AWP

In principle, the acoustical wave propagator $e^{-(t-t_0)\hat{\mathbf{H}}}$ given by Eq. (4) maps exactly the initial state vector $\Phi(x,t_0)$ to the state vector $\Phi(x,t)$ at a later time and thus provides complete information of the system under study at all times. However, in practice, this exponential propagator needs to be approximated as a finite polynomial expansion so that its operation on the wave function can be evaluated. Different approximations to the propagator along with the technique used to calculate the spatial derivatives lead to different time-evolution schemes. For example, the simplest scheme is based upon the Taylor expansion known as Euler's method. Recent work on the time-dependent Schrödinger equation indicates that the Chebyshev polynomial expansion scheme has some distinct features.[14] First, the Chebyshev polynomial expansion allows the use of very long time steps; sometimes a single time step completes the calculation. Second, the expansion coefficients of Chebyshev polynomials decay exponentially when the order of the coefficient function is sufficiently larger than its argument.

The main differences between the quantum wave propagator and the acoustical wave propagator are (1) the former is complex while the latter is real; and (2) the latter is in a matrix form. Consequently, real Chebyshev polynomials are used in the expansion of the acoustical wave propagator, and matrix operator algebra is required for its implementation. The real Chebyshev polynomials are defined in the range of $[-1,1]$. To ensure that the Chebyshev expansion is convergent, it is found that the system operator $\hat{\mathbf{H}}$ needs to be normalized by

$$\hat{\mathbf{H}}'=\frac{\hat{\mathbf{H}}}{\sqrt{\lambda_{\max}}}, \tag{5}$$

where $\lambda_{\max}$ is the maximum eigenvalue of the system operator.

Denoting $R=(t-t_0)\sqrt{\lambda_{\max}}$, the acoustical wave propagator is expanded in Chebyshev polynomials of the first kind as

$$e^{-(t-t_0)\hat{\mathbf{H}}}=e^{-R\hat{\mathbf{H}}'}=\sum_{n=0}^{\infty} a_n(R)T_n(\hat{\mathbf{H}}'), \tag{6}$$

where $a_n(R)=2I_n(R)$ except for $a_0(R)=I_0(R)$ and $I_n(R)$ is the nth-order modified Bessel function of the first kind. The zero- and first-order Chebyshev polynomials are defined as $T_0(\hat{\mathbf{H}}')=\mathbf{I}$ and $T_1(\hat{\mathbf{H}}')=\hat{\mathbf{H}}'$, and the rest can be calculated by the following recursive relations:

$$T_{n+1}(\hat{\mathbf{H}}')=2\hat{\mathbf{H}}'T_n(\hat{\mathbf{H}}')-T_{n-1}(\hat{\mathbf{H}}'). \tag{7}$$

Equation (4) then becomes

$$\Phi(x,t)=e^{-(t-t_0)\hat{\mathbf{H}}}\Phi(x,t_0)=\sum_{n=0}^{\infty} a_n(R)T_n(\hat{\mathbf{H}}')\Phi(x,t_0). \tag{8}$$

The calculation therefore boils down to a series of calculation of the normalized system operator $\hat{\mathbf{H}}'$ acting on the initial state vector, which mainly involves the evaluation of spatial derivatives.

In this work, we adopted a Fourier transformation scheme to evaluate the spatial derivatives.[15] Since

$$F\left[\frac{\partial^n}{\partial x^n}\psi(x,t)\right]=(jk)^nF[\psi(x,t)], \tag{9}$$

where $F[\,]$ represents the Fourier transform and k is the wave number as in e^{jkx}, the Fourier scheme allows the calculation of the spatial derivative by

$$\frac{\partial^n}{\partial x^n}\psi(x,t)=F^{-1}\{(jk)^nF[\psi(x,t)]\}. \tag{10}$$

Given a sampling interval of Δx, the discrete Fourier expansion of a wave packet supports a maximum wave number of $\pi/\Delta x$. Therefore, for flexural waves in a thin beam, the normalization factor $R=[\gamma^4(\pi/\Delta x)^4]^{1/2}(t-t_0)$, while for the sound pressure in a one-dimensional duct, $R=c(\pi/\Delta x)(t-t_0)$. The sampling interval is chosen to represent the highest frequency component of interest in the medium, where the shortest wavelength is supported.

III. NUMERICAL EXAMPLES AND DISCUSSION

A. Fourier method for spatial derivatives

Before implementing the Chebyshev polynomial expansion scheme to solve time-dependent acoustic wave equations, we first examine the accuracy of the Fourier transformation method for evaluating spatial derivatives. The usual practice is to calculate the spatial derivatives (e.g., a second-order derivative) by the five-point finite-difference method (FD5), in which the second derivative of a data set $(\psi_1,\ldots,\psi_m,\ldots,\psi_N)$ is given by

$$\left.\frac{d^2\psi}{dx^2}\right|_m=\frac{1}{12\Delta x^2}(-\psi_{m-2}+16\psi_{m-1}-30\psi_m+16\psi_{m+1}-\psi_{m+2}). \tag{11}$$

If higher accuracy is required, the seven-point finite-difference method (FD7) is applied,

$$\left.\frac{d^2\psi}{dx^2}\right|_m=\frac{1}{180\Delta x^2}(2\psi_{m-3}-27\psi_{m-2}+270\psi_{m-1}-490\psi_m+270\psi_{m+1}-27\psi_{m+2}+2\psi_{m+3}). \tag{12}$$

However, the finite-difference methods are based on local approximations to the derivative operator, which brings about error. Also, the convergence with decreasing grid spacing Δx is slow. To illustrate the advantages of the Fourier scheme, Eqs. (10), (11), and (12) are used to calculate the nth derivatives of a Gaussian function $G(x)=e^{-(x-64)^2}$. We observed that when the grid spacing Δx is reasonably small, the Fourier scheme is 8 and 7 orders of magnitude more accurate than the FD5 and FD7 methods, respectively. Two- and three-dimensional Gaussian functions have been used for testing. Again, the Fourier scheme was found to be around 8 orders of magnitude more accurate than the finite-difference methods.

An error analysis is given below to show how the Fourier scheme yields more accurate derivatives than the finite-difference methods. The finite-difference formulas are derived based on Taylor expansions and the associated truncation errors, in the FD5 method for example, are of the order

$$\delta^{\mathrm{FD5}}(x_m)=O\left[\frac{47}{4320}f^{(6)}(x_m)\Delta x^4\right]. \tag{13}$$

Obviously such errors are highly localized, depending upon the sixth derivative $f^{(6)}(x)$ at $x=x_m$.

On the other hand, the numerical errors in the Fourier scheme come from the discrete approximation of two continuous integrals [see Eq. (10)], which can be grouped together as a double integral in a discrete form

$$\psi^{(2)}(x_m)=\sum_{i=0}^{N-1}e^{j2\pi k_i x_m}\sum_{m'=0}^{N-1}G(x_m',k_i)\Delta x\Delta k, \tag{14}$$

where $G(x_m',k_i)=(j2\pi k_i)\psi(x_m')e^{-j2\pi k_i x_m'}$ and $\Delta k=1/(N\Delta x)$. The associated errors are estimated[16] as

$$\begin{aligned}\delta^{\mathrm{FFT}}(x_m)\approx&\mathrm{Re}\left[\frac{\Delta x^2\Delta k}{2}\sum_{i=0}^{N-1}e^{j2\pi k_i x_m}\sum_{m'=0}^{N-1}\frac{d}{dx'}G(x',k_i)\right|_{x'=x_m}\\&+\frac{\Delta x\Delta k^2}{2}\sum_{i=0}^{N-1}e^{j2\pi k_i x_m}\sum_{m'=0}^{N-1}\frac{d}{dk}G(x_m,k)\Big|_{k=k_i}\\&+\frac{\Delta x^2\Delta k^2}{4}\sum_{i=0}^{N-1}e^{j2\pi k_i x_m}\sum_{m'=0}^{N-1}\frac{d^2}{dx'\,dk}\\&\times G(x',k)\Big|_{k=k_i,x=x_m}\Bigg].\end{aligned} \tag{15}$$

In contrast to the localized errors in the finite-difference methods, the errors in the Fourier scheme are determined by a global error summation over the x' and k domains. For functions examined above, namely being Gaussian, $G(x',k)$ is symmetrical with respect to both x' and k, i.e., $G(x_0+x',k)=G^*(x_0-x',k)$ and $e^{j2\pi kx'}G(x',k)=[e^{-j2\pi kx'}\times G(x',-k)]^*$. As a result, the real parts of the three derivatives in Eq. (15) are antisymmetrical. Consequently, the positive and negative values in the summands cancel each other except for two end points at $i=0$ and $i=N/2$. This leads to very small values of $\delta^{\mathrm{FFT}}(x_m)$. Note that $G(x',k)$ is always symmetrical with respect to k, regardless of the explicit form of function $\psi(x')$, and cancellations in the summation will give rise to small errors in general.

B. Wave-packet evolution in a flexible beam

To demonstrate the application of the acoustical wave propagator developed in the previous section, we study the time evolution of a flexural wave packet in an infinite beam with initial conditions of

$$M|_{t=0}=e^{-x^2/4a^2}\left[\frac{x^2}{a^2}-2\right],\quad \dot{W}|_{t=0}=0. \tag{16}$$

The initial condition for the bending moment is equivalent to an initial bending displacement $W|_{t=0}=e^{-x^2/4a^2}$.

The exact analytical solution of the propagating wave packet in an infinite beam due to the initial displacement disturbance is available[17]

$$W=\left(1+\frac{\gamma^4t^2}{a^4}\right)^{-1/4}e^{a^2x^2/[4(a^4+\gamma^4t^2)]}\times\cos\left[\frac{\gamma^2x^2t}{4(a^4+\gamma^4t^2)}-\frac{1}{2}\tan^{-1}\frac{\gamma^2t}{a^2}\right]. \tag{17}$$

The exact expression of the bending moment can be determined by $M=-\rho A\gamma^4(\partial^2W/\partial x^2)$. This allows an assessment of the prediction accuracy of the Chebyshev–Fourier scheme developed in the previous sections. Also used for comparison is the Euler scheme, which keeps only the first two terms in the Taylor expansion of AWP, i.e.,

$$\Phi(x,t+dt)=(1-\hat{\mathbf{H}}\,dt)\Phi(x,t). \tag{18}$$

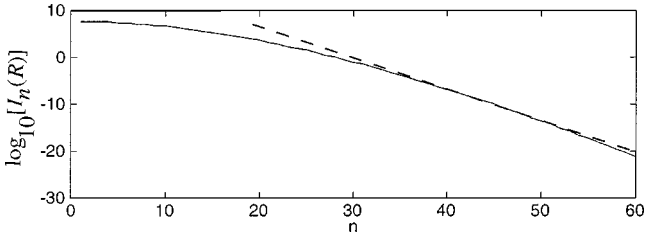


FIG. 1. The convergence properties of $I_n(R)$ for $R=20$.

The Fourier transform method is used for evaluating the spatial derivatives in the Euler scheme. Parameters used for numerical calculation are $E_B=21.6\times10^{10}\,\mathrm{N/m^2}$, $h=0.002\,\mathrm{m}$, $\rho_B=7800\,\mathrm{kg/m^3}$, $A=4\times10^{-5}\,\mathrm{m^2}$, and $a=1$.

The expansion coefficients of Chebyshev polynomials are characterized by the modified Bessel functions of the first kind. Their convergence properties are illustrated in Fig. 1. The coefficients decay quickly with increasing n. For a given R, the contribution of $I_n(R)$ to the convergence of the Chebyshev polynomials is limited by a dynamic range, which depends on the numerical precision defined for the calculation. For example, a double precision gives rise to a dynamic range of 10^{16}. As far as summation is concerned, if A is 10^{16} times larger than B, B will not contribute to $A\pm B$ because B is below the dynamic range of the significant numbers.

As shown in Fig. 1, $I_n(R)$ decreases monotonically with the increase of n. An $n_{\min}$ exists when $I_0(R)/I_{n_{\min}}(R)$ reaches the order of 10^{16}. Table I shows a list of R and $n_{\min}$ with corresponding $I_0(R)/I_{n_{\min}}(R)$ and $I_0(R)$. Small R corresponds to a small $n_{\min}$. Since the terms with $n\geqslant n_{\min}$ in the Chebyshev polynomial expansion have values below the significant digits of $I_0(R)$, they do not contribute to the final result and the sum of the polynomials converges to the order of $I_{n_{\min}}(R)$.

Figure 2 shows that increasing R is accompanied by the increased prediction error in the calculation of e^{-Rx} at $x=1$ using the Chebyshev polynomial expansion scheme. The prediction error is due to the limit of the dynamic range of the expansion coefficients as depicted in Fig. 3. Figure 3(a) shows the magnitudes of the first 60 terms of the Chebyshev polynomials of e^{-Rx} at $x=1$ and $R=30$. The decay characteristics are similar to that shown in Fig. 1. All the terms are superimposed to give the approximated value of e^{-30}. Figure 3(b) shows the approximated value from the successive differences of the first n terms for $n<n_{\min}$. These figures indicate that positive and negative terms with large magnitudes are canceled successively. The superimposed result converges as n increases. If the computer has a large dynamic

TABLE I. Dynamic range of the expansion coefficients of Chebyshev polynomials.

R	$n_{\min}$	$I_0(R)/I_{n_{\min}}(R)$	$I_0(R)$
5	25	3.3×10^{16}	2.4×10^{1}
10	32	1.4×10^{16}	2.8×10^{3}
20	43	2.7×10^{16}	4.4×10^{7}
30	50	1.8×10^{16}	7.8×10^{11}

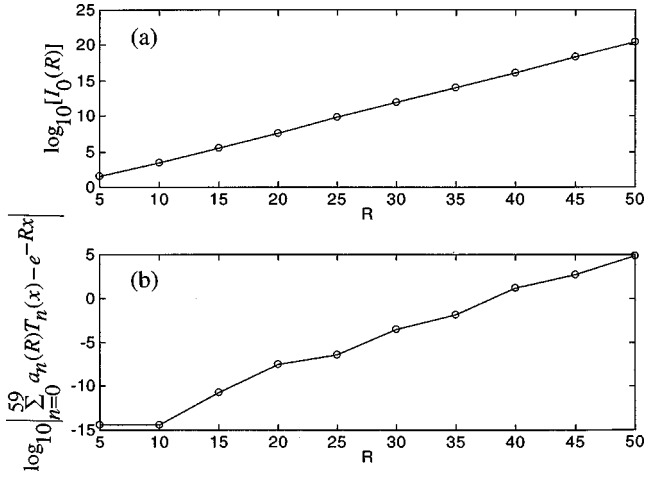


FIG. 2. (a) magnitude of the zero-order modified Bessel function of the first kind as a function of R; (b) prediction error by the Chebyshev expansion scheme for calculation of e^{-Rx} at $x=1$.

range, those terms with $n\geqslant n_{\min}$ will further contribute to the result. However, due to the limit of the dynamic range ($a_0(30)=I_0(30)T_0=7.8\times10^{11}$), the term $a_{51}(30)=2I_{51}(30)T_{51}=2.9\times10^{-4}$ is below the last significant digit of $a_0(30)$. Therefore, the sum of the polynomials converges to a magnitude of 10^{-4} as shown in Fig. 3(b).

To increase the accuracy of the approximation, we use smaller R by splitting the acoustical wave propagator in Eq. (6) into multiple components,

$$e^{-R\hat{\mathbf{H}}'}=e^{-\sum_{i=1}^{N_R}(R/N_R)\hat{\mathbf{H}}'}, \tag{19}$$

such that each propagator $e^{-(R/N_R)\hat{\mathbf{H}}'}$ preserves sufficient prediction accuracy when the Chebyshev polynomial expansion is used. Figure 4(a) shows the time evolution of a wave packet from the bending moment corresponding to an initial Gaussian displacement distribution. The dash-dotted curve is the initial bending moment and the solid curve is the moment at $t=6$ s. In this simulation the grid size N is 512, and the spatial sampling interval Δx is 0.7812 m.

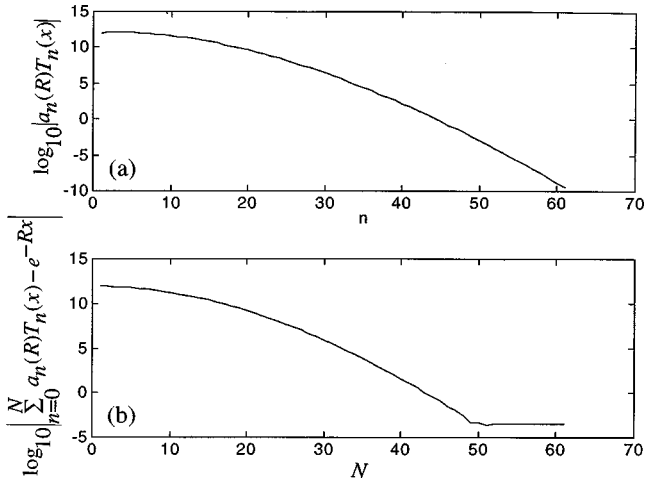


FIG. 3. (a) magnitude of the first 60 terms of the Chebyshev polynomial expansion of a function e^{-Rx} at $x=1$ and $R=30$; (b) convergence of the Chebyshev polynomial expansion.

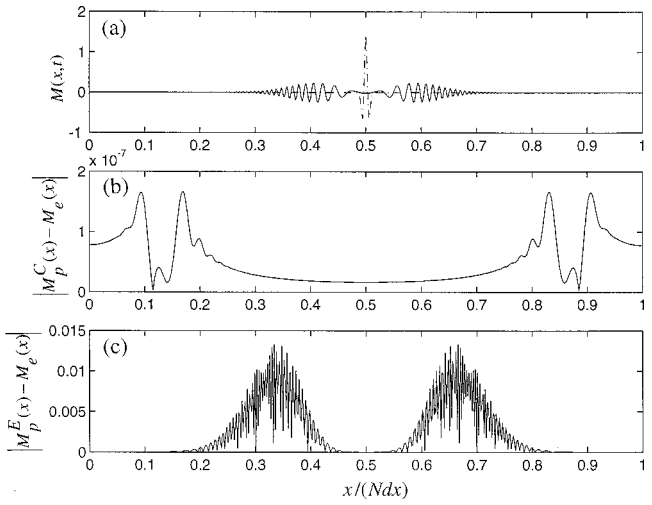


FIG. 4. Predicted wave packet in an infinite beam. (a) initial Gaussian wave packet (–·–) and distributed flexural wave displacement (———) at $t=6$ s using the Fourier and Chebyshev schemes ($N=512$, $\Delta x=0.7812$ m, $dt_C=0.2$ s); (b) absolute error between prediction by the Chebyshev scheme and exact solution; (c) absolute error between the predictions of the Euler scheme ($dt_E=0.0025$ s) and the exact solution.

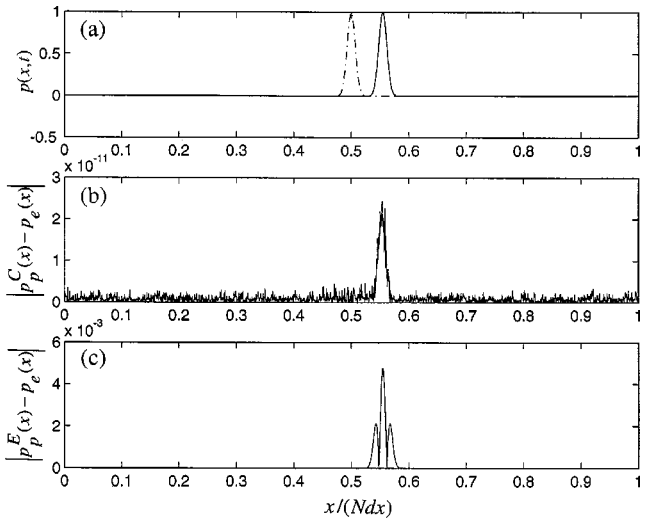


FIG. 5. (a) predicted sound-pressure wave packet propagation (———) from initial position (–·–) towards the positive x direction for 0.0032 s using Fourier and Chebyshev schemes ($dt_C=0.0001$ s; $N_R=32$; $dx=0.0098$ m, $a=0.1$; $N=2048$); (b) absolute error between prediction by the Chebyshev scheme and the exact solution; (c) absolute error between the predictions of the Euler scheme ($dt_E=0.000\,005$ s) and the exact solution.

The prediction error using Eqs. (8) and (19) is shown in Fig. 4(b). For this figure, the time step is $dt_C=0.2$ s. Sixty terms of Chebyshev polynomials were used in the expansion. Total $N_R=30$ propagators are used successively to give the displacement at $t=6$ s. The overall error is 6.635×10^{-8}, which is defined as

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_e-M_p)^2},$$

where M_e and M_p are, respectively, the exact bending moment and that predicted by either the Chebyshev or Euler scheme.

Figure 4(c) shows the absolute error by using the Euler scheme with a time step of $dt_E=0.0005$ s. For this case the overall error is 3.9×10^{-3}. Further decrease in the time step will slowly reduce the overall error. For example, the overall error was reduced to 3.5863×10^{-4} when a smaller time step of $dt_E=0.00005$ s was used. The finite-difference schemes based on the Taylor expansion of the time-evolution operator, even to fairly high orders, require extremely small time steps to ensure their convergence. Conversely, the Chebyshev scheme allows, in principle, time steps of any length. Details and discussion on the various schemes applied to quantum physics and quantum chemistry can be found, for example, in Ref. 13.

From a computational point of view, the fast Fourier transformation (FFT) algorithm is the most time-consuming part of the AWP operation. The numerical efficiency of the Chebyshev and Euler schemes are estimated below, based on the example shown in Fig. 4. For the Euler scheme, the number of calls of the FFT routine is

$$N_{\mathrm{FFT}}^{E}=\frac{4t}{dt_E}, \tag{20}$$

where t is total time of wave evolution and dt_E is the time step used in the Euler scheme. The overall error of 3.5863 $\times10^{-4}$ corresponds to $N_{\mathrm{FFT}}^{E}=4.8\times10^{5}$. On the other hand, the Chebyshev scheme involves the following number of calls of FFT:

$$N_{\mathrm{FFT}}^{C}=\frac{4t}{dt_C}N_C, \tag{21}$$

where dt_C is the time step in the Chebyshev scheme and N_C the number of Chebyshev polynomials used in the expansion. For the results shown in Fig. 4, an overall error of 6.635×10^{-8} was achieved by $N_{\mathrm{FFT}}^{C}=7.2\times10^{3}$. This comparison shows a significant increase of numerical efficiency when the Chebyshev scheme is used.

C. Sound wave packet evolution in a duct

With an initial distribution of sound pressure and air particle velocity

$$p|_{t=0}=e^{-x^2/a^2}, \tag{22}$$

$$v|_{t=0}=\frac{1}{\rho_o c_o}e^{-x^2/a^2}, \tag{23}$$

only one propagating wave packet

$$p(x,t)=e^{-[x-ct]^2/a^2} \tag{24}$$

exists in the duct.

Sound pressure at $t=0$ (dash-dotted curve) and $t=0.0032$ s (solid curves) are shown in Fig. 5(a). The Chebyshev scheme was used while the AWP is applied to an initial state vector specified by Eqs. (22) and (23). Parameters used are $a=0.2$; $c=344$ m/s; $dx=0.0098$, and $N=2048$. The prediction errors by the Chebyshev and Euler schemes are shown in Figs. 5(b) and 5(c), respectively.

When the Chebyshev scheme was used, the time interval was $t=0.0001$ s, and 60 Chebyshev polynomial functions were used for the expansion. The propagation operation was

divided into 32 substeps. With the Euler scheme, the time interval $dt_E = 0.000\,000\,5$ s. Once again, the Chebyshev scheme provides high accuracy with overall error of 2.5704×10^{-12} and $N_{\mathrm{FFT}}^{C} = 7680$. On the other hand, the prediction based on the Euler scheme has an overall error of 4.6165×10^{-4} and $N_{\mathrm{FFT}}^{E} = 25\,600$.

D. Time-domain boundary conditions

In the acoustical wave propagator, the traditional boundary of waves is described in terms of change of material or wave propagation properties of the media. For example, the solid boundary of a one-dimensional sound wave can be modeled by considering the elastic pressure (i.e., stress) wave in the solid. The longitudinal stress (positive in compression) in the boundary material can be described by Eq. (3) with

$$\hat{\mathbf{H}} = \begin{bmatrix} 0 & \rho_s c_L^2 \dfrac{\partial}{\partial x} \\ \dfrac{1}{\rho_s}\dfrac{\partial}{\partial x} & 0 \end{bmatrix}, \tag{25}$$

where $c_L = \sqrt{E/\rho_s}$. For this case, the speed of sound in the acoustical wave propagator and density are defined as

$$c(x) = \begin{cases} c_o, & x < L \\ c_L, & x \geqslant L \end{cases} \text{ and } \rho(x) = \begin{cases} \rho_o, & x < L \\ \rho_s, & x \geqslant L \end{cases}, \tag{26}$$

where $x = L$ is the location of the boundary between air and solid. The discontinuity at $x = L$ should be smoothed by the following convolution with an Gaussian function $G(x)$. For example,

$$\overline{c}(x) = \frac{1}{2\pi}\int C(k)G(k)e^{-jkx}\,dk$$
$$= F^{-1}\{F[c(x)]F[e^{-(\zeta - x)^2/b^2}]\}. \tag{27}$$

If very sharp boundaries are present, large b will be used and more grid points would be required to represent the correct derivatives.

Figure 6 shows the sound wave packet in air and at the boundary ($c_L = 1.5c_o$ and $\rho_s = 3\rho_o$). The boundary is at $L = 21$ m and the initial disturbance described by Eqs. (22) and (23) is at $x = 20$. The grid size and sampling interval are, respectively, $N = 2^{12}$ and $\Delta x = 0.0098$ m. The predicted magnitudes of the sound wave packet during and after its interaction with the boundary agree with the result from the frequency-domain analysis.

When the wave packet reaches the boundary, its magnitude is $2\rho_s c_L/(\rho_s c_L + \rho_o c_o)$ times higher than that of the incident wave due to the constructive interference of the incident and reflective waves. The reflective wave packet has a reduced magnitude by a factor of $(\rho_s c_L - \rho_o c_o)/(\rho_s c_L + \rho_o c_o)$ because of the energy loss in the boundary. The magnitude of transmitted wave to the solid is $2\rho_s c_L/(\rho_s c_L + \rho_o c_o)$ times higher than the incident wave. The width of the transmitted wave packet also becomes broader. Figure 7 illustrates the mechanism of the width increase of the transmitted wave packet. Assuming a wave packet with width Δx_A propagates towards a solid boundary as shown in Fig. 7,

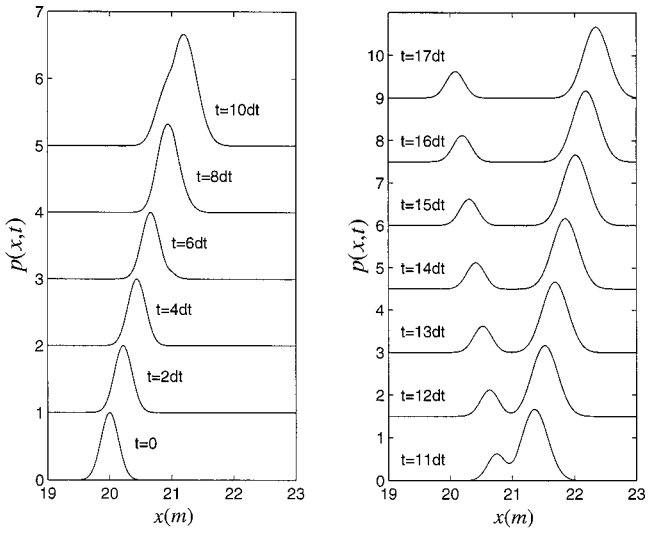


FIG. 6. Wave-packet evolution of pressure in a duct with solid boundary in $x \geqslant 21$. Each curve is offset from the one underneath by 1, and in (b) each is offset by 1.5.

the time duration for the wave packet to pass through a location $x = x_S$ in the solid boundary is between the following time range:

$$\frac{\Delta x_B}{c_o} + \frac{\Delta x_C}{c_L} \leqslant t \leqslant \frac{\Delta x_A + \Delta x_B}{c_o} + \frac{\Delta x_C}{c_L}. \tag{28}$$

Therefore, the incident wave takes time of $\Delta T = \Delta x_A/c_o$ to interact with the solid boundary. Within this time interval, the wavefront in the solid has reached

$$\Delta x_S = c_L \Delta T = \frac{c_L}{c_o}\Delta x_A, \tag{29}$$

which is the width of the transmitted wave packet in the solid. This explains that the width of the transmitted wave packet in the solid is c_L/c_o times broader than that of the incident wave packet in air.

IV. CONCLUSIONS

In this paper the acoustic wave propagator is used to describe the time-domain evolution of acoustical waves. A Chebyshev–Fourier scheme is implemented to carry out the operation of the acoustic wave propagator. This scheme is found to be highly accurate and computationally effective for the prediction of the time-domain evolution of acoustical waves. Numerical examples have been used to illustrate the implementation of AWP, with results in excellent agreement with well-known analytical solutions. The extension of the

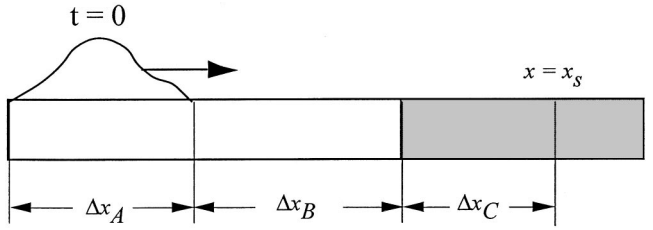


FIG. 7. Illustration of wave-packet propagation towards its solid boundary.

scheme to 2D and 3D will be straightforward.[18] The only technical difficulty involves large data arrays, which scale up exponentially with increased dimensions. However, since there is no matrix inversion required in the Chebyshev scheme, a 3D numerical grid of $100\times100\times100$ can be implemented in present desktop workstations.

[1] D. Botteldooren, ''Finite-difference time-domain simulation of low frequency room acoustic problems,'' J. Acoust. Soc. Am. **98**, 3302–3308 (1995).

[2] J. B. Schneider and O. M. Ramahi, ''The complementary operators method applied to acoustic finite-difference time-domain simulations,'' J. Acoust. Soc. Am. **104**, 686–693 (1998).

[3] F. Q. Hu, M. Y. Hussaini, and J. L. Manthey, ''Low-dissipation and low-dispersion Runge–Kutta schemes for computational acoustics,'' J. Comput. Phys. **124**, 177–191 (1996).

[4] K. Y. Fung, ''Time-domain computation of acoustics in confinements,'' Proceedings of the 5th International Congress on Sound and Vibration, 15–18 Dec. 1997, pp. 1839–1847.

[5] J. P. Coyette, ''Transient acoustics: evaluation of finite element and boundary element methods,'' Proceedings of ISMA 19, Leuven, Belgium, 1994, pp. 223–234.

[6] A. R. Atilgan and D. H. Hodges, ''Space-time mixed finite elements for rods,'' J. Sound Vib. **192**(3), 731–739 (1996).

[7] A. Asker and A. S. Cakmak, ''Explicit integration method for the time-dependent Schrödinger equation for collision problems,'' J. Chem. Phys. **68**, 2794–2798 (1978).

[8] M. S. Pindzola and D. R. Schults, ''Time-dependent close-coupling method for electron-impact ionization of hydrogen,'' Phys. Rev. A **53**, 1525–1536 (1996).

[9] M. D. Feit, J. A. Fleck, and A. Steiger, ''Solution of the Schrödinger equation by a spectral method,'' J. Comput. Phys. **47**, 412–433 (1982).

[10] E. A. McCullough and R. E. Wyatt, ''Dynamics of collinear H+H2 reaction I. Probability density and flux,'' J. Chem. Phys. **54**, 3578–3583 (1971).

[11] N. Balakrishnan, C. Kalyanraman, and N. Sathyamurthy, ''Time-dependent quantum mechanical approach to reactive scattering and related processes,'' Phys. Rep. **280**, 79–144 (1997).

[12] H. Tal-Ezer and R. Kosloff, ''An accurate and efficient scheme for propagating the time-dependent Schrödinger equation,'' J. Chem. Phys. **81**, 3967–3971 (1984).

[13] J. B. Wang and T. T. Scholz, ''Time-dependent approach to scattering by Chebyshev-polynomial expansion and the fast-Fourier-transform algorithm,'' Phys. Rev. A **57**, 3554–3559 (1998).

[14] C. Leforestier *et al.*, ''A comparison of different propagation schemes for the time domain dependent Schrödinger equation,'' J. Comput. Phys. **94**, 59–80 (1991).

[15] D. Gottlieb and S. A. Orszao, *Numerical Analysis of Spectral Methods: Theory and Applications* (SIAM, Philadephia, 1977).

[16] M. L. James, G. M. Smith, and J. C. Wolford, *Applied Numerical Methods for Digital Computation*, 3rd ed. (Harper & Row, New York, 1995).

[17] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).

[18] J. B. Wang and S. Midgley, ''Quantum waveguide theory: A direct solution to the time-dependent Schrödinger equation,'' Phys. Rev. B **60**, 13668–13675 (1999).

Approximations for modal coupling in scattered fields from orifices

J. L. Horner
Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough LE11 3TU, United Kingdom

R. Lyons
Department of Civil and Building Engineering, Loughborough University, Loughborough LE11 3TU, United Kingdom

B. A. T. Petersson
Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough LE11 3TU, United Kingdom

(Received 8 January 2000; revised 20 April 2000; accepted 24 April 2000)

Previous investigations have used Hankel transforms to establish the velocity potentials of the wave fields resulting from arbitrary angle plane wave impingement on a circular orifice in a rigid, thick wall. The scattered field from the orifice is examined, in particular the modal contributions to the amplitude of its velocity potential. For each *m,n* mode the amplitude is dependent upon the amplitude of the in-orifice waves and a driving term unique to each *m,n* mode. In establishing the amplitudes of the in-orifice waves, the effects of modal coupling are also considered. In this work these two components of the scattered wave amplitude are investigated on a modal basis and approximations given for coupling effects. These approximations are then used to calculate the scattered field and the results compared with conventional solutions that use full modal coupling.
 [S0001-4966(00)00708-6]

PACS numbers: 43.20.Fn [ANN]

INTRODUCTION

There are many engineering applications that require the transfer of fluid from one spatial domain to another through various shapes and sizes of openings. One example of this is the ventilation of living and working spaces where air is to be transported either between rooms or from the outside environment to an occupied space to replenish oxygen levels or to remove pollutants. In many cases the airflow supplying such a space has a low mean velocity. Common to this and other similar applications is the fact that some form of device will separate the two fluid domains, and this device may have several different objectives. Included in such objectives is the reduction of sound transmission or ''coloring'' of the sound to alter the frequency content. The device will determine the interface between two differing acoustic fields such as a free, outdoor sound field and a reverberant indoor sound field. In addition, the device may be situated in a short or a long duct, which will involve yet more field parameters.

In order to obtain a full solution over the whole wave-number domain for the reflected, scattered, in-orifice and transmitted fields, it is necessary to determine the effect of modal coupling.[1,2] Alternatively, approximate solutions have also been developed for the scattering which agree well with measured data. For example, Wilson and Soroka[3] gave an approximation to the diffraction of a plane wave incident normally on a circular aperture in a rigid wall of finite thickness. This was obtained by postulating rigid, massless, infinitely thin, plane pistons in each end of the aperture, whose motions simulate the movement of air particles at these pistons under acoustic excitement. Jun and Eom[4] have used a Hankel transform approach to obtain a more rigorous solution to the problem. A matrix formulation of the problem was used to estimate the velocity potentials of the four fields and this was validated against experimental data. In this study, the expressions which describe the modal energy in the scattered field are investigated in order to establish the contribution of higher-order modes to the field. The objective of the study is to determine the error incurred by taking a limited modal approach to the transmission of sound over a wide wave-number spectrum, through an orifice in a rigid wall.

For the plane wave or small Helmholtz number scenario, transmission line theory may be invoked and an orifice system described by an impedance matrix. In the practical case where the dimensions of the orifice grow large in comparison with the wavelength in the upper frequency range, the impedance approach is no longer directly applicable. This is due to the presence of higher-order modes in the duct, and some extension is necessary to accommodate for their contributions. Such an extension is enabled through an expansion of the duct field in terms of its cross-sectional modes and, by virtue of their orthogonality, component-wise modal impedances can be developed. The apparent simplicity of a one-to-one correspondence of the modal components, however, is hampered by the scattering and thence potential cross coupling between various modes. Although the formulation can be developed rigorously, the preservation of the apparent simplicity for engineering practice is highly desirable and it is thus justified to investigate the implications of this cross coupling.[5]

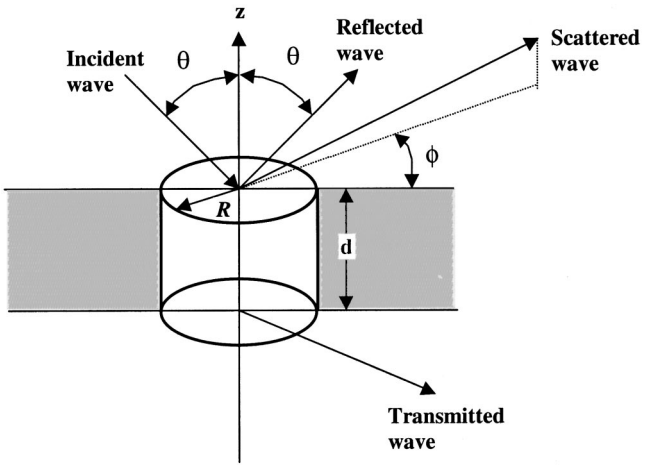


FIG. 1. Coordinate system for circular orifice.

I. DESCRIPTION OF ACOUSTIC FIELD

Consider a circular orifice of radius R in a rigid wall (Fig. 1). Following Ref. 4, the velocity potential for an incident wave of arbitrary angle, θ, may be given by

$$\Phi^i(r,\phi,z)=e^{-ikz\cos\theta}\sum_{m=-\infty}^{\infty} i^m J_m(kr\sin\theta)e^{im\phi}. \tag{1}$$

A time base of $e^{-i\omega t}$ is suppressed throughout for brevity. The resulting reflected and scattered velocity potentials are given by

$$\Phi^r(r,\phi,z)=e^{ikz\cos\theta}\sum_{m=-\infty}^{\infty} i^m J_m(kr\sin\theta)e^{im\phi},$$

$$\Phi^s(r,\phi,z)=\sum_{m=-\infty}^{\infty} e^{im\phi}\int_0^{\infty}\tilde{\Phi}^{s_m}(\zeta)J_m(\zeta r)e^{i\kappa z}\zeta\,d\zeta, \tag{2}$$

where the scattered wave number, $\kappa=\sqrt{k^2-\zeta^2}$ and $r<R$. Inside the orifice the field is given over the depth, d, by

$$\Phi^d(r,\phi,z)=\sum_{m=-\infty}^{\infty} e^{im\phi}\sum_{n=1}^{\infty}[b_n^m\sin k_z(z+d) + c_n^m\cos k_z(z+d)]J_m(k_n r), \tag{3}$$

where the wave number in the z direction $k_z=\sqrt{k^2-k_n^2}$. Assuming a rigid wall surrounding the orifice, the velocity potential gradients may be summed for the orifice area and regions outside. Applying a Hankel transform to the two expressions determined from the summed velocity potential gradients yields the following expressions:

$$\tilde{\Phi}^{s_m}(\zeta)=\frac{-i}{\kappa}\sum_{n=1}^{\infty} k_z(-c_n^m\sin k_z d+b_n^m\cos k_z d)I^m(k_n,\zeta), \tag{4}$$

in which

$$I^m(k_n,\zeta)=\begin{cases}\dfrac{R}{k_n^2-\zeta^2}[k_nJ_{m+1}(k_nR)J_m(\zeta R)-\zeta J_m(k_nR)J_{m+1}(\zeta R)] & k_n\neq\zeta,\\ \dfrac{R^2}{2}[J_m^2(k_nR)-J_{m-1}(k_nR)J_{m+1}(k_nR)] & k_n=\zeta.\end{cases} \tag{5}$$

It is of interest to consider the term $I^m(k_n,\zeta)$, Eq. (5), in more detail. Equation (4) describes the amplitude of the scattered field, which depends on the recoil wave number, ζ. As b_n^m and c_n^m are the amplitudes of the waves in the orifice, the term given by Eq. (5) may be considered to propagate the scattered field. As it depends on both m and n, each mode will give a different value of $I^m(k_n,\zeta)$. It should be noted that b_n^m and c_n^m are calculated by considering the transmitting side of the orifice and solving at all boundaries for continuity of velocity potentials and particle velocities. This will result in a set of n simultaneous equations whose solution yields values for b_n^m and c_n^m which include modal contributions from all modes considered in the calculation.[4]

II. DISCUSSION OF $I^M(k_n,\zeta)$ FUNCTION

In Figs. 2–4, $I^m(k_n,\zeta)$ is plotted against ζ/k_n, where k_n is the modal wave number of the mode of the interest. Thus, the scattered field will only propagate when ζ/k_n is greater than unity. Figure 2 shows $I^m(k_n,\zeta)$ for the m,n modes (0,2) and (0,3). It is not possible to show the plane wave mode (0,1) in this normalized form as k_n for the plane wave mode is zero. Although, for clarity, only two $(0,n)$ modes are shown, all higher $(0,n)$ modes have a maximum at ζ/k_n=unity and reduce in magnitude as n increases. The maximum value is given by Eq. (5) and the relationship between k_n and the maximum value for the $(0,n)$ modes, using simple curve fitting, can be estimated as

$$I^m(k_n,k_n)\approx 0.3032Rk_n^{-0.983}. \tag{6}$$

Figure 3 shows $I^m(k_n,\zeta)$ for the modes (1,1), (1,2), and (1,3). Here, it may be observed that all the modes have a maximum at $\zeta/k_n=1$, apart from the (1,1) mode. This pattern repeats for higher values of (m,n) with $n=1$ modes having a maximum above unity and $n>1$ modes having a maximum at unity. Again, the relationship between k_n for modes where $n>1$ and the maximum value of $I^m(k_n,k_n)$ can be described by Eq. (6). For the modes when $n=1$, the relationship is estimated as

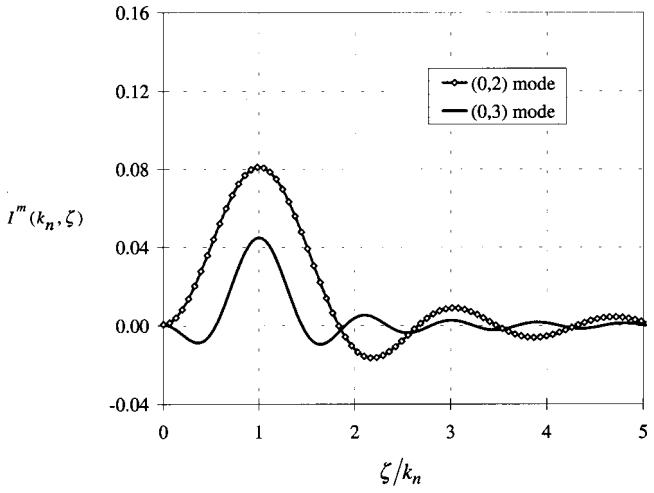


FIG. 2. Values of the function $I^m(k_n,\zeta)$ for modes (0,2) and (0,3).

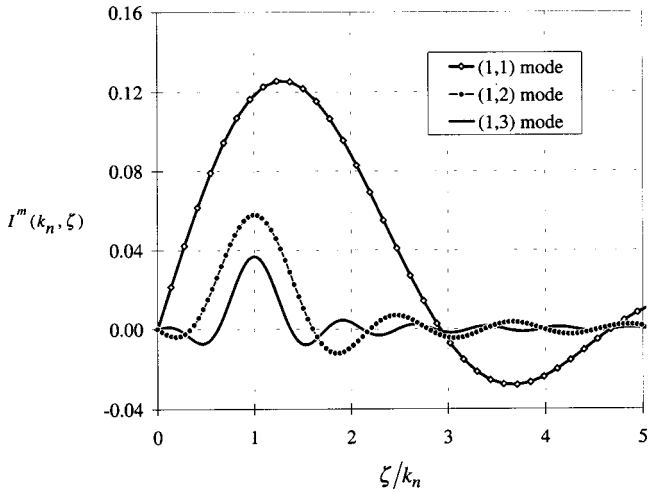


FIG. 3. Values of the function $I^m(k_n,\zeta)$ for (1,1) mode, (1,2) mode, and (1,3) mode.

$$I^m(k_n,\zeta)|_{\max} \approx 0.25Rk_n^{-1.1}. \tag{7}$$

Thus, Eq. (6) is only valid when $n>1$. As stated above, modes where $n=1$ have a peak value of $I^m(k_n,\zeta)$ at a point where $\zeta/k_n>1$. Figure 4 shows $I^m(k_n,\zeta)$ for the modes (1,1), (2,1), (3,1), (4,1), and (5,1). It shows that as m increases, the value of k_n/ζ at which $I^m(k_n,\zeta)$ is a maxium decreases. When the function $I^m(k_n,\zeta)$ is a maximum, it may be assumed that coincidence has occurred and that the wave number in the duct equals the wave number of the scattered field. Thus, if $I^m(k_n,\zeta)$ has a maximum at $\zeta/k_n>1$, it follows that the wave number k_n in the duct is reduced for those modes where $n=1$, the set of modes with radial nodal lines, but no circumferential nodal lines, as the wave number, k_n, is determined from the expression $J'_m(k_nR)=0$.

For the wave number to decrease, the radius of the duct, R, must increase for this set of modes. Any effective increase in R would imply a change in effective length of the orifice and hence an added-mass effect in the scattered field. The function $I^m(k_n,\zeta)$ is also present in the expression for the transmitted field,[4] implying that the effect would occur in both half spaces. For the set of modes, when $n\neq 1$, it could be considered that the circumferential modes have circular regions of positive and negative potential which negate the added-mass effect.

It should also be noted that the driving term, $I^m(k_n,\zeta)$ exhibits orthogonality, with the function for a particular value of k_n being zero at all other higher natural frequencies. This will result in only a single mode contributing to the peak values at cut on, in Eq. (4), the driving term for the scattered field. It implies also, at cut on, that modal coupling effects will occur due to the coupling of the ''in-orifice'' waves b_n^m and c_n^m rather than due to the n-mode summation in the expression. The determination of the amplitudes b_n^m and c_n^m and the effects of modal coupling on these terms will be described later. For modes where $n>1$, the first zero in $I^m(k_n,\zeta)$ occurs at $\zeta=k_n(m,1)$, not at the greater than 1 value observed in the peak of $I^m(k_n,\zeta)$.

In order to assess the contribution of each mode to the scattered field, the integral of $I^m(k_n,\zeta)$ had to be evaluated. Using the integrated value of $I^m(k_n,\zeta)$ versus modal wave number k_n, two straight-line relationships were observed. One line running through the $(m,1)$ data points and the other through the $(m,n>1)$ data points. Simple curve fitting provided the following expressions for the integral of $I^m(k_n,\zeta)$:

$$\int I^m(k_n,\zeta)d\zeta = ak_n^b;$$

$$\begin{aligned} &m,1 \quad a=0.6318 \quad b=-1.8819,\\ &m,n>1 \quad a=1.1939 \quad b=-2.0683. \end{aligned} \tag{8}$$

If $I^m(k_n,\zeta)$ is considered to drive the scattered field, then the square of $I^m(k_n,\zeta)$ may be considered proportional to the energy in the field. Figure 5 shows the integrated value of the square of $I^m(k_n,\zeta)$ versus the modal wave number. Again, two straight-line regions may be observed if logarithmic axes are used. These correspond to the $(m,1)$ and $(m,n>1)$ lines observed earlier. The following simple expression for the square of $I^m(k_n,\zeta)$ was determined:

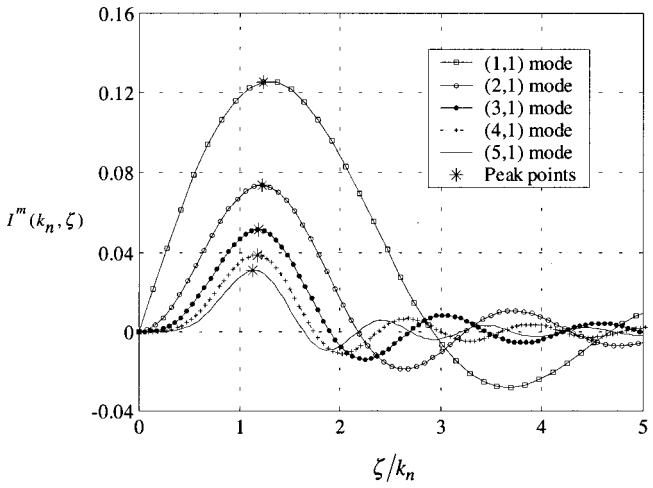


FIG. 4. Values of the function $I^m(k_n,\zeta)$ for the modes (1,1), (2,1), (3,1), (4,1), and (5,1).

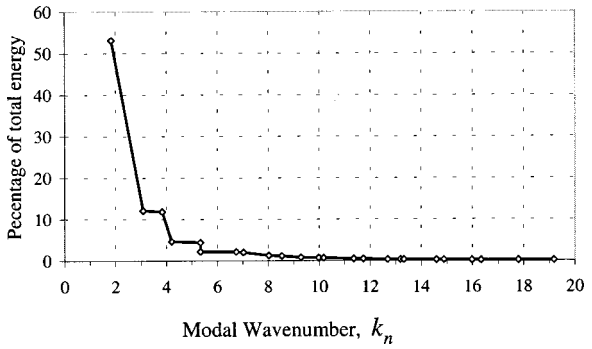


FIG. 5. Percentage of total energy in the first 24 modes from integrated value of the square of $I^m(k_n,\zeta)$, indicating energy in the field.

$$\int (I^m(k_n,\zeta))^2 d\zeta = a k_n^b;$$

$$\begin{aligned} m,1 \qquad & a=0.1380 \quad b=-2.9796, \\ m,n>1 \quad & a=0.2593 \quad b=-2.9379. \end{aligned} \tag{9}$$

Using Eqs. (8) and (9), it is possible to estimate the modal coupling contributions to any (m,n) mode of interest. Obviously, from Eq. (9), for any m set of modes the relationship between the value of the integral of the squared quantity for the first circumferential $(n=1)$ mode and any other n mode may be determined. For example, the (1,2) mode contribution is 8% of the (1,1) mode contribution and the (1,3) is 2% of the (1,1) mode.

III. DETERMINATION OF IN-ORIFICE AMPLITUDES

The velocity potential amplitudes, b_n^m and c_n^m, may be established by considering the conditions for continuity at the two faces of the orifices. Following the analysis in Ref. 4 yields the following set of equations to be solved for b_n^m and c_n^m, where n is the mode of interest and p represents any mode of the system:

$$\begin{aligned} &(c_p^m \cos k_z d + b_p^m \sin k_z d) I^m(k_p,k_p) \\ &\quad + \sum_{n=1}^{\infty} ik_z(-c_n^m \sin k_z d + b_n^m \cos k_z d) I_{np}^m \\ &\quad = 2 i^m I^m(k \sin\theta, k_p), \end{aligned} \tag{10}$$

where

$$I_{np}^m = \int_0^{\infty} \kappa^{-1} I^m(k_n,\zeta) I^m(k_p,\zeta)\zeta d\zeta, \tag{11}$$

and

$$c_p^m I^m(k_p,k_p) - \sum_{n=1}^{\infty} ik_z b_n^m I_{np}^m = 0. \tag{12}$$

It is then possible to arrange Eqs. (10) and (12) as a matrix of dimension $p_m \times p_{m'}$ where p_m is the maximum number of modes to be used in the calculation.[4] As can be seen from the above equations, those terms with a suffix p represent the coupling of a mode, n, with any mode, p. If p is set to equal n, then the equations will yield a solution with zero coupling effects. It should also be noted that the driving term $I^m(k \sin\theta, k_p)$ is identical to $I^m(k_n,\zeta)$ for any (m,n) mode if plotted against a normalized wave number. For the θ plane, the data should be normalized using the quantity $k \sin\theta/k_n$ where, again, $k \sin\theta/k_n = 1$ for a particular mode indicates cut on for that mode. Also, the maximum excitation for a mode can be predicted using Eqs. (6) and (7).

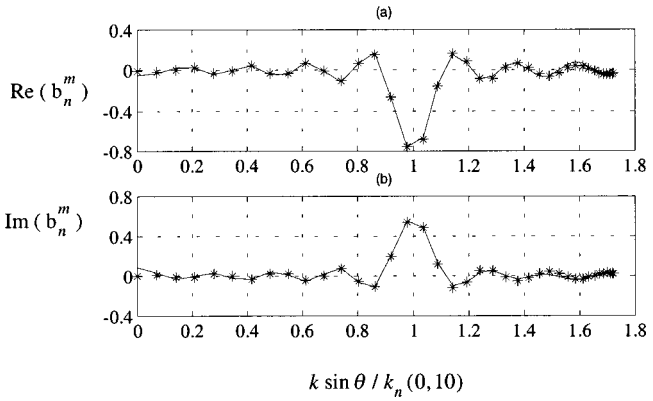


FIG. 6. (a) Real and (b) imaginary parts of b_n^m for the mode (0,10). (—— Full solution, * Single-mode solution.)

Therefore, it is possible to determine the amplitudes b_n^m and c_n^m in both the coupled and uncoupled form. If a mode is uncoupled, b_n^m and c_n^m can be calculated simply from

$$b_n^m = \frac{I^m(k_n,k_n)\gamma_n^m}{\Delta}, \quad c_n^m = \frac{ik_z I_{nn}^m \gamma_n^m}{\Delta}, \tag{13}$$

where $\Delta = (k_z^2 I_{nn}^{m\,2} + (I^m(k_n,k_n))^2)\sin k_z d + 2ik_z I_{nn}^m I^m(k_n,k_n) \times \cos k_z d$, and $\gamma_n^m = 2i^m I^m(k \sin\theta, k_n)$.

Figure 6 shows the real and imaginary parts of b_n^m for the mode (0,10), calculated using 15 n modes and a single uncoupled mode. A unit amplitude was assumed to impinge on the interface. Note that for both the uncoupled and coupled forms of solution, the integrals [Eqs. (5) and (11)] must be evaluated numerically. The difference between the two calculations is small, with zero error at cut on, $k \sin\theta/k_n = k_n(0,10)$. The zero error at cut on was expected, as the orthogonality demonstrated by the function $I^m(k_n,k_n)$ means no other mode contributes at that wave number. The most significant difference between the two calculations occurs at zero wave number, where the coupled calculation has a nonzero amplitude and the uncoupled calculation has a zero amplitude. This is due to the coupled calculation containing a contribution from the (0,1) mode, which is the only mode of a duct to have nonzero amplitude at zero wave number. Figure 7 shows the c_n^m wave for the same case. Again, the same observations may be made regarding differences at zero wave number and cut on. It should be noted that the error between the two calculations is less than that for the b_n^m wave. In comparing the differences for the real and imaginary parts for both the b_n^m and c_n^m waves, the largest differences occur in the imaginary part of the b_n^m wave. Although only a single mode is shown, the same differences were observed in other $(0,n)$ modes. For the plane wave mode, (0,1), a single mode gave an exact representation. Figure 8 shows how the difference between the two calcula-

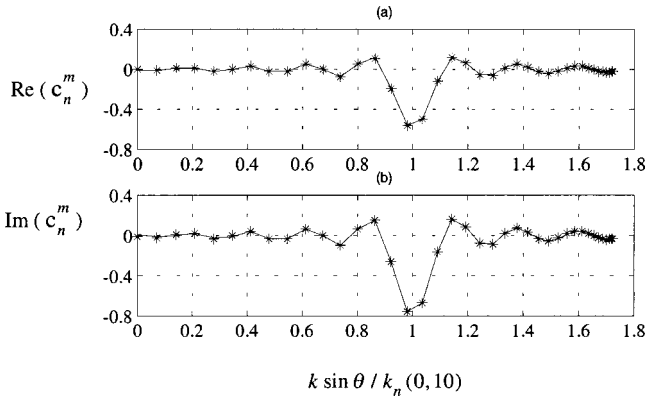


FIG. 7. (a) Real and (b) imaginary parts of c_n^m for the mode (0,10). (—— Full solution, * Single-mode solution.)

tions' changes with both normalized wave number and mode number $(0,n)$ for the imaginary part of the b_n^m wave. From the previous figures, this part of the wave amplitudes exhibited the greatest difference between the coupled and uncoupled calculations. From Fig. 8, it can again be observed that the largest error occurs at zero frequency, and the magnitude of the error drops with increasing wave number. It is also observed that the greater the value of n, the greater the region of error in the wave-number domain. It is clear from Fig. 8 that the gradient of the error increases with increasing values of n, but decreases with increasing wave number, shown normalized. As cut on of the mode is represented by $k \sin\theta/k_n=1$, then the error is greatest below the cut on wave number. The error is also cyclic in the wave-number domain indicating a phase shift after cut on. Before cut on of the mode, the two calculations are in phase. Similar observations may be made for the real part of b_n^m and both parts of c_n^m.

Consideration of higher sets of m modes indicates a similar pattern, but with no error at zero wave number as all modes other than (0,1) have zero amplitude at zero wave number. In sets of higher m modes, it may be observed that the greatest error in the wave-number domain, between the coupled and uncoupled calculations, occurs approximately in the region of $k \sin\theta/k_n=0$–0.2. For example, for the $(1,n)$ set of modes the approximation is within 5% of the coupled solution, in that region, for the set of n modes, $n=4$ to 7. This is the result of the uncoupled calculation giving near-zero amplitude in that region. After the $k \sin\theta=0.2k_n$ point, in the wave-number domain, the two calculations have approximately a 3% relative error, being exact at cut on.

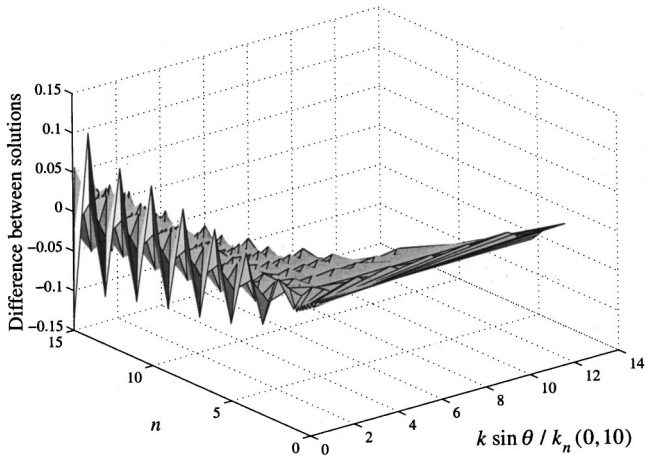


FIG. 8. Difference between coupled and single-mode solutions against normalized wave number and mode number $(0,n)$ for the imaginary part, b_n^m, of the wave ($k=50$, $R=1$ m, $d=0.1$ m, $n_{\max}=15$).

Using Eq. (9), it is possible to compare the energy in each of the $(1,n)$ modes. Summing the energy for the first 15 modes shows that 89% of the energy is in the (1,1) mode and 7% in the (1,2) mode. The fact that the approximate uncoupled prediction is within 5% of the coupled solution after the first two modes is expected, since the majority of the energy exists in those modes. Although the above discussion has concentrated on the $m=0$ and $m=1$ set of modes, similar observations were made using other values of m. These two sets of m values were examined as the $m=0$ set contained the plane wave.

IV. DISCUSSION AND CONCLUSIONS

The problem of decoupling the higher modes in an orifice in a rigid wall has been investigated with the two objectives of establishing the effect of coupling between modes and the best prediction of the amplitude at cut on for an uncoupled mode. The investigation was based on a Hankel transform approach to the problem, which allowed each set of m modes to be considered uniquely. In order to describe the scattered field due to such an orifice, it was necessary to examine the driving term for the field as well as the modal coupling of the in-duct walls. A simple expression, based on the modal wave number k_n, has been determined to establish the peak excitation at cut on. An anomaly was noted, in that the peak driving excitation for the $n=1$ set of modes occurred at a wave number greater than the cut-on wave number for the duct. It was concluded that this set of modes allowed for an added-mass effect, not present in those modes with circumferential nodes. As the integral of the square of the excitation term is proportional to the energy in each mode, it was investigated for a dependency on the modal wave number. Again, a simple relationship was found between this quantity and the modal wave number, which leads to an approximate method for comparing the energy in each mode of a system.

By observing orthogonality properties of the excitation term, $I^m(k_n,\zeta)$, the coupled contributions from other modes, at cut on, were introduced in the calculation of the amplitude of the two in-duct waves. By comparing the results of calculations for a coupled system and an uncoupled system, it was observed that at cut on the uncoupled calculation gave the same value of complex amplitude as the coupled calculation. This meant that an uncoupled calculation could be used to predict the scattered wave amplitude at cut on. If an uncoupled calculation was used, it was noted that the maximum error in wave amplitude occurred below the cut-on wave number. Above the cut-on wave number the error was smaller than below cut on, but resulted in a slight phase change between the two predictions.

In conclusion, it is proposed that a good approximation to the in-duct wave amplitudes, at cut on, may be obtained using just the uncoupled mode of interest. If the in-duct wave

amplitudes are established, it is then possible to predict the scattered field amplitude. At wave numbers other than the cut-on wave number, relative errors of less than 10% are introduced into the magnitude of the in-duct wave amplitude by using a single-mode calculation. It should be noted that at wave numbers above the cut-on wave number, a slight phase shift is also introduced into the predicted amplitude compared to the fully coupled solution. This phase shift is not observed at wave numbers below the cut-on wave number. Therefore, it may be concluded that it is possible to decouple the modes of the orifice system and yet obtain an answer that will give, at worst, 90% of the answer obtained using a fully coupled solution. As a decoupled scheme has been established, it allows such techniques as transmission line theory to be applied to ducts. Therefore, rather than calculating the response of the orifice system as a coupled system, it is possible to calculate transmission effects in a similar manner to truly orthogonal acoustic systems.

[1] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1986), Chap. 9.

[2] F. P. Mechel, *Schallabsorber, Bund I, Äussere Schallfelder; Wechselwirkungen* (Hirzel, Leipzig, 1989), Chap. 5.

[3] G. P. Wilson and W. W. Soroka, ''Approximation to the diffraction of sound by a circular aperture in a rigid wall of finite thickness,'' J. Acoust. Soc. Am. **37**, 286–297 (1965).

[4] K. H. Jun and H. J. Eom, ''Acoustic scattering from a circular aperture in a thick hard screen,'' J. Acoust. Soc. Am. **98**, 2324–2327 (1995).

[5] B. A. T. Petersson, J. L. Horner, and R. Lyons, ''An engineering approach for the acoustic characterization of aperture devices,'' Proc. Inst. Acoust. **21**(3), 131–138 (1999).

Sound propagation over layered poro-elastic ground using a finite-difference model

Hefeng Dong[a)]
Department of Telecommunications/Acoustics, Norwegian University of Science and Technology, Trondheim, Norway

Amir M. Kaynia and Christian Madshus
Norwegian Geotechnical Institute, Oslo, Norway

Jens M. Hovem
Department of Telecommunications/Acoustics, Norwegian University of Science and Technology, Trondheim, Norway

(Received 20 October 1999; revised 1 May 2000; accepted 15 May 2000)

This article presents an axisymmetric pressure-velocity finite-difference formulation (PV-FD) based on Biot's poro-elastic theory for modeling sound propagation in a homogeneous atmosphere over layered poro-elastic ground. The formulation is coded in a computer program and a simulation of actual measurements from airblast tests is carried out. The article presents typical results of simulation comprising synthetic time histories of overpressure in the atmosphere and ground vibration as well as snapshots of the response of the atmosphere–ground system at selected times. Comparisons with the measurements during airblast tests performed in Haslemoen, Norway, as well as the simulations by a frequency-wave number FFP formulation are presented to confirm the soundness of the proposed model. In particular, the generation of Mach surfaces in the ground motion, which is the result of the sound speed being greater than the Rayleigh wave velocity in the ground, is demonstrated with the help of snapshot plots.
[S0001-4966(00)04008-X]

PACS numbers: 43.20.Gp, 43.28.En, 43.28.Js [DEC]

INTRODUCTION

This article presents a new finite-difference calculation model for sound propagation in seismo-acoustic media. It also presents synthetic time histories of overpressure above ground and associated ground vibration for an instrumented airblast site in Haslemoen, Norway. The airblast tests, which were performed in June 1994, were part of an extensive field test program entitled *Blast Propagation through Forest*. One of the objectives of this program has been to use the collected data in order to verify existing noise/vibration prediction models and develop new calculation models.

Over the years, researchers have used different approaches to study the propagation of airborne acoustic pulses above a uniform flat ground. The applied methods have included complex impedance ground representation (e.g., Ref. 1), rigid-porous approximation,[2,3] viscoelastic approach (e.g., Ref. 4), and frequency-wave-number FFP (Fast Field Program) method (e.g., Ref. 5). The last model has been used by Chotiros[6] for studying sound propagation in water-saturated sand in the 10–100-kHz frequency band and by Hole *et al.*[7] for the simulation of low-frequency impulse noise and ground vibration in the Norwegian airblast tests in Haslemoen. A similar model has been used by Tooms *et al.*[8] to predict sound propagation at single frequencies and to study the effect of ground elasticity on atmospheric sound propagation.

A different computational technique based on pressure-velocity finite-difference formulation (PV-FD) is used in this study to simulate the propagation of sound over a layered poro-elastic ground. This model, which is formulated in an axisymmetric coordinate system, is an extension of the model originally developed for simulating acoustic wave propagation in borehole logging and seismics.[9] The model can handle a refracting atmosphere, although this is not considered in the simulations reported here. The advantage of this formulation is its flexibility in dealing with nonhomogeneities such as topographic features and range-dependent ground parameters. The present version of the model, however, does not consider body wave attenuation in the ground and the pore fluid.

The preceding model is used to simulate the dynamic overpressures and ground vibration measured at the airblast test site in Haslemoen, Norway. The simulations are also compared with those by the frequency-wave-number FFP model OASES.[7] In addition to presenting typical synthetic time histories of various response quantities, snapshots of the overpressure in the atmosphere and ground particle velocities are presented to highlight the overall features of sound propagation over porous media. In particular, the generation of Mach surfaces in the ground due to superseismic propagation of sound waves is demonstrated.

[a)]Permanent address: Institute of Theoretical Physics Northeast Normal University, Changchum, People's Republic of China. Electronic mail: dong@tele.ntnu.no

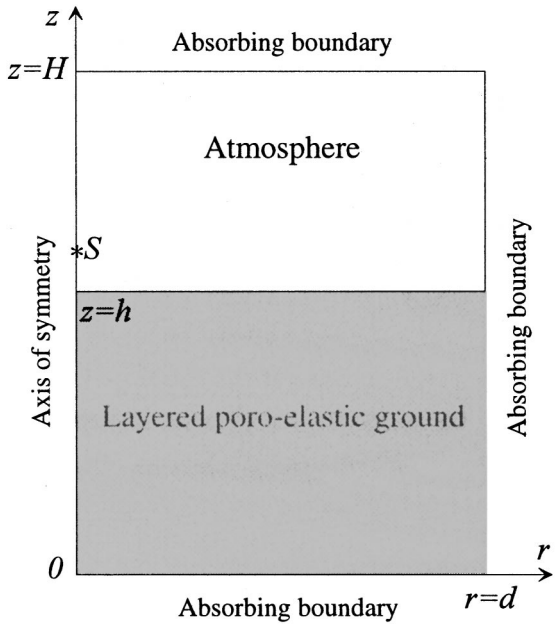


FIG. 1. Model geometry used for the finite-difference simulation. A point source is located at the z axis above ground.

I. THEORY AND METHOD

A. Theoretical formulations

Figure 1 shows a homogeneous atmospheric half-space above a layered poro-elastic ground which satisfies Biot's theory.[10,11] Under azimuth symmetry (i.e., wave field being independent of θ) the equations governing the pressure in the atmosphere, $P(r,z,t)$, and the velocities in the poro-elastic ground, $\vec{u}(r,z,t)$ and $\vec{W}_\phi(r,z,t)$, due to a point source in the atmosphere can be written as

$$\begin{aligned}\partial_{tt}P(r,z,t)/(\rho_0 V_0^2)&=\nabla\cdot(\nabla P(r,z,t)/\rho_0)\\&\quad+2g(t)\,\delta(r)\,\delta(z-z_0)/r,\\&\quad 0\leqslant r<+\infty,\ \ h<z<+\infty,\ \ t>0,\end{aligned}\tag{1}$$

$$\begin{aligned}\rho_1\partial_{tt}\vec{u}+e\rho_1\partial_{tt}\vec{W}_\phi+b\,\partial_t\vec{W}_\phi&=\vec{D}_1(\vec{u};\alpha M\phi)\\&\quad+\vec{D}_1(\vec{W}_\phi;M\phi^2),\\\rho\partial_{tt}\vec{u}+\rho_1\partial_{tt}\vec{W}_\phi&=\vec{D}_1(\vec{u};K_b-2\mu_b/3+\alpha^2M)\\&\quad+\vec{D}_1(\vec{W}_\phi;\alpha M\phi)+\vec{D}_2(\vec{u};\mu_b),\\&\quad 0\leqslant r<+\infty,\ \ -\infty<z<h,\ \ t>0,\end{aligned}\tag{2}$$

where $\delta(r)$ is the Kronecker delta, $g(t)$ defines the time variation of the source S, z_0 specifies the position of the source, and V_0 and ρ_0 are the sound speed in the atmosphere and mass density of air, respectively. The vector $\vec{u}$ denotes the velocity vector of the solid frame (in the ground), $\vec{W}=\phi\cdot(\vec{U}-\vec{u})$ is the velocity vector of the pore fluid relative to the solid frame and $\vec{W}_\phi=\vec{W}(r,z,t)/\phi$, $\vec{U}$ represents the velocity vector of the pore fluid, ϕ is porosity and e denotes tortuosity. The poro-elastic mass parameters are $\rho_1=\phi\rho_f$ and $\rho=\phi\rho_f+(1-\phi)\rho_s$ where ρ_s and ρ_f are the mass densities of the solid grains and the pore fluid, respectively. The energy dissipation in the porous medium is reflected by the parameter $b=\phi^2\eta/\kappa$, where κ denotes the permeability and η is the dynamic viscosity of the pore fluid. The stiffness attribute of the porous medium is represented by the shear modulus of the solid frame μ_b, $\alpha=1-K_b/K_s$ and $M=(\phi K_f+(\alpha-\phi)/K_s)^{-1}$, where K_b, K_s and K_f are the bulk moduli of the solid frame, solid grains and the pore fluid, respectively. Finally, the differential operators $\vec{D}_1$ and $\vec{D}_2$ are defined by

$$\vec{D}_1(\vec{f};A)=A\nabla\nabla\cdot\vec{f}+\nabla A\nabla\cdot\vec{f},\tag{3}$$

$$\begin{aligned}\vec{D}_2(\vec{f};B)&=B\nabla\nabla\cdot\vec{f}+B\nabla^2\vec{f}+\nabla B\times(\nabla\times\vec{f})\\&\quad+2(\nabla B\cdot\nabla)\vec{f}.\end{aligned}\tag{4}$$

For a homogeneous ground, the wave equations reduce to

$$\begin{aligned}\rho_f\partial_{tt}\vec{u}+\rho_c\partial_{tt}\vec{W}+\frac{\eta}{\kappa}\partial_t\vec{W}&=\alpha M\nabla(\nabla\cdot\vec{u})+M\nabla(\nabla\cdot\vec{W}),\\\rho\partial_{tt}\vec{u}+\rho_f\partial_{tt}\vec{W}&=(H-\mu_b)\nabla(\nabla\cdot\vec{u})+\mu_b\nabla^2\vec{u}+\alpha M\nabla(\nabla\cdot\vec{W}),\end{aligned}\tag{5}$$

where $\rho_c=e\rho_f/\phi$ and $H=K_b+4/3\mu_b+\alpha^2M$.

In the one-phase elastic limit, the porosity approaches zero and the coefficients in Biot's equations reduce to $K_b-2/3\mu_b=\lambda_b$ and $K_s=K_b$ where λ_b and μ_b are Lamé constants for the solid. Furthermore $\rho_1=b=0$, therefore the first part of Eq. (2) becomes redundant and the second part becomes the wave equation for elastic media:

$$\begin{aligned}\rho_s\partial_{tt}\vec{u}&=\vec{D}_1(\vec{u};K_b-2\mu_b/3)+\vec{D}_2(\vec{u};\mu_b)\\&=(\lambda_b+\mu_b)\nabla(\nabla\cdot\vec{u})+\mu_b\nabla^2\vec{u}+\nabla\lambda_b(\nabla\cdot\vec{u})\\&\quad+\nabla\mu_b\times(\nabla\times\vec{u})+2(\nabla\mu_b\cdot\nabla)\vec{u}.\end{aligned}\tag{6}$$

This means that Eq. (2) can be used for an elastic ground by setting porosity to zero and keeping the shear modulus finite.

On the ground $z=h$, the interface conditions of pressure and velocities result in

$$\begin{aligned}-\frac{1}{\rho_f}\partial_z P&=\partial_t u_z+\partial_t W_z,\\-\partial_t P&=\alpha M\nabla\cdot\vec{u}+M\nabla\cdot\vec{W},\\-\partial_t P&=H\nabla\cdot\vec{u}+\alpha M\nabla\cdot\vec{W}+2\mu_b(\partial_z u_z-\nabla\cdot\vec{u}),\\0&=\mu(\partial_z u_r+\partial_r u_z).\end{aligned}\tag{7}$$

To simulate the infinite lateral and vertical extent of the media, the far-end boundary as well as the top and bottom boundaries of the model are equipped with dashpots, denoted as absorbing boundary in Fig. 1. More details on these boundaries are given in the next section. In a layered poro-elastic ground one has to impose the conditions of stress equality (both shear and normal) and velocity compatibility at each layer interface. The equations can be readily established from those listed above.

TABLE I. Geodynamic parameters of the soil profile at Haslemoen. The values in the parentheses are used in the finite difference models.

Layer no.	H (m)	ρ_s (kg/m^3)	V_p (m/s)	V_s (m/s)
1	1.0	1500	250	130
2	1.5	1500	260	140
3	2.5	1500	280	150
4	5.0	1600	300	160
5	10.0	1700	1500 (400)	180 (200)
6	∞	1800	1500 (500)	250

B. Finite-difference formulation

The finite-difference approximations of Eq. (1) and the treatment of point source have been given by Dong *et al.*[9] and are incorporated in Appendix A for completeness. In a layered poro-elastic ground, the finite difference solution of Eq. (2) can be simplified by using the finite-difference operators defined with the following form,

$$\vec{\Omega}_j(\vec{f};A)=F_d\cdot\{\vec{D}_j(\vec{f};A)\}, \tag{8}$$

where $j=1,2$ and $F_d\cdot\{\ \}$ indicates a finite-difference representation of the variable inside the brackets. Using this notation, Eq. (2) can be expressed as

$$\begin{aligned}
&F_d\left\{\rho_1\partial_{tt}\vec{u}+e\rho_1\partial_{tt}\vec{W}_\phi+\phi^2\frac{\eta}{\kappa}\partial_t\vec{W}_\phi\right\}\\
&\quad=\vec{\Omega}_1(\vec{u};\alpha M\phi)+\vec{\Omega}_1(\vec{W}_\phi;M\phi^2)=\vec{\Phi}_1^k,\\
&F_d\{\rho\partial_{tt}\vec{u}+\rho_1\partial_{tt}\vec{W}_\phi\}\\
&\quad=\vec{\Omega}_1(\vec{u};\lambda_b+\alpha^2M)+\vec{\Omega}_1(\vec{W}_\phi;\alpha M\phi)+\vec{\Omega}_2(\vec{u};\mu_b)\\
&\quad=\vec{\Phi}_2^k,
\end{aligned} \tag{9}$$

where $\vec{\Phi}_1^k$ and $\vec{\Phi}_2^k$ contain only spatial derivatives that need to be evaluated at the current time $k\Delta t$ in the context of a time marching algorithm. Using the standard second-order explicit finite difference approximations,

$$\begin{aligned}
\frac{du}{dx}&=\frac{u(x+\Delta x)-u(x-\Delta x)}{2\Delta x},\\
\frac{d^2u}{dx^2}&=\frac{u(x+\Delta x)-2u(x)+u(x-\Delta x)}{\Delta x^2},
\end{aligned} \tag{10}$$

one can derive the finite-difference schemes for $\vec{\Phi}_1^k$ and $\vec{\Phi}_2^k$. Using the central finite difference representation in time on the left-hand side of Eq. (9) and rearranging the terms, one can write

$$\begin{aligned}
a_{11}\vec{u}^{k+1}+a_{12}\vec{W}_\phi^{k+1}&=\vec{b}_1+(\Delta t)^2\vec{\Phi}_1^k,\\
a_{21}\vec{u}^{k+1}+a_{22}\vec{W}_\phi^{k+1}&=\vec{b}_2+(\Delta t)^2\vec{\Phi}_2^k.
\end{aligned} \tag{11}$$

The vectors $\vec{u}^{k+1}$ and $\vec{W}^{k+1}$ can be obtained from

$$\vec{u}^{k+1}=\vec{\Delta}_1/\Delta,\quad \vec{W}_\phi^{k+1}=\vec{\Delta}_2/\Delta. \tag{12}$$

Expressions for a_{ij}, $\vec{b}_i$, $\vec{\Delta}_i$ $(i,j=1,2)$ and Δ are given in Appendix B. The velocity of the pore fluid $\vec{U}^{k+1}$ at the next time step can be obtained from Eq. (12):

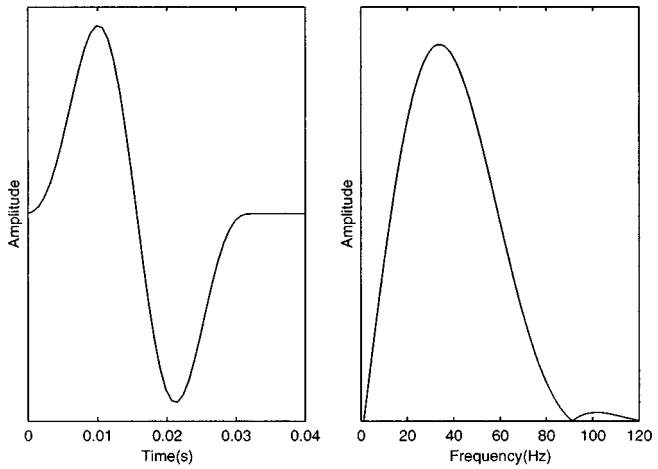


FIG. 2. Source pulse and its Fourier amplitude spectra used in present study. Central frequency is 30 Hz.

$$\vec{U}^{k+1}=\vec{u}^{k+1}+\vec{W}_\phi^{k+1}. \tag{13}$$

For the points on the ground surface, the interface equations between pressure and velocities were treated by Dong *et al.*[12] with a stable integration technique. The procedure is given in Appendix C.

For the points on the artificial boundaries ($r=d$, $z=0$ and $z=H$) the absorbing boundary conditions are used. Since the slow P-wave in a poro-elastic ground is strongly attenuated, its reflection from the artificial boundaries can be neglected. Therefore, only the reflections of the fast P-wave and S-wave from the artificial boundaries are considered. In the present study the Reynolds' absorbing boundary condition[13] was used. Numerical experiments showed that this method provides a good approximation.

The system of finite-difference equations is stable provided that the selected time step satisfies the following:[14]

$$\Delta t\leqslant\frac{\Delta x}{\sqrt{2}V_{\max}}, \tag{14}$$

where $V_{\max}$ is the maximum body wave velocity on the grid and Δx is the minimum grid size in both the r and z directions. In addition, Δx should not be taken greater than the following value,

$$\Delta x=\frac{\lambda_{\min}}{G}, \tag{15}$$

where $\lambda_{\min}=V_{\min}/f_{\up}$ is the minimum wavelength in the model, $V_{\min}$ is the minimum body wave velocity on the grid and $f_{\up}$ is the upper half-power frequency of the source;[15] this frequency is 2.5 times the central frequency of the source in this article. G is the number of grid points per minimum wavelength and should be taken at least equal to

TABLE II. Parameters of the poro-elastic layer.

Layer no.	H (m)	ρ_f (kg/m^3)	K_f (10^6 Pa)	ρ_s (kg/m^3)	K_s (10^6 Pa)	K_b (10^6 Pa)	μ_b (10^6 Pa)	ϕ
1	0.5	1.2	0.13	2700	9060	60.7	25.5	0.44

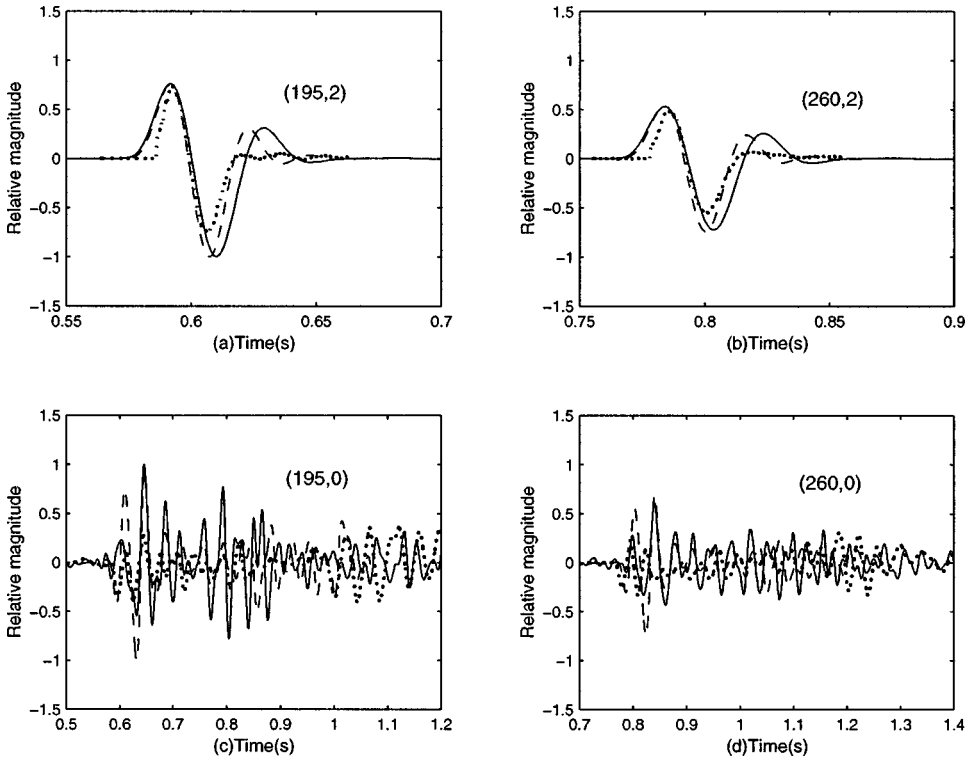


FIG. 3. Comparison between poroelastic finite-difference model *PORAC* (solid line), *OASES* (dashed line) and experimental data (dotted line). The upper two panels for pressure and lower panels for vertical velocity. (a) Overpressure at 195 m, (b) overpressure at 260 m, (c) vertical particle velocity of ground surface at 195 m, and (d) vertical particle velocity of ground surface at 260 m.

10 in order to reduce the grid dispersion. A value of $G=15$ was used in this study.

II. NUMERICAL RESULTS

The preceding pressure-velocity finite-difference formulation was coded in the computer program *PORAC*. This code was then used to simulate the overpressure in the air and the corresponding particle velocity in the ground that was measured at the Norwegian test site in Haslemoen, Norway in June 1994. The same measurements were also used by Hole *et al.*[7] in assessing the performance of the frequency-wave-number FFP code OASES.[5] The results from the present work are compared with both the measurements and the simulations by Hole *et al.*[7]

The Norwegian airblast tests were carried out in the forests of Haslemoen in June of 1994. The measurement setup comprised five microphones for registration of the impulse noise and three seismometers (one horizontal and two vertical) for recording ground vibration. The microphones were installed on two masts at elevations 2, 4, 8, 16 and 30 m above ground and the seismometers were positioned on the ground surface close to the masts. A total of 35 airblasts were carried out in which charges of 1, 8 and 64 kg were detonated at 2 m above ground at various distances ranging from 190 to 1400 m from the recording station. In addition, extensive meteorological and geotechnical/seismic tests were performed to characterize the atmospheric and ground conditions. A detailed account of the test program and measurements has been reported by Kerry[16] and Hole *et al.*[7] Based on these tests it was believed that an average uniform atmosphere with a sound speed of 340 m/s could provide an adequate model for numerical simulations.[7] *In situ* tests by the *Spectral Analysis of Surface Waves* method[17] and (Troxler) radioactive sound resulted in the soil profile given in Table I.

The time function describing the acoustic source is the same as that used by Hole *et al.*,[7]

$$g(t)=A[\sin(2\pi f_c t)-0.5\sin(4\pi f_c t)], \quad 0<t<1/f_c, \tag{16}$$

where A is a constant with dimension (s^{-2}), t denotes time

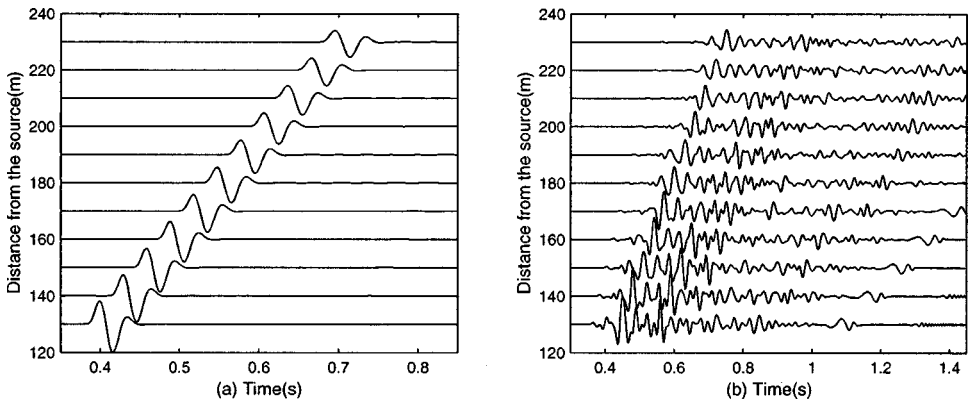


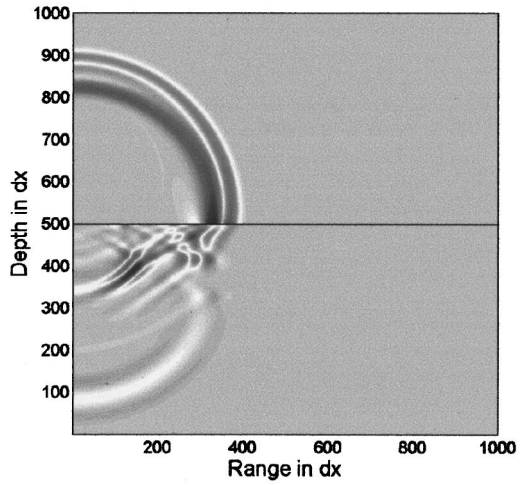
FIG. 4. Stacked plots of overpressure at 2 m above ground (a) and vertical particle velocity on the ground surface (b) at different distances from the source ranging from 130 to 230 m with spacing of 10 m.

and f_c is the central frequency. The above source and its Fourier spectrum are shown in Fig. 2. The central frequency of the sources is 30 Hz and $f_{up}=75$ Hz. The source positioned at the symmetry axis and 2 m above the ground is in accordance with the location of the C4 explosive charges in the tests. Examination of noise propagation with distance had shown a marked decrease in the dominant frequency of the noise with distance, which was stronger than could be explained by damping alone. This observation led Hole *et al.*[7] to conclude that the lengthening of the noise signals was due to the presence of a top porous soil layer. An extensive simulation study by Hole *et al.*,[7] using various combinations of permeability and thickness in the assumed surficial layer, revealed that a poro-elastic layer with a thickness of 0.5 m and permeability of $\kappa=10^{-8}$ m^2 gave a good match with the measured noise over a wide range. In this study the same thickness and permeability of the poro-elastic layer are used. The dynamic viscosity of the pore air $\eta=1.74\times10^{-5}$ kg/ms and tortuosity $e=1.25$ were used; the other parameters of the porous layer are listed in Table II. In order to reduce grid dispersion and CPU runtime the parameters $V_p=400$ and 500 and $V_s=200$ and 250 m/s were used in layers 5 and 6 instead of those in Table I. The density and sound speed in the atmosphere were taken equal to $\rho_0=1.2$ kg/m^3 and $V_0=340$ m/s, respectively. For the parameters of the system, Eqs. (14) and (15) give $\Delta x=0.115$ m and $\Delta t=1.33\times10^{-4}$ s.

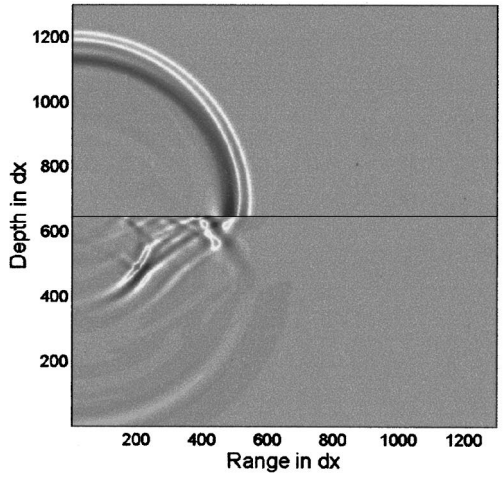
Figure 3 displays synthetic time histories of overpressure at 2 m above ground and associated particle velocity on the ground surface computed by *PORAC* at distances of 195 and 260 m from the blast. The figure also displays the corresponding results by the frequency-wave-number FFP model *OASES*[7] and the experimental data. Both the present results and *OASES* results were scaled to produce the same peak overpressure at 195 m. Examination of the plots for the various pressure data shows a fairly good accord in the amplitudes and arrival times between the cases. Moreover, the simulations reveal a decrease in the dominant frequency of the sound with distance as observed in the experimental data. For the ground response, the arrivals and amplitudes of the first peaks are almost the same in the two calculation models. However, the amplitudes of the subsequent peaks (resulting from reflected and refracted waves in the ground) are different. The difference could be partly due to absence of body wave attenuation in the ground.

Figure 4(a) displays a stacked plot of the simulated overpressures at different distances from the source ranging from 130 to 230 m with 10-m spacing. Figure 4(b) shows the corresponding ground motion stacked seismograms for the same receiver locations as in Fig. 4(a). The direct wave and surface wave can be recognized in this figure. The velocities of these waves can be calculated from the distance between the receivers and the corresponding difference of the arrival times (straight lines drawn through these arrivals provide a convenient way to calculate the velocities). The figure also displays the oscillatory nature of the ground motion that is a result of waves generated by air and refraction and multiple reflections at layer interfaces.

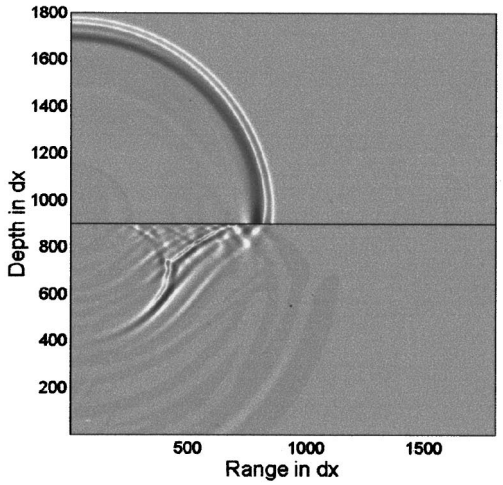
A global picture of the sound propagation and its interaction with the ground can be obtained by snapshot plots of the spatial variation of overpressure in the atmosphere and particle velocity in the ground. Figures 5(a)–5(c) present a set of such plots normalized by overpressures at times



(a) $t=0.1397$ s



(b) $t=0.1929$ s



(c) $t=0.2993$ s

FIG. 5. Snapshots of overpressure in the atmosphere and vertical particle velocity of the ground normalized by pressure at $t=0.1397$ s (a), 0.1929 s (b) and 0.2293 s (c).

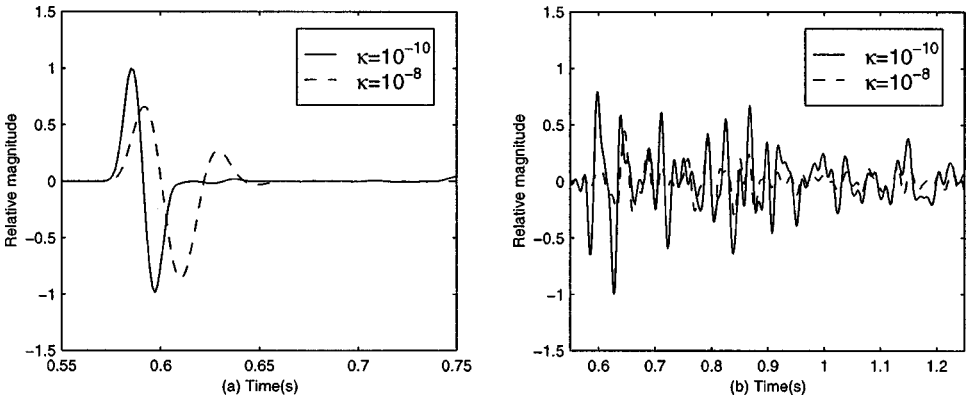


FIG. 6. Comparison between the results with different permeability (solid line for $\kappa=10^{-10}\,\mathrm{m}^2$ and dashed line for $\kappa=10^{-8}\,\mathrm{m}^2$) at 195 m from airblast position: (a) overpressure at 2 m above the ground and (b) vertical particle velocity of the ground surface.

0.1397, 0.1929 and 0.2293 s, respectively. The snapshots show a complicated deformation pattern in the ground, which results from the reflection and refraction of P- and S-waves as well as the surface waves. A careful examination of these plots shows the presence of Mach surfaces in the ground (although not quite straight due to ground layering). Because the velocities of P- and S-waves in the upper layers of the ground are lower than the sound speed, a super-seismic condition, characterized by two Mach surfaces associated with the two body waves, is realized. This visual feature is one of the most powerful attributes of the finite difference model *PORAC* that allows one to gain insight into the mechanism of generation and propagation of waves in the ground.

To discern the influence of permeability of the porous layer on the sound absorption and the induced ground vibration, different simulations were carried out by decreasing the permeability by 100 times, that is to $\kappa=10^{-10}\,\mathrm{m}^2$. Figure 6 shows a comparison between the results of overpressure at 2 m above ground [Fig. 6(a)] and vertical vibration on the ground surface [Fig. 6(b)] for different permeabilities. These results corroborate the conclusions by Hole *et al.*[7] in that the amplitude of overpressure and the peak particle velocity decrease with permeability. Furthermore, the dominant frequencies in both overpressure and ground vibration decrease as permeability increases.

Figures 7(a) and 7(b) show synthetic time histories by *PORAC* of the vertical velocities of pore air (solid line) and solid frame (dashed line) on the ground surface at 195 m from the source for different permeabilities [$\kappa=10^{-8}\,\mathrm{m}^2$ in Fig. 7(a) and $\kappa=10^{-10}\,\mathrm{m}^2$ in Fig. 7(b)]. All of the plots are normalized by the magnitude of the velocity of pore air with permeability $\kappa=10^{-8}\,\mathrm{m}^2$ [solid line in Fig. 7(a)]. In Fig. 7(a) the particle velocity of the solid frame is scaled up by a factor of 250 relative to the velocity of pore air. In Fig. 7(b) the velocities of pore air and solid frame are scaled up by factors of 4.2 and 110 relative to the velocity of pore air with $\kappa=10^{-8}\,\mathrm{m}^2$. Because there is only one poroelastic layer in the ground there is no refraction or reflection in the velocity of the pore air and the arrivals occur simultaneously with the sound wave. The multiple reflections can, however, be seen in the plots of the solid frame velocity. As expected, the amplitude of the velocity of pore air reduces as permeability decreases.

Figure 8 presents the snapshots of overpressure in the atmosphere and vertical particle velocity in the ground normalized by the pressure computed with the lower permeability of $\kappa=10^{-10}\,\mathrm{m}^2$. For direct comparison, Figs. 8(a) and 8(b) are plotted for the same times as in Figs. 6(a) and 6(b). The same general patterns are displayed in Figs. 8 and 6.

III. CONCLUSION

In this article a pressure-velocity finite-difference technique (*PORAC*) was developed for the simulation of sound propagation over layered porous media. Numerical simulations were carried out for the Norwegian airblast tests in Haslemoen. The top layer was taken as a poro-elastic material. Typical results of the simulations of the overpressure in the atmosphere and the velocities of the solid frame and pore air were presented. The results for the former two parameters were compared with the measurements and the results calculated by Hole *et al.* using a frequency-wave-number formulation. In addition, a number of snapshots of overpressure

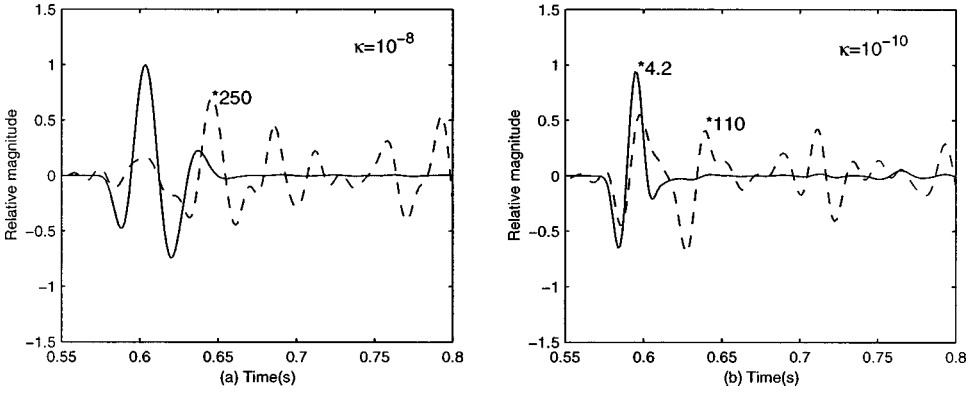


FIG. 7. Comparison of vertical velocities on the ground surface at 195 m from the airblast position (solid line for the velocity of pore air and dashed line for the velocity of solid frame). (a) $\kappa=10^{-8}\,\mathrm{m}^2$ (magnitude of the velocity of solid frame is amplified 250 times relative to that of pore air). (b) $\kappa=10^{-10}\,\mathrm{m}^2$ [magnitudes of the velocity of pore air and solid frame are amplified 4.2 and 110 times relative to the magnitude of the velocity of pore air in (a)].

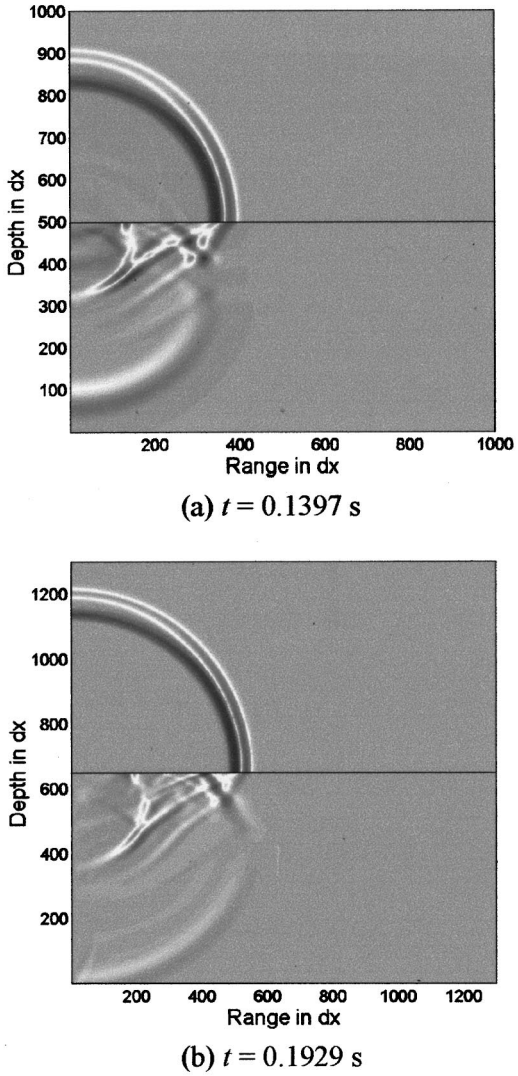


FIG. 8. Snapshots of overpressure in the atmosphere and vertical particle velocity of the ground normalized by pressure with permeability $\kappa = 10^{-10}\,\mathrm{m}^2$ and the rest of the parameters are same as those in Fig. 6 at $t = 0.1397$ s (a) and 0.1929 s (b).

and solid frame velocity were presented at different time steps. Among the attractive features of the presented model are its capability to present the results visually together with its versatility in dealing with arbitrary terrain conditions and range-dependent properties. Work is underway to extend *PORAC* to incorporate these features.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Lars R. Hole of the Norwegian Defense Construction Services (NDCS) and Professor Ulf Kristiansen of the Norwegian University of Science and Technology for their helpful discussions and suggestions. Partial financial support from Norwegian Defense Construction Service (NDCS) and the Norwegian Research Council under the program SIP-4 at Norwegian Geotechnical Institute (NGI) are gratefully acknowledged.

APPENDIX A: TREATMENT OF THE SOURCE AND FINITE-DIFFERENCE SCHEMES OF EQ. (1)

Integrating Eq. (1) over the domain G containing the singular point source in cylindrical coordinate system one gets

$$\int_G [\partial_{tt} P/(\rho_0 V_0^2)]dV = \int_G \nabla\cdot(\nabla P/\rho_0)dV + \int_G [2g(t)\delta(r) \times \delta(z-z_0)/r]dV. \tag{A1}$$

With the integral theorem of mean value, Green's formula, and the property of $\delta(\vec{r})$ function in Eq. (A1), one can easily obtain

$$\partial_{tt} P/(\rho_0 V_0^2)\Delta V = \oint_\Sigma (\partial_n P/\rho_0)dS_n + 4\pi g(t)\delta_{0r}\delta_{z_{0z}}, \tag{A2}$$

where Σ is the surface of the volume G. The finite-difference approximation of Eq. (A2) can be expressed as

$$\begin{aligned} P(0,j,k+1) &= 2P(0,j,k) - P(0,j,k-1) \\ &\quad + \gamma V_0^2\{4[P(1,j,k) - P(0,j,k)] \\ &\quad + P(0,j+1,k) - 2P(0,j,k) + P(0,j-1,k) \\ &\quad + (16\rho_0 g(k)/\Delta r\,\delta_{0i}\delta_{j_0 j})\}, \\ &\quad J<j<N, \quad 0<k<L. \end{aligned} \tag{A3}$$

The finite-difference formulation of Eq. (1) without source is

$$\begin{aligned} P(i,j,k+1) &= 2P(i,j,k) - P(i,j,k-1) + \gamma^2 V_0^2 \\ &\quad \times\{P(i+1,j,k) - P(i,j,k) + P(i-1,j,k) \\ &\quad + [P(i+1,j,k) - P(i-1,j,k)]/(2i) \\ &\quad + P(i,j+1,k) - 2P(i,j,k) \\ &\quad + P(i,j-1,k)\}, \\ &\quad 0<i<M, \ J<j<N, \ 0<k<L, \end{aligned} \tag{A4}$$

where the space grids are given by $r = i\Delta r$ and $z = j\Delta z$ and the time grid is given by $t = i\Delta t$; $P(r,z,t)$ is expressed by $P(i,j,k)$; and $M = a/\Delta r$, $J = h/\Delta z$, $N = H/\Delta z$, and $L = T/\Delta t$ where h is the ground surface and T the maximum measure time. In this article, we assume that $\Delta z = \Delta r$, and $\gamma = \Delta t/\Delta r$.

APPENDIX B: EXPRESSIONS FOR a_{ij}, $\vec{b}_i$, $\vec{\Delta}_i$ AND Δ IN EQS. (10) AND (11)

$$\begin{aligned} &a_{11} = \rho_1, \ a_{12} = E\rho_1(1 + b\Delta t/(2E\rho_1)), \\ &a_{21} = \rho, \ a_{22} = \rho_1, \\ &\vec{b}_1 = \rho_1(2\vec{u}^k - \vec{u}^{k-1}) \\ &\qquad + E\rho_1[2\vec{W}_\phi^k - (1 - b\Delta t/(2E\rho_1))\vec{W}_\phi^{k-1}], \\ &\vec{b}_2 = \rho(2\vec{u}^k - \vec{u}^{k-1}) + \rho_1(2\vec{W}_\phi^k - \vec{W}_\phi^{k-1}), \end{aligned} \tag{B1}$$

$$\vec{\Delta}_1 = a_{22}[(\Delta t)^2\vec{\Phi}_1^k + \vec{b}_1] - a_{12}[(\Delta t)^2\vec{\Phi}_2^k + \vec{b}_2],$$
$$\vec{\Delta}_2 = a_{11}[(\Delta t)^2\vec{\Phi}_2^k + \vec{b}_2] - a_{21}[(\Delta t)^2\vec{\Phi}_1^k + \vec{b}_1], \tag{B2}$$
$$\Delta = a_{11}a_{22} - a_{12}a_{21}.$$

APPENDIX C: TREATMENT OF BOUNDARY CONDITIONS ON GROUND SURFACE $z=h$

Let us consider a symmetrical grid block containing the interface (ground surface $z=h=J\Delta z$). Integrating Eqs. (1) and (2) over domains V_1 and V_2, respectively, we get

$$\int_{V_1} \frac{1}{\rho_0 V_0^2}\partial_{tt} P dV = \int_{V_1} \nabla\cdot\left(\frac{1}{\rho_0}\nabla P\right) dV, \tag{C1}$$

$$\int_{V2}[\rho_1\partial_{tt}u_j + e\rho_1\partial_{tt}W_{\phi j} + b\partial_t W_{\phi j}]dV = \int_{V2}[D_{1j}(\vec{u};\alpha M\phi) + D_{1j}(\vec{W}_\phi; M\phi^2)_\phi]dV, \tag{C2}$$

$$\int_{V_2}[\rho\partial_{tt}u_j + \rho_1\partial_{tt}W_{\phi j}]dV = \int_{V_2}[D_{1j}(\vec{u};K_b - 2\mu_b/3 + \alpha^2 M)]dV + \int_{V_2}[D_{1j}(\vec{W}_\phi;\alpha M\phi) + D_{2j}(\vec{u};\mu_b)]dV, \tag{C3}$$

where $j=1,2$ and

$$u_j = \begin{cases} u_r, & \text{for } j=1, \\ u_z, & \text{for } j=2, \end{cases} \quad W_{\phi j} = \begin{cases} W_{\phi r}, & j=1, \\ W_{\phi z}, & j=2. \end{cases} \tag{C4}$$

With the integral theorem of the mean value, the divergence theorem and the boundary conditions in Eq. (7) we can get a set of equations for $P(r,z,t)$, $u_j(r,z,t)$ and $W_{\phi j}(r,z,t)$. The difference schemes can be obtained by center difference and one-side difference. For $j=2$,

$$\begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix} \cdot \begin{pmatrix} P \\ u_z \\ W_{\phi z} \end{pmatrix}(i,J,k+1) = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix}, \tag{C5}$$

where

$$m_{11} = \frac{1}{\rho_0 V_0^2}, \quad m_{12} = -\gamma, \quad m_{13} = -\phi\gamma,$$
$$m_{21} = \phi\gamma, \quad m_{22} = \rho_1, \quad m_{23} = e\rho_1 + \frac{b\Delta t}{2}, \tag{C6}$$
$$m_{31} = \gamma, \quad m_{32} = \rho, \quad m_{33} = \rho_1,$$

$$d_1 = \frac{1}{\rho_0 V_0^2}(2P(i,J,k) - P(i,J,k-1)) - \gamma[u_z(i,J,k-1) - \phi W_{\phi z}(i,J,k-1)] + \gamma^2\frac{1}{\rho_0}\Bigg[2(P(i,J+1,k) - P(i,J,k)) + \frac{P(i+1,J,k) - P(i-1,J,k)}{2i} + P(i+1,J,k) - 2P(i,J,k) + P(i-1,J,k)\Bigg], \tag{C7}$$

$$d_2 = \gamma\phi P(i,J,k-1) + \rho_1[2u_z(i,J,k) - u_z(i,J,k-1)] + e\rho_1[2W_{\phi z}(i,J,k) - W_{\phi z}(i,J,k-1)] + \frac{b\Delta t}{2} \times W_{\phi z}(i,J,k-1) - \gamma^2\{[QR(i,J) + QR(i,J-1)]Z_1 + [QM(i,J) + QM(i,J-1)]Z_2\}, \tag{C8}$$

$$d_3 = \gamma P(i,J,k-1) + \rho(2u_z(i,J,k) - u_z(i,J,k-1)) + \rho_1(2W_{\phi z}(i,J,k) - W_{\phi z}(i,J,k-1)) - \gamma^2\{[QA(i,J) + QA(i,J-1)]Z_1 + [QR(i,J) + QR(i,J-1)]Z_2\} - 2\gamma^2[\mu_b(i,J) + \mu_b(i,J-1)] \times [u_z(i,J,k) - u_z(i,J-1,k)], \tag{C9}$$

where $QR = \alpha\phi M$, $QM = \phi^2 M$, $QA = K_b - \frac{2}{3}\mu_b + \alpha^2 M$ and

$$Z_1 = [u_r(i+1,J,k) + u_r(i+1,J-1,k) - u_r(i-1,J,k) - u_r(i-1,J-1,k)]/4 + [u_r(i,J,k) + u_r(i,J-1,k)]/(2i) + u_z(i,J,k) - u_z(i,J-1,k), \tag{C10}$$

$$Z_2 = [W_r(i+1,J,k) + W_r(i+1,J-1,k) - W_r(i-1,J,k) - W_r(i-1,J-1,k)]/4 + [W_r(i,J,k) + W_r(i,J-1,k)]/(2i) + W_z(i,J,k) - W_z(i,J-1,k). \tag{C11}$$

For $j=1$,

$$\begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} \cdot \begin{pmatrix} u_r \\ W_{\phi r} \end{pmatrix}(i,J,k+1) = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \tag{C12}$$

where

$$n_{11} = \rho_1, \quad n_{12} = e\rho_1 + \frac{b\Delta t}{2}, \tag{C13}$$
$$n_{21} = \rho, \quad n_{22} = \rho_1,$$

$$f_1 = \rho_1[2u_r(i,J,k) - u_r(i,J,k-1)] + e\rho_1[2W_{\phi r}(i,J,k) - W_{\phi r}(i,J,k-1)] + \frac{b\Delta t}{2}W_{\phi r}(i,J,k-1) + \gamma^2[QR(i,J)Z_3 + QM(i,J)Z_4], \tag{C14}$$

$$f_2 = \rho[2u_r(i,J,k) - u_r(i,J,k-1)] + \rho_1[2W_{\phi r}(i,J,k) - W_{\phi r}(i,J,k-1)] + \gamma^2[QA(i,J)Z_3 + QR(i,J)Z_4 + Z_5], \tag{C15}$$

$$Z_3 = [u_r(i+1,J,k) - 2u_r(i,J,k) + u_r(i-1,J,k)] + [u_r(i+1,J,k) + u_r(i,J,k)]/(2i+1) - [u_r(i-1,J,k) + u_r(i,J,k)]/(2i-1) + [u_z(i+1,J,k) - u_z(i+1,J-1,k) - u_z(i-1,J,k) + u_z(i-1,J-1,k)]/2, \tag{C16}$$

$$Z_4=[W_r(i+1,J,k)-2W_r(i,J,k)+W_r(i-1,J,k)]$$
$$+[W_r(i+1,J,k)+W_r(i,J,k)]/(2i+1)$$
$$-[W_r(i-1,J,k)+W_r(i,J,k)]/(2i-1)$$
$$+[W_z(i+1,J,k)-W_z(i+1,J-1,k)-W_z(i-1,J,k)$$
$$+W_z(i-1,J-1,k)]/2, \quad \text{(C17)}$$

$$Z_5=2\mu_b(i,J)[u_r(i+1,J,k)-2u_r(i,J,k)+u_r(i-1,J,k)]$$
$$+2\mu_b(i,J)\frac{u_r(i+1,J,k)-u_r(i-1,J,k)}{2i}$$
$$-2\mu_b(i,J)\frac{u_r(i,J,k)}{i^2}+[\mu_b(i,J)+\mu_b(i,J-1)]$$
$$\times\{u_r(i,J,k)-u_r(i,J-1,k)+[u_z(i+1,J,k)+u_z$$
$$\times(i+1,J-1,k)-u_z(i-1,J,k)-u_z(i-1,J-1,k)]/4\}. \quad \text{(C18)}$$

[1] F. G. Don and A. J. Cramond, ''Impulse propagation in a neutral atmosphere,'' J. Acoust. Soc. Am. **81**, 1341–1349 (1987).
[2] K. Attenborough, ''Ground parameter information for propagation modeling,'' J. Acoust. Soc. Am. **92**, 418–427 (1992).
[3] D. G. Albert and A. Orcutt, ''Acoustic pulse propagation above grassland and snow: Comparison of theoretical and experimental waveforms,'' J. Acoust. Soc. Am. **87**, 93–100 (1990).
[4] L. R. Hole, ''An experimental and theoretical study of propagation of acoustic pulses in a strongly refracting atmosphere,'' Appl. Acoust. **53**, 77–94 (1998).
[5] H. Schmidt, ''*OASES* version 2.0, application and upgrade notes,'' Massachusetts Institute of Technology, Department of Ocean Engineering (1997).
[6] N. P Chotiros, ''Biot model of sound propagation in water-saturated sand,'' J. Acoust. Soc. Am. **97**, 199–214 (1995).
[7] L. R. Hole, A. M. Kaynia, and C. Madshus, ''Measurement and simulation of low frequency impulse noise and ground vibration from airblasts,'' J. Sound Vib. **214**(2), 309–324 (1998).
[8] S. Tooms, S. Taherizadeh, and K. Attenborough, ''Sound propagation in a refracting fluid above a layered fluid-saturated porous elastic material,'' J. Acoust. Soc. Am. **93**, 173–181 (1993).
[9] H. Dong and K. Wang, ''Finite-difference numerical simulation of elastic wave field in the elastic solid formation-fluid borehole,'' Acta Geophysica Sinica (CHN) Suppl. 1 **38**, 205–215 (1995).
[10] M. A. Biot, ''Theory of elastic waves in a fluid-saturated porous rock: I. Low-frequency range,'' J. Acoust. Soc. Am. **28**, 168–178 (1956).
[11] M. A. Biot, ''Theory of elastic waves in a fluid-saturated porous rock: II. High-frequency range,'' J. Acoust. Soc. Am. **28**, 179–191 (1956).
[12] H. Dong, K. Wang, and J. M. Hovem, ''A stable integration technique for the interface conditions between two different media in the finite-difference method,'' in Proceedings 22nd Scandinavian Symposium on Physical Acoustics, Dept. of Telecommunications, NTNU, Norway (1999), pp. 47–48.
[13] A. C. Reynolds, ''Boundary conditions for the numerical solution of wave propagation problems,'' Geophysics **43**(6), 1099–1110 (1978).
[14] Z. S. Alterman and D. Loewenthal, ''Seismic waves in a quarter and three-quarter plane,'' Geophys. J. R. Astron. Soc. **20**, 101–126 (1970).
[15] R. M. Alford, K. K. Kelly, and D. M. Boore, ''Accuracy of finite-difference modelling of the acoustic wave equation,'' Geophysics **39**, 834–842 (1974).
[16] G. Kerry, ''An overview of the long range impulse sound propagation measurements made in Norway,'' Proc. Inter-noise **96**, 583–588 (1996).
[17] S. Nazarian and K. H. Stokoe, ''*In situ* shear wave velocities from spectral analysis of surface waves,'' in Proceedings 8th World Conf. Earthquake Engineering, San Francisco, III (1984), pp. 31–38.

Transport parameters for an ultrasonic pulsed wave propagating in a multiple scattering medium

Arnaud Tourin, Arnaud Derode, Aymeric Peyre, and Mathias Fink
Laboratoire Ondes et Acoustique, ESPCI, Université Paris VII, UMR 7587, 10 rue Vauquelin, 75005 Paris, France

(Received 20 April 1999; accepted for publication 4 April 2000)

A set of ultrasonic experimental methods was developed to characterize a multiple scattering medium in terms of l_s, l^*, l_a, respectively, the elastic, transport, and absorption mean free paths and D the diffusion constant. Actually, these quantities are the key parameters for a wave propagating in a disordered medium. Although they are widely used in optics, they are less common in acoustics. The underlying model is based on the expansion of the average solution for the heterogeneous Green's function equation. To validate this theoretical approach, a sample made of randomly located steel rods was used as a prototype. Through time-resolved measurements of the transmitted amplitude, the difference between the ballistic and the coherent wave is highlighted. In varying the sample thickness, l_s is determined, the coherent and diffusive regime are distinguished, and the transition from one to the other is followed. Furthermore, as a limit to a description of the average intensity based on the diffusion approximation, the existence of a coherent backscattering effect is shown. This latter gives a method to estimate D and l^*. These quantities being determined, it becomes possible to infer l_a using average time-resolved intensity measurements. Finally, some applications to coarse-grain stainless steels are discussed.
[S0001-4966(00)04007-8]

PACS numbers: 43.20.Gp, 43.20.Fn, 43.20.El [DEC]

INTRODUCTION

Multiple scattering is widely studied in physics. Especially, in solid-state physics (electron propagation in a random potential)[1] or in optics (propagation of light in turbid media),[2,3] it is usual to characterize experimentally the transport properties in a disordered loss-less medium in terms of l_s, the elastic mean free path, l^*, the transport mean free path, and D, the diffusion constant.

In this paper, we show that we can transpose with success this approach to the case of ultrasonic waves propagating in random media. It was already used in the past, to study elastic wave propagation in discrete random media. Especially, phase velocity and coherent attenuation were studied by use of the **T** matrix for one scatterer in conjunction with the quasicrystalline or Percus–Yevick approximations in order to take the correlations in the scatterers positions into account.[4,5] More recently, Turner and Weaver proposed a study of the incoherent intensity propagation in polycrystalline materials based on the radiative transfer equation.[6,7] However, a few experimental data do exist. One of the reasons is that only a few materials actually exhibit multiple scattering at the usual working frequencies. But, another lies in the technology necessary to study the frequency, time and space dependence of the acoustic speckle, i.e., the complicated figure produced by the interfering multiple echoes coming out of the multiple scattering medium. Today, the advent of piezoelectric transducers arrays brings new perspectives. Furthermore, the need in nondestructive evaluation of materials which exhibit multiple scattering, such as coarse-grain stainless steel which is used to build pipes in nuclear power plants, requires new approaches.

When an ultrasonic wave is propagating through a disordered loss-less medium, the regime of propagation highly depends on the ratios l_s/L and λ/l_s, where L is the medium thickness, λ the wavelength, and l_s the elastic mean free path. The latter gives the characteristic length of the exponential decay of the coherent wave, i.e., the part of the wave which, in average, propagates in the initial direction. By average, we mean a statistical average on all possible configurations of the scatterers. Mathematically, this coherent wave is denoted by $\langle G \rangle$, i.e., the part of the solution of the Green's function equation in the heterogeneous medium which remains after averaging on disorder. What is lost from this coherent part of the wave is transferred to secondary diffusive waves.

This coherent wave is often referred to as the ballistic beam (defined on one realization of disorder), i.e., the part of the wave which seems not to have been scattered at all by the heterogeneities in the medium.[8] In Sec. I, we highlight experimentally the difference between the ballistic and the coherent waves.

On a few mean free paths, the sample can be probed by this coherent wave and treated as an homogeneous medium in which the sound speed is just renormalized. This leads to the concept of effective medium, which is also theoretically and experimentally discussed in Sec. I.

Beyond several mean free paths, almost all energy has been transferred to the diffusive waves so that $\langle G \rangle \rightarrow 0$. Thus, to probe our medium, it is necessary to consider higher moments of the field amplitude, such as $\langle GG^* \rangle$, the average intensity. We show in Sec. II that the diffuse beam $\langle GG^* \rangle$ obeys a diffusion equation. At this stage, we introduce three other quantities of interest, the transport mean free path l^*, the diffusion constant D, and the absorption mean free path

l_a. l^* is the distance over which the wave has ''lost the memory'' of its initial direction, whereas D is related to the growth speed of the diffusive halo in the disordered medium. Finally, for a medium with losses, l_a is the characteristic distance of the intensity exponential decay due to absorption. In this approach, the average intensity is treated as a collection of particles subjected to a random walk and all coherence seems to be lost.

In Sec. III, we show the limits of this approach. Indeed, despite disorder, coherence is preserved in the initial direction of illumination, which manifests itself in the coherent backscattering effect. We take advantage of it to give new measurements of l^* and D which are not absorption dependent.

In these three sections, we support our discussion with results obtained on a disordered sample consisting of randomly located steel rods immersed in water.

Finally, in Sec. IV, we discuss some potential applications to real materials such as coarse grain stainless steels.

I. THE COHERENT WAVE

A. Theoretical background

In a ''liquid–liquid'' model, the heterogeneities can be described by a spatial dependency of both their compressibility $\chi(\mathbf{r})$ and density $\rho(\mathbf{r})$. The Green equation then becomes for a monochromatic scalar wave and a point source located in $\mathbf{r}'$[9]

$$\Delta G(\omega,\mathbf{r},\mathbf{r}')+k^2(\omega,\mathbf{r})G(\omega,\mathbf{r},\mathbf{r}') = -(\nabla \log \rho(\mathbf{r})\cdot\nabla)G(\omega,\mathbf{r},\mathbf{r}')+\delta(\mathbf{r}-\mathbf{r}'), \quad (1)$$

with $k^2(\omega,\mathbf{r})=\omega^2/c^2(\mathbf{r})$, $c(\mathbf{r})$ being the spatial dependent celerity.

Thus, the heterogeneities manifest themselves not only in the spatial dependency of the celerity but also in a new source term. If one defines the potential operator by $V(\mathbf{r})=k^2(\omega,\mathbf{r})-k_0^2(\omega)-(\nabla \log \rho(\mathbf{r})\cdot\nabla)$ where $k_0=\omega/c_0$ is the wave number for the homogeneous medium, the solution of Eq. (1) becomes

$$G(\omega,\mathbf{r},\mathbf{r}')=G_0(\omega,\mathbf{r}-\mathbf{r}')+\int G_0(\omega,\mathbf{r}-\mathbf{r}_1) \times V(\mathbf{r}_1)G(\omega,\mathbf{r}_1,\mathbf{r}')d\mathbf{r}'. \quad (2)$$

$G_0(\omega,\mathbf{r}-\mathbf{r}')$ is the solution of the Green equation in the homogeneous medium which writes in $\mathbf{k}$ space

$$G_0(\omega,\mathbf{k})=\frac{1}{k_0^2(\omega)-k^2}. \quad (3)$$

Equation (2) simply tells that, given a point source in $\mathbf{r}'$, the field amplitude at position $\mathbf{r}$ is the sum of the incident wave and the scattered wave. This latter results from the sum of all the waves scattered by the heterogeneities in $\mathbf{r}_1$. The difficulty lies in the fact that the wave incident on each $\mathbf{r}_1$ already includes scattering events.

The closed exact Eq. (2) can be expanded replacing the Green's function on the right-hand side by its expression on the left one. Then, for many applications involving ultrasound, it is sufficient to limit this expansion to the first order (''first Born approximation''), which means that the wave incident on each scatterer is the incident field G_0. Especially, it is true, at the usual frequency ($\approx$5 MHz), for biological media[10] or materials[11–14] where the single diffusion approximation holds. On the contrary, for media whose typical structures have the same size as the wavelength, multiple scattering must be taken into account. It is, for example, the case for some kinds of steels when the wavelength is of the same order as the grain size.[15]

Taking the average on disorder of Eq. (2), one obtains the Dyson equation[3]

$$\langle G(\omega,\mathbf{r}-\mathbf{r}')\rangle=G_0(\omega,\mathbf{r}-\mathbf{r}')+\int G_0(\omega,\mathbf{r}-\mathbf{r}_1)\Sigma(\mathbf{r}_1-\mathbf{r}_2) \times\langle G(\omega,\mathbf{r}_2-\mathbf{r}_0)\rangle d\mathbf{r}_1\, d\mathbf{r}_2, \quad (4)$$

where the self-energy Σ (analogous to the mass operator in quantum electrodynamics) is a nonlocal operator (depending only on the distance $|\mathbf{r}_2-\mathbf{r}_1|$) in which all information on the multiple scattering process has been hidden. In the $\mathbf{k}$ domain, the solution is

$$\langle G(\omega,\mathbf{k})\rangle=\frac{1}{k_0^2(\omega)-\Sigma(\omega,\mathbf{k})-k^2}. \quad (5)$$

If the scatterers are not too big compared to the wavelength and completely uncorrelated, the self-energy can become independent of $\mathbf{k}$ for a certain range of frequencies. It is the independent scattering approximation.[2] Note that this ''dilute medium'' approximation is different from a single scattering approximation like the first-order Born approximation. Indeed, it concerns an ''average medium'' and the self-energy contains information about multiple scattering. When the independent scattering approximation is made, Σ can be reintroduced in a new effective wave number k_e such that

$$k_e^2(\omega)=k_0^2(\omega)-\Sigma(\omega). \quad (6)$$

Then, the solution (5) has the same form as (3) in homogeneous medium with $k_0^2(\omega)$ replaced by $k_e^2(\omega)$. This introduces the concept of effective medium. On average, up to a few mean free paths, the ultrasonic wave can be described as a wave propagating in homogeneous medium with a renormalized velocity and an exponentially decaying amplitude. The imaginary part of k_e accounts for this decay of this coherent wave due to elastic scattering. It implies that the amplitude transmission coefficient of the coherent wave varies as

$$T_A=\exp\left(-\frac{L}{2l_s}\right), \quad (7)$$

where the characteristic distance of decaying l_s defines the elastic mean free path given by $l_s=1/(2\ \mathrm{Im}\, k_e)$. It is *a priori* frequency dependent.

In the independent scattering approximation, l_s can be estimated through the relation[2]

$$l_s=\frac{1}{n\sigma_T}, \quad (8)$$

where n is the density of scatterers and σ_T the total scattering cross section of one scatterer defined by

$$\sigma_T = \int_{4\pi} \frac{d\sigma(\theta,\varphi)}{d\Omega} d\Omega, \tag{9}$$

where $[d\sigma(\theta,\varphi)]/d\Omega$, the differential scattering cross section, represents the energy scattered in the direction (θ,φ), in spherical coordinates, in the solid angle $d\Omega$, when the scatterer is illuminated by a plane wave of unity intensity.

Relation (8) gives a method to estimate l_s if the density of scatterers and the scattering cross sections are known.

Experimentally, l_s can be inferred in measuring the coherent wave amplitude as a function of the sample thickness, as we will show. A priori, this requires an ensemble averaging. In optics of turbid media, the measured intensity is already an average on positions of the scatterers thanks to their Brownian motion. On the contrary, in acoustics, one has access to a single realization of disorder. In order to achieve an averaging on disorder to measure the coherent wave, we have to move our sample. Experimentally, we will highlight two contributions to the coherent wave, namely the ballistic and the forward-scattering ones.

B. Experiments

Up to now, only a few experiments concerning multiple scattering media have been reported in acoustics. Precise group velocity measurements have especially been performed by the group of Page and Sheng[16] in strongly scattering media. In these experiments, transducers large when compared to the wavelength were used as emitter and receiver. In our new experiments, we take advantage of ultrasonic arrays of subwavelength transducers to characterize the spatial distribution of the multiply scattered field. Furthermore, using very short pulses, we also have access to time-resolved measurements. As we will see, the results we obtained lead us to distinguish two components for the coherent wave, namely the ballistic and the forward-scattering contributions which can be separated in time.

The scatterers we consider are steel rods whose diameter is 0.8 mm. A modal decomposition of the pressure field[17] scattered by such a cylindrical object for our working center frequency gives access to the differential scattering cross section (Fig. 1) and thus to an estimation of $l_s = 4.2$ mm through relations (8) and (9). Of course, in our particular case, the differential scattering cross section depends on the angle θ only. The samples we use consist of collections of such steel rods, randomly located in the plane (Fig. 2). The thickness L is varied between 2.4 and 80 mm.

An ultrasonic source transmits a short pulse (3.2 MHz, 5/2 cycles) into the medium. The transmitted signals are recorded on an array of 128 transducers (pitch 0.4 mm), which gives access to the spatial distribution of the time-dependent pressure field. In order to perform an average on disorder, the sample is translated parallel to the array and the experiment is repeated for 80 positions (Fig. 3). The distance between source and receiver is kept constant. To change the thickness, we simply add more rods to the sample.

For sufficiently thin samples (experimentally actually up to $L < 15 l_s$) what we receive first, on one realization of disorder, is a front which looks like the front transmitted in water and which propagates at the same speed as if it ''did not see'' the medium (Fig. 4). It is what we choose to call the ballistic wave.[18]

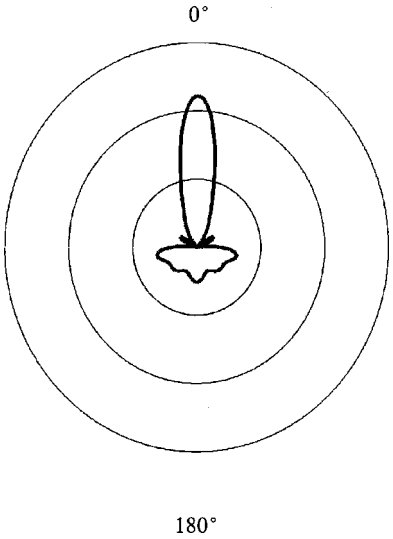


FIG. 1. Polar representation of the differential scattering cross section of a cylindrical steel rod calculated numerically using a modal decomposition of the scattered pressure. Frequency: 3.2 MHz, rod diameter 0.8 mm, $C_L = 5.9$ mm/μs, $C_T = 3.2$ mm/μs, density 7.85 g/cm^3. Angle $\theta = 0°$ corresponds to forward scattering.

For the thinnest samples, $L = 7$ mm in the example of Fig. 5, averaging on disorder ''does not change anything:'' the ballistic front is self-averaging and thus can be confused with the coherent wave. For $L = 30$ mm (Fig. 6), the ballistic wave is more difficult to distinguish. On each transducer, we always receive a first echo which has propagated at the speed of sound in water. But, from transducer to transducer, the amplitudes are greatly varying. Averaging is needed to make the coherent wave really emerge.

Figure 7 shows that the ballistic wave is not the only contribution to the coherent wave. The coherent wave is the part of the transmitted wave which cannot be averaged out. The ballistic wave contributes to the coherent wave but multiple scattering events can also add up to it. Actually, the coherent wave consists of the ballistic wave plus all scattering contributions emerging out of the sample with the same direction as the initial wave. That is the reason why, in Fig. 6, other fronts appear behind the first average front arriving at the ballistic time. What remains after averaging is the

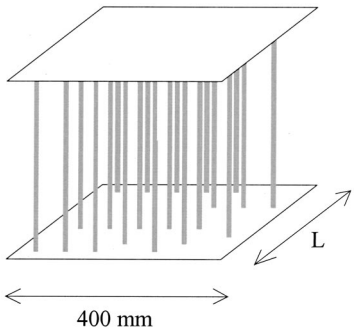


FIG. 2. The samples we use are made of randomly located steel rods. The average spacing between two rods is 2.4 mm. There are about 19 rods per cm^2.

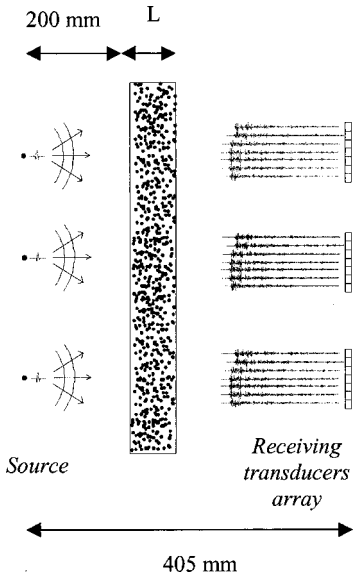


FIG. 3. Experimental setup. An ultrasonic source (3.2 MHz) transmits a broadband pulse into the multiple scattering sample whose thickness is L. The multiply scattered signals are recorded on the 128 transducers of the array. The experiment is repeated for 80 positions of the sample.

ballistic front plus the forward-scattering contribution. In the case we are interested in, these additional contributions can be related to a resonance of the scatterers which manifests in the curve of the total scattering cross section versus frequency as a dip around 2.8 MHz (Fig. 7). A corresponding ''dwell time''[2] can be associated with this resonance. It can be evaluated in a first approximation as the inverse width of the resonance. Its value (between 1 and 2 μs) explains well the time shift observed between the different fronts in Fig. 6. A spectral analysis of the coherent wave confirms that its two contributions do not belong to the same physics (see Fig. 8). The time-Fourier transforms of the wave in water and of the

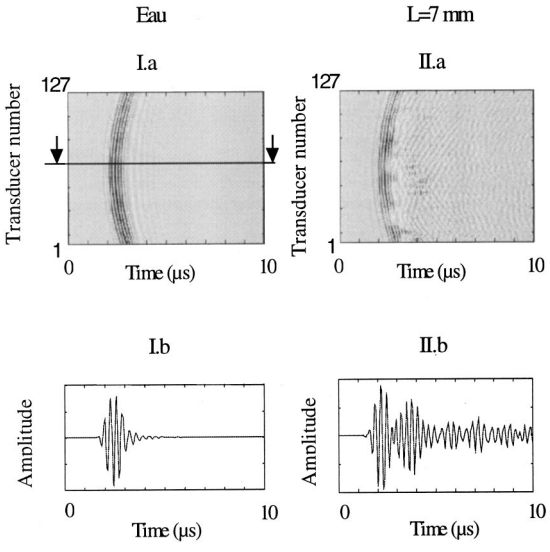


FIG. 4. (I a) and (I b) Wave transmitted in water and recorded on the transducer array. (a) *B*-scan representation. (b). *A*-scan representation: signal received on transducer #64. (II a) and (II b) Wave transmitted through a 7-mm-thick sample. (a) We receive a ballistic front followed by the first scattered waves. (b) *A*-scan representation: we clearly distinguish the ballistic pulse from the scattered signals arriving later.

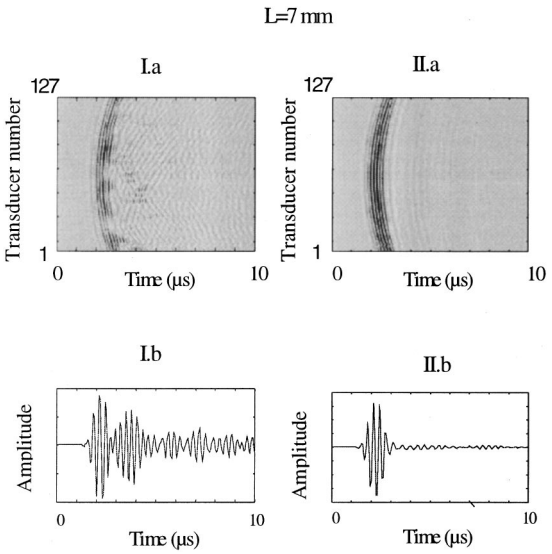


FIG. 5. $L = 7$ mm. (I a) and (I b) Transmitted wave on one realization of disorder. (II a) and (II b) Only the ballistic pulse remains after averaging on disorder.

ballistic wave are about the same. All are like the ballistic wave did not see the medium, even if it is not rigorously true. On the contrary, the forward-scattering contributions to the coherent wave, which appear only after averaging, are affected by dispersion. Their spectrum is shifted around 2.7 MHz, i.e., near the scatterer frequency resonance.

As we work in a pulse mode, the first idea to characterize the medium is to use the decay of the ballistic pulse amplitude to estimate the elastic mean free path, assuming that the inferred value is defined for the center frequency of the initial pulse. If defining the ballistic transmission coefficient as the ratio of the amplitudes received on transducer #64, respectively, through the medium and in water, we notice in a semilogarithmic scale that the linear fit works only up to $L = 20$ mm (Fig. 9). However, as pointed out before, in practice, the transmitted amplitude is actually often averaged

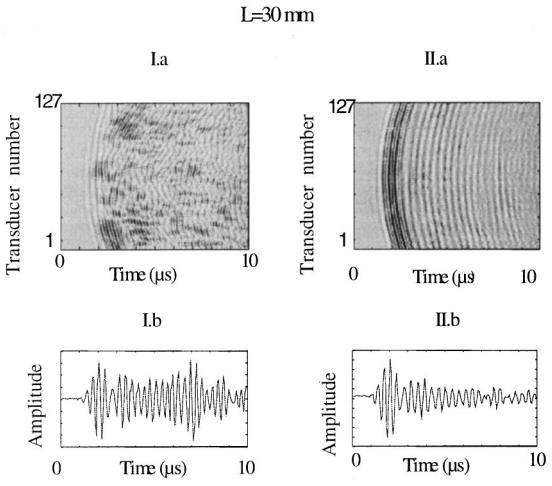


FIG. 6. $L = 30$ mm (I a) and (I b) Transmitted wave on one realization of disorder. The ballistic wave still exists but is more difficult to distinguish. The amplitude fluctuates a lot along the transmitted front. (II a) and (II b) After averaging on disorder, a coherent front emerges at the ballistic time but is followed by other contributions arriving later.

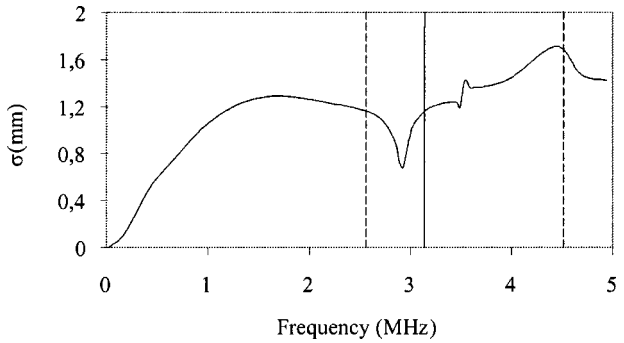


FIG. 7. Evolution of the total scattering cross section versus frequency for one steel rod. This curve was computed using a modal decomposition of the scattered pressure field. The center frequency used for the experiment is indicated by the full vertical line. The dotted lines define the bandwidth at −20 dB of our transducers.

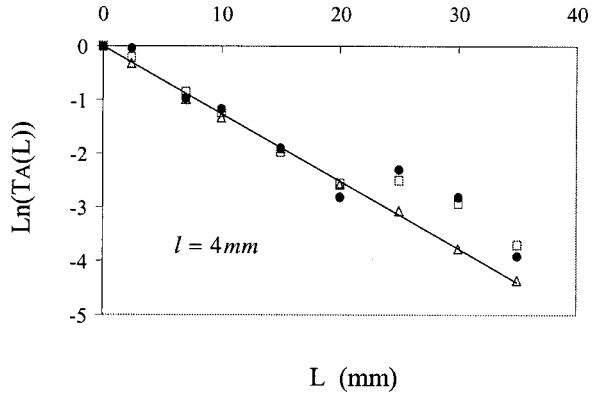


FIG. 9. Evolution of the transmitted amplitude versus sample thickness in a semilogarithmic scale. Full circles: ballistic amplitude received on transducer #64. Squares: ballistic amplitude averaged on the 127 transducers. Triangles: ballistic amplitude averaged on disorder.

on a transducer whose size is big compared to the wavelength[16] which already gives an estimation for the coherent wave. It would require us to average the ballistic amplitude on the whole array aperture. But, even if this definition is adopted, the fit is not really better. On the contrary, the transmission coefficient for the ballistic wave averaged on disorder is well fitted by a straight line even for the biggest thickness, which is consistent with the effective medium theory discussed below. From the slope, we obtain $l = 4$ mm, in good accordance with the value theoretically inferred through Eqs. (8) and (9). The slight difference can be easily explained by the presence of absorption, as we will show further. We include in this term all loss mechanisms others than scattering; especially, a vertical spreading of the wave along the rods cannot be completely neglected. Thus, the value of the mean free path we estimate actually depends on both scattering and absorption, i.e., is such that

$$1/l = 1/l_a + 1/l_s . \tag{10}$$

Finally, this study in a pulse mode highlights a danger of the confusion between the ballistic wave and the coherent wave. The ballistic wave is self-averaging only for a small thickness. Some experimentalists could see in the break in the linear fit beyond $L = 20$ mm the sign of a new regime. Actually, it is only due to insufficient statistics in the estimation of the coherent wave. But, above all, the confusion masks another danger, the fact that the ballistic wave is not the only contribution to the coherent wave, a fact that we have for the moment neglected.

At this stage, we indeed defined our transmission coefficients in a pulse mode so that we completely neglected the forward-scattering contribution to the coherent wave. When taking the time-Fourier transform of the whole coherent wave and defining the amplitude transmission coefficient for each frequency contained in the initial pulse, we notice in Table I that the mean free path can actually vary from the simple to the double. More precisely, we find for the center frequency the same value as the one inferred in a pulse mode. But, near the resonance, the mean free path is about 8.3 mm which is consistent with the fact that the total scattering cross section is about twice as big for this frequency (Fig. 7).

Finally, the ballistic wave is one component of the coherent wave but not necessarily the only one. On one realization of disorder, the experimentalist can observe a ballistic wave only if the sample consists of a basic homogeneous medium in which impurities are included with a relatively

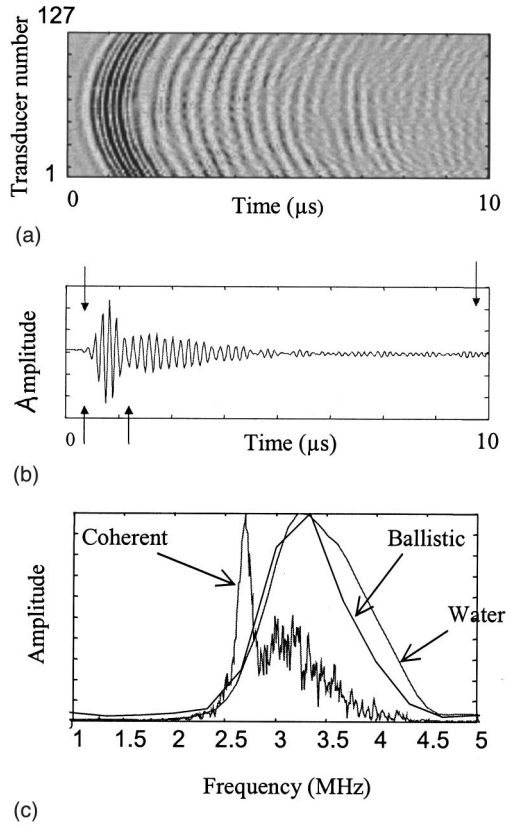


FIG. 8. (a) *B*-scan representation of the coherent wave transmitted through a 30-mm-thick sample. (b) Sum of the 127 transducers after phase readjustment to take the front curvature into account. (c) Time-Fourier spectra of the average ballistic pulse and of the coherent wave compared to the spectrum in water.

TABLE I. Frequency dependence of the mean free path.

F (MHz)	2.5	2.7	3	3.2	3.5	4	4.5
l (mm)	5.1	8.3	4.6	4	3.5	3.6	3.1

low density, for example, a collection of steels rods immersed in water. Of course, this ballistic wave remains after averaging but other contributions resulting from forward-scattered waves can also add up to it, as we noticed. In the case of granular media, the ballistic wave makes no sense anymore. However, in this case, there is a forward-scattered wave propagating along an ''effective contact medium'' which is the only contribution to the coherent wave.[18]

II. BEYOND THE COHERENT WAVE

A. Theoretical background

After several mean free paths, the coherent wave is no longer visible since almost all energy has been transferred to the diffusive waves and $\langle G\rangle \to 0$. To probe our sample, it becomes necessary to study the average intensity in the medium. It can be shown[3] that the diffusive energy $\langle I\rangle = \langle GG^*\rangle$ obeys an equation formally analogous to (4), the Bethe–Salpeter equation

$$\begin{aligned}\langle G(\mathbf{r}',\mathbf{r}_0')G^*(\mathbf{r}'',\mathbf{r}_0'')\rangle &= \langle G(\mathbf{r}',\mathbf{r}_0')\rangle\langle G^*(\mathbf{r}'',\mathbf{r}_0'')\rangle + \int \langle G(\mathbf{r}',\mathbf{r}_1)\rangle \\ &\times \langle G^*(\mathbf{r}'',\mathbf{r}_2)\rangle U(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4) \\ &\times \langle G(\mathbf{r}_3,\mathbf{r}_0')G^*(\mathbf{r}_4,\mathbf{r}_0'')\rangle d\mathbf{r}_1\, d\mathbf{r}_2\, d\mathbf{r}_3\, d\mathbf{r}_4. \qquad (11)\end{aligned}$$

In this expression, the four-point function $U(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4)$ is called the Vertex function.

In the ''Boltzmann approximation,'' (11) leads to the well-known radiative transfer equation (RTE)[19] which had first been derived phenomenologically by astrophysicists.[20]

Today, efficient methods exist to solve the RTE numerically. But, one can also use the diffusion approximation, which treats the propagation of intensity in the medium as a random walk. In this case, it has been shown that the time-dependent average intensity $\langle I(\mathbf{r},\mathbf{r}';t,t')\rangle = \langle |G(\mathbf{r},\mathbf{r}';t,t')|^2\rangle$ obeys the diffusion equation (2)

$$D\Delta\langle I(\mathbf{r}-\mathbf{r}',t)\rangle - \frac{\partial\langle I(\mathbf{r}-\mathbf{r}',t)\rangle}{\partial t} = \delta(\mathbf{r}-\mathbf{r}')\,\delta(t), \qquad (12)$$

where D is the diffusion constant. It is a dynamic parameter representing an area swept by time units and is related to the growth speed of the diffusive halo in the medium. It is related to the transport mean free path l^* by

$$D = \frac{1}{d} V_E l^*, \qquad (13)$$

where d is the dimension. V_E, the transport speed, is neither the phase velocity nor the group velocity. It measures the energy flow in the diffusive regime and is called transport velocity.[21]

The transport mean free path is sometimes confused with the elastic mean free path introduced above. Actually, the two mean free paths are related by an expression in which the anisotropy of individual scattering appears

$$l^* = \frac{l_s}{1-\overline{\cos(\theta)}}. \qquad (14)$$

$\overline{\cos(\theta)}$ represents the average cosine of the scattering angle by one scatterer given for a cylindrical rod by

$$\overline{\cos(\theta)} = \frac{1}{2\pi}\int_0^{2\pi} \cos(\theta)\,\frac{d\sigma(\theta)}{d\theta}\,d\theta.$$

For the steel rods, $\overline{\cos\theta}$ equals 0.13 and the transport mean free path is thus estimated to be $l^* = 4.8$ mm.

In d dimensions ($d=2$ for our sample geometry), the solution of (12) in free space is written as

$$\langle I(\mathbf{r},\mathbf{r}',t)\rangle = \frac{1}{(4\pi Dt)^{d/2}}\exp\left(-\frac{|\mathbf{r}-\mathbf{r}'|^2}{4Dt}\right). \qquad (15)$$

For a semi-infinite slab between 0 and L, the intensity transmittance for an incident plane wave on a slab is written[22]

$$T(t;L) = \frac{e^{-L/l_a}}{(4\pi Dt)^{1/2}}\sum_{-\infty}^{\infty}\{e^{-[(2n-1)L-2a]^2/4Dt} - e^{-[(2n-1)L]^2/4Dt}\}, \qquad (16)$$

with $a = 5/3l^*$. $T(t;L)$ gives the probability for a given incident pulse of intensity to travel during a time t before leaving the medium. Thus, in literature, it is often referred to as the ''time-of-flight (TOF) distribution''. The position of the maximum of this curve is given by

$$t_m \approx 0.1\,\frac{L^2}{D}. \qquad (17)$$

Furthermore, the tail of the distribution is well fitted by a single exponential whose decay rate is

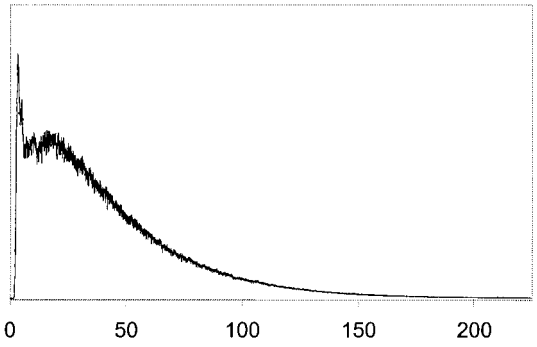
$$\frac{1}{\tau} = \pi^2 D/L^2. \qquad (18)$$

Nevertheless, the coherent and incoherent waves can experimentally coexist in a large domain of ratio l_s/L. Thus, one part of the energy, which would be in the diffusive part given by (16) in a purely multidiffusive regime, remains in the coherent part. Thus, the diffusion approximation is only valid for the thickest samples. Furthermore, the TOF highly depends on absorption in the medium. The estimation of D through Eq. (16) is therefore not certain and the theoretical time of flight cannot be directly used to fit the curve of the transmitted intensity for a given thickness, except if the diffusion constant is already known. In this case, the absorption mean free path is the only adjustable parameter.

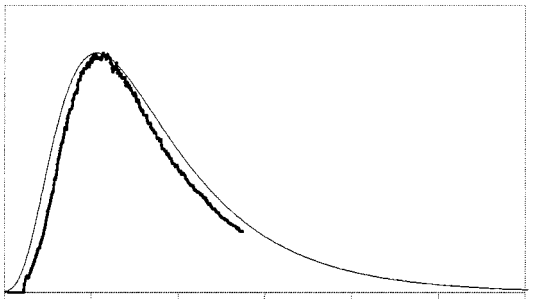
B. Experiment

The experimental setup is the same as before (Fig. 3). The average transmitted energy is calculated by integrating the square amplitude on 80 realizations of disorder and on the receiving 128 transducers.

We present in Fig. 10 two time-of-flight distributions: in the first case ($L = 35$ mm), the coherent and diffusive waves clearly coexist. Behind the coherent peak, the solution of the diffusion equation starts building. For 80 mm, no coherent part remains and the TOF is well fitted by Eq. (16), assuming that the diffusion constant is known and equal to D



a)



b)

FIG. 10. Experimental time-of-flight distributions for two different thickness: (a) $L=35$ mm, (b) $L=70$ mm. For $L=35$ mm, the coherent part still exists. For $L=70$ mm, the regime of propagation turns out to be purely diffusive. The curve has been smoothed by integrating the intensity on small time windows whose duration is the same as the initial pulse. The theoretical fit based on Eq. (16) gives access to the absorption mean free path if the diffusion constant is known.

$=3.2\ \mathrm{mm}^2/\mu\mathrm{s}$ and taking $l_a=375$ mm. We will show in the following section that D can be measured through the recording of the coherent backscattering effect.

If D can thus be inferred through a dynamic measurement, the transport mean free path l^* can be theoretically estimated with a stationary one. Indeed, when integrating the time-dependent solution of the diffusion equation (16), one finds an algebraic dependency of the transmission coefficient in intensity[22]

$$T(L)\propto\frac{l^*}{L}. \tag{19}$$

Experimentally, the problem is that the regime is not purely diffusive, at least up to 70 mm as pointed before. But beyond, the noise cannot be neglected anymore when one records the transmitted intensity. Thus, we do not have enough measurements to plot the transmission coefficient versus thickness and determine l^*. So, this method is not reliable to determine l^*. In the third section, we will see that the coherent backscattering effect gives a means to measure l^* without needing to make measurements on different sample thickness.

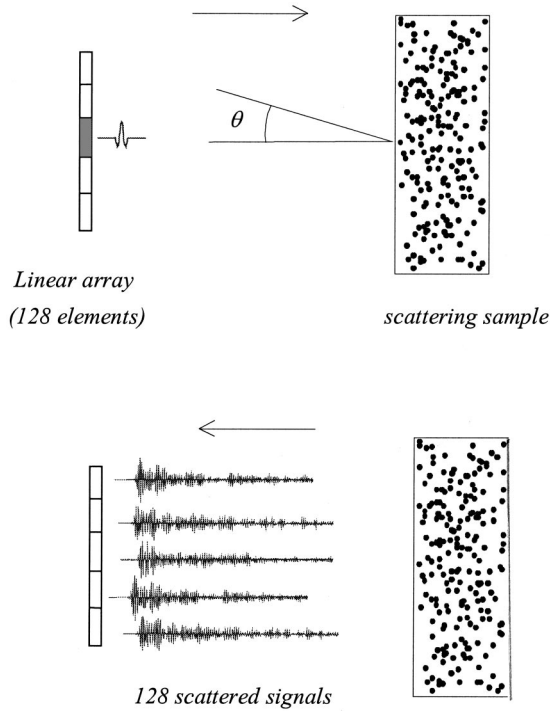


FIG. 11. Recording of the coherent backscattering effect. One of the elements of a 128-transducer array transmits a pulsed wave into the medium (15 cycles). The 128 backscattered signals are recorded on the array as a function of time. The number of the receiving transducer is converted into the angle θ.

III. THE COHERENT BACKSCATTERING EFFECT

In 1985, the coherent backscattering effect (CBE) was first observed in light scattering.[23,24] It was the first experimental evidence that the Boltzmann theory, and especially the equation of radiative transfer, fails in describing completely the propagation of light in random media. Indeed, this effect proves that coherent interference effects can survive even in strongly disordered media.

In the past, numerous works in optics were published on this topic. We present here experimental results with acoustic waves which give a new method to determine accurately both l^* and D.[25]

Our experiment consists of transmitting an ultrasonic pulse in a random sample made of 6000 steel rods (thickness 80 mm). The backscattered signals are recorded on a 128-elements array (Fig. 11). From these signals, we measure either the stationary intensity (as it is usually done in optics) or the time-dependent intensity. In the first case, we just integrate the square amplitude over the whole signal excluding, however, the first echoes to get rid of the single scattering contribution. In the second case, we integrate the square amplitude over small time windows (5 μs) taken at different positions (Fig. 12). Ensemble averaging is performed by translating the array along the sample. After averaging on 80 realizations of disorder, we do not observe an isotropic average intensity distribution as one could have naively expected (Fig. 13).

When the pulsed wave is transmitted into the multiple scattering medium and the backscattered intensity is recorded, several peaks occur at arbitrary random reflection angles. But, if an average on configurations is now applied,

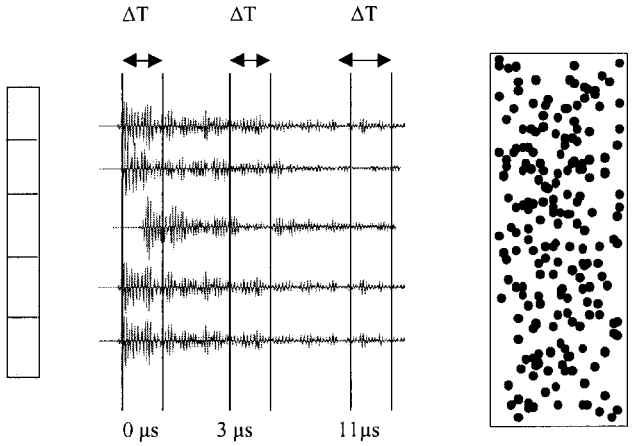


FIG. 12. On each transducer, we calculate the time-resolved intensity integrating the squared amplitude on small time windows ($\Delta T \approx 5\ \mu$s).

all these peaks are expected to vanish: if disorder is strong enough, the intensity should be uniformly distributed. However, there is one particular peak which exists for all realizations of disorder. Indeed, whatever the scatterers configuration is, a given ray path and its reciprocal counterpart always interfere constructively in the backscattering direction, since they are necessarily in phase, under the assumption of reciprocity of the medium. That is why this coherent backscattering peak emerges after configuration averaging.

In the stationary case, using a diagrammatic approach, Tourin *et al.* have shown that in 2D the angular half-width at half maximum is related to the transport mean free path by the relation[25]

$$\delta = 1.08 \frac{1}{kl^*}. \tag{20}$$

This yields experimentally $l^* = 4.8$ mm, in perfect agreement with the estimation based on the knowledge of l_s and the relation (14).

Thus, the stationary CBE gives us an efficient means to evaluate l^*. In the case where the medium is *a priori* unknown, especially concerning the scattering cross sections of its scatterers, the determination of both mean free paths l_s and l^* gives an estimation of the coefficient of anisotropy previously defined (14).

In the dynamic case, the width of the cone is found to decrease with time as shown in Fig. 14, which can be related

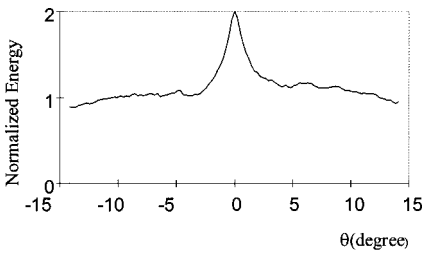


FIG. 13. Stationary coherent backscattering effect. The energy has been calculated integrating the square amplitude received on each transducer on the whole signal.

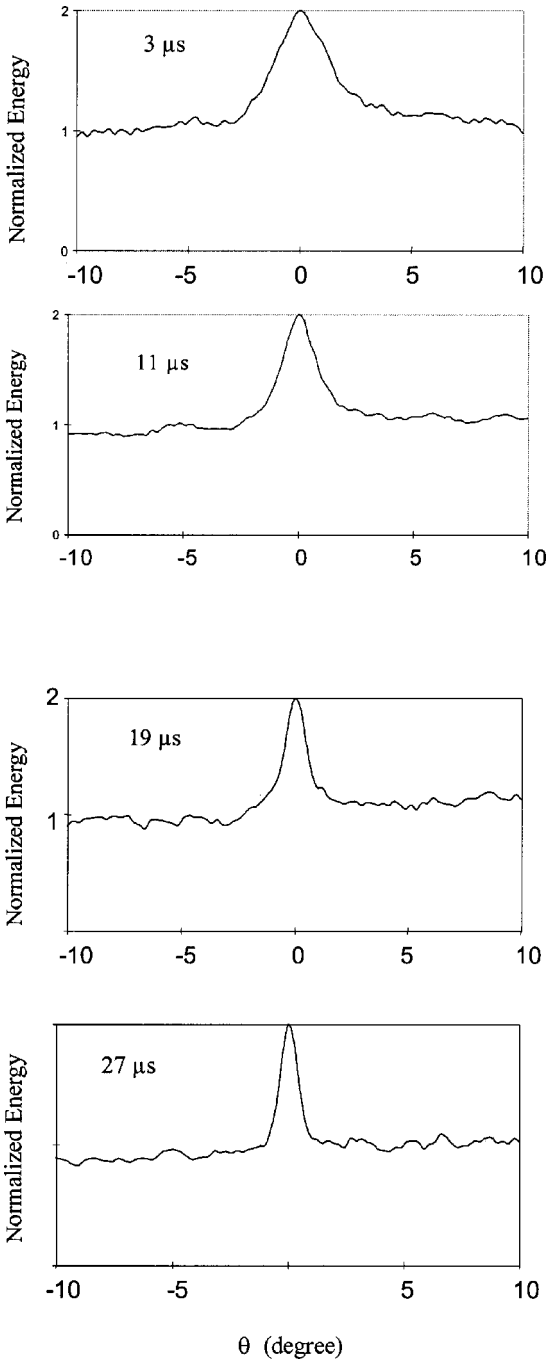


FIG. 14. Dynamic coherent backscattering effect. We present here the results obtained by integrating the squared amplitude over small windows (width 5 μs) taken at 3, 11, 19, and 27 μs and averaging over 80 realizations.

to the growth of the diffusive halo in the medium. It has been shown[25] that the angular half-width at half maximum is related to the diffusion constant D by

$$\delta = 1.12 \frac{1}{k\sqrt{Dt}}. \tag{21}$$

The plot of the width versus time in a log–log scale is well fitted by a line whose slope is 1/2 (Fig. 15). The intercept gives $D = 3.2\ \text{mm}^2/\mu\text{s}$.

It has been shown that this measurement of D is not absorption dependent.[26] Thus, knowing this value, we can

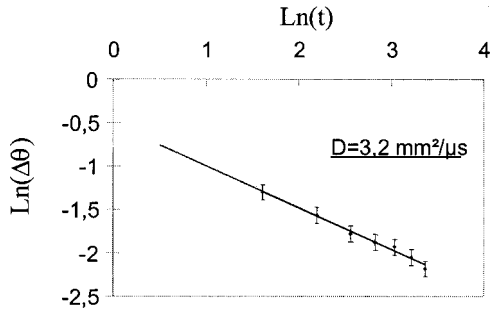


FIG. 15. Half-width at half maximum versus time in a log–log scale.

deduce from the TOF (16) the absorption mean free path as pointed out before and finally the elastic mean free path through relation (10) which enables us to distinguish between elastic and absorption effects in the wave attenuation.

Finally, the coherent backscattering effect gives us an accurate and elegant way to determine the transport properties of an ultrasonic wave in a multiple scattering medium. Completed by measurements in a transmit mode, it enables to determine the four key parameters for a wave propagating in a disordered medium, namely l_s, l^*, D, and l_a. We present in the next paragraph an application of this method to characterization of coarse-grain stainless steels.

IV. APPLICATION TO COARSE-GRAIN STAINLESS STEELS

Some varieties of coarse-grain stainless steels have grain sizes comparable to the wavelengths used to test them. We present here results concerning a parallelepiped block, immersed in water, whose dimensions are $120\times120\times60$ mm^3. When we illuminate this latter with a pulsed wave of 1.5-MHz center frequency, a long signal looking like these observed in our prototype appears behind the interface echo (Fig. 16). Especially, these multiply scattered echoes explain why it is very difficult to detect a defect in such a material.[15]

To record the coherent backscattering effect, we have used the same configuration as the one we presented in Sec. III. The sample thickness is 60 mm. We chose our time-window integration to get rid of the specular and the single scattering contributions. After averaging, a peak centered on the initial source emerges (Fig 17).

This experiment gave us hope for characterizing the mean grain size of such multiple scattering materials by an ultrasonic measurement performed in a backscattering configuration. To this goal, we have to extend our theory to take into account the refraction at the steel–water interface.

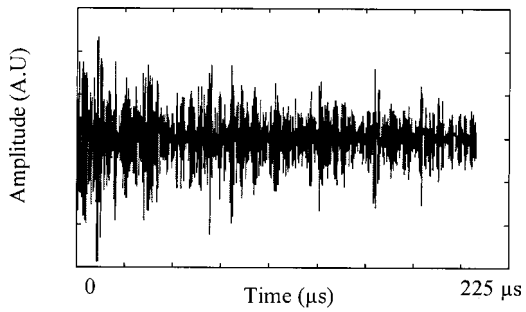


FIG. 16. A linear array of 128 transducers transmits a pulsed plane wave (3/2 cycles at 1.5 MHz, i.e., a duration of 1 μs) into the sample. We record on each transducer the backscattered signals. We present as an example the signal received after the interface echo on transducer #10.

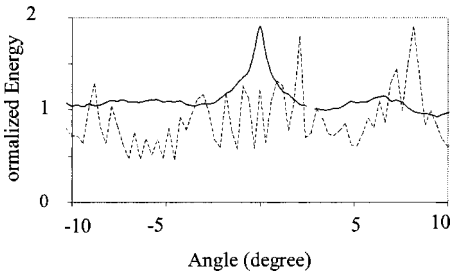


FIG. 17. Dotted line: stationary intensity recorded on a single realization of disorder. Full line: stationary coherent backscattering obtained after averaging on 200 realizations.

V. CONCLUSION

We have proposed a set of experiments, in relation to the underlying theoretical concepts, in order to determine the transport parameters of an ultrasonic wave propagating in a random medium. Although the multiple scattering theory that we use is based on some ''dilute medium approximations,'' we found that it enables us to fit our experimental results quite well.

For samples whose thickness is not too large compared to the elastic mean free path, we can probe the medium with the coherent wave, i.e., the average field amplitude, which propagates in an effective medium, decaying exponentially with the distance. We have shown that this coherent wave includes both ballistic and forward-scattering contributions. The elastic mean free path l_s can be measured from the variation of the coherent amplitude with the sample thickness.

Beyond several mean free paths, the transport of energy in the medium is diffusive and characterized by two parameters: a stationary one, the transport mean free path, which gives the length over which the wave has lost the memory of its initial direction; a dynamic one, the diffusion constant, which is related to the growth speed of the diffusive halo in the medium. D and l^* can be estimated through the coherent backscattering effect. l_a can be then inferred using the time-of-flight distribution (time-resolved transmitted intensity).

These kinds of measurements are quite common in optics. But, contrary to optics, we have shown that the coherent wave and the incoherent wave can actually coexist in a large domain of ratio l_s/L, which highlights a new intermediate regime of propagation.

[1] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).
[2] A. Lagendijk and B. A. van Tiggelen, Phys. Rep. **270**, 143 (1996).
[3] P. Sheng, *Introduction to Wave Scattering, Localization, and Mesoscopic Phenomena* (Academic, New York, 1995).
[4] L. Tsang, J. A. Kong, and T. Habashy, ''Multiple scattering of acoustic waves by a random distribution of discrete spherical scatterers with the quasicrystalline and Percus–Yevick approximation,'' J. Acoust. Soc. Am. **71**, 552 (1982).
[5] V. K. Varadan, T. Ma, and V. V. Varadan, ''A multiple scattering theory for elastic wave propagation in discrete random media,'' J. Acoust. Soc. Am. **77**, 375 (1985).

[6] J. A. Turner and R. L. Weaver, ''Radiative transfer of ultrasound,'' J. Acoust. Soc. Am. **96**, 3654 (1994).
[7] J. A. Turner and R. L. Weaver, ''Radiative transfer and multiple scattering of diffuse ultrasound in polycrystalline media,'' J. Acoust. Soc. Am. **96**, 3675 (1994).
[8] POAN Research Group, *New Aspects of Electromagnetic and Acoustic Wave Diffusion*, edited by G. Höler, Springer Tracts in Modern Physics, Vol. 144 (Springer, Berlin, 1998), pp. 6, 7.
[9] L. A. Chernov, *Wave Propagation in a Random Medium* (McGraw-Hill, New York, 1960).
[10] K. K. Shung and G. A. Thieme, *Ultrasonic scattering in Biological Tissue* (CRC Press, Boca Raton, FL, 1992).
[11] F. J. Margetan, T. A. Gray, and R. B. Thompson, ''A technique for quantitatively measuring microstructurally induced ultrasonic noise,'' in *Review of Progress in Quantitative NDE*, Vol. 10, edited by D. O. Thompson and D. E Chimenti (Plenum, New York, 1991), pp. 1721–1728.
[12] J. H. Rose, ''Ultrasonic backscatter from microstructure,'' in *Review of Progress in Quantitative NDE*, Vol. 11, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1992), pp. 1677–1684.
[13] M. D. Russel and S. P. Neal, ''Grain noise power spectrum estimation for weak scattering polycrystalline materials using experimentally estimated backscatter coefficients: Normal incidence,'' Ultrasonics **32**, 163 (1994).
[14] M. D. Russel and S. P. Neal, ''Grain noise power spectrum estimation for weak scattering polycrytalline materials using experimentally estimated backscatter coefficients: Oblique incidence,'' Ultrasonics **32**, 173 (1994).
[15] J. M. Bordier, M. Fink, A. Le Brun, and F. Cohen-Tenoudji, ''The Influence of Multiple Scattering in Incoherent Ultrasonic Inspection of Coarse Grain Stainless Steel,'' Proceedings of the 1991 Ultrasonics Symposium, Vol. 2, pp. 803–808.
[16] J. H. Page, P. Sheng, H. P. Scriemer, I. Jones, X. Jing, and D. A. Weitz, Science **271**, 634 (1996).
[17] N. Gespa, *La diffusion acoustique par des cibles élastiques de forme géométrique simple* (Cedocar, 1987).
[18] X. Jia, C. Caroli, and B. Velicky, Phys. Rev. Lett. **82**, 1863 (1999).
[19] G. C. Papanicolaou and R. Burridge, J. Math. Phys. **16**, 2074 (1975).
[20] S. Chandrasekhar, *Radiative Transfer* (Oxford University Press, London, 1950).
[21] M. P. van Albada, B. A. van Tiggelen, A. Lagendijk, and A. Tip, ''Speed of propagation of classical waves in strongly scattering media,'' Phys. Rev. Lett. **66**, 3132 (1991).
[22] A. Z. Genack, in *Scattering and Localization of Waves in Random Media*, edited by P. Sheng (World Scientific, Singapore, 1990), pp. 207–311.
[23] M. P. van Albada and A. Lagendijk, Phys. Rev. Lett. **55**, 2692 (1985).
[24] E. Wolf and G. Maret, Phys. Rev. Lett. **55**, 2696 (1985).
[25] A. Tourin, A. Derode, P. Roux, B. A. van Tiggelen, and M. Fink, Phys. Rev. Lett. **79**, 3637 (1997).
[26] A. Tourin, ''Multiple scattering and time-reversal of ultrasonic waves,'' Ph.D. dissertation, The University Denis Diderot at Paris, 1999 (in French).

Approximate expressions for viscous attenuation in marine sediments: Relating Biot's "critical" and "peak" frequencies

Altan Turgut
Acoustics Division, Naval Research Laboratory, Washington, DC 20375

(Received 12 October 1998; accepted for publication 3 April 2000)

Simple approximate relations are proposed for the viscous attenuation per cycle of the fast compressional and shear waves in the low-to-intermediate frequency range. Corresponding closed-form formulas are derived for frequencies at which maximum viscous attenuation per cycle occurs according to the Biot–Stoll theory of elastic wave propagation in marine sediments. In the new formulas, Biot's approximation [M. A. Biot, J. Acoust. Soc. Am. **34**, 1254–1264 (1962)] for the frequency-dependent viscosity correction factor $F(f)$ and the assumption of relatively low specific loss ($Q^{-1}<0.2$) [J. Geertsma and D. C. Smith, Geophysics **26**(2), 169–181 (1962)] are used to provide an accurate representation of the fast compressional and shear wave attenuation from low frequencies through a transition region extending to two or three times Biot's critical frequency f_c. The approximate viscodynamic behavior of marine sediments for the fast compressional and shear waves shows similarities to that of a ''homogeneous relaxation'' process for an anelastic linear element [A. M. Freudenthal and H. Geiringer, *Encyclopedia of Physics* (Springer-Verlag, 1958), Vol. 6].
[S0001-4966(00)03307-5]

PACS numbers: 43.20.Mv, 43.30.Ma [DLB]

INTRODUCTION

In the original Biot theory[1,2] of elastic wave propagation in a fluid-filled porous medium, wave attenuation is attributed primarily to the viscous losses due to relative motion between the pore fluid and the skeletal frame. Biot examined the corresponding attenuation and phase velocity dispersion curves for fast and slow compressional waves and shear waves as a function of frequency and provided asymptotic expressions of attenuation and velocity at both low and high frequencies. He defined a characteristic frequency f_c at which viscous forces and inertial forces are of the same order of magnitude and stated that the maximum attenuation per cycle (specific attenuation) for both fast compressional and shear waves occurs at a ''transition'' or ''peak'' frequency f_p where f_p/f_c is near unity. Stoll[3] incorporated the inelasticity of the frame into the Biot theory in order to account for various forms of energy dissipation that occur at the grain contacts (i.e., friction relaxation, local viscous losses, etc.). He pointed out that while losses in the skeletal frame dominate at very low and very high frequencies, global viscous losses (due to relative motion between the skeletal frame and pore fluid) might become predominant near the transition frequency.

The transition frequency defines a boundary between low- and high-frequency attenuation regimes and provides a potential for the accurate determination of permeability coefficient from the acoustic measurements. While the transition frequency is most strongly dependent on permeability, it also depends on other sediment parameters such as the mass-coupling coefficient. The complex dependency of the transition frequency on several parameters makes it difficult to estimate the permeability coefficient accurately. To simplify the estimation, several approximate expressions were proposed in the literature to estimate maximum specific attenuation and the location of transition frequency. Badiey *et al.*[4] used $f_p/f_c \approx 1$ in their calculation to evaluate maximum specific attenuation in the Biot–Stoll model. Yamamoto[5] proposed a poroviscous frequency number $N_f=(f/f_c)/2\pi$ and assumed that the maximum specific attenuation occurs at $N_f=0.1$ which corresponds to $f_p/f_c=0.628$. Using Biot's suggested values for mass-coupling factor and structure factor, White[6] indicated that Biot's critical frequency f_c is 127 Hz whereas the peak frequency f_p is about 40 Hz for water-saturated gravel. Turgut and Yamamoto[7] noticed a dependency of the f_p/f_c ratio on the mass-coupling factor and proposed a simple relation for transition frequency of the fast compressional wave by assuming that the viscosity correction factor F is substantially real and almost equal to unity near the transition frequency.

In this paper, more accurate approximate formulas are derived to relate the transition frequencies of both fast compressional and shear waves to Biot's critical frequency by using an approximate form of the viscosity correction factor.[8] The accuracy of the new and previously proposed formulas are compared for different types of sediments and superiority of the new formulas is demonstrated. If the transition frequencies can be identified from the attenuation and sound speed measurements, such simple expressions would have significant practical value for estimating the permeability and also help to provide additional constraints for the estimation of other most relevant Biot parameters. Inversion of the Biot parameters from the acoustic data is beyond the scope of this work and will be addressed in a separate paper.

I. APPROXIMATE FORMULAS FOR SPECIFIC ATTENUATION AND PEAK FREQUENCIES

In this section, an approximate solution for shear-wave specific attenuation is derived using an approximate form of

viscosity correction factor F and assuming relatively low values of the specific attenuation ($Q^{-1} \approx 0.2$ or lower). The approximate solution for the compressional-wave specific attenuation is also given by incorporating the approximate form of F in the derivations given in Greetsma and Smit[9] or Turgut and Yamamoto.[7] Then, corresponding peak frequencies for both shear and compressional waves are introduced. For simple harmonic rotational motion, the coupled dynamical equations are[10]

$$\begin{aligned} \omega^2(\rho\Psi_1 - \rho_f\Psi_2) &= \mu l^2 \Psi_1, \\ \omega^2(-\rho_f\Psi_1 + m\Psi_2) &= \frac{i\omega F \eta}{k_s}\Psi_2, \end{aligned} \tag{1}$$

where Ψ_1 and Ψ_2 are vector potentials of the displacements of the skeletal frame and fluid, respectively, μ is the shear modulus, ρ is the bulk density, ρ_f is the density of fluid, m is the mass-coupling factor, l is the complex wave number, ω is the angular frequency, η is the dynamic viscosity of the pore fluid, F is a the viscosity correction factor which incorporates the frequency dependence of viscous resistance, and k_s is the permeability.

The determinant of the coefficients in Eq. (1) must be zero to obtain a solution that is different from the trivial solution. The resulting frequency equation for the admissible solutions to Ψ_1 and Ψ_2 is

$$\begin{vmatrix} \rho\omega^2 - \mu l^2 & -\rho_f\omega^2 \\ -\rho_f\omega^2 & m\omega^2 - i\omega F\eta/k_s \end{vmatrix} = 0. \tag{2}$$

In the case of homogeneous plane waves, the root of the above equation gives the complex wave number of the shear wave as

$$l^2 = \frac{\rho\omega^2(m - \rho_f^2/\rho - iF\eta/k_s\omega)}{\mu(m - iF\eta/k_s\omega)}. \tag{3}$$

Considering that not all of the pore fluid moves in the direction of the macroscopic pressure gradient due to the shape and orientation of the pores, Biot defined the mass-coupling parameter m as

$$m = C_\alpha \frac{\rho_f}{\beta} \quad (C_\alpha > 1), \tag{4}$$

where C_α is the mass-coupling coefficient and β is the porosity. The mass-coupling coefficient C_α has a theoretical value of 3 for a skeletal frame with isotropic porosity.[3] Biot also derived analytical expressions for the viscosity correction factor F based on a study of different microvelocity fields within the pore channels with simple geometries. The resulting complex correction factor is given as a function of a dimensionless parameter κ, which is defined as the ratio of the pore size parameter a to the oscillatory boundary layer thickness, $\kappa = a[\omega/(\eta/\rho_f)]^{1/2}$. The pore size parameter a can be related to porosity and permeability as

$$a = (C_s k_s/\beta)^{1/2}, \tag{5}$$

where C_s (corresponding to δ in Ref. 2) is the structural factor which includes the effects of both arbitrary pore shape and sinuosity. The structural factor takes a value of 8 for parallel capillary tubes with a radius a or a value of 3 for the case of a two-dimensional duct with a spacing of $2a$. Since the frequency dependence of the viscosity correction factor does not depend strongly on the shape of the pores, analytical expressions derived for both cases can also be used for arbitrary pore shapes with a properly chosen structural factor C_s. In fact, as Biot indicated, a C_s value of 8 for the parallel capillary tubes and a value of $\frac{16}{3}$ for two-dimensional ducts provides almost identical frequency dependence of the viscosity correction factor.

The parameter κ can be written by using Biot's[4] definition of critical frequency, $f_c = \beta\eta/2\pi\rho_f k_s$, as

$$\kappa = (C_s f/f_c)^{1/2}. \tag{6}$$

Using variational and Lagrangian methods, Biot has shown that without introducing any significant error, the viscosity correction factor for the two-dimensional duct model can be approximated as $F[(C_s f/f_c)^{1/2}] \approx 1 + iC_s(f/f_c)/15$ (Eq. 3.18 of Ref. 8) in the frequency range $0 < f/f_c < 12/C_s$. By using normal viscodynamic coordinates, he obtained a similar expression for frequencies up to two or three times the critical frequency by retaining the first term of a rapidly convergent series. He gave as the physical reason for this rapid convergence that, at a given frequency, predominant excitation occurs only for the viscodynamic modes whose characteristic relaxation times are larger than the period of the oscillation. He also noted that expansion of exact expressions for the viscosity correction factor should give similar results by limiting the expansion to order κ^2. If the pore space is modeled as two-dimensional ducts, the exact expression of the viscosity correction factor is given as

$$F(\kappa) = \frac{1}{3}\frac{i^{1/2}\kappa \tanh i^{1/2}\kappa}{1 - (1/i^{1/2}k)\tanh i^{1/2}\kappa}. \tag{7}$$

Using $\tanh z = z - z^3/3 + 2z^5/15 - \cdots$, the above expression can be approximated to order κ^2 as

$$F(\kappa) \approx \frac{1 - i\kappa^2/3}{1 - 2i\kappa^2/5} \approx 1 + \frac{i\kappa^2}{15}. \tag{8}$$

In Fig. 1, approximate and exact forms of the viscous correction factor from the two-dimensional model are plotted as a function of frequency for soft and hard sediments with the properties adopted from Stoll and Kan[10] and Badiey *et al.*[4] The approximate and exact values are virtually identical and differences for both soft and hard sediments can not be distinguished below two or three times the critical frequency. Similarly, if the pore space is modeled as parallel capillary tubes, the exact expression of the viscosity correction factor is given as

$$F(\kappa) = \frac{1}{4}\frac{\kappa K(\kappa)}{1 - (2/i\kappa)K(\kappa)}, \tag{9}$$

where

$$K(x) = \frac{\mathrm{ber}'x + \mathrm{bei}'x}{\mathrm{ber}x + \mathrm{bei}x}, \tag{10}$$

in which ber and bei are the real and imaginary parts of the Kelvin functions and ber′ and bei′ are their derivatives. Using power series expansion of Kelvin functions,[11] we obtain

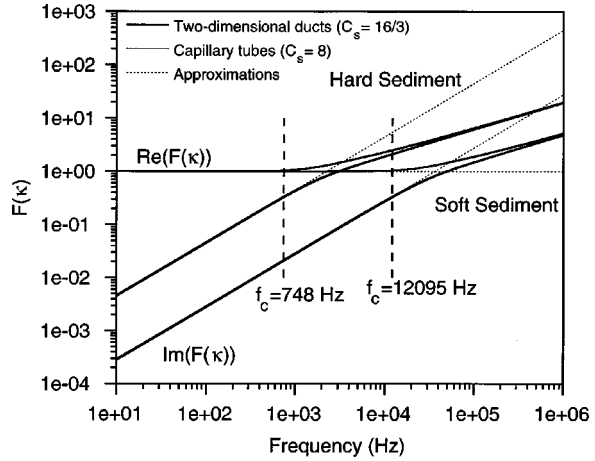


FIG. 1. Approximate and exact viscosity correction factor F for hard and soft sediments with $C_\alpha=2.0$. The exact viscosity correction factor is calculated for both two-dimensional model with $C_s=\frac{16}{3}$ and capillary tube model with $C_s=8$. The results are so nearly equal that they are not distinguishable. Approximate viscosity correction factor is calculated using Eq. (8) with $C_s=\frac{16}{3}$. Vertical dashed lines indicate the characteristic frequencies for hard and soft sediments.

$K(x)=ix/2+x^3/16-ix^5/384+\cdots$, which provides an approximate form of Eq. (9) to order κ^2 as

$$F(\kappa)\approx\frac{1}{1-i\kappa^2/24}\approx 1+\frac{i\kappa^2}{24}. \tag{11}$$

For comparison, the exact viscous correction factor for the capillary tube model is also plotted in Fig. 1. In the calculation of the exact viscous correction factors, the structural factors $C_s=\frac{16}{3}$ and $C_s=8$ are used for the two-dimensional and capillary tube models, respectively. The results are sufficiently similar that they are not distinguishable (see the superimposed thick and thin solid lines in Fig. 1). Note that the agreement is slightly better at the high-frequency region since Biot obtained these values from the high-frequency approximations. In low-frequency region, the two-dimensional model with $C_s=5$ further reduces the slight discrepancy with the capillary model (not shown) as our approximate expressions for both models [Eqs. (8) and (11)] suggest.

In the following development, Eq. (8) is used to compare the approximate and exact shear-wave specific attenuation since a two-dimensional duct model was used in the calculation of the exact viscous correction factor [Eq. (7)]. By inserting the approximate form of the viscosity correction factor [Eq. (8)] in Eq. (3), the wave number l can be approximated for the frequency range $0<f/f_c<12/C_s$ as

$$\begin{aligned} l^2&\approx\frac{\rho(2\pi f)^2}{\mu}\,\frac{\tilde m-\rho_f^2/\rho-i\rho_f/(\beta f/f_c)}{\tilde m-i\rho_f/(\beta f/f_c)}\\ &=\frac{\rho(2\pi f)^2}{\mu}\\ &\quad\times\frac{1+(\beta^2/\rho_f^2)\tilde m(\tilde m-\rho_f^2/\rho)(f/f_c)^2-i(\rho_f^2/\rho)(f/f_c)}{1+(\beta^2/\rho_f^2)\tilde m^2(f/f_c)^2}, \end{aligned} \tag{12}$$

where $\tilde m=\rho_f(C_\alpha+C_s/15)/\beta$ is the approximate virtual mass coefficient. Note that if we replace $\tilde m$ by m in the above equation, it corresponds to Biot's original formulation for low-frequency ranges ($0<f/f_c<0.15$) where the Poiseuille flow assumption is valid (see Eqs. 7.5 and 7.6 in Ref. 1). Both Biot's low-frequency formulation and Eq. (12) for the wave number l show similar behavior to that of an anelastic linear element. The complex modulus for an anelastic linear element is given by[12]

$$M(\omega)=M_r\left(\frac{1+\tau_u\tau\omega^2+i\omega(\tau_u-\tau)}{1+\tau^2\omega^2}\right), \tag{13}$$

where M_r is the relaxed elastic modulus, τ_u is the relaxation time of strain at constant stress, and τ is the relaxation time of stress at constant strain. Then, the square of the wave number l for the anelastic linear element can be written as

$$l^2=\frac{\rho\omega^2}{M(\omega)}=\frac{\rho\omega^2}{M_r}\left(\frac{1+\tau_u\tau\omega^2-i\omega(\tau_u-\tau)}{1+\tau_u^2\omega^2}\right). \tag{14}$$

By comparing Eqs. (12) and (14), the corresponding viscodynamic relaxation times can be identified as

$$\tau_u=\frac{\beta\tilde m}{2\pi f_c\rho_f}\quad\text{and}\quad\tau=\frac{\beta}{2\pi f_c\rho_f}\left(\tilde m-\frac{\rho_f^2}{\rho}\right). \tag{15}$$

It is not surprising to observe the similar frequency dependence in the above equations considering that the viscous coupling is also a relaxation process.[13]

For low Q^{-1} values, the attenuation per cycle is approximated as[9]

$$Q^{-1}\approx-\mathrm{Im}(l^2)/\mathrm{Re}(l^2). \tag{16}$$

We obtain specific attenuation Q_3^{-1} for the shear wave from Eq. (12) as

$$Q_3^{-1}\approx\frac{(\beta\rho_f/\rho)(f/f_c)}{(\tilde m\beta^2/\rho_f^2)(\tilde m-\rho_f^2/\rho)(f/f_c)^2+1}, \tag{17}$$

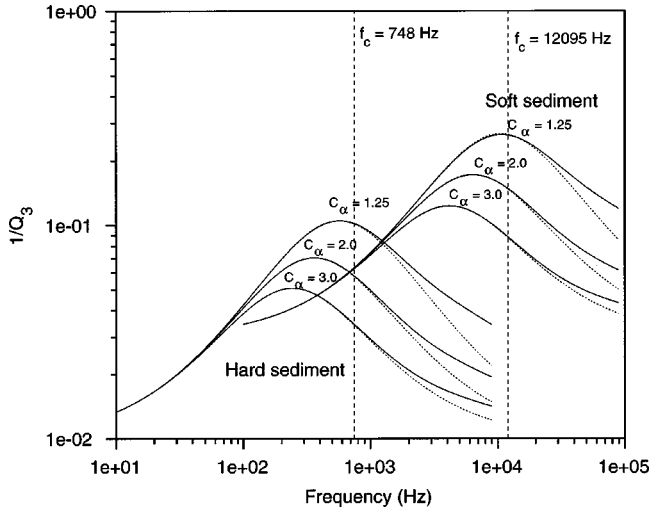


FIG. 2. Approximate and exact shear specific attenuation for hard and soft sediments with $C_s=3.0$ and $C_\alpha=1.25$, 2.0, and 3.0. Solid lines indicate the exact shear specific attenuation. Dotted lines indicate approximate shear specific attenuation calculated from Eq. (22). Vertical dashed lines indicate the characteristic frequencies for hard and soft sediments.

TABLE I. Input sediment parameters (from Refs. 4 and 10).

Parameters	Symbols (units)	Sediment type	
		Hard	Soft
Grain density	ρ_r (kg/m^3)	2650	2650
Fluid density	ρ_f (kg/m^3)	1000	1000
Sediment density	ρ (kg/m^3)	1875	1396
Grain bulk modulus	K_r (N/m^2)	3.6×10^{10}	3.6×10^{10}
Fluid bulk modulus	K_f (N/m^2)	2.25×10^{9}	2.25×10^{9}
Frame bulk modulus	K_s (N/m^2)	4.36×10^{7}	3.69×10^{7}
Shear modulus	μ (N/m^2)	2.61×10^{7}	2.21×10^{7}
Dynamic viscosity of pore fluid	η (kg/m/s)	1.0×10^{3}	1.0×10^{3}
Permeability	k_s (m^2)	1.0×10^{-10}	1.0×10^{-11}
Porosity	β	0.47	0.76
Shear specific loss	δ_s	0.01	0.03
Frame volumetric specific loss	δ	0.0075	0.0225

and the specific attenuation for an anelastic linear element from Eq. (14) as

$$Q^{-1}=\frac{\omega(\tau_u-\tau)}{\tau_u\tau\omega^2+1}. \tag{18}$$

In general, the frequency $f_p=1/(2\pi\sqrt{\tau_u\tau})$ maximizes Q^{-1} in both Eqs. (17) and (18), and readily gives the maximum viscous attenuation per cycle as $Q(f_p)^{-1}=(\tau_u-\tau)/2\sqrt{\tau_u\tau}$. By using the specialized expressions in Eq. (15), we obtain our approximation to the peak frequency of the specific attenuation for the shear wave as

$$f_{p3}=f_c\frac{\rho_f}{\beta}\left(\frac{\rho}{\tilde{m}(\rho\tilde{m}-\rho_f^2)}\right)^{1/2}. \tag{19}$$

By taking a similar approach, and following the derivation given in Eq. (17) of Turgut and Yamamoto[14] with the assumption that the bulk modulus of the grain is considerably greater than the bulk moduli of the skeletal frame and the pore fluid (i.e., Biot elastic moduli H, M, and C are almost equal, which is equivalent to the assumption of suspension-like sediments), the viscous specific attenuation for the fast compressional wave can be approximated as

$$Q_1^{-1}\approx\frac{(\beta/\rho_f)(\rho_f^2/\rho+\rho-2\rho_f)f/f_c}{[(\tilde{m}+\rho-2\rho_f)\beta^2/\rho_f^2](\tilde{m}-\rho_f^2/\rho)(f/f_c)^2+1}, \tag{20}$$

and corresponding peak frequency is obtained as

$$f_{p_1}=f_c\frac{\rho_f}{\beta}\left(\frac{\rho}{(\tilde{m}+\rho-2\rho_f)(\rho\tilde{m}-\rho_f^2)}\right)^{1/2}. \tag{21}$$

Note that f_{p_1} is usually greater than f_{p_3} since the inequality $\rho<2\rho_f$ holds for most marine sediments.

II. EXAMPLES FOR HARD AND SOFT SEDIMENTS

In this section, the approximate solution of the specific attenuation given in Eq. (17) is compared with the exact solution for soft and hard sediments. The peak frequencies are calculated for different values of mass-coupling coefficients by using Eqs. (19) and (21). Errors are then evaluated for the new formulas and those previously used in the literature.[4,5]

In the Biot–Stoll theory,[3] attenuation due to intergranular friction is included by assuming a near-elastic skeletal frame, resulting in the Biot elastic moduli H, C, M, and μ becoming complex constants $\overline{H}$, $\overline{C}$, $\overline{M}$, and $\overline{\mu}$. After replacing the shear modulus μ by $\overline{\mu}=\mu_r+i\mu_i$ in Eq. (12), the specific attenuation given in Eq. (17) is modified by the addition of the shear logarithmic decrement[3] $\delta_s=\mu_i/\mu_r$ in the formulas since $\mu_r\gg\mu_i$, yielding

$$Q_3^{-1}\approx\delta_s+\frac{(\beta_{\rho_f}/\rho)(f/f_c)}{(\tilde{m}\beta^2/\rho_f^2)(\tilde{m}-\rho_f^2/\rho)(f/f_c)^2+1}. \tag{22}$$

In Fig. 2, the specific attenuation given in the above equation is compared with the Biot–Stoll theory predictions for the hard and soft sediments (see Table I). As expected from the range of validity of the approximation for the viscosity correction factor used above, at both low-frequency and transition frequency regions ($f/f_c<12/C_s$), the agreement between the exact and approximate predictions is sufficiently good that the approximate relations can be conveniently used to predict the peak frequencies. The discrepancy for higher frequencies shows the breakdown of the approximation to the viscosity correction factor. In Table II, the peak frequen-

TABLE II. Comparison of estimated and Biot–Stoll peak frequencies of the fast compressional and shear waves for the hard and soft sediments with different mass-coupling coefficient values. Percent error for each estimation is given in parentheses.

	Biot–Stoll peak frequencies (Hz)		Estimated peak frequencies (Hz)			
C_α	f_{p3}	f_{p1}	f_{p3}	f_{p1}	$f_p/f_c=1$	$f_p/f_c=0.628$
			Hard sediment ($C_s=3.0$)			
1.25	574.7	595.4	567.3(1.28)	579.2(2.72)	748.0(−30.15)(−25.63)	470.0(18.22)(21.06)
2.00	362.9	371.3	361.2(0.47)	366.2(1.37)	748.0(−106.12)(−101.45)	470.0(−26.58)(−29.51)
3.00	244.0	247.8	243.5(0.20)	245.8(0.81)	748.0(−206.56)(−201.86)	470.0(−92.62)(−89.67)
			Soft sediment ($C_s=3.0$)			
1.25	10 750	13 480	10 550(1.86)	12 768(5.28)	12 095(−12.51)(10.27)	7600(29.30)(43.62)
2.00	6375	7272	6337(0.60)	7124(2.04)	12 095(−89.72)(−66.32)	7600(−19.22)(−4.51)
3.00	4159	4526	4149(0.24)	4483(0.95)	12 095(−190.82)(−167.23)	7600(−82.74)(−67.92)

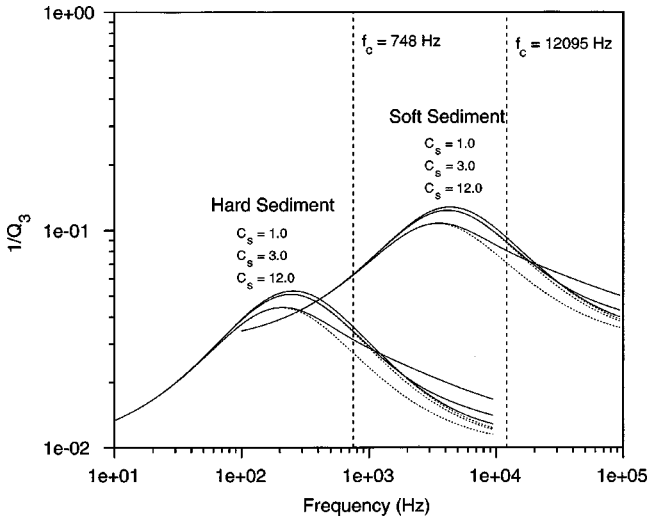


FIG. 3. Approximate and exact shear specific attenuation for hard and soft sediments with $C_\alpha = 3.0$ and $C_s = 1.0$, 3.0, and 12.0. Different structural factor values. Solid lines indicate the exact shear specific attenuation. Dotted lines indicate approximate shear specific attenuation calculated from Eq. (22). Vertical dashed lines indicate the characteristic frequencies for hard and soft sediments.

cies of fast compressional and shear waves calculated from the proposed formulas are compared with exact values for hard and soft sediments. The exact peak frequencies are determined numerically by fine frequency sampling of the Biot–Stoll theory predictions in the transition region. In parentheses we show the percent errors which were calculated as $100\times(f_{p_{ex}}-f_{p_{es}})/f_{p_{ex}}$, where $f_{p_{ex}}$ and $f_{p_{es}}$ are exact and estimated peak frequencies, respectively. The percent errors indicate that the new formulas estimate the peak frequencies within 5% error for fast compressional waves and within 2% for the shear waves.

The estimation errors are primarily due to the assumption of low Q^{-1} values for the shear waves and the assumption of suspensionlike sediments for the fast compressional waves. The estimation errors for both fast compressional and shear waves become larger as the mass-coupling coefficient is decreased. This increase in error is due to the increase in specific attenuation as the virtual mass $\tilde{m}$ is reduced in Eq. (17). In this case, higher values of specific attenuation might violate the assumption of low Q^{-1}, especially for the shear waves (e.g., see Fig. 2 for the case of $C_\alpha = 1.25$ for the soft sediment).

The Biot characteristic frequencies and the peak frequencies calculated using $f_{p_3}/f_c \approx 1.0$ and $f_{p_3}/f_c \approx 0.628$ are also given in Table II. These relations predict a single peak frequency for both fast compressional and shear waves and the percent errors can be as much as 200% for different values of the mass-coupling coefficient. Figure 3 shows the dependency of shear specific attenuation on the structural factor C_s which also can be determined from measured dispersion and attenuation curves. Again, the agreement between the exact and approximate predictions is satisfactory within the frequency range $f < 12\, f_c/C_s$, which is expected to include the peak frequencies for typical C_s values for marine sediments.[15] In Table III the peak frequencies of fast compressional and shear waves calculated from the proposed formulas are compared with Biot–Stoll peak frequency values for $C_s = 1.0$ and $C_s = 12.0$. The estimation error increases but still remains within 5% as C_s is increased from 1.0 to 12.0. The estimation errors for the two previously used formulas also increase as C_s increases. For example, $f_{p_3}/f_c \approx 1.0$ gives 256% and $f_{p_3}/f_c \approx 0.628$ gives 124% estimation errors for the hard sediment with a mass-coupling coefficient of $C_\alpha = 3.0$ and a structural factor of $C_s = 12$.

III. CONCLUDING REMARKS

Approximate expressions for viscous attenuation per cycle for both fast compressional and shear waves in marine sediments are derived and corresponding peak frequencies are given by two closed-form formulas. These new formulas are based on Biot's observation[8] that the viscosity correction factor may be represented by a truncated power series expansion with excellent accuracy for the frequency range of $0 < f/f_c < 12/C_s$. The estimated peak frequencies are within 5% error for both hard and soft sediments considered here. In both formulas, the peak frequencies are related to Biot's characteristic frequency by a correction term which depends on the mass-coupling coefficient, structural factor, and densities of sediment and pore fluid. Biot's characteristic frequency depends on three basic parameters of the sediment namely porosity, permeability, and viscosity of the pore fluid. In principle, most of these parameters and mass-coupling coefficient can be estimated from measured phase dispersion and attenuation data by means of geoacoustic inversion techniques (see, e.g., Ref. 14). For example, one can estimate the mass-coupling coefficient from the ratio of the

TABLE III. Comparison of estimated and Biot–Stoll peak frequencies of the fast compressional and shear waves for the hard and soft sediments with different structural factor values. Percent error for each estimation is given in parentheses.

	Biot–Stoll peak frequencies (Hz)		Estimated peak frequencies (Hz)			
C_s	f_{p3}	f_{p1}	f_{p3}	f_{p1}	$f_p/f_c=1$	$f_p/f_c=0.628$
			Hard sediment ($C_\alpha=3.0$)			
1.0	254.6	258.7	254.6(0.00)	257.0(0.66)	748.0(−193.79)(−189.14)	470.0(−84.60)(−81.68)
12.0	209.9	212.8	203.7(2.95)	205.3(3.52)	748.0(−256.36)(−251.50)	470.0(−123.92)(−120.86)
			Soft sediment ($C_\alpha=3.0$)			
1.0	4350	4752	4349(0.03)	4716(0.76)	12 095(−178.05)(−154.52)	7600(−74.71)(−59.93)
12.0	3556	3835	3439(3.29)	3668(4.35)	12 095(−240.13)(−215.38)	7600(−113.72)(−98.17)

high- and low-frequency fast-compressional speed values, and porosity and density from the low-frequency fast-compressional speed using Gassmann's equations and an empirical relationship between porosity shear modulus.[7,16] Then, the permeability can be estimated from the peak frequency by relating the permeability and the structural factor using the Cozeny–Karman relation.[15] We also note that the agreement between approximate and exact specific attenuation for the frequency range of $0<f/f_c<12/C_s$ indicates a potential of using a simpler model which mimics the Biot–Stoll theory from low to intermediate frequencies. These frequencies cover any of the applications in bottom-interacting underwater acoustics. For high-frequency applications, Biot's high-frequency asymptotic formulas of the sound speed and attenuation (Eqs. 5.23, 5.31, 6.13, 6.17 in Ref. 2) can be easily modified by including proper losses for the skeletal frame.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. The author thanks Stephen Wolf for the helpful suggestions in the preparation of this paper.

[1] M. A. Biot, ''Theory of propagation of elastic waves in a fluid-saturated porous solid, I. Low frequency range,'' J. Acoust. Soc. Am. **28**, 168–178 (1956).
[2] M. A. Biot, ''Theory of propagation of elastic waves in a fluid-saturated porous solid, II. High frequency range,'' J. Acoust. Soc. Am. **28**, 179–191 (1956).
[3] R. D. Stoll, ''Acoustic waves in saturated sediments,'' in *Physics of Sound in Marine Sediments*, edited by L. Hampton (Plenum, New York, 1974).
[4] M. Badiey, A. H.-D. Cheng, and Y. Mu, ''From geology to geoacoustics—Evaluation of Biot-Stoll sound speed and attenuation for shallow water acoustics,'' J. Acoust. Soc. Am. **103**, 309–320 (1998).
[5] T. Yamamoto, ''Propagator matrix for continuously layered porous seabeds,'' Bull. Seismol. Soc. Am. **73**, 1599–1620 (1983).
[6] J. E. White, *Underground Sound* (Elsevier, The Netherlands, 1983), p. 102.
[7] A. Turgut and T. Yamamoto, ''Measurements of acoustic wave velocities and attenuation in marine sediments,'' J. Acoust. Soc. Am. **87**, 2376–2383 (1990).
[8] M. A. Biot, ''Generalized theory of acoustic propagation in porous dissipative media,'' J. Acoust. Soc. Am. **34**, 1254–1264 (1962).
[9] J. Geertsma and D. C. Smit, ''Some aspects of elastic wave propagation in fluid-saturated porous solid,'' Geophysics **26**(2), 169–181 (1961).
[10] R. D. Stoll and T. K. Kan, ''Reflection of acoustic waves at a water-sediment interface,'' J. Acoust. Soc. Am. **70**, 149–164 (1981).
[11] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (U.S. Government Printing Office, Washington, DC, 1972), p. 379.
[12] A. M. Freudenthal and H. Geiringer, ''The mathematical theories of the inelastic continuum,'' in *Encyclopedia of Physics* (Springer-Verlag, Berlin, 1958), Vol. 6, p. 267.
[13] D. D. Jackson and D. L. Anderson, ''Physical mechanisms of seismic-wave attenuation,'' Rev. Geophys. Space Phys. **8**(1), 1–63 (1970).
[14] A. Turgut and T. Yamamoto, ''Synthetic seismograms for marine sediments and determination of porosity and permeability,'' Geophysics **53**(8), 1056–1067 (1988).
[15] J. M. Hovem and G. D. Ingram, ''Viscous attenuation of sound in saturated sand,'' J. Acoust. Soc. Am. **66**, 1807–1812 (1979).
[16] T. Yamamoto, M. Trevorrow, M. Badiey, and A. Turgut, ''Determination of the seabed porosity and shear modulus profiles using a gravity wave inversion,'' Geophys. J. Int. **98**, 173–182 (1989).

Bessel beams of finite amplitude in absorbing fluids

Kevin B. Cunningham and Mark F. Hamilton
Department of Mechanical Engineering, The University of Texas at Austin, Austin, Texas 78712-1063

(Received 17 February 2000; accepted for publication 7 May 2000)

Second-harmonic generation in a Bessel beam is investigated analytically, with emphasis on the effect of absorption. It is shown that absorption has a strong influence on the far-field beam profile. Numerical results are presented for higher harmonics and for waveform distortion in a Bessel beam that forms a shock. © *2000 Acoustical Society of America.* [S0001-4966(00)04208-9]

PACS numbers: 43.25.Lj [FD]

INTRODUCTION

A small-signal Bessel beam possesses a transverse amplitude profile given by $J_0(\rho)$, where J_0 is the zeroth-order Bessel function of the first kind and ρ is a normalized distance from the axis of the beam. The remarkable property of the beam is that it propagates indefinitely without diffraction. A theoretical description of optical Bessel beams published in 1987 by Durnin[1] precipitated a surge of papers on the subject, including applications of the principle to acoustical Bessel beams.[2–4] A brief survey of papers on Bessel beams in both optics and acoustics is provided by Stepanishen and Sun.[5]

Only recently have nonlinear effects been considered in Bessel beams. Two papers on acoustical second-harmonic generation appeared in 1996. In one, Du *et al.*[6] derived a solution for the near field of the second harmonic, where a beam profile given approximately by $J_0^2(\rho)$ was obtained. In the other, Ding and Lu[7] derived a solution for the far field and obtained $J_0(2\rho)$ for the beam profile. In neither analysis was absorption taken into account.

The present paper contains a more general analysis of second-harmonic generation in a Bessel beam, first by investigating solutions for lossless propagation in greater detail, and second by including absorption. It is shown that the far-field beam profile is $J_0(2\rho)$ only for very small values of a characteristic absorption parameter. As the absorption parameter is increased, the beam profile evolves toward a distribution given approximately (i.e., except in the neighborhood of minima) by $J_0^2(\rho)$. Numerical results are presented for higher harmonics and for waveform distortion in a Bessel beam that forms a shock.

I. SECOND-HARMONIC GENERATION

The sound pressure $p(r,z,t)$ in the plane $z=0$ is assumed to be

$$p(r,0,t)=p_0 J_0(r/b)\sin\omega t, \tag{1}$$

where p_0 is the peak source pressure, J_n is the nth-order Bessel function of the first kind, r is distance from the beam axis, b is the characteristic radius of the source, and ω is its angular frequency. To investigate second-harmonic generation in the case of weak nonlinearity we express the sound pressure as $p=p_1+p_2$, where $|p_2|\ll|p_1|$,

$$p_n(r,z,t)=\frac{1}{2j}q_n(r,z)e^{jn(\omega t-kz)}+\text{c.c.},\quad n=1,2, \tag{2}$$

z is distance from the source plane, $k=\omega/c_0$ is the wave number at the source frequency, and c_0 is the sound speed. We thus have

$$q_1(r,0)=p_0 J_0(r/b),\quad q_2(r,0)=0, \tag{3}$$

for the source conditions on q_n.

It is convenient to express the solutions for q_n using the following dimensionless notation:

$$P_n=q_n/p_0,\quad \rho=r/b,\quad \sigma=z/z_0,\quad a_n=\alpha_n z_0,\quad N=\beta k z_0 p_0/\rho_0 c_0^2, \tag{4}$$

where $z_0=kb^2/2$ is a diffraction length, α_n the absorption coefficient at frequency $n\omega$, β the coefficient of nonlinearity, and ρ_0 the ambient density of the fluid. At this stage we do not impose any restriction on the frequency dependence of the absorption coefficient. Dispersion may be taken into account by adding an imaginary part to α_n.

Integral solutions for P_1 and P_2 are obtained from the KZK nonlinear parabolic wave equation,[8]

$$P_1(\rho,\sigma)=\int_0^\infty P_1(\rho',0)G_1(\rho,\sigma|\rho',0)\rho'\,d\rho', \tag{5}$$

$$P_2(\rho,\sigma)=\frac{N}{2}\int_0^\sigma d\sigma' \times\int_0^\infty P_1^2(\rho',\sigma')G_2(\rho,\sigma|\rho',\sigma')\rho'\,d\rho', \tag{6}$$

where

$$G_n(\rho,\sigma|\rho',\sigma')=\frac{j2ne^{-a_n(\sigma-\sigma')}}{\sigma-\sigma'}J_0\left(\frac{2n\rho\rho'}{\sigma-\sigma'}\right)\times\exp\left(-\frac{jn(\rho^2+\rho'^2)}{\sigma-\sigma'}\right). \tag{7}$$

Evaluation of the integral in Eq. (5) with $P_1(\rho,0)=J_0(\rho)$ yields

$$P_1(\rho,\sigma)=J_0(\rho)e^{-a_1\sigma+j\sigma/4}, \tag{8}$$

substitution of which into Eq. (2) gives

$$p_1(r,z,t)=p_0 J_0(r/b)e^{-\alpha_1 z}\sin[\omega(t-z/c_{\text{ph}})], \tag{9}$$

where

$$c_{\rm ph}=\frac{c_0}{1-1/2k^2b^2}. \tag{10}$$

The primary beam thus propagates at phase speed $c_{\rm ph}$ without diffraction. We note that the KZK equation is based on the assumption of large kb.

To calculate the second harmonic it proves helpful to follow Ding and Lu[7] and make use of the identity

$$J_0^2(\rho')=\frac{2}{\pi}\int_0^{\pi/2}J_0(2\rho'\sin\theta)d\theta, \tag{11}$$

after substituting Eq. (8) in (6). The integrals over ρ' and σ' can then be evaluated to obtain

$$P_2(\rho,\sigma)=\frac{2N}{j\pi}e^{-a_2\sigma+j\sigma/2}\times\int_0^{\pi/2}\left(\frac{e^{(a_2-2a_1)\sigma}-e^{-j(\sigma/2)\cos^2\theta}}{\cos^2\theta-j2(a_2-2a_1)}\right)\times J_0(2\rho\sin\theta)d\theta. \tag{12}$$

Although one may set $\rho=0$ in Eq. (12) to investigate the axial field, an alternative integral representation can be useful in this case. Substitute Eq. (8) in (6), set $\rho=0$, perform the integration over ρ', and let $y=\sigma-\sigma'$ to obtain

$$P_2(0,\sigma)=\tfrac{1}{2}Ne^{-2a_1\sigma+j\sigma/2}\times\int_0^{\sigma}e^{-(a_2-2a_1)y-jy/4}J_0(y/4)dy. \tag{13}$$

A. Lossless propagation

In the absence of absorption ($a_n=0$) the integral in Eq. (13) for the axial field is known,[9]

$$P_2(0,\sigma)=\tfrac{1}{2}N\sigma[J_0(\sigma/4)+jJ_1(\sigma/4)]e^{j\sigma/4}. \tag{14}$$

The magnitude of $P_2(0,\sigma)$, normalized by N, is plotted as the solid lines in Fig. 1. The weak oscillations, noted by Du *et al.*[6] in their numerical analysis, extend far from the source. The periodicity of the oscillations is $\Delta\sigma\simeq4\pi$, i.e., approximately half the spatial periodicity of the Bessel functions in Eq. (14).

The field off axis is investigated by setting $a_n=0$ in Eq. (12),

$$P_2(\rho,\sigma)=\frac{2N}{j\pi}e^{j\sigma/2}\times\int_0^{\pi/2}\left(\frac{1-e^{-j(\sigma/2)\cos^2\theta}}{\cos^2\theta}\right)J_0(2\rho\sin\theta)d\theta. \tag{15}$$

This is the expression obtained by Ding and Lu.[7] To examine evolution of the beam in the near field, keep only terms through $O(\sigma^2)$ in the expansion of the exponential in the integrand, and the resulting integrals can be evaluated to obtain

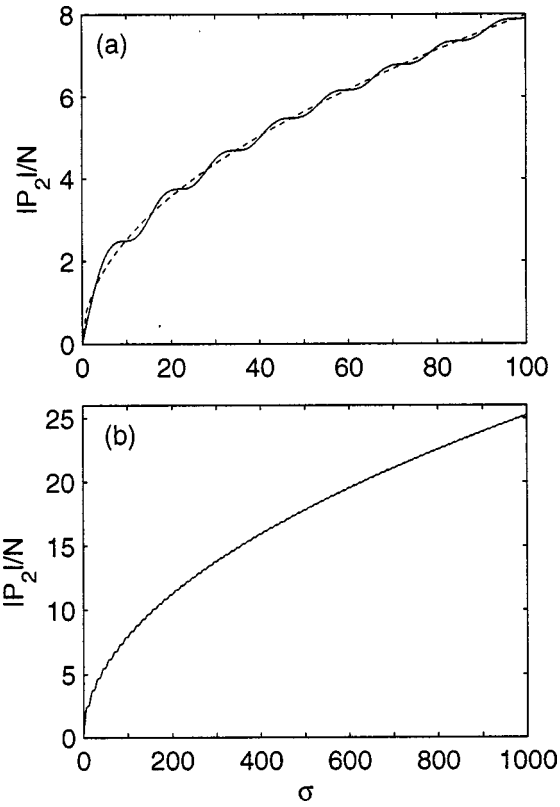


FIG. 1. Axial amplitude of the second harmonic in a lossless fluid. Solid lines: Eq. (14). Dotted line in (a): far-field approximation given by Eq. (18).

$$P_2(\rho,\sigma)\simeq\tfrac{1}{2}N\{\sigma J_0^2(\rho)-j\tfrac{1}{8}\sigma^2[J_0^2(\rho)+J_1^2(\rho)]\}e^{j\sigma/2},\quad \sigma\lesssim1. \tag{16}$$

Away from the zeros of $J_0(\rho)$ the $O(\sigma^2)$ term is small in comparison with the $O(\sigma)$ term, such that a reasonable approximation of Eq. (16) is

$$P_2(\rho,\sigma)\simeq\tfrac{1}{2}N\sigma J_0^2(\rho)e^{j(\sigma/8)[3-J_1^2(\rho)/J_0^2(\rho)]},\quad \sigma\lesssim1. \tag{17}$$

The dependence of the amplitude on $J_0^2(\rho)$ in the near field, seen here to be approximate but accurate away from minima in the beam profile, was obtained by Du *et al.*[6] via another approach. The dependence of the phase on ρ reveals the onset of diffraction, despite the nondiffracting nature of the primary beam.

In the far field, asymptotic analysis of Eq. (15) yields at leading order

$$P_2(\rho,\sigma)\simeq N(2\sigma/\pi)^{1/2}J_0(2\rho)e^{j(\sigma/2-\pi/4)},\quad \sigma\gtrsim100. \tag{18}$$

This result was obtained by Ding and Lu.[7] In Fig. 2, beam profiles obtained from Eqs. (17) (dashed lines) and (18) (dotted lines) establish the regions in which these equations are valid via comparisons with numerical evaluation of Eq. (15) (solid lines). The profiles begin almost equal to $J_0^2(\rho)$ in the near field and evolve to $J_0(2\rho)$ in the far field. The fact that $\sigma\simeq1$ proves to be a transition point for the beam profiles reveals that z_0 is an appropriate diffraction length associated with second-harmonic generation. Because absorption has not been taken into account and the field structure of the primary beam remains constant with propagation distance,

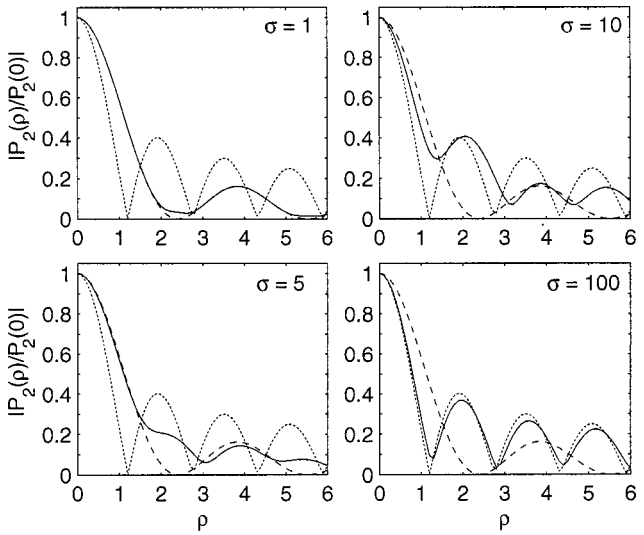


FIG. 2. Beam profiles for the second harmonic in a lossless fluid as a function of distance from the source. Solid lines: Eq. (15). Dashed lines: $J_0^2(\rho)$. Dotted lines: $|J_0(2\rho)|$.

the predicted growth of the second harmonic in the far field is unbounded and proportional to $\sigma^{1/2}$. The magnitude of Eq. (18) along the axis of the beam, $N(2\sigma/\pi)^{1/2}$, is shown as the dotted line in Fig. 1(a). Although it does not account for the oscillations, this asymptotic result provides a reasonable estimate of the entire axial field.

Evolution of the second-harmonic beam profile from approximately $J_0^2(\rho)$ in the near field to $J_0(2\rho)$ in the far field is due to the combined effects of nonlinearity and diffraction. The profile of the primary beam is $J_0(\rho)$ throughout the entire field, and because the square of this field generates the second harmonic, it behaves as a volume distribution of virtual sources with amplitude shading proportional to $J_0^2(\rho)$. Radiation from a source plane with a $J_0^2(\rho)$ amplitude profile diffracts as it propagates. Diffraction causes a wave radiated from a given virtual source plane to eventually diverge, decrease in amplitude, and shift slightly in phase relative to the primary beam. It is the superposition of the diffracting radiation from virtual source planes distant from the observation point, and the second-harmonic components generated close to the observation point and therefore having profiles proportional to $J_0^2(\rho)$, that produces the $J_0(2\rho)$ beam profile in the far field. In contrast, a spherically spreading sound beam in a lossless fluid always generates a second harmonic with a far field directivity proportional to the square of the primary-beam directivity, because the far-field contribution always dominates the near-field contribution at sufficiently large distances.[8] What makes Bessel beams different is that the collimation of the field establishes a balance between these near-field and far-field contributions at the observation point.

It is shown below that absorption disrupts this balance, and the far-field beam profile generally differs from $J_0(2\rho)$.

B. Effects of absorption

It is convenient to introduce the absorption parameter

$$\hat{a} = a_2 - 2a_1. \tag{19}$$

We restrict our attention to the case $\hat{a} > 0$, which corresponds to a fluid for which the absorption coefficient increases with frequency raised to a power greater than unity ($\alpha_\omega \propto \omega^m$, $m > 1$). This criterion includes, for example, thermoviscous fluids ($m = 2$) and soft tissue ($m \simeq 1$). Although dispersion is taken into account by simply augmenting the analysis that follows to include complex absorption coefficients, subsequent discussion is based on the assumption that a_n is real. The integrals in Eqs. (12) and (13) have not been evaluated in closed form with $\hat{a} \neq 0$, and we therefore consider their asymptotic properties.

An analytic approximation of Eq. (12) can be obtained for large values of $\hat{a}$. In this case, the fraction under the integral may be expanded in powers of $\hat{a}^{-1}$. Evaluation of the integrals resulting from the first two terms in the expansion yields

$$\begin{aligned} P_2(\rho,\sigma) \simeq (N/2\hat{a}) e^{-2a_1\sigma + j\sigma/2} \{ & (1 - e^{-\hat{a}\sigma}) J_0^2(\rho) \\ & - (j/4\hat{a})[1 - (1 + \hat{a}\sigma) e^{-\hat{a}\sigma}] \\ & \times [J_0^2(\rho) + J_1^2(\rho)] \}, \quad \hat{a} \gtrsim 1, \end{aligned} \tag{20}$$

which is valid for all σ. The restriction $\hat{a} \gtrsim 1$ was deduced from comparisons with Eq. (12). In the far field, Eq. (20) becomes

$$\begin{aligned} P_2(\rho,\sigma) \simeq (N/2\hat{a}) e^{-2a_1\sigma + j\sigma/2} \{ & J_0^2(\rho) \\ & - (j/4\hat{a})[J_0^2(\rho) + J_1^2(\rho)] \}, \quad \hat{a} \gtrsim 1, \quad \hat{a}\sigma \gtrsim 3. \end{aligned} \tag{21}$$

Away from the zeros of $J_0(\rho)$ it is reasonable to approximate Eq. (20) by

$$\begin{aligned} P_2(\rho,\sigma) \simeq (N/2\hat{a}) J_0^2(\rho) (e^{-2a_1\sigma} - e^{-a_2\sigma}) e^{j[\sigma/2 - \phi(\rho,\sigma)]}, \\ \hat{a} \gtrsim 1, \end{aligned} \tag{22}$$

where

$$\phi(\rho,\sigma) = \frac{1}{4\hat{a}} \left(1 - \frac{\hat{a}\sigma}{e^{\hat{a}\sigma} - 1} \right) \left(1 + \frac{J_1^2(\rho)}{J_0^2(\rho)} \right). \tag{23}$$

Like Eq. (20), Eq. (22) is valid for all σ. For lossless fluids, Eqs. (20) and (22) reduce to Eqs. (16) and (17), respectively, for the near field.

A large value of $\hat{a}$ indicates that a_2 is large in relation to $2a_1$. As a result, second-harmonic radiation from the near field is suppressed in relation to second-harmonic generation closer to far-field observation points, because the former is attenuated as $e^{-a_2\sigma}$, the latter at the slower rate $e^{-2a_1\sigma}$ (i.e., proportional to P_1^2). For $\hat{a} \gtrsim 1$ the imbalance is sufficiently great that the field structure of the second harmonic is determined completely by the local behavior of the primary beam. Specifically, the beam profile is proportional to $J_0^2(\rho)$ (away from minima), and the amplitude decreases as $e^{-2a_1\sigma}$. As seen below, the far-field attenuation rate $e^{-2a_1\sigma}$ is obtained for all $\hat{a} > 0$.

We now proceed to the far-field limit of Eq. (12),

$$P_2(\rho,\sigma)\simeq \frac{N}{j\pi}e^{-2a_1\sigma+j\sigma/2} \times\int_0^{\pi}\frac{J_0(2\rho\sin\theta)d\theta}{\cos^2\theta-j2\hat{a}},\quad \hat{a}\sigma\gtrsim 3, \tag{24}$$

where the upper integration limit was increased from $\pi/2$ to π (with the integral multiplied by 1/2). An analytic approximation of the integral can be obtained for small $\hat{a}$. In this case it is convenient to employ the following representation of the Dirac delta function,

$$\delta(x)=\lim_{\epsilon\to 0}\frac{\epsilon/\pi}{x^2+\epsilon^2}. \tag{25}$$

Setting $x^2=\cos^2\theta$ and $\epsilon^2=-j2\hat{a}$ in Eq. (24) yields

$$P_2(\rho,\sigma)\simeq \frac{NJ_0(2\rho)}{\sqrt{2\hat{a}}}e^{-2a_1\sigma+j(\sigma/2-\pi/4)},\quad \hat{a}\sigma\gtrsim 3,\quad \hat{a}\lesssim 10^{-2}. \tag{26}$$

Note that the condition $\hat{a}\lesssim 10^{-2}$, deduced from comparisons of $J_0(2\rho)$ with beam profiles calculated with Eq. (12), requires σ to be sufficiently large that the inequality $\hat{a}\sigma\gtrsim 3$ remains satisfied. The dependence on $J_0(2\rho)$ occurs because as $\hat{a}\to 0$, attenuation of second-harmonic radiation from the near field ($e^{-a_2\sigma}$) becomes almost identical to attenuation of virtual source planes in the far field ($e^{-2a_1\sigma}$), in which case the balance that yields the profile $J_0(2\rho)$ in Eq. (18) for lossless fluids is virtually unaltered. The condition $\hat{a}\ll 1$ means only that $a_2\simeq 2a_1$, and not that a_1 and a_2 individually are small. This observation is relevant to tissue, for which the dependence of absorption on frequency is nearly linear.

Investigation of beam profiles for arbitrary values of $\hat{a}$ requires numerical analysis. We therefore introduce the following far-field directivity function based on Eq. (24):

$$D(\rho)=|I(\rho)/I(0)|, \tag{27}$$

where

$$I(\rho)=\int_0^{\pi/2}\frac{J_0(2\rho\sin\theta)d\theta}{\cos^2\theta-j2\hat{a}},\quad \hat{a}\sigma\gtrsim 3. \tag{28}$$

From Eqs. (26) and (21) it is seen that

$$D(\rho)\simeq|J_0(2\rho)|,\quad \hat{a}\lesssim 10^{-2}, \tag{29}$$

$$\simeq\left(\frac{J_0^4(\rho)+(4\hat{a})^{-2}[J_0^2(\rho)+J_1^2(\rho)]^2}{1+(4\hat{a})^{-2}}\right)^{1/2},\quad \hat{a}\gtrsim 1. \tag{30}$$

As noted above, away from the minima in the beam profile a good approximation for large $\hat{a}$ is

$$D(\rho)\simeq J_0^2(\rho),\quad \hat{a}\gtrsim 1. \tag{31}$$

Figure 3 illustrates the evolution of $D(\rho)$ (solid lines) from $|J_0(2\rho)|$ (dotted lines) to roughly $J_0^2(\rho)$ (dashed lines) with increasing $\hat{a}$. No significant changes occur outside the range of $\hat{a}$ shown. For simplicity, Eq. (31) was used in Fig. 3 as the reference for large $\hat{a}$, rather than Eq. (30).

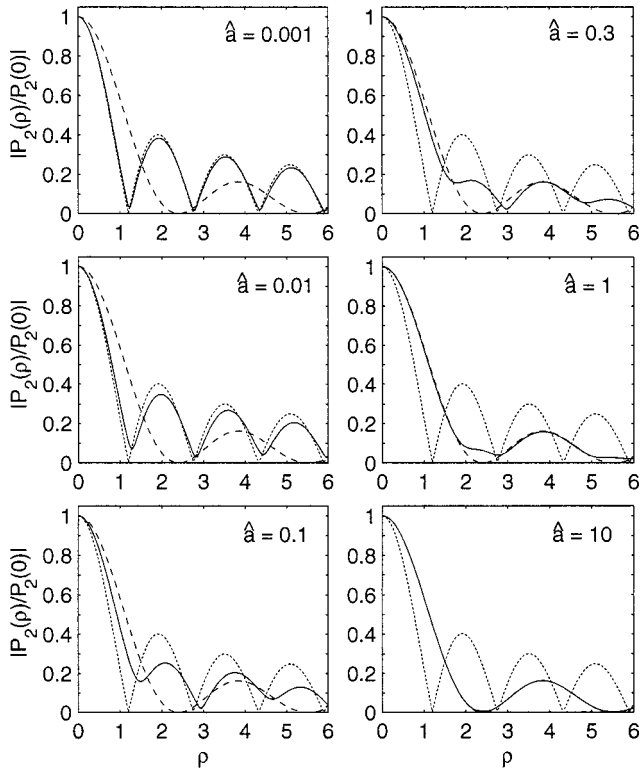


FIG. 3. Far-field beam profiles as a function of absorption. Solid lines: Eq. (27). Dashed lines: $J_0^2(\rho)$. Dotted lines: $|J_0(2\rho)|$.

We conclude this section by deriving a far-field axial solution for all $\hat{a}>0$. Begin by setting $\rho=0$ in Eq. (24) and extending the upper integration limit to 2π,

$$P_2(0,\sigma)\simeq\frac{N}{2\pi j}e^{-2a_1\sigma+j\sigma/2}\int_0^{2\pi}\frac{d\theta}{\cos^2\theta-j2\hat{a}},\quad \hat{a}\sigma\gtrsim 3. \tag{32}$$

Let $\zeta=e^{j\theta}$ to obtain

$$P_2(0,\sigma)\simeq-\frac{2}{\pi}Ne^{-2a_1\sigma+j\sigma/2}\int_C\frac{\zeta d\zeta}{\zeta^4+2(1-j4\hat{a})\zeta^2+1}, \tag{33}$$

where the integral is performed around the unit circle C defined by $|\zeta|=1$, within which the integrand has two simple poles,

$$\zeta_{\pm}=\pm j\sqrt{1-j4\hat{a}+j2\sqrt{j2\hat{a}+(2\hat{a})^2}}. \tag{34}$$

Use of residue theory to evaluate Eq. (33) yields

$$P_2(0,\sigma)\simeq\frac{Ne^{-2a_1\sigma+j\sigma/2}}{\sqrt{j2\hat{a}+(2\hat{a})^2}},\quad \hat{a}\sigma\gtrsim 3. \tag{35}$$

Equation (35) reduces for large $\hat{a}$ to the axial value of Eq. (21), and for small $\hat{a}$ to the axial value of Eq. (26). Combining Eqs. (27) and (35), one may construct the following expression for the amplitude of the sound pressure in the far field:

$$|P_2(\rho,\sigma)|\simeq\frac{ND(\rho)e^{-2a_1\sigma}}{[(2\hat{a})^2+(2\hat{a})^4]^{1/4}},\quad \hat{a}\sigma\gtrsim 3. \tag{36}$$

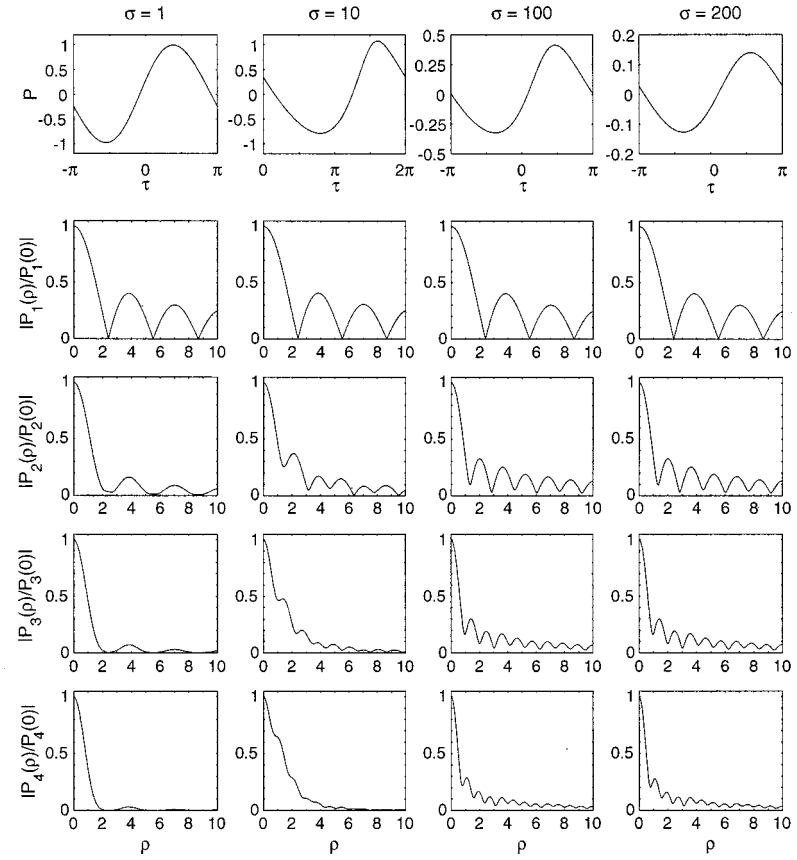


FIG. 4. Simulations obtained from Eq. (37) with $A = 0.01$ and $N = 0.1$. First row: axial waveforms. Second through fifth rows: beam profiles for $n = 1,2,3,4$, respectively.

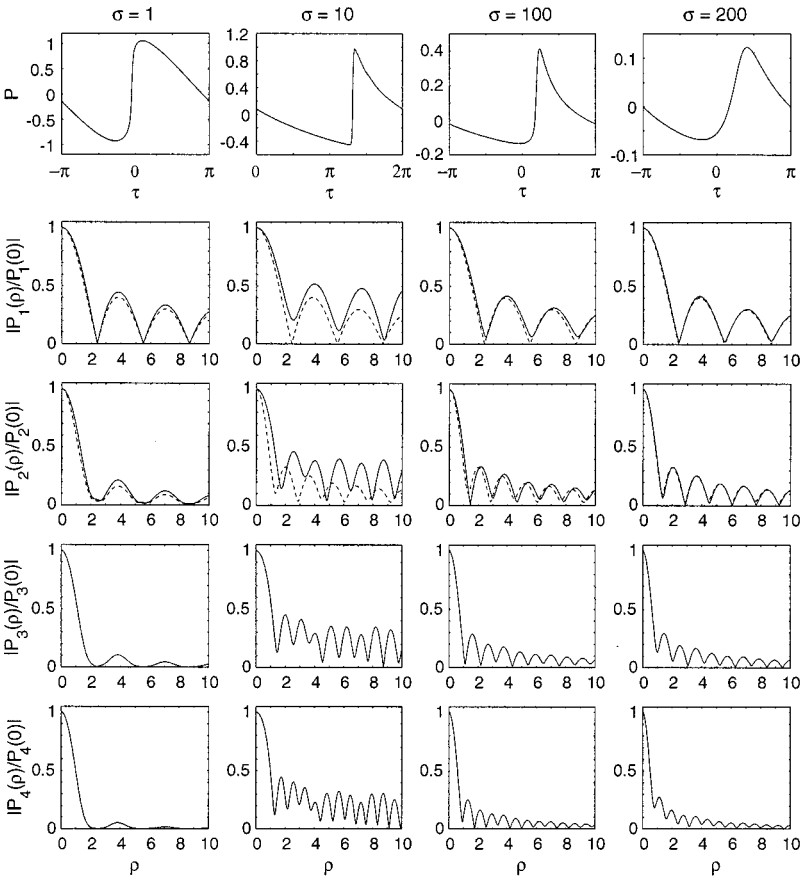


FIG. 5. Simulations obtained from Eq. (37) with $A = 0.01$ and $N = 1$, presented in same format as Fig. 4. Dashed lines in second row are $|J_0(\rho)|$, and in third row they are obtained from Eq. (12).

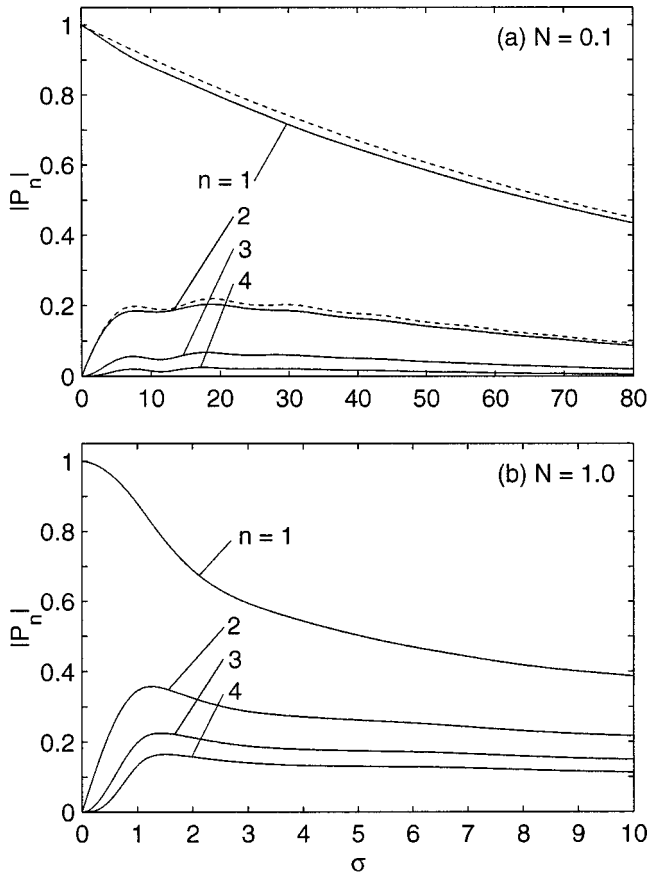


FIG. 6. Axial propagation curves obtained from Eq. (37) with $A=0.01$: (a) $N=0.1$; (b) $N=1$. Dashed lines in (a): $e^{-A\sigma}$ for $n=1$; Eq. (13) for $n=2$.

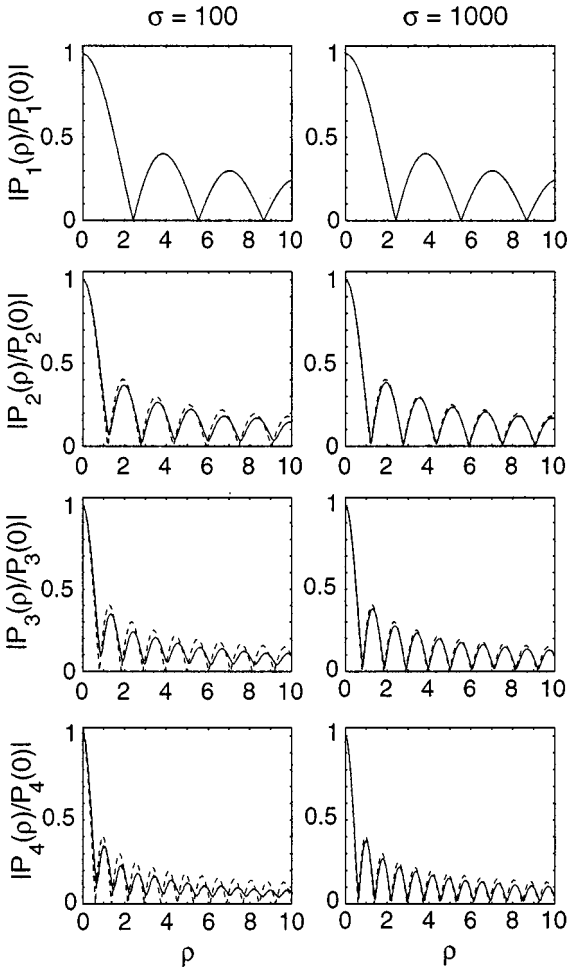


FIG. 7. Beam profiles obtained from Eq. (37) with $A=0$ and $N=0.01$. Dashed lines: $|J_0(n\rho)|$.

II. NUMERICAL RESULTS

In this section we consider harmonics higher than the second, and circumstances for which substantial waveform distortion, including shock formation, occurs during propagation. The source condition of interest remains Eq. (1) but we employ numerical solutions of the KZK equation, written here in the form[10]

$$\frac{\partial P}{\partial \sigma}=\frac{1}{4}\int_{-\infty}^{\tau}\left(\frac{\partial^2 P}{\partial \rho^2}+\frac{1}{\rho}\frac{\partial P}{\partial \rho}\right)d\tau'+A\frac{\partial^2 P}{\partial \tau^2}+NP\frac{\partial P}{\partial \tau}, \quad (37)$$

where $P(\rho,\sigma,\tau)=p/p_0$, $\tau=\omega(t-z/c_0)$, $A=\alpha_1 z_0$, and the remaining parameters are defined in Eqs. (4). The absorption term in Eq. (37) corresponds to thermoviscous fluids, for which $\alpha_n=n^2\alpha_1$ for all harmonics in the beam. In terms of the notation used in Sec. I we have $a_1=A$, $a_2=4A$, and $\hat{a}=2A$. Numerical solutions were obtained with the algorithm described in Refs. 10 and 11. The spectral representation of P that is consistent with Eq. (2), and which is used for plotting harmonic propagation curves and beam profiles, is given by

$$P(\rho,\sigma,\tau)=\frac{1}{2j}\sum_{n=1}^{\infty}P_n(\rho,\sigma)e^{jn\tau}+\text{c.c.} \quad (38)$$

We begin by setting $A=0.01$, for which small-signal absorption is negligible over the diffraction length z_0, and the criterion $\hat{a}\sigma\gtrsim 3$ identified in Sec. I as the far field of the second harmonic corresponds to $\sigma\gtrsim 150$. Two values of the nonlinearity parameter are considered, $N=0.1$ (weak nonlinearity) in the first simulation, and $N=1$ (for which a shock is formed) in the second.

The results for $N=0.1$ are presented in Fig. 4, where the four columns correspond to the distances $\sigma=1$, 10, 100, 200. Axial waveforms are shown in the first row. Only very slight waveform distortion is observed, most noticeable at $\sigma=10$, and by $\sigma=200$ absorption has nearly returned the waveform to its original sinusoidal shape. The beam profiles for P_1 (second row) are given everywhere by linear theory, $J_0(\rho)$ (the comparisons are indistinguishable and not shown), as expected for weak nonlinearity. Likewise, the beam profiles for P_2 (third row) are indistinguishable from calculations based on Eq. (12). The fourth and fifth rows show the beam profiles for P_3 and P_4, respectively. In the near field, $\sigma\lesssim 1$, the profile for P_n is approximately $J_0^n(\rho)$, as expected for weak nonlinearity if beam diffraction is not taken into account. In the far field, the profile for P_n possesses n times as many sidelobes as there are for P_1.

The results for $N=1$ are presented in Fig. 5. The waveforms reveal that the shock formation distance is $\sigma\simeq 1$, as predicted for a plane wave with $A/N\ll 1$. A steep shock front is observed at $\sigma=10$, and it is sustained out to $\sigma\simeq 100$. By $\sigma=200$ the shock has been eroded by absorption, and the

waveform thereafter becomes increasingly sinusoidal in appearance. The beam profiles for P_1 and P_2 (solid lines) are in reasonably good agreement with the dashed lines corresponding to $J_0(\rho)$ and Eq. (12), respectively, except at $\sigma=10$. As in Fig. 4, the beam profiles for P_n contain n times as many sidelobes as there are for P_1, even in the shock region. Note that at $\sigma=10$ the sidelobes in the profiles shown in Fig. 5 for P_3 and P_4 are considerably more developed than in Fig. 4.

Axial propagation curves obtained from Eq. (37) for $N=0.1$ and $N=1$ are presented in Figs. 6(a) and 6(b), respectively, again with $A=0.01$. The dashed lines in Fig. 6(a) are given by $e^{-A\sigma}$ for P_1 [i.e., by Eq. (8)] and by Eq. (13) for P_2. In Fig. 6(b), Eqs. (8) and (13) (not shown) diverge from the solutions for P_1 and P_2 beyond the shock formation distance at $\sigma\simeq 1$.

We conclude with an observation about far-field beam profiles when nonlinearity is weak and absorption can be ignored entirely. Ding and Lu[12] performed an asymptotic analysis that predicts the far-field beam profile for P_n to be $J_0(n\rho)$ in this case, a result that may be compared with numerical solutions of Eq. (37). Beam profiles calculated at $\sigma=100$ and $\sigma=1000$ for $A=0$ and $N=0.01$ are shown as the solid lines in Fig. 7, where the dashed lines correspond to $J_0(n\rho)$. The profiles indeed tend toward $J_0(n\rho)$ in this particular case, with the distance for convergence increasing with n. For σ sufficiently large, however, the assumption that both nonlinearity is weak and absorption is negligible must be abandoned, in which case the beam profiles may be expected to deviate from $J_0(n\rho)$.

III. SUMMARY

Integral expressions obtained for the second harmonic in a Bessel beam reveal that its field structure is determined by the dimensionless absorption parameter $\hat{a}=(\alpha_2-2\alpha_1)z_0$. As $\hat{a}$ increases from zero to unity, the far-field beam profile for the second harmonic evolves from $J_0(2\rho)$ to, away from minima, $J_0^2(\rho)$. Numerical solutions of the KZK equation reveal that when shock formation occurs, the integral expressions provide good approximations for the second-harmonic beam profiles outside the shock region.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research.

[1] J. Durnin, ''Exact solutions for nondiffracting beams. I. The scalar theory,'' J. Opt. Soc. Am. A **4**, 651–654 (1987).

[2] D. K. Hsu, F. J. Margetan, and D. O. Thompson, ''Bessel beam ultrasonic transducer: Fabrication method and experimental results,'' Appl. Phys. Lett. **55**, 2066–2068 (1989).

[3] J.-y. Lu and J. F. Greenleaf, ''Ultrasonic nondiffracting transducer for medical ultrasonics,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**, 438–447 (1990).

[4] J. A. Campbell and S. Soloway, ''Generation of a nondiffracting beam with frequency-independent beamwidth,'' J. Acoust. Soc. Am. **88**, 2467–2477 (1990).

[5] P. R. Stepanishen and J. Sun, ''Acoustic bullets: Transient Bessel beams generated by planar apertures,'' J. Acoust. Soc. Am. **102**, 3308–3318 (1997).

[6] G. Du, Y. Zhang, and Z. Zhu, ''Nonlinear distortion of a non-diffraction ultrasonic field,'' in *Nonlinear Acoustics in Perspective: Proceedings of the 14th International Symposium on Nonlinear Acoustics*, edited by R. J. Wei (Nanjing University Press, Nanjing, 1996), pp. 189–194.

[7] D. Ding and Z. Lu, ''The second harmonic component in the Bessel beam,'' Appl. Phys. Lett. **68**, 608–610 (1996). See further elaboration by D. Ding and Z. Lu, *ibid.* **71**, 723–724 (1997).

[8] M. F. Hamilton, ''Sound beams,'' in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, Boston, 1998), Chap. 8.

[9] Y. L. Luke, ''Integrals of Bessel functions,'' in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1965), Chap. 11.

[10] Y.-S. Lee and M. F. Hamilton, ''Time-domain modeling of pulsed finite-amplitude sound beams,'' J. Acoust. Soc. Am. **97**, 906–917 (1995).

[11] S. J. Younghouse, ''Acoustic streaming at high Reynolds numbers produced by focused sound beams with shocks in real fluids,'' Master's thesis, The University of Texas at Austin, 1998.

[12] D. Ding and J.-y. Lu, ''Higher-order harmonics of limited diffraction Bessel beams,'' J. Acoust. Soc. Am. **107**, 1212–1214 (2000).

Calculations of internal-wave-induced fluctuations in ocean-acoustic propagation

Stanley M. Flatté and Galina Rovner
Physics Department and Institute of Tectonics, University of California at Santa Cruz, Santa Cruz, California 95064

(Received 1 February 1999; accepted for publication 18 April 2000)

Variability in the ocean sound-speed field on time scales of a few hours and horizontal spatial scales of a few kilometers is often dominated by the random, anisotropic fluctuations caused by the internal-wave field. Results have been compiled from analytical approaches and from numerical simulations using the parabolic approximation into an efficient set of algorithms for calculating approximations to internal-wave effects on temporal and spatial coherences, coherent bandwidths, and regimes of acoustic fluctuation behavior. These approximate formulas account for the background, deterministic, sound-speed profile and the anisotropy of the internal-wave field, and they also allow for the incorporation of experimentally determined profiles of sound speed, buoyancy frequency, and sound-speed variance. The algorithms start from the geometrical-acoustics approximation, in which the field transmitted from a source can be described completely in terms of rays whose characteristics are determined by the sound speed as a function of position. Ordinary integrals along these rays provide approximations to acoustic-fluctuation quantities due to the statistical effects of internal waves, including diffraction. The results from the algorithms are compared with numerical simulations and with experimental results for long-range propagation in the deep ocean. [S0001-4966(00)00208-3]

PACS numbers: 43.30.Cq, 43.30.Re, 43.30.Qd [DLB]

INTRODUCTION

It is now well known that variability in the ocean sound-speed field on time scales of a few hours and horizontal spatial scales of a few kilometers is often dominated by the random, anisotropic variations caused by the internal-wave field. Internal-wave variability combined with the effects of the background deterministic sound-speed ocean waveguide, determine important temporal and spatial fluctuations of an acoustic signal transmitted in the ocean.[1,2]

In the geometrical-acoustics approximation, the field transmitted from a source can be described completely in terms of rays, whose characteristics are determined by the sound speed as a function of position (the ocean acoustic waveguide).

We have developed a code we call ''CAFI'' (Computation of Acoustic Fluctuations from Internal waves). This code first calculates geometry of the rays and the behavior of ray separations (the deterministic focusing and defocusing of ray bundles in the ocean waveguide). Then, the code provides estimates for acoustic-fluctuation quantities due to the statistical effects of internal waves. These estimates involve numerical integrals along rays, and they take into account first-order diffraction effects by means of path-integral techniques.[1] The Garrett–Munk internal-wave spectrum[3] is built into the code.

Section I describes our model of internal waves, including their spectrum and correlation lengths. Section II defines the acoustic-fluctuation quantities that will be calculated in this paper. Section III describes the geometry of rays in the ocean sound channel, including the behavior of ray separation functions. Section IV describes the different regimes of acoustic-fluctuation behavior in terms of two general parameters calculable from the statistics of the internal-wave field: Φ, the strength parameter and Λ, the diffraction parameter. Sections V, VI, VII, and VIII give results for the calculations of spatial scales, frequency scales, temporal scales, and intensity higher moments. Section IX describes two limits on the validity of our path-integral algorithms: one due to anisotropy and the other due to inhomogeneity. The anisotropy limits can be expressed in terms of lower limits on the acoustic frequency. That is, calculations for frequencies below these limits may violate conditions of validity. The inhomogeneity criteria are expressed in terms of numbers that should be less than unity for validity. Finally, Sec. X compares the calculations from our efficient algorithms to computationally intensive numerical simulations and to experimental observations in two long-range, deep-ocean experiments.

We remark here that the algorithms have been compared with other experiments over the years with some success. Those other experiments have ranged in frequency from a few tens to a few thousand Hertz; in range from a few kilometers to a few megameters, and in water depth from a few hundred to a few thousand meters. Unfortunately, despite the extent of these ranges in parameters, the number of comparisons to experiments has been so small, and the accompanying environmental knowledge has been so limited, that we cannot make very strong statements about the accuracy of these expressions.

The code has at least two major limitations; first, it assumes a range-independent deterministic sound channel (on top of which the internal waves vary) in order to achieve much of its efficiency. Because of range independence, many parts of the calculations can be done over a range of one double loop (about 50 kilometers), rather than the range of the experiment (which might be several thousand kilome-

ters), thus saving large amounts of computer time. This limitation is of course less and less important the shorter the range of interest, but even for very long ranges, this approximation does not seem to introduce major error in statistical quantities.

The second limitation is the assumption of the Garrett–Munk (GM) internal-wave spectrum.[3] Long-range, integrating acoustic experiments must depend on internal consistency of measured acoustic variables to determine whether this assumption is valid in any particular case. We hope that enough such experiments will be done in the future to test this assumption in more detail.

We have tried to express most of the algorithms in terms of pulse propagation rather than single-frequency behavior. Given a specific frequency, it is possible to calculate the strength parameter Φ; the formulas we use depend on Φ being large compared with unity. For example, the application of these formulas to a very detailed short-range (35 km), high-frequency (3200 Hz), pulsed experiment (part of the Azores Fixed Acoustic Range (AFAR) experiment[4]) was quite successful. In that experiment, $\Phi \approx 50$. The observed pulse wander and pulse spread were both about 2 ms, as predicted. However, extrapolating that success to the long-range (3200 km), low-frequency (75-Hz) Acoustic Thermometry of Ocean Climate (ATOC) experiment[5] has not been as successful. In that experiment, $\Phi \approx 5$, and the observed and predicted pulse spreads were about 3 and 300 ms, respectively, thus differing by a factor of 100. The cause of this discrepancy is not known; it could be partially an effect of the low Φ value. We note that the transmitted pulse in AFAR had a Gaussian rms deviation of 0.5 ms, while the corresponding number for the transmitted pulse in the ATOC experiment was 10 ms. But, we don't know how to translate this knowledge into a change in our pulse-spread predictions. We are forced to conclude that at this time, with one success and one failure, our algorithm for calculating the spread of a single pulse has not been verified (as opposed to other quantities, such as the travel-time wander τ, and the rms wavefront tilts).

I. INTERNAL WAVES

The most important source of ocean variability on time scales from a few minutes to days is internal waves. If one ignores other effects, such as mesoscale eddies and fronts, then the field of sound speed can be expressed as

$$C(\mathbf{x},t) = C_0[1 + U_0(z) + \mu(\mathbf{x},t)], \tag{1}$$

where C_0 is a reference sound speed, $C(z) = C_0[1 + U_0(z)]$ is the depth-dependent average sound-speed profile, and $\mu(\mathbf{x},t)$ is a random, zero-mean fluctuation field caused by internal waves.

The dynamics of internal waves are controlled by the gradient of density in the ocean, which is conveniently measured by the buoyancy frequency $n(z)$ (often called the Brunt–Väisälä frequency). The buoyancy frequency typically varies from 3 cph near the surface to 0.2 cph near the bottom; the fractional rate of decrease diminishes with depth. We need the entire $n(z)$ profile for a complete characterization of internal waves.

Returning to Eq. (1), we will need the complete behavior of $\mu(\mathbf{x},t)$ in order to determine the acoustic effects due to internal waves. This behavior is determined from the displacement $\zeta(x,t)$ of an isodensity surface due to internal waves by the formula

$$\mu(\mathbf{x},t) = \zeta(x,t)[\partial_z U_0 - \gamma_A], \tag{2}$$

where γ_A is the adiabatic sound-speed gradient, which is a function of temperature, salinity, and pressure.

Internal waves in the ocean are distributed over modes, with different frequencies and vertical mode numbers. Each of these modes affects acoustic propagation in a slightly different way, and it requires integration over all the modes to predict any given acoustic-fluctuation effect.

Garrett and Munk have provided us with a model for the internal-wave field in the ocean.[3] Their model provides a depth profile of displacement variance, combined with a normalized distribution over frequency and vertical mode number.

Equation (2) provides us with a way to relate the variances of sound-speed fluctuation and internal-wave displacement

$$\langle \mu^2 \rangle = \langle \zeta^2 \rangle [\partial_z U_0 - \gamma_A]^2. \tag{3}$$

By use of the Wentzel-Kramers-Brillouin-Jordan (WKBJ) approximation for internal-wave vertical wave functions, one arrives at a depth dependence for ζ, which allows us to generate an explicit depth profile of $\langle \mu^2 \rangle$

$$\langle \mu^2 \rangle = \langle \zeta_0^2 \rangle \frac{n_a}{n(z)} [\partial_z U_0 - \gamma_A]^2, \tag{4}$$

and ζ_0 is a reference internal-wave displacement that is observed at a depth at which the buoyancy frequency is the reference value n_a, taken to be 3 cph. Garrett and Munk suggested that the reference ζ_0 be 7.3 m.

Equation (4) provides us with a way of calculating the profile of $\langle \mu^2 \rangle$ from the profiles of sound speed and buoyancy frequency, and values of ζ_0 and γ_A. However, in certain cases the $\langle \mu^2 \rangle$ profile is available from direct observation, and our calculations can incorporate such observations.

In order to calculate internal-wave effects on acoustics, we need more than the profiles of sound speed, buoyancy frequency, and $\langle \mu^2 \rangle$. We also need the distribution of internal waves over frequency and mode number. This is provided by the Garrett–Munk model:[3]

$$S(\omega, j) = \langle \mu^2 \rangle \frac{N_\omega \omega_i (\omega^2 - \omega_i^2)^{1/2}}{\omega^3} \frac{N_j}{(j^2 + j_*^2)}, \tag{5}$$

where ω_i is the inertial frequency (one cycle per day at 30-deg latitude), j_* is an empirical constant, usually taken to be 3, and N_j and N_ω are normalization factors such that the integral over frequency and the sum over j of S yield $\langle \mu^2 \rangle$.

Given the above information about internal waves, we can calculate important quantities associated with the effect of internal waves on sound propagation. For example, the correlation lengths of internal waves along different directions are internal-wave quantities of direct interest for acoustic propagation.[1,6]

In order to calculate an effective vertical correlation length for internal waves, an approximate dispersion relation is needed. This is provided by the following expression for the components of the wave vector of an internal wave relative to the direction of propagation (θ) of an acoustic ray:[1]

$$k_v = \alpha j n(z), \tag{6}$$

$$k_h = \alpha j(\omega^2 - \omega_i^2)^{1/2}, \tag{7}$$

$$k_y = \alpha j(\omega^2 - \omega_i^2 - n(z)^2 \tan^2\theta)^{1/2}, \tag{8}$$

where k_h is the horizontal wave-vector component in the direction of the propagation, k_y is the horizontal wave-vector component perpendicular to the propagation, and α is a constant that depends on the profile of buoyancy frequency

$$\alpha^{-1} = \frac{1}{\pi}\int_0^D n(z)dz. \tag{9}$$

In some of our formulas for acoustic-fluctuation quantities we will need averages over the spectrum of ω^2, k_v^2, or other quantities. Such averages are indicated by curly brackets, and are defined in the following way: The average of a general quantity, represented by the symbol a in this case, is given by

$$\{a\} = \frac{\Sigma_j \int d\omega\, a\,(2/k_y)\, S(\omega,j)}{\Sigma_j \int d\omega\,(2/k_y)\, S(\omega,j)}. \tag{10}$$

Thus, for example, we may define a vertical correlation length L_v by $L_v^{-2} = \{k_v^2\}$.

The correlation length $L_p(\theta,z)$ (''p'' refers to the correlation length parallel to the ray) of the sound-speed fluctuations along a ray oriented at angle θ and at depth z is needed for calculations of many statistical properties of the ray. We calculate L_p using an empirical formula that accounts for ray curvature as well as orientation[6]

$$L_p = L_{p0}\frac{[1-\exp[-(\sigma_c/\sigma)^p]]^q}{1+(\rho_s/\rho_c)^{q_a}(\sigma_c/\sigma)^{p_a}}, \tag{11}$$

where ρ_c, σ_c, p, p_a, q, q_a are empirical constants: $\rho_c = 3.5$, $\sigma_c = 0.0204$, $p = 0.385$, $p_a = 0.5$, $q = 1.3$, $q_a = 2.0$, and σ is a dimensionless parameter that represents the ratio between an internal-wave length scale r_i and the ray radius of curvature r at depth z

$$\sigma = \frac{r_i}{r}; \quad r_i = \frac{3}{4B\alpha^2\omega_i^2}, \tag{12}$$

where B is the scale height of the buoyancy-frequency profile ($\approx$1000 m). A typical value of r_i is 1000 km. The dimensionless parameter ρ_s is related to the anisotropy of internal-wave spectrum: $\rho_s = (n(z)\tan\theta)/\omega_i$.

The empirical formula for L_p applies for $\sigma \geqslant 0.3$. If $\sigma > 0.3$ then the values for $\sigma = 0.3$ are taken.

Our treatment of the ocean medium includes the effects of anisotropy, statistical inhomogeneity, the background sound channel, and the empirically observed internal-wave spectrum. The experimental information about the medium that is required to numerically evaluate the various expressions we give here consists of three depth profiles: sound speed, buoyancy frequency, and sound-speed variance. Figure 1 shows these three profiles for a typical midlatitude experiment (Slice89) in the North Pacific.[7] Our expressions also require ray-separation behavior in the form of Green's functions calculated from the geometry of the propagating rays.

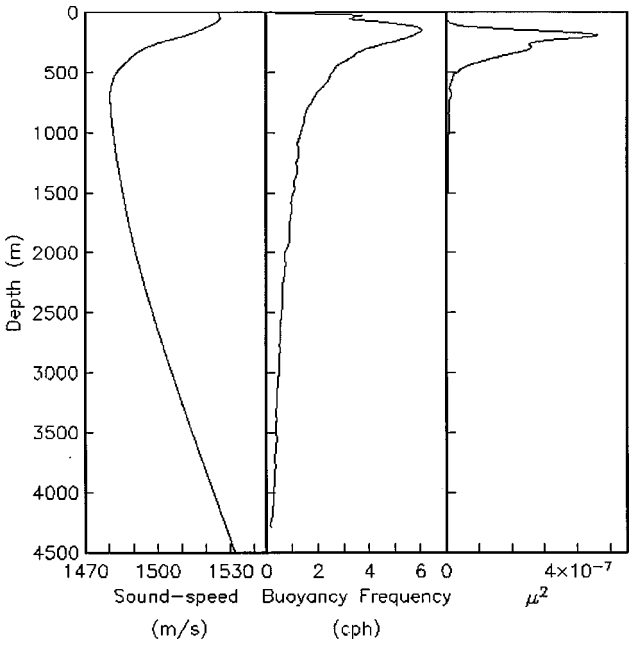


FIG. 1. Profiles of sound speed, buoyancy frequency, and fractional sound-speed variance for the Slice89 experiment.

II. STATISTICAL QUANTITIES OF ACOUSTICAL INTEREST

The general character of fluctuations in a wave field is controlled by two parameters: Φ representing strength and Λ representing size (spatial extent) of inhomogeneities. Thus, these two parameters depend on the internal-wave parameters, as well as the geometry of the acoustic propagation case under consideration and the acoustic frequency σ. The regime of sound transmission (e.g., unsaturated or saturated) is dependent on the values of Φ and Λ. (It is important to note that these quantities are single-frequency quantities. Some parts of broadband behavior are not well characterized by Λ and Φ at present.)

A. Field coherence scales

Let $\psi(0)$ be the observed acoustic field at a given reference time, position, and acoustic frequency. The coherence of the field is described by normalized second moments,[1,4] of which the most important are given in the following equations. The left-hand side is the moment of interest, while the right-hand side is the path-integral approximation. Some of the moments can be expressed more accurately in terms of the phase-structure function $D(\Delta t,\Delta z,\Delta y)$.[8] The Δ quantities here and below are separation variables from the reference values of space, time, and frequency. These are normalized second moments in the sense that the values at zero separation are unity because the mean intensity has been set to unity.

$$\langle \psi^*(\Delta z)\psi(0)\rangle = \exp[-\tfrac{1}{2}D(0,\Delta z,0)] \approx \exp\left[-\frac{1}{2}\left(\frac{\Delta z}{z_0}\right)^2\right], \quad (13)$$

$$\langle \psi^*(\Delta y)\psi(0)\rangle = \exp[-\tfrac{1}{2}D(0,0,\Delta y)] \approx \exp\left[-\frac{1}{2}\left(\frac{\Delta y}{y_0}\right)^{3/2}\right], \quad (14)$$

$$\langle \psi^*(\Delta t)\psi(0)\rangle = \exp[-\tfrac{1}{2}D(\Delta t,0,0)] \approx \exp\left[-\frac{1}{2}\left(\frac{\Delta t}{t_0}\right)^2\right], \quad (15)$$

$$\langle \psi^*(\Delta\sigma)\psi(0)\rangle \approx \exp[-\tfrac{1}{2}(\Delta\sigma\tau)^2] \times \exp(i\Delta\sigma\tau_1 - \tfrac{1}{2}(\Delta\sigma\tau_0)^2). \quad (16)$$

The approximations to these moments represented by the right-hand-side terms depend on approximating the phase-structure functions for small argument. These small-argument approximations break down if the Δ quantities are larger than Φ times their characteristic sizes (such as z_0, or t_0). But, the moments have decreased to very small values at those points, as long as Φ is large.

Note that τ in Eq. (16) is the rms travel-time fluctuation of an impulse traveling along the unperturbed ray. The term $\exp[-1/2(\Delta\sigma\tau)^2]$ is equal to $\langle\psi(\sigma)\psi^*(\sigma+\Delta\sigma)\rangle$ if all the microrays were to lie exactly on the unperturbed ray. The additional terms are due to each microray trajectory being displaced from the unperturbed ray; τ_1 represents a shift of the mean arrival time of a pulse; τ_0 describes the pulse spreading due to the differences in arrival times of the different microrays. We remark here that the travel-time fluctuation τ and the strength parameter Φ are very simply related: $\Phi=\sigma\tau$.

B. Intensity coherence time

We describe one fourth moment of the field, which is a second moment of intensity $I=|\psi^2|$

$$\frac{\langle I(\Delta t)I(0)\rangle}{\langle I^2\rangle} = 1+\exp\left[-\left(\frac{\Delta t}{t_I}\right)^2\right]. \quad (17)$$

C. Microray focusing parameter

The higher moments of intensity are approximately given by

$$\langle I^n\rangle \approx N!\langle I\rangle^N\left[1+\frac{N(N-1)}{2}\gamma\right], \quad (18)$$

where γ is called the microray focusing parameter.[9]

D. Validity of path integral

We must examine the validity of two assumptions made during the path-integral derivations. The first is homogeneity: we assumed that quantities like $\langle\mu^2(z)\rangle$ were uniform over a Fresnel-zone cross section. This breaks down for very long ranges and low frequencies, and it may also break down due to a highly structured sound-speed or buoyancy-frequency profile. The second assumption is isotropy: we assume that $L_p(\theta,z)$ represents the effective correlation length for all paths within a Fresnel length of each other at that point along the ray. This breaks down if the bundle of possible paths around the equilibrium ray experiences smaller effective L_p. We estimate the fractional inhomogeneity errors (which we call ϵ_{inhom}) in our calculations of the strength parameter Φ, the coherence time t_0, and the vertical coherence length z_0, and require these parameters to be less than unity for validity of the appropriate path-integral formulas. We estimate anisotropy errors by calculating the frequency f_v, below which an effective parameter qL_V^2/L_E is less than unity. We calculate three such frequencies in which the weighting of this effective parameter is like Φ, t_0, or z_0, respectively.

Thus, the detailed statistics of acoustic signal behavior are provided by the calculation of the quantities in the above expressions: Φ, Λ, τ, τ_0, τ_1, coherence lengths (z_0 and y_0), coherence times (t_0 and t_I), γ, and the various weighted values of ϵ_{inhom} and f_v.

III. RAY GEOMETRY

Ray characteristics are determined by the sound speed as a function of position. The general equation of ray acoustics for a ray in a depth-dependent sound-speed profile may be expressed as

$$\partial_x\theta = -C^{-1}\partial_z C, \quad (19)$$

where θ is the angle of the ray with the horizontal x coordinate and $\partial_x z=\tan\theta$.

Because the sound speed $C(z)$ never varies by more than a few percent from the reference C_0, fully refracted rays within the ocean volume never attain angles of more than $\approx 15°$ from the horizontal, a fact important to the parabolic-equation approximation.[1,2] In the ray approximation this is equivalent to setting $\tan\theta=\theta$, so we can replace (11) by

$$\partial_{xx}z = -C^{-1}\partial_z C. \quad (20)$$

We use Eq. (19) to numerically integrate a ray from an initial starting depth (source depth) and initial angle (θ_0), and assuming the environment to be range independent.

Figure 2 shows an example of such a ray, with definitions of many important quantities that characterize it. For example, R_D is the double loop length: the horizontal distance in meters between the acoustic source and the next point on the ray $z(x)$ at which $z=z_0$ and $\theta=\theta_0$; $z_{\min}$ is the z coordinate of the upper turning point or apex; and $z_{\max}$ is the z coordinate of the lower turning point.

We calculate travel time tt using Snell's Law:[1]

$$tt = \frac{C(z_s)}{\cos\Theta_0}\int dx/C^2(z_{\text{ray}}(x)). \quad (21)$$

A. Ray separations

The integrals to be performed in our algorithms require knowledge of the behavior of two rays traversing the internal-wave field. The difference between the two rays will appear in the integrals. If the two rays are close by, their

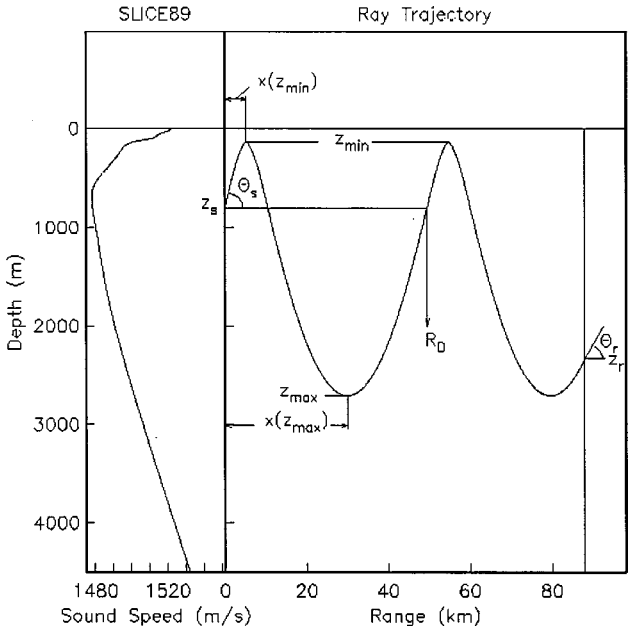


FIG. 2. Ray geometry for a sample sound-speed profile.

differences can be expressed by differentials. Two rays that are separated in time but not in space are particularly easy to treat. Two rays that are separated by horizontal distances are not subject to the sound channel, and their differences are just straight lines. The vertical separation of two rays is affected by the sound channel.

The vertical separation between the unperturbed ray $z_{\text{ray}}(x)$ and a nearby ray $z(x)$ is defined by

$$\xi(x)=z(x)-z_{\text{ray}}(x). \tag{22}$$

If $\xi(x)$ is small, its behavior is determined by the behavior of the second derivative of U_0.[1,2]

$$\partial_{xx}\xi(x)+U_0''\xi(x)=0, \tag{23}$$

where the primes in U_0'' mean the derivative with respect to z, and it is understood that U_0'' is evaluated at $z_{\text{ray}}(x)$.

We define the four solutions of Eq. (23) by the following boundary conditions:
$\xi_1(x)$ is defined by: $\xi_1(0)=0$, $\xi_1(R)=1$.
$\xi_2(x)$ is defined by: $\xi_2(0)=1$, $\xi_2(R)=0$.
$S_1(x)$ is defined by: $S_1(0)=0$, $S_1'(0)=1$.
$S_2(x)$ is defined by: $S_2(0)=1$, $S_2'(0)=0$.

S_1 and S_2 can be calculated by a marching algorithm, because we have their initial conditions. Once we have them at the receiver range R, we can calculate ξ_1 and ξ_2. The value of S_1 at the receiver is used to calculate transmission loss TL in the geometrical-acoustics approximation

$$\text{TL}=-10\log_{10}|S_1(R)R|. \tag{24}$$

B. Green's functions

The Green's function $g(x,x')$ represents two rays that start together at the source but with slightly different vertical angles, and separate from each other until range x is reached, whereupon one of the rays shifts its vertical angle such that subsequent propagation of both rays brings them together at the receiver. Such a situation is represented by the differential equation

$$\partial_{x'x'}g(x,x')+U_0''g(x,x')=\delta(x'-x), \tag{25}$$

with $g(x,0)=g(x,R)=0$. The solution for $g(x,x')$ can be given in terms of the ray separation functions $\xi_i(x)$ as[2]

$$g(x,x')=\begin{cases}\xi_1(x')\xi_2(x)S_1(R) & \text{on } (0,x)\\ \xi_1(x)\xi_2(x')S_1(R) & \text{on } (x,R).\end{cases} \tag{26}$$

C. Caustics

Past a few tens of kilometers from the acoustic source the sound-speed field has a rather complicated structure. Several rays from a single source may arrive at a receiver from different directions, giving rise to multipath. An envelope tangent to a family of rays is called a caustic surface. The Green's function $g(x,x)$ goes through infinity at caustics. This happens when $S_1(x)=0$ so that the boundary conditions cannot be satisfied.

Caustics are those locations at which geometrical acoustics predicts infinite intensity. It is of importance to avoid placing receivers very near caustics because our various expressions for acoustic-fluctuation quantities break down there. However, other approaches can be used in those regions.[10]

We find locations of caustics of rays by looking for zeros of $S_1(x)$.

IV. WAVE PROPAGATION REGIMES

Sound-speed variability can be described by the strength parameter Φ and the diffraction parameter Λ. The regions of different sound-fluctuation behavior can be defined in terms of regions in $\Lambda-\Phi$ space.[1,11,2]

The saturated region is the region of the sound fluctuation $(\Lambda-\Phi)$ diagram where a large number of micropaths are generated; the unperturbed ray is breaking into many rays. In the fully saturated regime these new rays are spread vertically over a region larger than a correlation length $L_{\mathbf{v}}$; this is the regime $\Lambda>1$ and $\Phi>1$, or $\Lambda<1$ and $\Lambda\Phi>1$. In the partially saturated regime they are spread over a region smaller than $L_{\mathbf{v}}$; this is the regime $\Lambda\Phi^2>1$, $\Lambda\Phi<1$, $\Phi>1$. The paths that are separated by less then $L_{\mathbf{v}}$ have some correlation with each other; the paths separated by more than a correlation length are statistically independent.

The strength parameter, Φ, can be expressed as

$$\Phi^2=q_0^2\int_0^R dx\langle\mu^2(z_{\text{ray}})\rangle L_p(\theta,z_{\text{ray}}), \tag{27}$$

so each increment in range (dx) contributes to the strength of the signal fluctuations an amount proportional to the variance of μ at this point of the ray and to the correlation length of the sound-speed fluctuations along the direction of the ray at that point L_p.[11,2]

The diffraction parameter, Λ, is a weighted average along the ray of $\{k_v^2\}(R_F)^2/2\pi$, where R_F is the Fresnel zone radius (given in terms of the Green's function).

$$\Lambda = \frac{q_0}{\Phi^2} \int_0^R dx \langle \mu^2(z_{\text{ray}}) \rangle L_p(\theta, z_{\text{ray}}) \{k_{\mathbf{v}}^2\} |g(x,x)|, \quad (28)$$

where the integral is taken along the equilibrium ray, $g(x,x)$ is the Green's function, and q_0 is the acoustic wave number.[2]

V. SPACE SCALES

A. Vertical coherence length

The vertical coherence length of the acoustic field z_0 that results from the approximations leading to Eq. (13) is defined by

$$z_0^{-2} = q_0^2 \ln \Phi \int_0^R dx \langle \mu^2 \rangle L_p \{k_{\mathbf{v}}^2\} [\xi_1(x)]^2, \quad (29)$$

where q_0 is the acoustic wave number, $\{k_{\mathbf{v}}^2\}^{1/2}$ is an appropriately averaged internal-wave vertical wave number, $\langle \mu^2 \rangle$ is the variance of sound-speed fluctuations, L_p is the sound-speed fluctuation correlation length along the ray tangent, and $\xi_1(x)$ is a particular ray-separation function.[12]

B. rms vertical-angle spread

An approximate rms vertical arrival angle variation $\theta_{\mathbf{v}}$ can be calculated from z_0 and the acoustic wave number q:[12]

$$\theta_{\mathbf{v}} = \frac{1}{(q \cdot z_0)}. \quad (30)$$

C. Horizontal coherence length

In the horizontal there is no sound channel, so that a ray-separation function can be defined as

$$\Delta y' = \Delta y_s + \frac{x}{R}(\Delta y_r - \Delta y_s), \quad (31)$$

where Δy_s and Δy_r are horizontal separation variables at the source and receiver, respectively.[12] For most purposes it is sufficient to set $\Delta y_s = 0$ and $\Delta y_r = 1$ km, so that $\Delta y'$ is always less than the horizontal correlation length of internal waves (about 10 km).

The horizontal correlation length y_0 that results from the approximations leading to Eq. (14) is defined by

$$y_0^{-3/2} = 2q_0^2 \int_0^R dx \langle \mu^2 \rangle L_p \left[\frac{x}{R} + \frac{\Delta y_s}{\Delta y_r} \left(1 - \frac{x}{R} \right) \right]^{3/2} y_h^{-3/2}, \quad (32)$$

and the rms horizontal wavefront tilt θ_h is

$$\theta_h = (\Delta y_r q_0^4 y_0^3)^{-1/4}. \quad (33)$$

VI. FREQUENCY SCALES (COHERENT BANDWIDTH)

The quantities τ_1 and τ_0 of Eq. (16) are calculated as follows:[4,13,12]

$$\tau_1 = \frac{\ln \Phi}{2C_0} \int_0^R dx \langle \mu^2 \rangle L_p \{k_{\mathbf{v}}^2\} g(x,x), \quad (34)$$

$$\tau_0^2 = \frac{1}{2} \left(\frac{\ln \Phi}{C_0} \right)^2 \int_0^R dx \langle \mu^2 \rangle L_p \{k_{\mathbf{v}}^2\} \cdot \int_0^R dx' \langle \mu^2 \rangle \times L_p \{k_{\mathbf{v}}^2\} \cdot [g(x,x')]^2. \quad (35)$$

The reciprocal of τ_0 is the coherent bandwidth.

VII. TIME SCALES

A. Pulse rms travel time

The quantity τ is the rms deviation of the travel time from the value it would have if there were no internal-wave fluctuation in the ocean. Also, the coherence function for small frequency separation [Eq. (16)] depends on τ:[2,12]

$$\tau^2 = C_0^{-2} \int_0^R dx \langle \mu^2(z) \rangle L_p(\theta, z) = (\Phi/\sigma_0)^2. \quad (36)$$

B. Field coherence time

$$t_0^{-2} = q_0^2 \int_0^R dx \langle \mu^2 \rangle L_p \{\omega^2\}, \quad (37)$$

where $\{\omega^2\}^{1/2}$ is an appropriately averaged internal-wave frequency that is dependent on local depth and position of the ray $z_r(x)$.[4]

Thus, t_0 is the effective overall acoustic phase time scale, and Φt_0 is the effective internal-wave time scale.

C. Intensity decoherence time

The intensity decoherence time t_I is calculated from:[12]

$$t_I^{-2} = q_0^2 \ln \beta_0 \ln \Phi \int_0^R dx \langle \mu^2 \rangle L_p \{k_{\mathbf{v}}^2\} \{\omega^2\} \quad (38)$$

$$\cdot \int_0^R dx' \langle \mu^2 \rangle L_p \{k_{\mathbf{v}}^2\} [g(x,x')]^2, \quad (39)$$

where

$$\ln \beta_0 = \begin{cases} \ln \dfrac{1}{2\Lambda \Phi \ln \Phi} & \text{if } \dfrac{1}{2\Lambda \Phi \ln \Phi} > e \\ 1 \quad \text{if } \dfrac{1}{2\Lambda \Phi \ln \Phi} \leq e. \end{cases} \quad (40)$$

VIII. MICRORAY FOCUSING PARAMETER

The calculation of the microray focusing parameter γ defined in Sec. II has been described and shown to be useful in comparing with data.[9] However, when comparing with a short-range experiment like AFAR, the results are very sensitive to the exact shape of the sound-speed profile. Comparison of the calculation with long-range experiment has not been done. Thus, this calculation at present must be regarded as a best guess to be evaluated at some future time.

In terms of the Garrett–Munk internal wave spectrum,[3] the microray focusing parameter γ is expressed by[2,9]

$$\gamma = 2q_0^2 \int_0^R dx \langle \mu^2 \rangle L_p \{P(j,x)\}, \quad (41)$$

where $P(j,x)$ can be approximately expressed in terms of two functions of x, called α and β

$$P(j,x)=(1-\cos\beta j^2)\exp(-\alpha j^2), \tag{42}$$

$$\alpha=M_j\ln\Phi k_{\mathbf{v}}^2\int_0^R dx'\langle\mu^2\rangle L_p k_{\mathbf{v}}'^2[g(x,x')]^2, \tag{43}$$

$$\beta=\frac{k_{\mathbf{v}}^2}{q_0}g(x,x), \tag{44}$$

$$M_3^{-1}=\sum_j\frac{1}{j(j^2+j_*^2)}, \tag{45}$$

$$\{P(j,x)\}\equiv M_j\sum_{j=1}^{\infty}\frac{P(j,x)}{j(j^2+j_*^2)}. \tag{46}$$

For very small $\alpha<10^{-3}$, we may do an asymptotic expansion to obtain

$$\{P(j,x)\}=\frac{1}{4}\frac{\beta^2}{\alpha}. \tag{47}$$

For $10^{-3}<\alpha<10^{-1}$, we approximate the sum by an integral which is valid for small α

$$\{P(j,x)\}\approx\frac{M_j}{2j_*^2}\left[\frac{1}{2}\ln\left(1+\frac{\beta^2}{\alpha^2}\right)+e^{\alpha j_*^2}E_i(-\alpha j_*^2)-\mathrm{Re}\{e^{(\alpha+i\beta)j_*^2}E_i[-(\alpha+i\beta)j_*^2]\}\right], \tag{48}$$

where $Ei(z)=\gamma_E+\ln(-z)+\Sigma_{j=1}^{\infty}z^j/(j\cdot j!)$, $|\arg(-z)|<\pi$, $\gamma_E=0.577\,21...$ is Euler's constant, or

$$\{P(j,x)\}\approx\frac{M_j}{2j_*^2}\cdot e^{\alpha j_*^2}\left\{[1-\cos(\beta j_*^2)]\gamma_E-\ln[\alpha^2 j_*^4+\beta^2 j_*^4]^{1/2}\cdot[\cos(\beta j_*^2)-e^{-\alpha j_*^2}]+(1-e^{-\alpha j_*^2})\ln(\alpha j_*^2)+\sin(\beta j_*^2)\arctan(\beta/\alpha)+\sum_{j=1}^{n}\frac{(\alpha\cdot j_*^2)^j-(\alpha^2 j_*^4+\beta^2 j_*^4)^{j/2}\cos[j(\arctan(\beta/\alpha)+\beta j_*^2)]}{j\cdot j!}\right\}, \tag{49}$$

where $n=3$ for $10^{-3}<\alpha\leqslant 10^{-2}$; $n=6$ for $10^{-2}<\alpha\leqslant 10^{-1}$; and for $\alpha>10^{-1}$ we directly sum Eq. (46) with $P(j,x)$ expressed by Eq. (42) with a decreasing number of terms as α becomes large, from six terms for $0.1<\alpha\leqslant 0.13$ to one term for $\alpha<5$.

IX. VALIDITY OF PATH INTEGRAL

A. Anisotropy

We calculate the validity frequency, below which the theoretical treatment of anisotropy is not justified, with three different weightings (weighted by Φ, t_0, or z_0)

$$f_v(\Phi)=\frac{\int_0^R dx\langle\mu^2\rangle L_p^2 q^{-1}\{k_{\mathbf{v}}^2\}}{\int_0^R dx\langle\mu\rangle^2 L_p}\cdot\sigma_0, \tag{50}$$

$$f_v(t_0)=\frac{\int_0^R dx\langle\mu^2\rangle L_p^2\{\omega^2\}q^{-1}\{k_{\mathbf{v}}^2\}}{\int_0^R dx\langle\mu^2\rangle L_p\{\omega^2\}}\cdot\sigma_0, \tag{51}$$

$$f_v(z_0)=\frac{\int_0^R dx\langle\mu^2\rangle L_p^2\{k_{\mathbf{v}}^2\}\cdot[\xi_1(x)]^2\cdot q^{-1}\{k_{\mathbf{v}}^2\}}{\int_0^R dx\langle\mu^2\rangle L_p\{k_{\mathbf{v}}^2\}[\xi_1(x)]^2}\cdot\sigma_0. \tag{52}$$

B. Inhomogeneity

A semiquantitative estimate of the effect of inhomogeneity is given by the average fractional change of the integrand in question over the height of a Fresnel zone. We simply evaluate such a change directly, knowing the Fresnel length, and using the three different weightings we have chosen.[12]

Let $z_r(x)$ be the vertical coordinate of the unperturbed ray at x; then, a position that is vertically displaced by a Fresnel-zone radius is $z_{rf}=z_r(x)+[(2\pi/q_0)g(x,x)]^{1/2}$. Then, the validity parameters are

$$\epsilon_{\mathrm{inhom}}(\Phi)=\frac{\int_0^R dx\langle\mu^2(z_{rf})\rangle L_p(z_{rf})-\int_0^R dx\langle\mu^2(z_r)\rangle L_p(z_r)}{\int_0^R dx\langle\mu^2(z_r)\rangle L_p(z_r)}, \tag{53}$$

$$\epsilon_{\mathrm{inhom}}(t_0)=\frac{\int_0^R dx\langle\mu^2(z_{rf})\rangle L_p(z_{rf})\{\omega^2(z_{rf})\}}{\int_0^R dx\langle\mu^2(z_r)\rangle L_p(z_r)\{\omega^2(z_r)\}}-1, \tag{54}$$

$$\epsilon_{\mathrm{inhom}}(z_0)=\frac{\int_0^R dx\langle\mu^2(z_{rf})\rangle L_p(z_{rf})\{k_{\mathbf{v}}^2(z_{rf})\}\cdot[\xi_1(x)]^2}{\int_0^R dx\langle\mu^2(z_r)\rangle L_p(z_r)\{k_{\mathbf{v}}^2(z_r)\}[\xi_1(x)]^2}-1. \tag{55}$$

X. RESULTS

We present here comparisons between the calculations from CAFI algorithms, PE numerical simulations, and experimental results.

The first and most stable acoustic fluctuation quantity that can be calculated and observed is the travel-time fluctuation variance for a well-identified timefront. Figure 3 shows such variances for the 1000-km, 250-Hz, Slice89 experiment, as a function of the upper-turning-point (UTP) depth of the rays making up each timefront.[14] The CAFI calculations and the numerical simulations were carried out with an internal-wave strength at one-half the GM reference level. It is seen that over the UTP depth region of 100–300 m there is reasonable agreement between the numerical

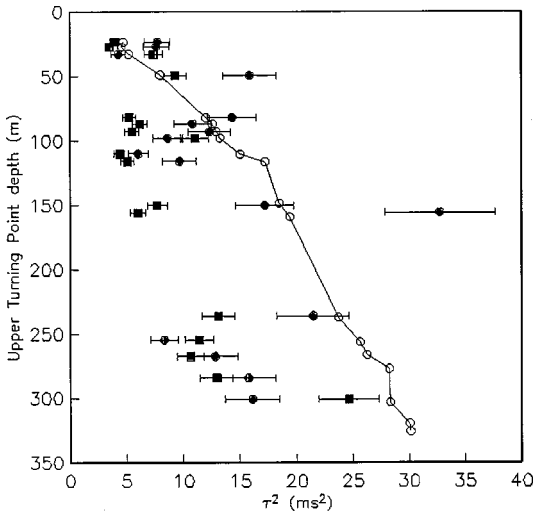


FIG. 3. Travel-time variance τ^2 as a function of the upper-turning-point (UTP) depth for the Slice89 experiment. Open circles are CAFI calculations for wavefronts of ID -35 to -45 and 36 to 45. Closed circles are results from simulated wave propagation through random realizations of internal waves using the parabolic equation approximation. Closed squares are experimental results from each identifiable wavefront. In the simulation and CAFI cases, the sound-speed fluctuations due to internal waves were taken to be at 0.5 of the standard Garrett–Munk reference level (Ref. 1).

simulations and the experimental results, with values between 5 and 25 ms^2. The CAFI calculations differ from the simulations by factors of the order of 2.

Figure 4 shows a similar comparison for a much longer-range experiment: the Acoustic Engineering Test (AET) carried out just before the ATOC experiment went on-line. This experiment had a range of 3.5 Mm and a frequency of 75 Hz.[5] In this case the CAFI calculations are based on *in situ* measurements of $\langle \mu^2 \rangle$ that correspond to internal-wave strengths between 0.5 and 1.0 times the standard GM reference level. We see that the comparison between the CAFI calculations and the experimental observations differ by factors of the order of 1.5.

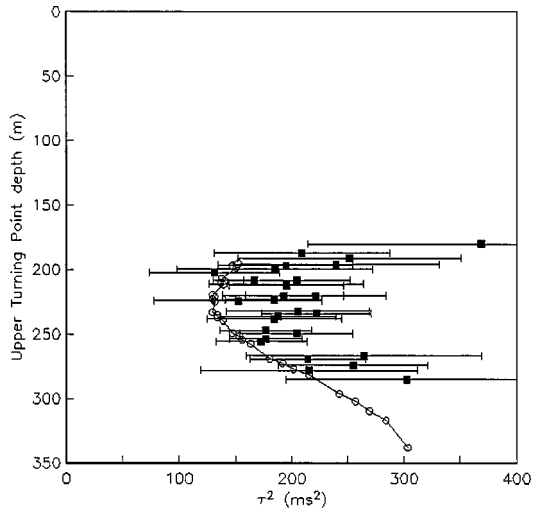


FIG. 4. Travel-time variance τ^2 as a function of the upper-turning-point (UTP) depth for the AET experiment. Open circles are CAFI calculations and closed squares are experimental results from each identifiable wavefront. In the CAFI calculations the sound-speed fluctuations due to internal waves were taken from *in situ* measurements; they correspond to strengths between 0.5 and 1.0 of the standard Garrett–Munk reference level (Ref. 1).

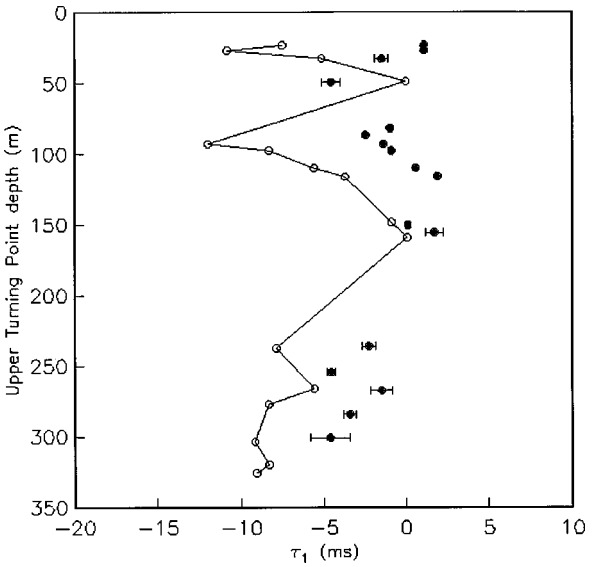


FIG. 5. Internal-wave bias τ_1 for Slice89. CAFI calculations (open circles) and PE simulations (closed circles).

A surprising feature of pulse propagation through a random medium is the early arrival (bias) of pulses averaged over a medium ensemble.[15] Figure 5 shows a comparison between CAFI estimates with PE numerical simulations for the conditions of the Slice89 experiment. The CAFI estimates seem to have roughly the same behavior with depth as the simulations, but they differ in magnitude by factors of the order of 2.

The predictions of pulse spread (not shown) for the very long-range low-frequency ATOC experiment[5] are in gross disagreement with the observations, in contrast with previous agreement reached in the AFAR experiment.[4] The origin of this discrepancy is unknown at this time.

Figures 6 and 7 show CAFI predictions for the rms arrival-angle fluctuations for the Slice89 configuration. The

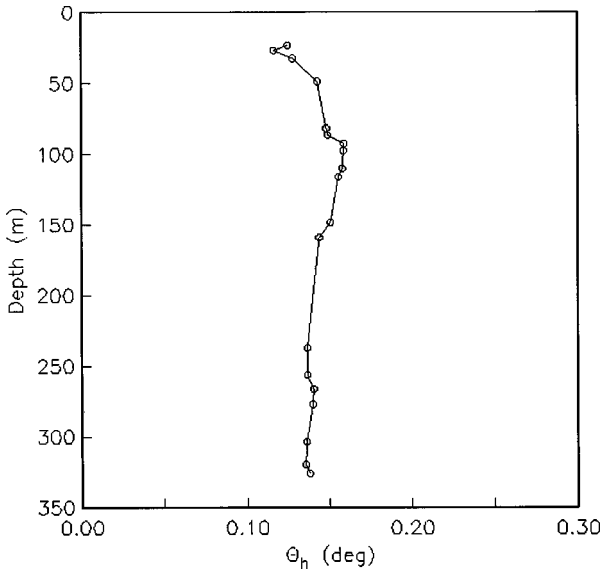


FIG. 6. rms horizontal angle spread θ_h: CAFI calculations with Slice89 parameters.

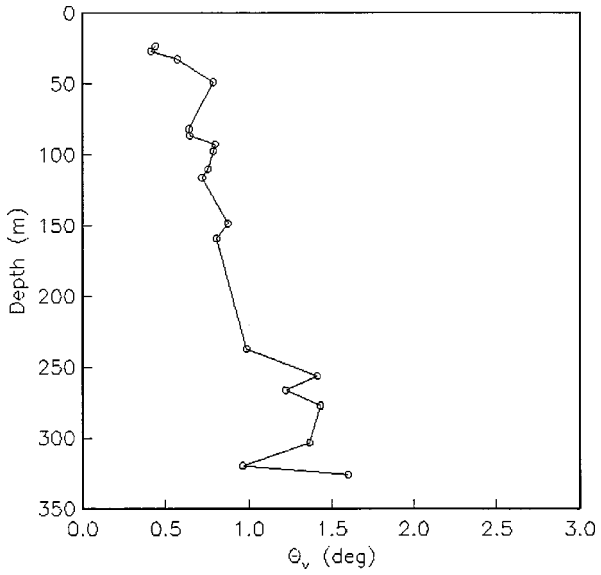


FIG. 7. rms vertical angle spread θ_v: CAFI calculations with Slice89 parameters.

vertical and horizontal arrival angles have standard deviations of about 1 and 0.15 deg, respectively.

Finally, we note that values of the anisotropy validity frequencies for Slice89 and AET are roughly in the regions of 100 and 50 Hz, respectively, while the inhomogeneity parameters have values between 0.3 and unity. The microray focusing parameters have values of the order of 0.1.

The above results indicate that for these particular cases, CAFI algorithms provide predictions of travel-time variances with accuracies of about a factor of 2. On the other hand, pulse-spread comparisons have been unsuccessful. There have been so few comparisons for other quantities such as wavefront coherence in space or time that no strong conclusions can be drawn. Past comparisons with other experiments at different ranges and frequencies have been previously summarized.[2] These algorithms (and the CAFI code in particular, if so desired) remain a choice that provides results without inordinate computer requirements for the purposes of designing experiments or for predicting internal-wave effects for experiments that will be carried out for other purposes. In order to establish the accuracy of CAFI algorithms beyond these somewhat qualitative statements, indeed to improve that accuracy, a major program of comparison with parabolic-equation simulations would be necessary, involving the variation of source–receiver geometry, acoustic frequency, and sound-speed profile.

We have assumed a GM spectrum in our calculations, and our predictions are limited to cases in which this is appropriate. Long-range, integrating acoustic experiments must depend on internal consistency of measured acoustic variables to determine whether this assumption is valid in any particular case. We hope that enough such experiments will be done in the future to test this assumption.

The question of what to do about shallow-water situations frequently arises. For example, large areas of coastal seas may have depths of no more than 100 or 200 m. The measurements of internal waves in these regions sometimes show dramatic, non-GM effects such as solitons, etc. However, it is also true that many measurements in these regions show internal-wave fields that are not far from GM behavior, and in those cases calculations based on CAFI algorithms would be quite reasonable for acoustic frequencies in the 100-Hz regime or above. One will expect vertical coherence across the water column, since often the lowest vertical internal-wave mode dominates, but temporal and horizontal coherences, and travel-time variances, will be well approximated without modification. This is not to say that we have a reliable shallow-water model, only that using CAFI and GM with these caveats is a starting point.

The CAFI code may be downloaded from the World-wide Web at http://www.es.ucsc.edu/~smf/cafi/cafi.html

ACKNOWLEDGMENTS

This research was supported in part by the U.S. Office of Naval Research Ocean Acoustics Program, Code 321OA. This is Contribution Number 415 of the Institute of Tectonics.

[1] S. Flatté, R. Dashen, W. Munk, K. Watson, and F. Zachariasen, *Sound Transmission Through a Fluctuating Ocean* (a 300-page monograph published by the Cambridge University Press in their series on Mechanics and Applied Mathematics, 1979).
[2] S. Flatté, ''Wave propagation through random media: Contributions from ocean acoustics,'' Proc. IEEE **71**, 1267–1294 (1983).
[3] W. Munk and C. Garrett, ''Internal waves in the ocean,'' Annu. Rev. Fluid Mech. **11**, 339–369 (1979).
[4] R. Dashen, S. Flatté, and S. Reynolds, ''Path-integral treatment of acoustic mutual coherence functions for rays in a sound channel,'' J. Acoust. Soc. Am. **77**, 1716–1722 (1985).
[5] J. Colosi, E. Scheer, S. Flatté, B. Cornuelle, M. Dzieciuch, W. Munk, P. Worcester, B. Howe, J. Mercer, R. Spindel, K. Metzger, T. Birdsall, and A. 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).
[6] S. Flatté and G. Rovner, ''Path-integral expressions for fluctuations in acoustic transmission in the ocean waveguide,'' in Methods of Theoretical Physics Applied to Oceanography, edited by P. Müller, Proceedings of the Ninth 'Aha Huliko'a Hawaiian Winter Workshop, 1997 (School of Ocean and Earth Science and Technology (SOEST) Special Publication), pp. 167–174.
[7] T. Duda, S. M. Flatté, J. Colosi, B. Cornuelle, J. Hildebrand, W. Hodgkiss, Jr., P. Worcester, B. Howe, J. Mercer, and R. Spindel, ''Measured wavefront fluctuations in 1000-km pulse propagation in the Pacific Ocean,'' J. Acoust. Soc. Am. **92**, 939–955 (1992).
[8] S. Flatté and R. Esswein, ''Calculation of the phase-structure function density from oceanic internal waves,'' J. Acoust. Soc. Am. **70**, 1387–1396 (1981).
[9] S. Flatté, D. Bernstein, and R. Dashen, ''Intensity moments by path integral techniques for wave propagation through random media, with application to sound in the ocean,'' Phys. Fluids **26**, 1701–1713 (1983).
[10] J. Simmen, S. M. Flatté, and G.-Y. Wang, ''Wavefront folding, chaos, and diffraction for sound propagation through ocean internal waves,'' J. Acoust. Soc. Am. **102**, 239–255 (1997).
[11] S. Flatté and R. Esswein, ''Calculation of the strength and diffraction parameters in oceanic sound transmission,'' J. Acoust. Soc. Am. **67**, 1523–1531 (1980).
[12] S. Flatté and R. Stoughton, ''Predictions of internal-wave effects on ocean acoustic coherence, travel-time variance, and intensity moments for very long-range propagation,'' J. Acoust. Soc. Am. **84**, 1414–1424 (1988).
[13] S. Flatté, S. Reynolds, and R. Dashen, ''Path-integral treatment of intensity behavior for rays in a sound channel,'' J. Acoust. Soc. Am. **82**, 967–972 (1987).
[14] J. Colosi, S. M. Flatté, and C. Bracher, ''Internal-wave effects on 1000-km oceanic acoustic pulse propagation: Simulation and comparison with experiment,'' J. Acoust. Soc. Am. **96**, 452–468 (1994).
[15] J. Codona, D. Creamer, S. Flatté, R. Frehlich, and F. Henyey, ''Average arrival time of wave pulses through continuous random media,'' Phys. Rev. Lett. **55**, 9–12 (1985).

Acoustic scattering by benthic and planktonic shelled animals

Timothy K. Stanton and Dezhang Chu
Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Peter H. Wiebe
Department of Biology, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Robert L. Eastwood and Joseph D. Warren
Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 14 September 1998; revised 27 September 1999; accepted 27 October 1999)

Acoustic backscattering measurements and associated scattering modeling were recently conducted on a type of benthic shelled animal that has a spiral form of shell (*Littorina littorea*). Benthic and planktonic shelled animals with this shape occur on the seafloor and in the water column, respectively, and can be a significant source of acoustic scattering in the ocean. Modeling of the scattering properties allows reverberation predictions to be made for sonar performance predictions as well as for detection and classification of animals for biological and ecological applications. The studies involved measurements over the frequency range 24 kHz to 1 MHz and all angles of orientation in as small as 1° increments. This substantial data set is quite revealing of the physics of the acoustic scattering by these complex shelled bodies and served as a basis for the modeling. Specifically, the resonance structure of the scattering was strongly dependent upon angle of orientation and could be traced to various types of rays (e.g., subsonic Lamb waves and rays entering the opercular opening). The data are analyzed in both the frequency and time domain (compressed pulse processing) so that dominant scattering mechanisms could be identified. Given the complexity of the animal body (irregular elastic shell with discontinuities), approximate scattering models are used with only the dominant scattering properties retained. Two models are applied to the data, both approximating the body as a deformed sphere: (1) an averaged form of the exact modal-series-based solution for the spherical shell, which is used to estimate the backscattering by a deformed shell averaged over all angles of orientation, and produces reasonably accurate predictions over all $k_1 a_{\mathrm{esr}}$ (k_1 is the acoustic wave number of the surrounding water and a_{esr} is the equivalent spherical radius of the body), and (2) a ray-based formula which is used to estimate the scattering at fixed angle of orientation, but only for high $k_1 a_{\mathrm{esr}}$. The ray-based model is an extension of a model recently developed for the shelled zooplankton *Limacina retroversa* that has a shape similar to that of the *Littorina littorea* but swims through the water [Stanton *et al.*, J. Acoust. Soc. Am. **103**, 236–253 (1998b)]. Applications of remote detection and classification of the seafloor and water column in the presence of shelled animals are discussed. © *2000 Acoustical Society of America.* [S0001-4966(00)01702-1]

PACS numbers: 43.30.Ft, 43.30.Sf, 43.30.Hw, 43.20.Fn [DLB]

LIST OF SYMBOLS

1,2	subscripts indicating medium ''1'' (surrounding fluid) and medium ''2'' (body medium)
a	radius of sphere
$\overline{a}$	average radius of an irregular sphere
a_{esr}	equivalent spherical radius—radius of sphere that has same volume as object of interest
α_L	related to dispersion of Lamb wave
b_m	modal-series coefficient for spherical shell
β_L	attenuation coefficient of Lamb wave on elastic shelled sphere
c_1, c_2	sound speed for medium ''1'' (surrounding fluid) and medium ''2'' (body medium)
c_L	sound speed of Lamb wave
$\overline{c_L}$	c_L evaluated for $k_1\overline{a}$
Δ_a	deviation in effective radius from mean radius of rough sphere
f	scattering amplitude
f_{bs}	scattering amplitude in backscattering direction
f_{spec}, $f_{\mathrm{op}}, f_{\mathrm{Lamb}}$	components of f_{bs} from front interface (''specular wave''), opercular opening of shell, and in association with Lamb wave, respectively
$\mathcal{F}$	packing factor for scatterers on the seafloor [=(area covered by scatterers)/(total area of seafloor)]
F_{spec}, F_{op}, F_L	empirical factors, ranging in value from 0 to 1, to account for loss of each respective wave (specular, opercular, Lamb) due to discontinuity in shell
g	ρ_2/ρ_1
G_L	coupling coefficient for combination of landing

	and launching of Lamb waves on shell
h	c_2/c_1
hc	h for compressional sound speed in elastic shell
hs	h for shear sound speed in elastic shell
i	$\sqrt{-1}$
k_1	acoustic wave number ($=2\pi/\lambda_1$) in medium ''1'' (surrounding fluid)
λ_1	acoustic wavelength in medium ''1'' (surrounding fluid)
p_{scat}	scattered pressure
P_{inc}	incident pressure at the object
Φ_L	phase shift of Lamb wave heuristically added for nonideal body
ρ_1, ρ_2	mass density in medium ''1'' (surrounding fluid) and medium ''2'' (body medium)
$\mathcal{R}_{12}$	plane wave/plane interface reflection coefficient [reflection off of medium ''2'' (body medium) due to incident beam in medium ''1'' (surrounding fluid)] $[=(\rho_2 c_2/\rho_1 c_1 - 1)/(\rho_2 c_2/\rho_1 c_1 + 1)]$
RTS	reduced target strength
S_A	area scattering strength
σ_r	$\langle \Delta_a^2 \rangle^{1/2}$ (roughness parameter of irregular shell)
σ_{bs}	differential backscattering cross section
TS	target strength
θ_g	grazing angle ($\theta_g = 90°$ corresponds to normal incidence)
θ_L	launch/land angle for Lamb wave
$\langle ... \rangle$	average over ensemble of statistically independent samples

INTRODUCTION

Marine organisms can sometimes be a significant source of acoustic scattering. The scattering properties of the animals need to be understood for two major reasons: (1) Sonar performance prediction. When designing a sonar system, all sources of reverberation, i.e., unwanted scattered signals, need to be modeled so that the performance of the system can be predicted in terms of the signal-to-reverberation ratio. The reverberation from marine life can sometimes cause false alarms in active sonar detection systems. (2) Bioacoustical oceanography. Biologists can take advantage of the fact that certain marine life scatters sound and use acoustic systems to survey the animals. The scattering needs to be modeled so that the echo data from the survey can be interpreted in terms of meaningful biological parameters such as length and numerical density.

There is a wide range of animals that live in the ocean, including fish and zooplankton that live in the water column and the animals that live in the benthic (seafloor) zone. The fish and planktonic animals will contribute to acoustic scattering from the water column while the benthic animals will contribute to scattering by the seafloor. There are thousands of species present in the ocean and it would not be practical to model the scattering by each animal on a species-by-species basis. It has been convenient to characterize the animals according to their gross anatomical group. For example, in a recent study, zooplankton were grouped according to whether they were weak scatterers, elastic shelled bodies, or gas-bearing (Stanton *et al.*, 1994, 1996, 1998a, 1998b). Fish are sometimes characterized according to whether or not they have a swimbladder (Foote, 1980).

Shelled animals are present both in the water column and on the seafloor. For example, pteropods are a group of gastropod molluscar zooplankton that swim throughout the water column and other gastropod groups reside on the seafloor. There has been significant evidence that these animals can influence or even dominate the volume and seafloor reverberation, respectively, when present (Wiebe *et al.*, 1996, 1997; Jackson *et al.*, 1986; Stanic *et al.*, 1989; Zhang, 1996). The main issue in modeling the acoustic scattering by the shelled bodies is that their shape is so complex. In addition, few measurements of the material properties are available for use in the modeling. There are no exact (analytical) mathematical solutions available to describe the scattering, so approximate analytical or numerical approaches must be used. Further complicating the issue is the fact that, even if a general solution were known, the shape of the shells is sufficiently complex that the characterization of the morphology is a challenge.

In a recent study, the acoustic scattering properties of several major zooplankton groups were studied and models developed (Stanton *et al.*, 1998b). One of these groups involved the planktonic shelled gastropod (*Limacina retroversa*). This is a thin-shelled snail that swims throughout the water column, can occur at numerical densities of up to thousands per cubic meter, and can (collectively) produce significant reverberation levels (Wiebe *et al.*, 1996, 1997). At-sea measurements of the scattering by the animals were made in a laboratory-style tank shortly after they were caught. Broadband signals were used to investigate the resonance scattering structure and scattering models were developed. Key scattering mechanisms were identified in this high $k_1 a_{\text{esr}}$ region: there was significant evidence that a subsonic Lamb wave was the source of some of the structure observed (Stanton *et al.*, 1998b) (k_1 is the wave number of the surrounding water and a_{esr} is the equivalent spherical radius of the irregular body, which is the radius of a sphere with the same volume as the irregular body). Furthermore, the structure and overall scattering levels changed with orientation, which indicated that the irregularities played a major role in the scattering. Significant progress was made toward the understanding of scattering by complex shelled bodies in this study. However, due to the time constraints of performing the research at sea, the experiments were naturally limited by the fact that the full range of orientations and acoustic frequencies could not be realized for all animals (it was essential to perform these experiments at sea as it allowed studies to be performed on freshly caught animals spanning a wide range of species not available near the coast). Furthermore, only a limited class of species was used. The next logical step in this research is to extend this type of study to a wider range of frequencies and orientations so that the physics of the scattering process can be understood over a wider range of

conditions. Also, it is important to examine other classes of animals within a gross anatomical group.

We have just completed an extensive set of acoustic backscattering measurements involving benthic shelled bodies. The experiments involved a wide range of acoustic frequencies spanning from $k_1 a_{\text{esr}}$ much less than unity to $k_1 a_{\text{esr}}$ much greater than unity and the full range of orientation angles in one plane. The animals were chosen to fit two criteria: (1) to be from the seafloor and (2) to have a shell with a morphology similar to that of the planktonic ones studied so that direct comparisons (or simple extrapolations) could be made. With this high-quality data set, dominant scattering mechanisms can be clearly identified with the scattering geometry. The high $k_1 a_{\text{esr}}$ scattering model initially developed for the planktonic ones described above is extended and applied to these benthic animals. Furthermore, an all-$k_1 a_{\text{esr}}$ model is applied to the benthic animals for data averaged over all angles of incidence. Similarities between the scattering by the two animal types have been determined and connections between the scattering model predictions and applications to volume and seafloor scattering when shelled bodies are present are summarized.

This paper is organized as follows. In Sec. I, basic definitions and scattering models are presented involving both the ray-based and modal-series-based models. Section II describes the experiment and Secs. III and IV present the results of the experiment and comparisons between the models and some of the results. Applications of the laboratory-based results to surveys in the ocean of the water column and seafloor are discussed in Sec. V.

I. THEORY

The targets under investigation are irregular shelled bodies with discontinuities. There are no exact analytical solutions that can describe the scattering by these bodies, therefore approximate or numerical solutions must be explored. After a section defining basic scattering terms, two approximate approaches are summarized. One involves an average of the exact modal-series solution to the spherical shell which has application to the problem for all $k_1 a_{\text{esr}}$, and the other involves a ray-based approach which is valid for only the high $k_1 a_{\text{esr}}$ region (i.e., $k_1 a_{\text{esr}}$ greater than unity).

A. Definitions

The far-field backscattered energy from a bounded body is commonly expressed in terms of the target strength TS, which is a logarithmic form of the scattering amplitude f evaluated in the backscattering direction. The scattering amplitude describes the efficiency with which an object scatters sound and is defined in terms of the incident and scattered pressures P_{inc} and p_{scat} as

$$p_{\text{scat}} = P_{\text{inc}} \frac{e^{ik_1 r}}{r} f, \tag{1}$$

where r is the distance between the object and receiver. This expression is associated with a single echo or realization from a given individual target. With the geometry and source/receiver terms listed separately in the above equation, the scattering amplitude contains all of the scattering physics related to the target itself. The amplitude depends upon the object size, shape, orientation, and material properties as well as the acoustic wavelength. Scattering modeling efforts result in derivations or descriptions of f.

The target strength is defined in terms of the backscattering amplitude f_{bs} as

$$\text{TS} = 10 \log |f_{bs}|^2 = 10 \log \sigma_{bs}, \tag{2}$$

where f_{bs} is f evaluated in the backscattering direction and the differential backscattering cross section σ_{bs} is defined as $\sigma_{bs} = |f_{bs}|^2$ (this differential cross section should not be confused with the commonly used backscattering cross section σ where $\sigma = 4\pi\sigma_{bs}$). The target strength is in units of decibels relative to $1 m^2$.

In order to compare the scattering by objects of the same shape but of different sizes, the target strength is normalized according to the square of some dimension. The normalized or reduced target strength (RTS) for the case of the ideal sphere is defined in terms of the TS and radius a of the sphere as

$$\text{RTS} = \text{TS} - 10 \log \pi a^2. \tag{3}$$

For the case of irregular shelled animals, the equivalent spherical radius, a_{esr}, can be used in Eq. (3).

For the case involving multiple targets or multiple realizations of the same target, it is useful to study the average echo energy. At high enough frequencies or short enough wavelengths, the phases of the echoes from the targets will be random with respect to each other and the average echo energy from the aggregation will be equal to the sum of the average echo energy from each individual (averaged over independent realizations). The equality holds true only when effects due to multiple scattering such as extinction and higher-order scattering (beyond first order) are negligible. Since the backscattering cross section is proportional to echo energy, the ''average'' target strength of a target is commonly expressed in terms of the average value of the backscattering cross section

$$\langle \text{TS} \rangle = 10 \log \langle \sigma_{bs} \rangle, \tag{4}$$

where the averaging process, denoted by the brackets $\langle ... \rangle$, was performed before the logarithm operation was taken. The averaging is over an ensemble of statistically independent realizations of one or more variables such as animal size and/or orientation.

B. Modal-series solution to spherical shell—all $k_1 a_{\text{esr}}$

Although there are no exact analytical solutions to the scattering by these irregular shelled bodies, there are exact solutions to objects with certain simple shapes. It is useful to explore exact solutions involving simple shapes in order to determine insight into the scattering physics as well as conditions under which they provide a reasonable approximation to the more complex shapes. There exists an exact solution to the spherical shell. Although the animals are somewhat elongated (ratio of length to width is approximately 1.6) and irregular with discontinuities, there is some utility to investigating this solution, especially in light of the fact that there

is not an exact analytical solution to the shape that more resembles the animals. There will be some common elements in the scattering processes of both the sphere and the animals, especially since most of the animal's shell is closed and continuous (as opposed to the open half shell of a clam). At the high values of k_1a (where a is the radius of the outer boundary of the spherical shell), there will be scattering from the front curved interface and surface elastic waves. Both the sphere and the animal will have both types of processes. There will, however, be distinct differences in the structure and overall values involving single realizations because of the variability of the local radius of curvature and shell thickness of the animal as well as the discontinuities. At the lower values of k_1a, the scattering will be more a function of the volume of the shell and there is more potential for applicability of the sphere solution.

The backscattering amplitude for the spherical shell is given in general form as

$$f_{bs}=\frac{-i}{k_1}\sum_{m=0}^{\infty}(-1)^m(2m+1)b_m. \tag{5}$$

The term b_m is the modal-series coefficient and can be expressed in terms of the ratio of two 6×6 determinants based upon the boundary conditions. The radius of the shell, a, is implicit in b_m. The solution was first derived in Goodman and Stern (1962) and is exact over all k_1a.

The modal-series solution to the sphere was transformed by others into a ray solution (Überall, 1973; Marston, 1992) in which the scattering amplitude was expressed in terms of an infinite series of rays. The amplitude and phase of each ray were determined through the Sommerfeld–Watson Transformation (SWT). One of those forms of the SWT was used in Stanton *et al.* (1998b) as a basis of modeling the scattering by planktonic shelled animals (see also discussion in Sec. I C). In that analysis, the ray solution was (coherently) averaged over an ensemble of sizes to emulate the effect of the multiple ray paths around the irregular body for a single realization of animal size. Using that same approach, the modal-series solution could be averaged to obtain a similar effect. The ray approach has great utility since various rays that appear explicitly in the formulation can be given different weighting according to the physical process specific to the irregular body [Stanton *et al.* (1998b)]. In contrast, these rays do not appear explicitly in the modal-series solution. The ray approach, however, is only valid in the high k_1a region. Thus although the modal-series approach may not be as useful in the high k_1a region as the ray approach, the solution provides predictions (of varying accuracy) over the whole range of k_1a.

In the application to the shelled bodies in this paper, the modal-series solution will be averaged over a range of sizes and shell thickness in order to estimate the scattering by the animals averaged over angle of orientation. The average increases the number of ray paths and reduces differences in structure of the predictions and data. This approach was also generally successful in applying an ensemble average of the exact solution for the solid elastic sphere to the scattering by individual sand grains and suspended sediment (Thorne *et al.*, 1993, 1995).

C. Approximate ray solution to irregular shell—$k_1a_{esr}\gtrsim1$ (geometric scattering region)

A convenient method for modeling the scattering by the irregular shelled bodies involves the use of a ray approach. The formulation can be constructed using the dominant scattered rays and can also be manipulated analytically. There is great potential with a ray solution to be able to predict the resonance structure of the scattering. The disadvantage is the fact that ray solutions are generally only valid in the high k_1a_{esr} region, or geometric scattering region.

There is a range of sophisticated generalized ray theories that can be used, including those described in Felsen and Lu (1989), Ho and Felsen (1990), Norris and Rebinsky (1994), Ho (1994), Yang *et al.* (1995), Rebinsky and Norris (1995), and Yang *et al.* (1996). Included in those formulations are arbitrary coupling coefficients between the incident wave and the circumferential waves, and arbitrary ray paths due to the irregular surface of the body.

In a recent paper, we derived a simplified formula based upon a limited set of data involving the planktonic gastropod *Limacina retroversa* (Stanton *et al.*, 1998b). The formula used a single coupling coefficient associated with an average radius of the body. The random phases of the circumnavigating Lamb waves were estimated through assignment of a randomized radius of curvature of the body. As an approximation, each wave was constructed so as to follow the path of a great circle.

Included in that formulation were two terms associated with rays caused by the scattering by the front interface of the body and by a Lamb wave circumnavigating the body. These terms were chosen as there was evidence that both existed sometimes with substantial energy. Our much more extensive set of data presented herein involving the benthic gastropod snail (*Littorina littorea*), which is morphologically similar to Limacina, indicate that for certain orientations, there is also a ray that propagates into the opercular opening and scatters back toward the receiver. We therefore extend the formula presented in Stanton *et al.* (1998b) to

$$f_{bs}\simeq f_{\text{spec}}+f_{\text{op}}+f_{\text{Lamb}}, \tag{6}$$

where the terms f_{spec} and f_{op} represent the scattering from the front interface and back of the opercular opening, respectively, and f_{Lamb} is the Lamb wave that circumnavigates the body.

This addition of the term f_{op}, in essence takes the place of one of the terms f_{tw} that is discussed in that paper and ignored. That term was originally intended to be a transmitted wave that passed through the shell, reflected off of various interfaces beyond the front interface, and then returned to the receiver. Since the shell is hard, the term was considered to be small. However, for the case in which the opercular opening is facing the sonar system, there is no boundary for the ray to pass through at the opening and the bundle of rays associated with the opening will pass into the shell unperturbed. Also, the specular wave f_{spec} and wave associated with the opercular opening f_{op} interfere in a manner consistent with waves reflecting off of two points separated by a distance comparable to the diameter of the body. There are other phenomena that can also give rise to such an effect, but

have not been included because they are generally of such low energy. They are the waves scattered by the tip and edge of the opercular opening. It is calculated that the tip scatters sound with a target strength of about −90 dB, which is 10's of dB lower than the specular return from the curved shell face (Bowman *et al.*, 1987). The backscattered echo from the edge at these very high frequencies is 10's of dB lower than the specular return when the sound is well off normal incidence from the local plane of the edge (Jebsen and Medwin, 1982).

The various terms in the above equation are given as

$$f_{\text{spec}} \simeq \frac{a_{\text{spec}}}{2} \mathcal{R}_{12} F_{\text{spec}} e^{-i2k_1 a_{\text{spec}}}, \tag{7}$$

$$f_{\text{op}} \simeq \tfrac{1}{2} a_{\text{op}} \mathcal{R}_{12} F_{\text{op}} e^{i2k_1 a_{\text{op}}}, \tag{8}$$

$$\begin{aligned} f_{\text{Lamb}} \simeq & -\tfrac{1}{2} G_L e^{i\Phi_L} \overline{a} e^{-2(\pi-\theta_L)\beta_L} \\ & \times e^{i2k_1\overline{a}[(c_1/\overline{c_L})(\pi-\theta_L)-\cos\theta_L]-i\pi/2} F_L \\ & \times \sum_{m=0}^{\infty} (-1)^m e^{-2\pi m\beta_L} e^{i2\pi m k_1 \overline{a} c_1/\overline{c_L}} e^{-(1/2)\gamma^2\sigma_r^2}, \end{aligned} \tag{9}$$

where

$$\theta_L = \sin^{-1}(c_1/c_L), \tag{10}$$

$$G_L \simeq 8\pi\beta_L c_1/c_L, \tag{11}$$

$$\overline{c_L}/c_1 \simeq k_1\overline{a}/(\alpha_L + 1/2), \tag{12}$$

and

$$\begin{aligned} \gamma = k_1\{2[(c_1/\overline{c_L})(\pi-\theta_L)-\cos\theta_L] + 2\pi m c_1/\overline{c_L} \\ + Bk_1\overline{a}[2(\pi-\theta_L)+2\pi m]\}. \end{aligned} \tag{13}$$

The terms f_{spec} and f_{op} contain a reflection coefficient $\mathcal{R}_{12}$ due to the interface they are associated with. Similarly, f_{Lamb} contains a coupling coefficient G_L that describes the efficiency with which the incident signal couples with the shell and reradiates. B accounts, in part, for dispersion according to the relation $c_1/c_L = c_1/\overline{c_L} + Bk_1\Delta_a$, where the mean speed $\overline{c_L}$ is c_L evaluated at $k_1\overline{a}$. The terms a_{spec} and a_{op} in Eqs. (7) and (8) for f_{spec} and f_{op} are the local radii of curvature that the incident wave sees, respectively. The $\overline{a}$ in Eq. (9) for f_{Lamb} is the average value of the radius of the irregular sphere. The ray paths along lines of constant meridional angle experienced deviations in path length due to deviations in effective radius from this mean. The attenuation coefficient β_L and dispersion term α_L of the Lamb wave are the imaginary and real parts of the complex root of the denominator of the modal-series coefficient for a fluid-filled elastic spherical shell. The term θ_L is the angle at which the Lamb wave launches and lands at the surface and Φ_L is an empirically determined phase shift due to the irregularity of the body. The term σ_r is the root-mean-square (rms) deviation of the shell radius from the mean value $\overline{a}$. Note that there are several ''a'' terms defined in the attempt to describe the irregularity of the shell. Although they differ in value, they are all comparable to a_{esr}.

Details of the above ray-based equations are discussed in Stanton *et al.* (1998b). One key aspect of this equation is the roughness-induced attenuation of the Lamb wave that was derived in that paper. In addition to the derived terms, there are empirically determined terms such as F_{spec}, F_{op}, and F_L which are used to weight the different scattered rays. For example, the opercular opening may be aimed away from the acoustic system, preventing a ray from traveling into the opening and back out toward the system. In that case, $F_{\text{op}} = 0$. For the case when the opening is aimed toward the system, then F_{op} is near or equal to unity. F_{spec} and F_L are also used to compensate the energy associated with the specular and Lamb waves with respect to the direction of the opercular opening.

The above ray formula presented in Eq. (6) is only valid for $k_1 a_{\text{esr}}$ greater than unity. However, in that region it has the potential for predicting the resonance structure at least qualitatively, if not also quantitatively. The formula is useful for both single realizations and averages over ensembles.

II. EXPERIMENTAL SETUP AND ANIMALS

There has been a series of acoustic scattering experiments involving both planktonic and benthic shelled animals. The experiments involving the planktonic gastropods (2-mm-long *Limacina retroversa*) were conducted in a laboratory-style tank setup on the deck of a ship using freshly caught live animals. The work conducted during two cruises (1993 and 1994), is described in Stanton *et al.* (1994, 1996, 1998a, and 1998b) and Chu and Stanton (1998). The recently completed experiments involving the benthic gastropods, periwinkles (*Littorina littorea*), were conducted during 1997 in a similar tank setup in a laboratory at the Woods Hole Oceanographic Institution. Because of the dominance of the shell in the scattering, the tissue of the benthic animals was removed (for experimental convenience) and the experiments were performed on the empty shells. Descriptions of the periwinkles and experimental setup are given below.

A. Animals

The periwinkles were collected during low tide off of rocks at the shoreline at Woods Hole, MA. This species was chosen, in part, because of its similarity in morphology to the planktonic gastropods studied earlier so that comparison of the scattering by the two different types of animals could be made. The periwinkle was also chosen because of the fact that this is a common shape of many benthic species. Studying this particular species could provide insight into the scattering by other ones of similar morphology.

The shape of both types of animals is in the form of a spiral (Fig. 1). More precisely, each animal is in the shape of a tube that is wound into a spiral. Each 360° rotation of the tube corresponds to a cycle of growth. The material properties of the shell of each type of animal are very similar to each other—the planktonic shell is composed of aragonite (Lalli and Gilmer, 1989) and the benthic is a combination of aragonite and calcite (Lowenstam, 1954). Aragonite and calcite are similar materials. Both minerals are composed of $CaCO_3$ (calcium carbonate) with their differences being in

their crystalline structure. Also, the density, compressional sound speed, and shear sound speed of calcite are 2.71 g/cc, 6530 m/s, and 3360 m/s, respectively, while the density and compressional sound speed of aragonite are 2.92 g/cc and 5820 m/s, respectively (the value of shear sound speed could not be found) (Carmichael, 1982). The differences in material properties will result in differences in scattering resonances for a given size and shape, but not overall scattering levels. The major acoustical difference between the two types of animals is due to the fact that the planktonic animals typically have much thinner shells than those of the benthic animals of the same size. The thinner shell allows the planktonic animal to swim throughout the water column.

Six periwinkles were used in the backscatter experiments ranging in size from 6.2 to 13.7 mm (Table I). The ratio of length to width of each animal is roughly constant for the range of sizes used. This is an important observation as it allows the use of a single shape for scattering predictions over a range of lengths. This quality of constant proportions involving outer dimensions as well as shell thickness is further illustrated in the studies involving a larger range of size of animals (Figs. 2 and 3).

B. Experimental setup

The experimental setup and associated aspects of system calibration and data acquisition are very similar to those described in detail in earlier papers (Stanton, 1990; Chu *et al.*, 1992) and summarized in more recent papers (Stanton *et al.*, 1994, 1998a) and will only be briefly summarized here. The principal difference between this setup and previously reported ones is that the acoustic transducers were aimed horizontally rather than being mounted on the bottom and looking up (Fig. 4). They were aimed in this direction to facilitate measurements of backscatter versus angle of orientation by use of a computer-controlled stepper motor (also a new addition) that rotated the shells suspended in the acoustic beam.

The system consists of an array of pairs of acoustic transducers that was facing in the horizontal direction in a 3.7-m-long by 2.4-m-wide by 1.5-m-deep tank. The tank was filled with filtered seawater. The transducers within each pair are closely spaced and identical. One of the two transducers is used as the transmitter and the other the receiver. By using two transducers, closer scattering ranges can be achieved, ringing of the transmitter is less of a problem, the system is linear, and the system is easy to calibrate. The transducers span the frequency range of 24 kHz to 1 MHz (although note that the 24-kHz transducers, new additions, are physically separate from the array). The frequencies used in this particular experiment were 24, 50, 75, 120, 165, 200, 250 (BB), 500 kHz (BB), and 1 MHz (BB) where broadband (BB) octave-bandwidth transducers were used in the upper frequency region. The transducers at frequencies of 200 kHz and below were narrow band.

The pulse-echo electronics were controlled by the same personal computer that controlled the stepper motor and included a programmable waveform generator and power amplifier at the transmission end of the system and a preamplifier, bandpass filter, and digital oscilloscope (for capture, display, and transfer of data to computer) at the receiving end of the system.

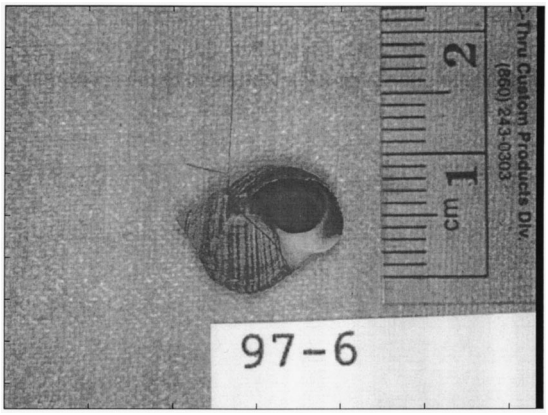


FIG. 1. Photo of periwinkle (*Littorina littorea*, animal No. 97-6) used in the 1997 acoustic scattering experiments at Woods Hole Oceanographic Institution. This is close to what would be a side view in the scattering measurements. The thin line is the tether (59 μm diameter). The opercular opening is facing the camera similarly to how it faced the acoustic transceiver at an orientation of about 240°.

The animals were suspended in the middle of the acoustic beams by a single thin monofilament line that was 59 μm in diameter and was acoustically transparent at these frequencies. Each animal was at a range of 51 cm from the transducer pairs spanning 50 kHz–1 MHz and at a range 73 cm from the 24 kHz transducers. These distances correspond to the far field of each transducer. Also, at this range, each animal was within the first Fresnel zone of the transceiver, making this a far-field scattering condition. The line was tied around the midsection of the body (Fig. 1) and no glue was used (the tying of the tether onto these small shells was done by first affixing the shell onto a piece of two-sided tape, then using tweezers to wrap the tether around the shell and tie the knots). The tether was tied to a clip that was connected to the shaft of the stepper motor so that the tether coincided with the center of the shaft. The shells were heavy enough so that the tether was fully extended throughout the experiment. Through use of the small torsional strength of the tether,

TABLE I. Outer dimensions and related quantities of the periwinkles used in the acoustic scattering measurements. The length is the measured maximum tip-to-tip distance. The width was measured between the plane containing the face of the opercular opening and the outer point of the shell on the opposite side of the shell. Because of the complexity of the animal shape and in order to be consistent, the axis along which the width was measured was not necessarily perpendicular to the axis along which the length was measured. The equivalent spherical radius a_{esr} is used for some simulations and is the radius of the sphere that has the same volume as that of the animal. The volume of the animal is calculated by assuming it has a prolate spheroidal shape with length and width as given in this table.

Animal ID number	Length (mm)	Width (mm)	a_{esr} (mm)	Length/width
97-1	6.2	3.9	2.3	1.6
97-2	6.6	4.2	2.4	1.6
97-3	8.8	5.6	3.3	1.6
97-4	10.0	6.7	3.8	1.5
97-5	10.6	7.0	4.0	1.5
97-6	13.7	8.4	4.9	1.6

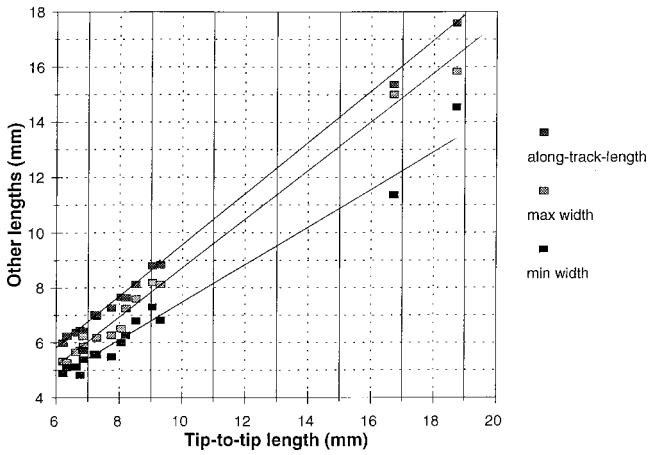


FIG. 2. Measurements of several different outer dimensions of various sized periwinkles plotted as a function of maximum tip-to-tip length. These were different animals (but same species caught at the same time and location) than the ones used in the acoustics experiment. The along-track length is the length of the animal (very close to the maximum length) along the track of the acoustic signal when aimed at the apex (i.e., end-on incidence). The widths were measured generally transversely with respect to the lengthwise axis. Given the complexity of the shape of the body, the crosswise axes weren't necessarily perpendicular to the lengthwise axis, although they were nearly so. The lines are least-square fits.

rotation of the stepper motor shaft could cause the shell to rotate the same amount (difficulties with this approach are discussed in Sec. II C).

C. Measurement procedure

The system was calibrated both before and after the entire series of scattering measurements (i.e., involving all periwinkles) was performed. Each scattering measurement involved a sequence of first measuring the background reverberation of the tank without the animal present in the beam, then placing the animal in the beam and measuring the backscatter over the range of orientation angles 0°–450° in

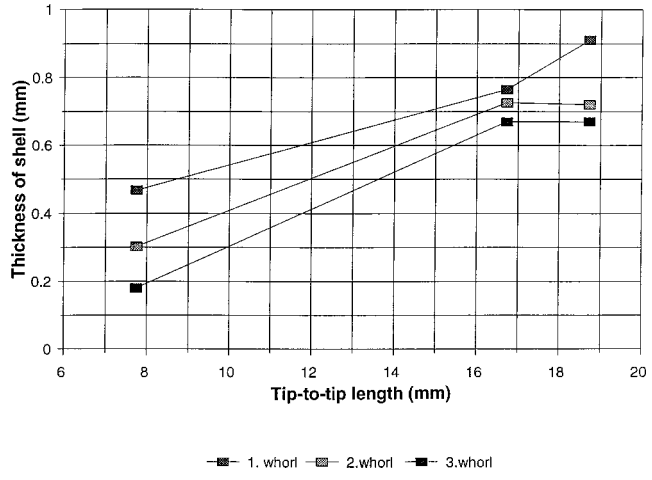


FIG. 3. Thickness of animal shell plotted as a function of tip-to-tip length of body. Because of the growth process of the animal, the shell was naturally of varying thickness (the older sections of the shell were thicker). Therefore, thickness as measured from several different sections or ''whorls'' within each body are plotted. Each whorl corresponds to one 360° rotation of the spiral shape of the body, which corresponds to one growth cycle. The apex of the shell is solid. The lines drawn were simply to connect data from the same class of whorl.

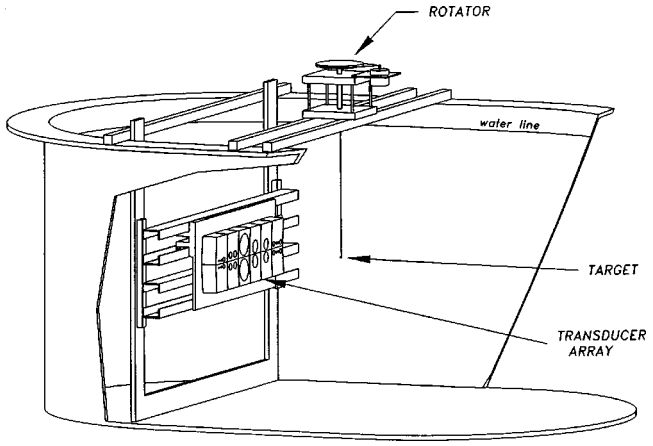


FIG. 4. Experimental setup for measurements of acoustic backscatter versus angle of orientation.

as small as 1°-increments. The scattering measurements were performed using each of the transducers listed above with one animal (No. 97-1) and using a subset of the transducers for the other five animals.

Calibration was performed by configuring the transducer elements so that the transmitter and receiver transducers of each pair were separated and facing each other. The separation distance was 198 cm for the 24 kHz transducers and 68 cm for all other frequencies. Separation was done by first placing the entire array on the bottom of the tank and then raising each subarray (i.e., the linear array that contains one transducer from each pair) so that the axis of each subarray is aligned vertically and the transducers are facing in the horizontal direction.

Since bubbles can greatly contaminate the quality of the measurements, great care was taken so that there were no bubbles present. Before calibration and the backscatter measurements, all transducers were soaped so that their surfaces would be in good contact with the water and buildup of bubbles would be retarded. Furthermore, the animal shells and tether were soaped before the backscatter measurements. Soaping the shells involved soaking them in soapy water and agitating them so that the soap could reach the interior.

In addition to the initial soaping, the tether was occasionally wiped so that any bubbles that may be building up but were not detected visually could be eliminated. Also, before each time a new shell was placed in the beam for a scattering measurement, a high-speed jet of water was aimed into the opercular opening so that bubbles and debris would be removed from the interior of the shell. The jet, which was produced by a laboratory fluid dispenser, was repeatedly pulsed for several minutes to insure that no bubbles remained inside the shell.

During the background reverberation measurements, the echoes were summed over many pings so that random noise would be eliminated and the resultant coherent echo was stored in the digital oscilloscope. Once the animal was placed in the acoustic beam for the animal scattering measurements, the oscilloscope subtracted the background signal from the echo in real time so that all that remained was the echo from the animal (plus random noise). This procedure was especially important for the lower frequencies where the

reverberation from various parts of the tank dominated the echoes. Because of the low signal-to-noise ratio for the 24-kHz measurements, the animal echo after subtraction was also summed from repeated transmissions so that the random noise in the echo could be reduced. The 50-kHz backscatter measurements would not have been possible without the subtraction procedure, and the 24-kHz measurements required both procedures.

During the backscatter measurements, great care was taken so that all animals were rotated through the same body plane and that the set of angles of rotation from each animal was the same (e.g., 0° was the same orientation relative to the transducers for each animal). The lengthwise axis of each animal was approximately horizontal. In order to align the animals in a consistent manner, the shaft was rotated manually immediately before the measurements began for a given animal so that each experiment would begin with the apex of the animal facing the transducer pair. More precisely, the animal was aligned so that the line connecting the outer tip of the opercular opening (opposite the apex) and the apex intersected the middle of the transducer pair. The opercular opening was aimed in the horizontal direction so that at approximately 240° of rotation, the opening would be facing the transducer. The alignment was performed visually and showed excellent repeatability. By use of the computer control, the scattering measurement involved a sequence of alternating single pings and rotations of 1–3° of the stepper motor.

One challenge involved the fact that there appeared to be torsional friction on the tether used to suspend the animals. The source of this friction is perhaps from the surface tension of the water. Because of this friction, the rotation of the animal didn't always correspond with the rotation of the stepper motor. In order to insure one-for-one correspondence between the rotation of the motor and animal, all experiments involved 450° of rotation. After each run, the polar plots of backscatter versus angle of orientation were visually checked. If there was not good agreement between the 0–90° data and the 360–450° data, the data set was rejected and the measurement was repeated. Sometimes the measurement needed to be repeated several times to produce satisfactory results. The basis for acceptance was straightforward. If any slippage due to the friction occurred, there would always be a substantial amount of it. Thus the rejected data sets would have slippage of the order of tens of degrees and would be obviously unacceptable visually. The acceptable data sets had a very close agreement between the 0–90° data and 360–450° data, as will be illustrated in the next section.

III. EXPERIMENTAL RESULTS

The acoustic backscatter data were analyzed in both the frequency and time domain as well as with respect to animal orientation. In the frequency domain, the target strength versus frequency is investigated for single pings at a fixed angle of orientation as well as averaged over a range of angles. In the time domain, the broadband signals are compressed via a cross-correlation method that resembles matched filtering. This temporal analysis, which is used in concert with the spectral analysis, is crucial in understanding the dominant scattering mechanisms.

A. Observations of dependence of scattering upon acoustic frequency, animal orientation, and animal size

There were strong dependencies observed of the acoustic backscattering upon acoustic frequency as well as size and orientation of the animals (Figs. 5–8). These dependencies are illustrated in plots of backscattering versus orientation at various fixed frequencies or bands of frequencies and fixed animal size (Figs. 5, 6); backscattering versus orientation at fixed frequency and different animal sizes (Fig. 7); and backscattering versus frequency at various fixed angles of orientations (Fig. 8).

Except for the lowest frequencies, the backscattering by the periwinkles was strongly dependent upon angle of orientation (Fig. 5). The orientation dependence also varied strongly with frequency. Generally, the data at the lowest frequencies showed that the backscattering was mildly dependent upon orientation. As frequency is increased, the lobes of the scatter pattern tend to become narrower as well as greater in number.

At the higher frequencies, there was much structure in the scatter patterns. In an attempt to eliminate much of the fine structure that was highly variable from frequency to frequency, the directivity patterns were investigated after an average across the band of frequencies from each of the broadband transducers was made (Fig. 6). Some of the fine structure was, indeed, eliminated, although some strong sidelobes still remained.

For a fixed frequency, the orientation dependence of the backscattering also varied with animal size (Fig. 7). At 50 kHz, the data show that the scatter pattern is nearly omnidirectional for the smallest animal and strongly directional for the largest animal.

At any given fixed angle of orientation of the animal, there is also significant dependence of the backscattering upon acoustic frequency (Fig. 8). At each angle over the entire range of orientations, there existed a very strong structure containing peaks and deep nulls. The peaks and nulls were generally not regularly spaced.

B. Compression of broadband echoes in time domain

In addition to the spectral analysis described above, the echoes from an individual periwinkle were analyzed in the time domain as well. All transmitted waveforms were sufficiently long (usually 200–400 μs) that individual features could not be resolved in the time domain without further processing. Because of the broadband nature of the echoes made possible by the octave-bandwidth transducers, the echoes using those transducers could be temporally compressed so that some of the features could be revealed.

The compression was performed by cross-correlating the echoes with the waveform received in the calibration. This ''replicate'' corresponded to the echo one would expect in the backscatter measurement if a target possessing a uniform

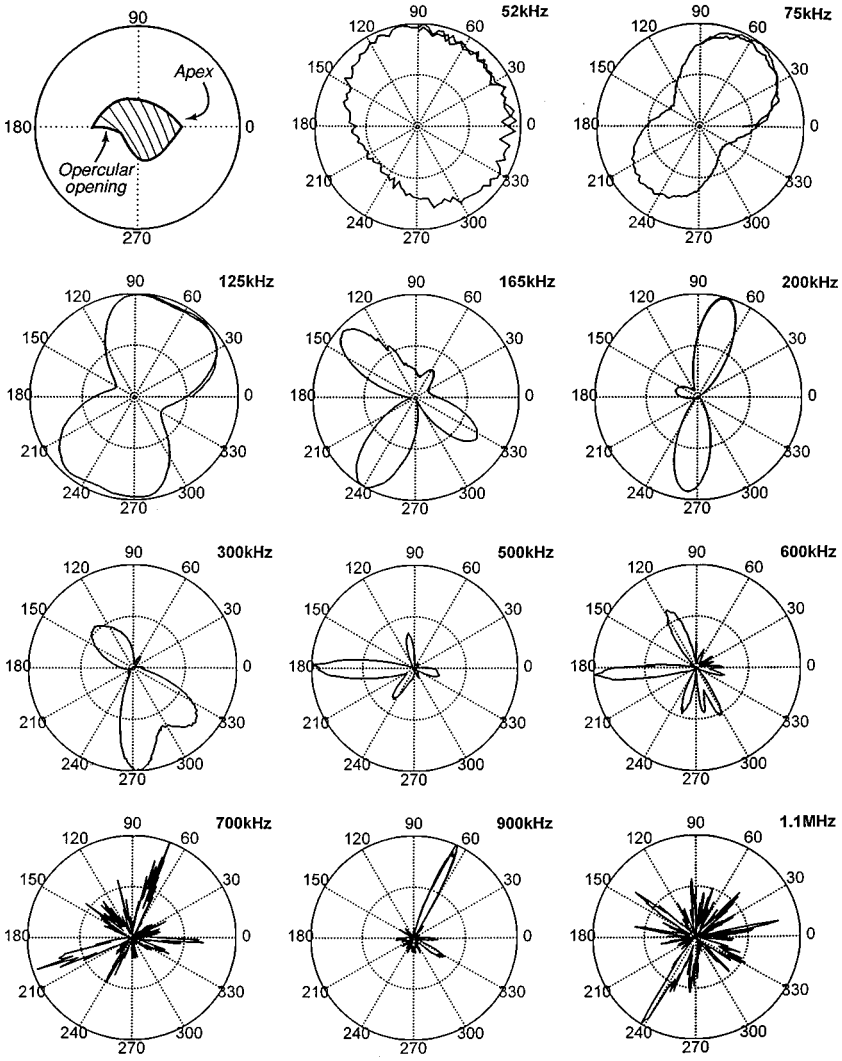


FIG. 5. Normalized differential backscattering cross section σ_{bs} (linear scale, not logarithmic) versus angle of orientation for a number of frequencies for an individual periwinkle animal (No. 97-1, 6.2-mm long). Each plot was normalized by its maximum value, resulting in a 0–1 range of normalized σ_{bs}. Bottom view of animal in experiment is sketched in upper left corner. The data were collected over the range of angles 0 to 450° in 1°–3° steps (1° steps for frequencies at or above 165 kHz, 2–3° steps for the lower frequencies). Data from all 450° are shown in order to illustrate reproducibility of the rotation of the animal. 0° corresponds to the apex of the shell aimed at the acoustic transducers (the apex was actually aimed slightly above the transducers—an amount that varied from animal to animal). The opercular opening faces the transducers at approximately 240° (the normal of the plane of the opening faced slightly above the transducers at 240°). The overlap between the main lobe at about 65° in the 900-kHz data that was observed in both the 0–90° data and in the 360–450° data illustrates the high degree of reproducibility with this tether arrangement. The 24-kHz data for this animal was only collected at 0° and is therefore not shown. $k_1 a_{\mathrm{esr}} = 1$ at about 100 kHz for this animal.

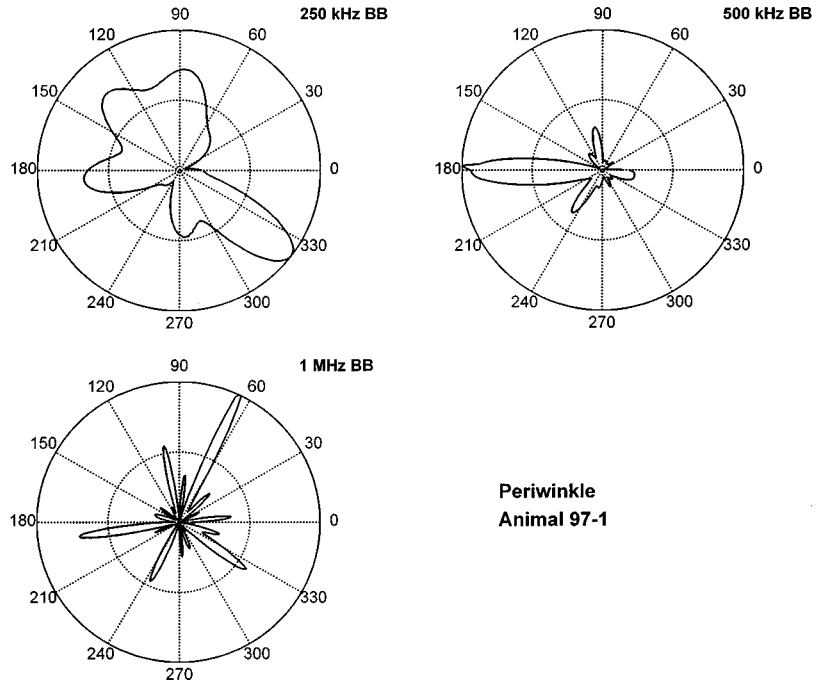


FIG. 6. Normalized backscattered energy versus angle of orientation for the three broadband measurements for a single animal (No. 97-1, 6.2-mm long). The energy (plotted on a linear scale, not logarithmic) was calculated by summing the square of the time series. Each plot is normalized by the maximum value, resulting in a 0–1 range in values. The center frequency of each (octave bandwidth) transducer is shown in the plot. Certain spectral components (Fourier bins) of this broadband data are shown in Fig. 5. $k_1 a_{\mathrm{esr}} = 1$ at about 100 kHz for this animal.

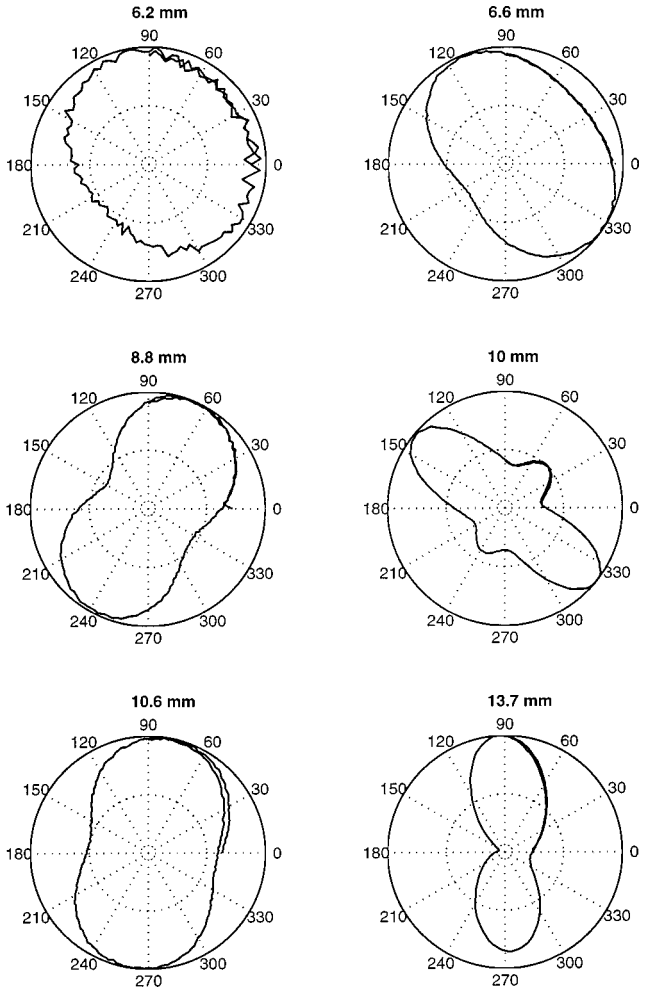


FIG. 7. Normalized differential backscattering cross section σ_{bs} (linear scale, not logarithmic) versus angle of orientation for six individual periwinkles of different sizes at 50 kHz. The animals ranged in length from 6.2 to 13.7 mm. $k_1 a_{esr} \cong 1$ for the 13.7-mm-long animal. Each plot was normalized by its maximum value, resulting in a 0–1 range of normalized σ_{bs}. Data collected in same manner as described in Fig. 5 caption.

frequency response was used. This cross-correlation process is formally known as a matched filter if the replicate incorporates the scattering properties of the target. Our replicate does not incorporate the scattering properties since it is those properties that are unknown and under investigation. Therefore, the pulse-compression process performed on these echoes generally was not a matched filter. However, the deviations of the cross-correlation process using the idealized (uniform frequency response) replicate from what one would expect from a matched filter output contains information (Stanton *et al.*, 1998a; Chu and Stanton, 1998). For example, if the scatterer is composed of a number of facets, the compressed pulse output using the calibration waveform for the replicate as described above will be composed of multiple main lobes, one per feature.

The compressed pulse output for the 500-kHz broadband data from animal 97-1 showed a structure in the (compressed) time series that depended upon orientation (Figs. 9 and 10, Table II). For many orientation angles, the output contained two or more major lobes. The separation or time delay between the lobes varied with angle. Some of the time delays corresponded to distances greater than what one would expect from two facets positioned along the direction of propagation of the signal and separated by a distance equal to the diameter of the body. Thus it is apparent that more than a simple facet or feature-based model is required to describe the scattering. These latter features are qualitatively consistent with those subsonic circumferential waves observed with planktonic gastropods (*Limacina retroversa*) in Stanton *et al.* (1998a, b) and Chu and Stanton (1998).

IV. COMPARISON OF SCATTERING PREDICTIONS WITH DATA

The structure in the scattering data (plotted versus frequency) presented in Sec. III indicates that there are interfer-

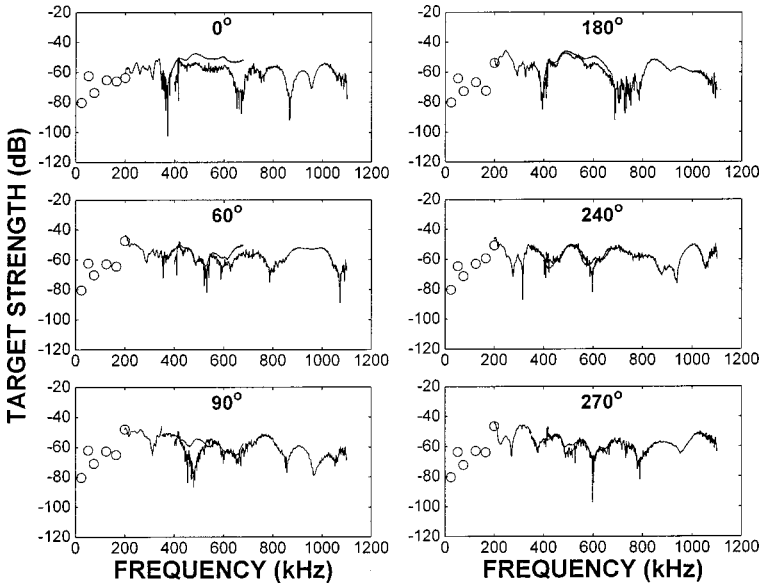


FIG. 8. Target strength versus frequency for six angles of orientation for an individual periwinkle (No. 97-1, 6.2-mm long). Four of the angles were chosen arbitrarily as they were multiples of 90°. The angles 60 and 240° were chosen because they corresponded to peaks in some of the scattering patterns, as illustrated in Fig. 5. $k_1 a_{esr} = 1$ at about 100 kHz for this animal. The discrete points are data from single-frequency transducers while the continuous curves are data from the broadband transducers. The overlapping curves are due to the fact that frequency ranges of the broadband transducers sometimes overlapped each other. Data collected in same manner as described in Fig. 5 caption. Since the scattering pattern was essentially uniform with respect to angle for the 24 kHz measurement, the TS value for 24 kHz made at 0° was used in each plot for that frequency.

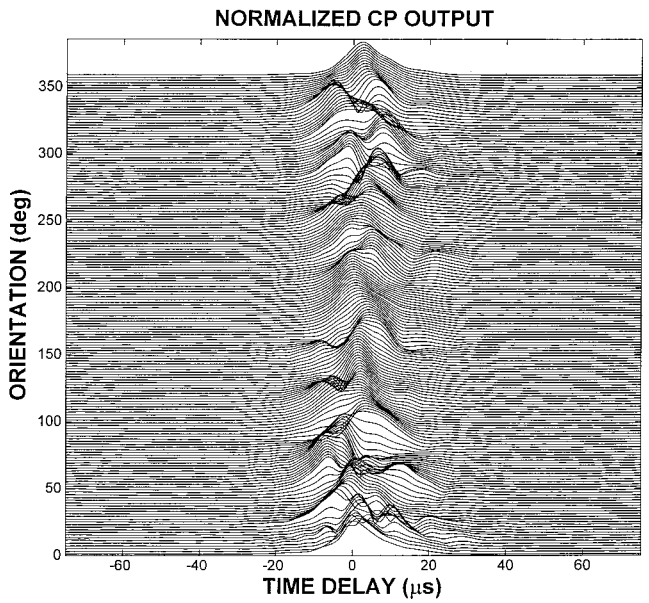


FIG. 9. Compressed pulse (CP) output for 500-kHz broadband echoes from individual periwinkle (animal 97-1, 6.2-mm long) as a function of orientation angle. Data collected as described in Fig. 5 caption. For the purposes of illustration, only a subset of the angles over the entire range 0 to 360° is shown. Each CP output is normalized to the same peak amplitude.

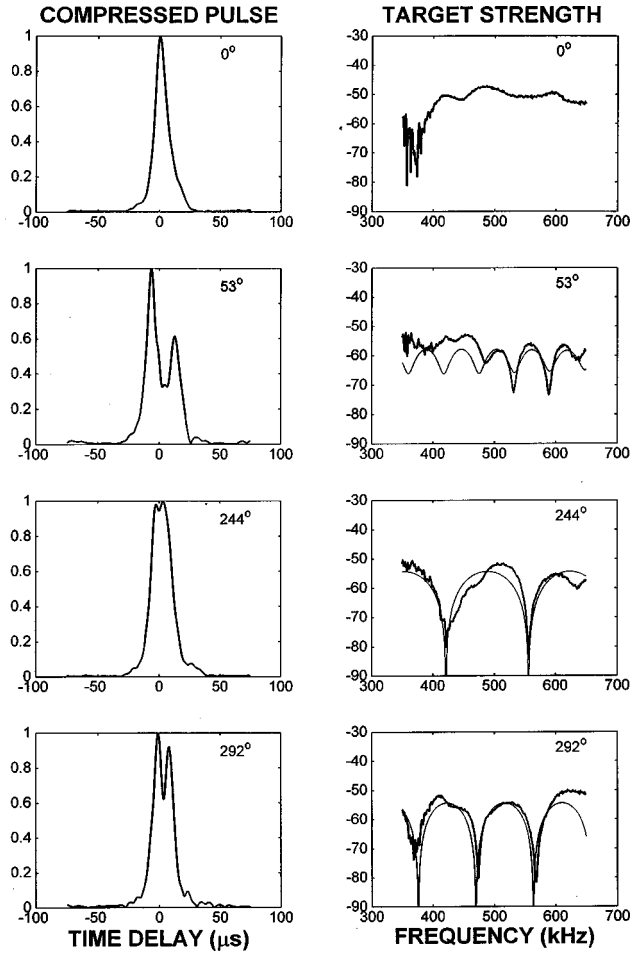


FIG. 10. Compressed pulse (normalized), frequency spectrum, and ray-based scattering predictions for several orientations of an individual periwinkle (animal No. 97-1, 6.2-mm long). Data (thick lines) from set described in Fig. 5. Theoretical predictions (thin lines) were produced as follows: 53° orientation: These calculations were similar to those for *Limacina retroversa* given in Fig. 2 of Stanton *et al.* (1998b) with the main difference being animal size, F_{spec}, and Lamb wave speed. Equation (6) of this paper was used with only f_{spec} and f_{Lamb} contributing ($f_{\text{op}}=0$). $\mathcal{R}_{12}=0.84$, $a_{\text{spec}}=\overline{a}=a_{\text{esr}}=2.28$ mm, $\sigma_r=0.025\overline{a}$ (inferred from the data), $\beta_L=0.002\ k_1\overline{a}$ (the $k_1\overline{a}$ functional dependence is based on the analytical Lamb wave model and the coefficient 0.002 is based upon a fit to the data), $\alpha_L=3k_1\overline{a}$ (the ''3'' is based on fit to data; the ''$k_1\overline{a}$'' is based on model), $\Phi_L=-\pi/2$ (inferred from data) for the subsonic wave, $\theta_L=\pi/2$ (based on model), $c_L/c_1=k_1\overline{a}/(\alpha_L+0.5)\simeq 1/3$ (this analytical relationship was predicted theoretically and the value of 1/3 was observed), $F_L=1$ (inferred), $F_{\text{spec}}=1$ (inferred), (the $\mathcal{R}_{12}=0.84$ is calculated for the coefficient expected for a semi-infinite planar half-space of calcite where the published values for calcite $g\equiv\rho_2/\rho_1=2.646$ and $h\equiv c_2/c_1=4.345$ (longitudinal waves) (from Carmichael, 1982), the Lamb wave series was truncated to include only the $m=0$ term, and $F_{\text{op}}=0$. For the 244 and 292° orientations: Equation (6) of this paper was used with f_{spec} and f_{op} contributing ($f_{\text{Lamb}}=0$). For both cases, g and h are the same as for the 53° orientation above, $F_{\text{spec}}=1$, $F_{\text{Lamb}}=0$, and the term $a_{\text{spec}}=a_{\text{op}}=a_{\text{esr}}=2.28$ mm in the magnitude portion of each term. The phase shifts, which were more sensitive to shape and size, required adjustment from the mean radius for a good fit to the data: for 244° orientation, $a_{\text{spec}}=a_{\text{op}}=1.22a_{\text{esr}}$ in the exponent and $F_{\text{op}}=e^{-i\pi/4}$; for 292° orientation, $a_{\text{spec}}=a_{\text{op}}=1.75a_{\text{esr}}$ in the exponent and $F_{\text{op}}=1$.

ence phenomena taking place. These phenomena are most likely due to multiple rays of different phases being scattered back toward the receiver and interfering with each other. The presence of multiple returns in the compressed pulse data is consistent with this hypothesis. The dependence of scattering versus angle of orientation of the animal is consistent with the fact that the shape of the shell is complex. The amplitude and phase of the various classes of scattered rays are most likely changing with orientation.

The structure of the scattering for high k_1a_{esr} is modeled on a ping-by-ping basis using a ray-based approach. In order to estimate the scattering over all k_1a_{esr}, a modal-series-based approach is used, but only over ensemble-averaged echoes. In all predictions, the material properties for calcite are used as an approximation to the shell material. As discussed earlier, the shell is composed of two very similar materials—aragonite and calcite. To the authors' knowledge, complete information on material properties is only available on calcite, but not aragonite or calcite-aragonite. Given the closeness of (known) material properties of calcite and aragonite, use of calcite-only parameters is a reasonable approximation.

A. Single-ping analysis

The structure of the patterns in the plots of target strength versus frequency for high k_1a_{esr} is not predictable using the exact modal-series solution to the elastic fluid-filled spherical shell. Given the complexity of the shape of the animal, this is not a surprise. In order to model the structure of the scattering by these complex shapes, an approximate ray approach is used. This approach is intuitive and physically tractable to various rays that will scatter geometrically and cause interference such as the ones observed in Fig. 10. For some angles and bands of frequency, the interference pattern is weak (top row, Fig. 10), indicating that only one ray is dominating the scattering for that orientation. For other angles, two or more rays contribute to the scattering.

The raylike behavior of the scattering at the high k_1a_{esr}

TABLE II. Time separation between main features of compressed pulse output from individual periwinkle for certain angles of orientation. Data from Fig. 10. The time delay from the 53° data is consistent with that of subsonic circumferential waves. The time delay from the 292° data is consistent with the sonic wave round trip between two facets separated by a distance comparable to the size of the body. At 0 and 292° orientation, the apex and opercular opening of the shell are aimed at, or in the general direction of, the transducer pair, respectively.

Orientation (degrees)	Time separation (μ sec)	Comments
0	···	single peak
53	19	subsonic wave
292	9	sonic wave

region is consistent with the type of scattering behavior observed with the limited set of data with the 2-mm-long planktonic gastropods (*Limacina retroversa*) in Stanton *et al.* (1998a, b). With those animals, there was an oscillatory behavior in the scattering that was dependent upon orientation. The oscillations were consistent with the presence of a subsonic Lamb wave with a speed of approximately 1/8 that of the surrounding water.

In these periwinkle data, both a subsonic and sonic wave are observed after the initial echo and both have been modeled (Table II, Fig. 10). The data containing the sonic wave are consistent with rays scattering off of two facets where the path of the acoustic wave was in the water (bottom two rows of Fig. 10). The round-trip time is consistent with distances comparable to the diameter of the body. In this latter case, Eq. (6) was used, using Eqs. (7) and (8) for the specular (front interface) and opercular opening rays. The Lamb wave contribution from Eq. (9) was not included (i.e., $F_L=0$).

The data containing the subsonic wave indicate that there are two waves—one from a facet on the body (a likely candidate being the front interface) and the other due to the antisymmetric zeroth-order Lamb wave which is subsonic and travels around the elastic shell (second row of Fig. 10). Here, Eq. (6) was used with Eqs. (7) and (9) to include the specular and Lamb wave contributions, but excluding the opercular ray (i.e., $F_{\mathrm{op}}=0$).

There is excellent comparison between the two-ray model and the data where the sonic waves are present. The orientations for these data correspond to ones where the opercular opening is facing in the general direction of the receiver. Hence it is reasonable to obtain two sonic waves, one from the front interface and one that travels into the opercular opening and back toward the receiver to interfere with the specular return.

For the data involving subsonic rays, there was reasonable agreement between the predictions and the data for the upper portion of the frequency band and poor agreement between the data and predictions for the low part of the band. The fact that there is agreement over only part of the band indicates that the scattering is more complex than modeled and other rays may exist that contribute significantly to the scattering. This is not a surprise given the simplifications leading up to the predictions. Only one of the large class of Lamb waves was included, and other waves were excluded as well.

What is remarkable about the portion of the frequency band in which the subsonic wave analysis is consistent with the data is the fact that the equation used was the same one used in the predictions of the scattering by the planktonic gastropods in Stanton *et al.* (1998b). In that analysis, a combination of measured parameters (e.g., animal size), inferred parameters (e.g., roughness-induced phase shift of Lamb wave), and parameters that were a compromise between inferred values and values obtainable through simplified theoretical predictions (e.g., Lamb wave coupling coefficient) were used. All of those parameters were used in this current analysis with the exception of animal size (the size of this periwinkle was used), F_{spec} (which is expected to vary with orientation), and Lamb wave speed. The inferred subsonic wave was 1/3 that of the sonic speed versus 1/8 for the *Limacina*. This difference is to be expected since this set of data is at a higher k_1a_{esr} value than for the *Limacina* as well as involving a disproportionately thicker shell, and the Lamb wave speed naturally increases with k_1a_{esr} and relative shell thickness in that region (Kargl and Marston, 1989; Kaduchak *et al.*, 1995).

This ray analysis illustrates some of the basic physics of the scattering process. At other angles, the structure is less regular, indicating that the scattering is more complex and would most likely require more rays in the predictions.

B. Average echoes

It is clear in the above single-ping analysis that the scattering process is very complex. It is so complex, that it may not be practical to be able to accurately predict the scattering for each angle of orientation. In field applications, observations of acoustic scattering usually involve aggregations of animals (both planktonic and benthic applications). These aggregations will involve a distribution of animal size and orientation whose ensemble-averaged levels will tend to have much less structure in the scattering pattern. Thus it is important to use a formulation that can, at the least, predict the averaged smoothed scattering levels.

One candidate for predicting the ensemble-averaged scattering is the exact modal-series solution for the fluid-filled elastic spherical shell. Although it cannot predict the structure of the scattering from single pings for these animals, it can be useful as an approximation in predicting the averaged scattering because it contains much of the underlying physics of the scattering process. Implicit in the solution are the specular wave, all classes of Lamb waves, internally transmitted waves, and Franz waves. Other waves, such as those entering the opercular opening, are not included. Also not included are effects of the irregularities of the shell on the various waves. However, by averaging the predictions over size and shell thickness for values of k_1a_{esr} above 0.5, the resultant variation in phase of the (implicit) waves due to the different sizes and shell thicknesses will tend to have a similar effect that the irregularities have on the waves and smooth out the structure.

The averaging process consisted of averaging the backscattering cross section, as predicted by the exact modal-series solution, over a range of sizes and shell thicknesses. The averaging was only performed for values of k_1a_{esr} above

0.5 and only over a small range of a_{esr} and shell thickness (the range of a_{esr} and shell thickness corresponded to measured irregularity of the shells). For values of $k_1 a_{esr}$ below 0.5, there was no averaging. In this region, there is no interference structure. Furthermore, since the wavelength is generally much greater than any feature on the object, the scattering is much less sensitive to irregularities and orientation. In fact, for the ideal spherical shell, the scattering in this region depends upon the volume of the shell (Goodman and Stern, 1962). The scattering by fluid objects (elastic bodies are fluidlike in this region) of any shape in this region also only depends upon the volume (not shape) of the body. In order to prevent any biases associated with the averaging, averaging was not performed in this lower $k_1 a_{esr}$ region.

Certainly, this approach of averaging over size and shell thickness is not rigorous and it is applied strictly to obtain the observed effect of smoothed out structure due to the averaging of data over orientation. However, there is some formal basis for the averaging over size and shell thickness. As shown in Stanton *et al.* (1998b) and summarized in Eqs. (6)–(9), the scattering has explicit components that depend upon the local radius of curvature of the curved surface facing the transceiver (f_{spec} and f_{op}). This radius depends (implicitly) upon orientation. Thus an average over a range of radii of curvature of the irregular object is, indeed, connected to the process of averaging over angle of orientation. A similar argument can be made for terms in Eqs. (6)–(9) that depend upon shell thickness which, in turn, depends upon orientation. For example, in a more rigorous treatment, $\mathcal{R}_{12}$ is replaced by a term that depends upon shell thickness (Marston, 1992). There are, however, many complexities in the scattering process that cannot be taken into account with this approach, such as orientation-dependent effects associated with the discontinuity of the shell. What is especially important in use of this averaging approach is recognizing that the return associated with the specular reflection from the front interface (f_{spec}) is relatively strong compared with the other ones. If that were the only wave returning from the target, then the averaging process over size and shell thickness would strongly resemble the average over orientation. Although there may be larger errors associated with the other terms, the variation over size in the averages varies the phases of the different terms and creates the desired effects of smoothing out the structure.

In addition to the data to be presented in this section, there is strong empirical evidence in other independent studies that the averaging over size and shell thickness is a reasonable approximation. For example, in a study by Thorne *et al.* (1995), an average over sizes of the exact modal-series solution for solid elastic spheres was compared with data involving irregular solid elastic objects over a wide range of $k_1 a_{esr}$ (Fig. 9 of that paper). The data from the objects involved an average over angle of orientation and sometimes over a distribution of size. There was excellent agreement between the approximate predictions and data. Their analysis demonstrated that, although the approach is not formally correct, it is a reasonable approach toward describing the average scattering for this very complex class of problems.

In this current analysis there was reasonable agreement between the averaged modal-series-based solution and the data averaged over all angles of orientation (Fig. 11). The model parameters were based either on direct measurements or published values (no parameters were adjusted to improve the fit to the data). There was reasonable agreement over much of the range of $k_1 a_{esr}$ (the Rayleigh and geometric region) including the position of the Rayleigh/geometric transition region. However, there were also some significant deviations. There was a strong resonance in the data near $k_1 a_{esr} = 0.5$ that could not be predicted with this scattering approach (Figs. 11 and 12). There were also some disagreements near the Rayleigh/geometric transition region and at the highest values of the $k_1 a_{esr}$. The latter two regions were very sensitive to the choice of scattering parameters. By varying average shell thickness and bulk material properties (neither were known accurately) in the simulations, different parts of the data could be predicted, but never all simultaneously (not shown). Hence the parameters used for Fig. 11 were published values of material properties and average measured shell thickness.

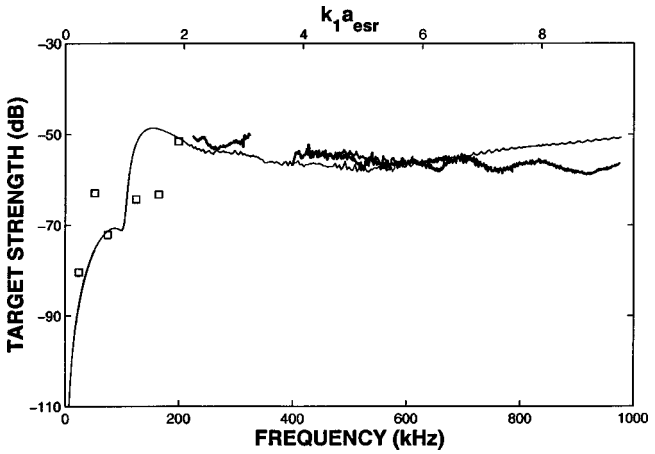


FIG. 11. Comparison between modal-series solution for fluid-filled spherical shell averaged over range of sizes and shell thicknesses with the measured backscattering by an individual periwinkle averaged over all angles of orientation. The average over size and shell thickness (done for $k_1 a_{esr} > 0.5$) is intended to partially emulate the range of radii of curvature and shell thicknesses experienced by the various scattered rays over all angles of orientation. Since the tissue of the body was removed from the shell, the interior fluid of the shell is modeled in this case to be the same as the exterior fluid. Same data as presented in Fig. 5 (animal 97-1, 6.2-mm long). Since the data did not vary significantly with orientation at 24 kHz, only the single datum collected at 0° orientation is plotted at that frequency. Spherical shell solution taken directly from Goodman and Stern (1962). Parameters for the predictions are $\bar{a} = a_{esr} = 2.28$ mm (that is, the average radius used in the predictions was set equal to the measured equivalent spherical radius of the animals; the a_{esr} shown in the plot are the two terms $\bar{a} = a_{esr}$), $g = 2.65$, $hc = 4.35$ (compressional sound speed contrast in shell), $hs = 2.24$ (shear sound speed contrast in shell), and shell thickness $= 0.2$ mm (from average measured thickness) (material properties from Carmichael, 1982). Here the fractional shell thickness $=$ (shell thickness)$/\bar{a} = 0.09$ (where $\bar{a}$ is the average radius used in the predictions). Both the radius and shell thickness were varied in the averages (Gaussian distributed) with a 10% standard deviation (s.d.) and spanning values within ± 2 s.d. of the means.

C. Modeling the low $k_1 a_{esr}$ resonance

As discussed above, there were strong scattering levels observed near $k_1 a_{esr} = 0.5$. The levels were observed with

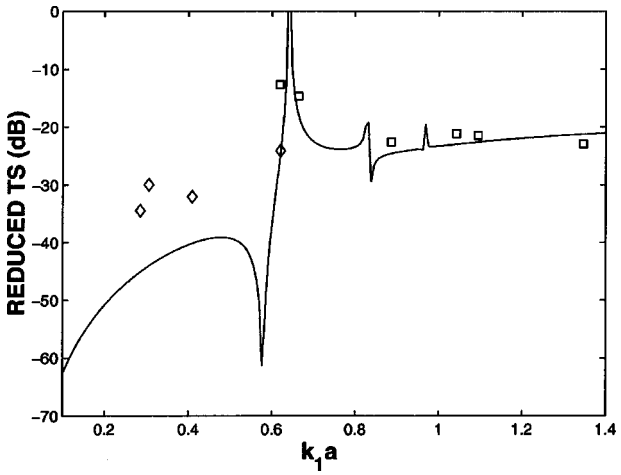


FIG. 12. Modeling study of the resonance structure measured near $k_1a_{esr}=0.5$ (which is at about $k_1a=0.6$ on this plot). Six individual periwinkles were used (Nos. 97-1 through 97-6) and two frequencies [24 kHz (diamond) and 50 kHz (square)]. In order to achieve a fit between the model predictions and the data in this exploratory study, the shell thickness in the predictions needed to be approximately 1/5 that of the average measured shell thickness, while the modeling radius ''*a*'' needed to be slightly larger than the equivalent spherical radius of the animal. A single realization of the modal-series solution described in Fig. 11 was used with $g=2.65$, $hc=4.35$, $hs=2.24$, and fractional shell thickness=(shell thickness)/$a=0.014$. The values of a_{esr} given in Table I for the animals were multiplied by 1.25 for use in k_1a and *RTS* in the plot of data. All data correspond to 0° orientation (i.e., the apex was aimed at the transducers).

repeated measurements of the same animal for replication as well as with a different sized animal. The levels could not be predicted using the modal-series-based approach used above with the average shell thickness values.

We explored the low k_1a_{esr} resonance with the modal-series approach by varying parameters outside of the measured range and found that, for a very thin shell and use of a radius that is slightly larger than the equivalent spherical radius, one could predict a resonance in that region (Fig. 12). In that study, the thickness of the simulated shell was 0.014 times the simulated radius of the shell. This corresponds to a shell that has a thickness that is roughly 1/5 that of the average thickness of the actual measured shell. This observation could lead to two possibilities: (1) Since the shell has a variable thickness, perhaps there are sections that are thin enough to produce such a resonance, (2) since the shell is not spherical and hollow, but rather has internal structure, the structure provides a mechanism for altering the resonance characteristics of the shell.

V. APPLICATION TO ACOUSTIC SCATTERING IN THE OCEAN

In spite of the fact that the scattering by the shelled benthic and planktonic animals is strong, there have been few quantitative studies of the acoustic scattering by such animals in the ocean. Below are brief summaries and discussions of applications presented in other papers of simplified versions of these acoustic scattering models to sound scattering by the seafloor and water volume when shelled animals are present.

A. Seafloor

The scattering by a seafloor covered with shelled animals has been predicted using the results from the periwinkle experiment described herein and compared with data collected in the ocean. The analysis is described in Stanton (2000) and will only be briefly summarized and discussed in the context of the results presented in this paper. In that paper, the key assumption was that first-order scattering from the top layer of shells dominated the scattering. Higher-order scattering, shadowing, and contributions from the seafloor were explicitly ignored which limited the analysis (although note that the Lambert's Law formula implicitly takes these effects into account to some extent). In spite of these limitations, the formulation provided insight into the scattering problem. With these assumptions, the area scattering strength S_A of the seafloor at high frequencies was expressed quite simply, in Lambert's Law form, as

$$S_A=\langle \mathrm{RTS}^{(m)}\rangle+10\log\mathcal{F}+10\log\sin^2\theta_g, \tag{14}$$

where $\langle \mathrm{RTS}^{(m)}\rangle$ is the reduced target strength based on the average scattering properties (measured or modeled) of an individual animal, $\mathcal{F}$ is the packing factor of the layer of animals (fraction of area of seafloor covered by the animals), and θ_g is the grazing angle ($\theta_g=90°$ corresponds to normal incidence). What is important about this equation is the fact that the area scattering strength is expressed in terms of the *reduced* target strength of the animals rather than the target strength. As defined before, the reduced target strength is the target strength normalized by size of the animal. For high k_1a_{esr}, the reduced target strength (based on the average scattering properties) is not only independent of size, but, on average, relatively independent of frequency. Thus for geometric scattering, the area scattering strength is relatively independent of acoustic frequency and size of feature on the seafloor. Because of this independence, to a first approximation, the data and modeling for one size of animal can possibly be applied to scattering by surfaces containing other sizes (provided that they too are in the geometric scattering region).

Equation (14) and geometric scattering data in this paper (Fig. 11) were used to make predictions for area scattering strength and compared with scattering data collected by Jackson *et al.* (1986) and Stanic *et al.* (1989). In those papers, data were presented involving sound scatter at 20–50 and 20–180 kHz, respectively, by sections of the seafloor that were covered with a dense layer of shells. Although the size distribution of shells was not documented in the Jackson *et al.* (1986) paper and the sizes reported in the Stanic *et al.* (1989) paper were different than those used in this experiment, our values of reduced target strength (based on average scattering properties) were used because of the independence of S_A, to first order, upon size. The predictions were within the range of values observed by Jackson *et al.* (1986) of area scattering strength near normal incidence and slightly higher than the normal incidence values extrapolated via Lambert's Law from the Stanic *et al.* (1989) data (in the geometric scattering region where the frequencies were above 60 kHz) which involved shallow grazing angles. This reasonable comparison showed that, in spite of the many assumptions

made in the surface scattering formulation, the reduced target strength data (or equivalently, the reduced target strength predictions) could robustly make the seafloor scattering predictions for a limited range of conditions. These results show promise of applying normalized scattering data or model from one type of (individual) benthic animal to seafloor regions containing aggregations of another type.

B. Volume

The scattering by planktonic shelled animals has been predicted using an earlier version of the models described here and compared with acoustic survey data. The analysis is presented in detail in Wiebe *et al.* (1996) and will only be briefly summarized below in the context of the results presented in this paper. In that paper, a survey using a 420-kHz echosounder over the Georges Bank (near Cape Cod, MA) is described. Gastropods with numerical densities of up to 1000's m^{-3} were observed and sometimes dominated the echoes (volume scattering strength of up to -56 dB). In the analysis, a ''high-pass'' model from Stanton *et al.* (1994) was used which incorporated the low k_1a limit based on a fluid sphere modal-series solution for low frequencies and a ray solution [specular component given in Eq. (7) of this paper] for the high frequencies. Since all data were in the geometric scattering region, only the high k_1a_{esr} portion was taken advantage of in the research. Acoustic scattering predictions were made based upon size and taxa of animals caught in net tows. The predictions were compared with data collected in the (420-kHz) acoustic surveys from the same body of water. There was very good agreement between predictions and observed volume scattering strengths. The slope of the plot of predicted values versus observed data was essentially identical to the idealized expected slope of unity, although there was roughly a 3.5-dB offset between predictions and observations. In some of the regions, it is estimated that the (2 mm) gastropods dominated the scattering, which is useful in validation of the modeling approach.

The near agreement between the predictions and observations using the very simple approach in that analysis was promising. The simple model, which has no structure in the high k_1a_{esr} region, apparently was reasonable in estimating the scattering by the aggregation of scatterers. As shown in Fig. 11 of this paper, when the scattering is averaged over a distribution of sizes and/or orientations, the high k_1a_{esr} portion of the scattering tends to become smooth. Thus some of the details that are brought out in the more complex model in Eq. (6) may not be very important in the aggregation average. The exception to that observation is that, although the structure may be washed out in the average over size and orientation, the various waves that give rise to the structure have substantial energy and that energy will contribute to the overall scattering levels.

Finally, since the data presented in Wiebe *et al.* (1996) were only at one (high k_1a_{esr}) frequency, the low k_1a_{esr} portion of the simplified model could not be tested. However, the results presented in Fig. 11 of this paper indicate that the averaged scattering near the Rayleigh/geometric transition region has enough structure that the simple model used in Wiebe *et al.* (1996) most likely will not suffice. A more general model that incorporates the elastic shelled properties of the body is required, such as the one presented in this paper.

VI. SUMMARY AND CONCLUSIONS

The backscattering data collected with both the benthic animals described here and the planktonic ones described in previous studies indicate the great complexity of the scattering processes involved. The fact that the scattering is so strongly dependent upon angle of orientation shows that the irregularity of the shell needs to be taken into account. Both the spectral and temporal (compressed pulse) analyses provided valuable information regarding the dominant scattering processes. In the high k_1a_{esr} region, there were typically two or more rays scattered back toward the receiver. One of the rays was the specular return off of the front interface while the other was either a ray from within the opercular opening, a subsonic Lamb wave, or other rays not quantified in the analysis. The Rayleigh and Rayleigh/geometric scattering regions were also complicated.

There was reasonable success in predicting the scattering by use of both ray-based and modal-series-based models. Depending upon orientation, the ray-based models used either the echo from the front interface and a Lamb wave or the echo from the front interface and a ray from the opercular opening. Other orientations involved scattering patterns that were more complex than a simple two-ray model could predict. The modal-series solution was averaged over a range of sizes and shell thicknesses to produce reasonable predictions of the data when averaged over angle of orientation. The average of the sizes and thicknesses seemed to at least partially account for the range of curvature and thicknesses seen by the incident acoustic waves over the range of angles. What was especially remarkable in the (periwinkle) analysis was that the ray model involving the Lamb wave used all but three of the parameters used in the analysis involving the smaller gastropods. The three parameters were animal size (which naturally changes with animal), F_{spec} (which was expected to change with orientation angle), and Lamb wave speed (which was expected to change with the higher k_1a_{esr} and relative shell thickness). Thus the parameters appropriate for the planktonic gastropods which have a similar shape, but disproportionately thinner shell, appear to be applicable to the benthic gastropods.

Application of the data and modeling involving the periwinkles to the problem of scattering by the seafloor were discussed. The results from another paper were summarized which used the measured reduced target strength of the periwinkles to predict scattering by shell-covered sections of the seafloor. The predictions were close to or comparable to the observations, which is a remarkable result considering the differences (or possible differences) between the periwinkles and the animals in the seafloor experiments. Given the fact that the area scattering strength at high frequencies is, to first order, independent of size of (volumetric) feature when a covering exists, this study showed the promise in using reduced target strength data or models such as those presented in this paper to help predict sound scattering by shell-covered seafloors.

Applications of the scattering models to planktonic shelled animals were also discussed. The results of an earlier paper were summarized, in which a greatly simplified version of the models presented in this work produced reasonable predictions of the scattering observed in an acoustic survey of a region in which planktonic gastropods dominated the scattering. However, that work involved only a single high frequency (geometric scattering) and scattering by ensembles of animals. Work involving lower frequencies and/or echoes involving single animals would require use of the more sophisticated models.

In conclusion, through analysis and modeling of the extensive data set presented herein, great progress has been made toward the understanding of the complex nature of the scattering by benthic and planktonic shelled animals. A better understanding of the dominant scattering mechanisms has been achieved for these complex bodies as well as scattering models developed or extended with conditions of validity determined.

ACKNOWLEDGMENTS

The authors are grateful to Erhan Mutlu of the Institute of Marine Sciences, Middle East Technical University, Erdemli, Icel, Turkey, for assisting in the measurement of the periwinkle shells during his visit to the Woods Hole Oceanographic Institution (WHOI) in 1997 and Shirley Barkley and Sheila Hurst of WHOI, Woods Hole, MA, for preparing the manuscript to this article. This work was supported by the U.S. Office of Naval Research Grant Nos. N00014-95-1-0287 and N00014-96-1-0878, and the MIT/WHOI Joint Graduate Education Program. This is contribution number 9794 for the Woods Hole Oceanographic Institution.

Bowman, J. J., Senior, T. B. A., and Uslenghi, P. L. E., Eds. (**1987**). *Electromagnetic and Acoustic Scattering by Simple Shapes* (Hemisphere, New York).

Carmichael, R. S., Ed. (**1982**). *CRC Handbook of Physical Properties of Rocks* (CRC, Boca Raton, FL), Vol. II.

Chu, D., and Stanton, T. K. (**1998**). ''Application of pulse-compression techniques to broadband acoustic scattering by live individual zooplankton,'' J. Acoust. Soc. Am. **104**, 39–55.

Chu, D., Stanton, T. K., and Wiebe, P. H. (**1992**). ''Frequency dependence of sound backscattering from live individual zooplankton,'' ICES J. Mar. Sci. **49**, 97–106.

Felsen, L. B., and Lu, I. T. (**1989**). ''Ray treatment of wave propagation on thin-walled curved elastic plates with truncations,'' J. Acoust. Soc. Am. **86**, 360–374.

Foote, K. G. (**1980**). ''Importance of the swimbladder in acoustic scattering by fish: A comparison of gadoid and mackerel target strengths,'' J. Acoust. Soc. Am. **67**(6), 2084–2089.

Goodman, R. R., and Stern, R. (**1962**). ''Reflection and transmission of sound by elastic spherical shells,'' J. Acoust. Soc. Am. **34**, 338–344. For modal-series coefficients involving spherical shells with an interior fluid different than that of the exterior fluid see, for example, Stanton, T. K. (**1990**). ''Sound scattering by spherical and elongated, shelled bodies,'' *ibid.* **88**, 1619–1633.

Ho, J. M. (**1994**). ''Geometrical theory of acoustic scattering by thin elastic shells,'' J. Acoust. Soc. Am. **96**, 3115–3125.

Ho, J. M., and Felsen, L. B. (**1990**). ''Nonconventional traveling wave formulations and ray-acoustic reductions for source-excited fluid-loaded thin elastic spherical shells,'' J. Acoust. Soc. Am. **88**, 2389–2414.

Jackson, D. R., Baird, A. M., Crisp, J. J., and Thomson, P. A. G. (**1986**). ''High-frequency bottom backscatter measurements in shallow water,'' J. Acoust. Soc. Am. **80(4)**, 1188–1199.

Jebsen, G. M., and Medwin, H. (**1982**). ''On the failure of the Kirchhoff assumption in backscatter,'' J. Acoust. Soc. Am. **72**, 1607–1611.

Kaduchak, G., Kwiatkowski, C. S., and Marston, P. L. (**1995**). ''Measurements and interpretation of the impulse response for backscattering by a thin spherical shell using a broad-bandwidth source that is nearly acoustically transparent,'' J. Acoust. Soc. Am. **97**, 2699–2708.

Kargl, S. G., and Marston, P. L. (**1989**). ''Observations and modeling of the backscattering of short tone bursts from a spherical shell: Lamb wave echoes, glory, and axial reverberations,'' J. Acoust. Soc. Am. **85**, 1014–1028.

Lalli, C. M., and Gilmer, R. W. (**1989**). *Pelagic Snails* (Stanford University Press, Stanford), Chap. 4.

Lowenstam, H. A. (**1954**). ''Factors affecting the aragonite: calcite ratios in carbonate-secreting marine organisms,'' J. Geol. **62**, 284–322.

Marston, P. L. (**1992**). ''Geometrical and catastrophe optics methods in scattering,'' in Physical Acoustics, edited by A. D. Pierce and R. N. Thurston (Academic, New York), Vol. 21.

Norris, A. N., and Rebinsky, D. A. (**1994**). ''Acoustic coupling to membrane waves on elastic shells,'' J. Acoust. Soc. Am. **95**, 1809–1829.

Rebinsky, D. A., and Norris, A. N. (**1995**). ''Benchmarking an acoustic coupling theory for elastic shells of arbitrary shape,'' J. Acoust. Soc. Am. **98**, 2368–2371.

Stanic, S., Briggs, K. B., Fleischer, P., Sawyer, W. B., and Ray, R. I. (**1989**). ''High frequency acoustic backscattering from a coarse shell ocean bottom,'' J. Acoust. Soc. Am. **85**(1), 125–136.

Stanton, T. K. (**1990**). ''Sound scattering by spherical and elongated shelled bodies,'' J. Acoust. Soc. Am. **88**, 1619–1633.

Stanton, K. T. (**2000**). ''On acoustic scattering by a shell-covered seafloor,'' J. Acoust. Soc. Am. **108**, 551–555.

Stanton, T. K., Chu, D., and Wiebe, P. H. (**1996**). ''Acoustic scattering characteristics of several zooplankton groups,'' ICES J. Mar. Sci. **53**, 289–295.

Stanton, T. K., Chu, D., Wiebe, P. H., Martin, L. V., and Eastwood, R. L. (**1998a**). ''Sound scattering by several zooplankton groups. I. Experimental determination of dominant scattering mechanisms,'' J. Acoust. Soc. Am. **103**, 225–235.

Stanton, T. K., Chu, D., and Wiebe, P. H. (**1998b**). ''Sound scattering by several zooplankton groups. II. scattering models,'' J. Acoust. Soc. Am. **103**, 236–253.

Stanton, T. K., Wiebe, P. H., Chu, D., Benfield, M., Scanlon, L., Martin, L., and Eastwood, R. L. (**1994**). ''On acoustic estimates of zooplankton biomass,'' ICES J. Mar. Sci. **51**, 505–512.

Thorne, P. D., Hardcastle, P. J., and Soulsby, R. L. (**1993**). ''Analysis of acoustic measurements of suspended sediments,'' J. Geophys. Res. **98**(C1), 899–910.

Thorne, P. D., Waters, K. R., and Brudner, T. J. (**1995**). ''Acoustic measurements of scattering by objects of irregular shape,'' J. Acoust. Soc. Am. **97**(1), 242–251.

Überall, H. (**1973**). ''Surface waves in acoustics,'' in *Physical Acoustics*, edited by W. B. Mason and R. N. Thurston (Academic, New York), Vol. 10, pp. 1–60.

Wiebe, P. H., Mountain, D. G., Stanton, T. K., Greene, C. H., Lough, G., Kaartvedt, S., Dawson, J., and Copley, N. (**1996**). ''Acoustical study of the spatial distribution of plankton on Georges Bank and the relationship between volume backscattering strength and the taxonomic composition of the plankton,'' Deep-Sea Res. II **43**(7–8), 1971–2001.

Wiebe, P. H., Stanton, T. K., Benfield, M. C., Mountain, D. G., and Greene, C. H. (**1997**). ''High-frequency acoustic volume backscattering in the Georges Bank coastal region and its interpretation using scattering models,'' IEEE J. Ocean Eng. **22**(3), 445–464.

Yang, Y., Norris, A. N., and Couchman, L. S. (**1995**). ''Acoustic scattering from fluid-loaded elastic shells: A Gaussian beam approach,'' J. Acoust. Soc. Am. **98**, 611–622.

Yang, Y., Norris, A. N., and Couchman, L. S. (**1996**). ''Ray tracing over smooth elastic shells of arbitrary shape,'' J. Acoust. Soc. Am. **99**, 55–65.

Zhang, L. (**1996**). ''High-Frequency Seafloor Acoustic Backscattering with the Presence of Turbulence in Seawater,'' Ph.D. dissertation, University of Southern Mississippi.

On acoustic scattering by a shell-covered seafloor

Timothy K. Stanton
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 14 September 1998; revised 27 September 1999; accepted 27 October 1999)

Acoustic scattering by the seafloor is sometimes influenced, if not dominated, by the presence of discrete volumetric objects such as shells. A series of measurements of target strength of a type of benthic shelled animal and associated scattering modeling have recently been completed (Stanton *et al.*, ''Acoustic scattering by benthic and planktonic shelled animals,'' J. Acoust. Soc. Am., this issue). The results of that study are used herein to estimate the scattering by the seafloor with a covering of shells at high acoustic frequencies. A simple formulation is derived that expresses the area scattering strength of the seafloor in terms of the average reduced target strength or material properties of the discrete scatterers and their packing factor (where the reduced target strength is the target strength normalized by the geometric cross section of the scatterers and the averaging is done over orientation and/or a narrow range of size or frequency). The formula shows that, to first order, the backscattering at high acoustic frequencies by a layer of shells (or other discrete bodies such as rocks) depends principally upon material properties of the objects and packing factor and is independent of size and acoustic frequency. Estimates of area scattering strength using this formula and measured values of the target strength of shelled bodies from Stanton *et al.* (this issue) are close to or consistent with observed area scattering strengths due to shell-covered seafloors published in other papers. [S0001-4966(00)02702-8]

PACS numbers: 43.30.Hw, 43.30.Sf [DLB]

INTRODUCTION

Acoustic scattering by the seafloor has long been studied in order to either predict the performance of sonar systems or to use sound to quantitatively map the seafloor. The scattering is influenced by the roughness of the interfaces between the water and bottom and subbottom layers as well as inhomogeneities (Medwin and Clay, 1998; Ogilvy, 1991; Urick, 1983; Jackson *et al.*, 1986a, b; Jackson and Briggs, 1992; Jackson and Ivakin, 1998; Stanic *et al.*, 1989; Tang *et al.*, 1994, 1995; Richardson and Briggs, 1996; Ivakin, 1998). There are both continuously varying inhomogeneities and discrete ones. Rocks, shells, and gas pockets are among the discrete inhomogeneities.

There is evidence that the presence of shells on the seafloor can influence, if not dominate, the scattering (Jackson *et al.*, 1986b; Stanic *et al.*, 1989; Zhang, 1996). Descriptions to date of the effects of the scattering by beds of shells have generally involved incorporating the shells as part of the continuously rough seafloor. This approach can produce reasonable estimates of the scattering provided that the bed of shells resembles a single-valued featureless surface. For other conditions, the discrete or volumetric nature of the shells can result in a multi-valued surface (e.g., a spherical shell lying on an interface is described by a multi-valued function). Scattering effects specific to a multi-valued surface may be important in the estimates for both dense and sparse distributions. Accounting for the discrete nature of scattering by shell-covered seafloors has been limited, in part, by the general lack of information on the scattering characteristics of individual shells (Zhang, 1996).

Recently, an extensive set of measurements of target strength has been performed on the scattering by a type of benthic shelled animal (Stanton *et al.*, 2000). This substantial data set served as a basis for a target strength model of the animals for a wide range of sizes and acoustical frequencies. In this paper, the model and data are used to estimate the levels of acoustic scattering at high acoustic frequencies that may be expected from a seafloor that is covered with shells. A simple approximate formula for scattering by a layer of discrete scatterers is derived in order that the estimates be made. A comparison of the estimate using the target strength data is made with seafloor scattering data presented in Jackson *et al.* (1986b) and Stanic *et al.* (1989).

I. TARGET STRENGTH MEASUREMENTS OF BENTHIC SHELLED ANIMALS

In a recent study, the scattering characteristics of periwinkles (*Littorina littorea*), a type of benthic shelled animal, were studied (Stanton *et al.*, 2000). Measurements of backscattering were made in free space—i.e., the animals were away from any boundaries. The backscattering was measured over parts or all of the range 24 kHz to 1 MHz for 0- to 360-degrees orientation in as small as 1-degree increments. The length of the six animals ranged from 6 to 14 mm. Discrete (narrow-band) frequencies were used over most of the frequency range and broadband signals were used at the higher frequencies. Both the spectral and temporal (pulse compression) characteristics of the data were examined and served, in part, as the basis of scattering modeling.

The scattering process of the animals was observed to be quite complex as the echoes were strongly dependent upon both frequency and angle of orientation. For example, at the high frequencies, dominant echoes were observed from the front interface as well as sometimes from the inside of the opercular opening and from circumferential waves (subsonic Lamb waves). Generally, the animals were found to behave

approximately as deformed elastic shelled spheres with discontinuities. A ray analysis was able to describe the scattering qualitatively at the higher frequencies for the single ping (single realization) analysis. However, in order to provide quantitative predictions of the scattering, an approximate all-frequency model was used which was based upon the modal series solution to the smooth elastic spherical shell. The solution was averaged over sizes and shell thickness in order to predict echoes from ensembles of randomly oriented shells. Many of the errors associated with the use of the modal series solution were eliminated as a result of the averaging process.

II. A VOLUMETRIC-BASED SEAFLOOR SCATTERING FORMULATION

Modeling of the scattering by a shell-covered seafloor is extremely challenging. A rigorous approach would be to use a formulation combining effects due to all boundaries (surficial roughness, interface roughness between layers, discrete scatterers) as well as smoothly varying inhomogeneities (see, for example, Ivakin, 1998). Such an approach is beyond the scope of this current analysis where the intention is to simply provide an estimate of the contribution of the scattering by the shells under very limited conditions (i.e., a layer on the surface of the seafloor near normal incidence that dominates the scattering).

In this simplified approach, the following assumptions are made:

(1) Scattering by the shell-covered bottom is modeled using only volume scattering considerations.
(2) The shapes of the shells do not deviate significantly from a sphere (i.e., not to be needlelike).
(3) Multiple scattering is neglected as a first approximation for these closely spaced scatterers except when Lambert's law is used to describe angular dependence of seafloor scattering and multiple scattering is implicit.
(4) High-frequency acoustics (i.e., geometric optics) approximations are made:
   (a) For single targets, $k_1 a_{\mathrm{esr}} \gtrsim 1$, where $k_1 (= 2\pi/\lambda_1)$ is the acoustic wave number in the surrounding water and λ_1 is the acoustic wavelength. The term a_{esr} is the equivalent spherical radius of the body, which is the radius of a sphere that has the same volume as that of the body.
   (b) For multiple targets, the phases from the individual scatterers are randomly and uniformly distributed over the range 0 to 360 degrees.
(5) The scatterers are randomly oriented so that the ensemble average backscattered cross section, normalized by the geometric cross section, is independent of $k_1 a_{\mathrm{esr}}$ at high $k_1 a_{\mathrm{esr}}$.
(6) The layer is dense enough so that it dominates the scattering.

In order to estimate the effects of the scattering (at best to first order), the floor is considered from a volume scattering viewpoint and is assumed to be a planarlike array of scatterers. Also, the phases of the echoes from the bodies are assumed to be randomly and uniformly distributed over the range 0 to 360 degrees. With this random phase approximation, the signals add incoherently; that is, the average energy from ensembles of the scatterers is equal to the sum of the energy of the individuals.

Although the sizes of shelled bodies present in a given seafloor study are not necessarily known, estimates of the scattering can still be made in the geometric scattering region through use of backscattering cross sections normalized by cross-sectional area of the scatterer. This allows measurements or models of the scattering by an object at one size to be scaled for applications involving objects of other sizes. The objects are required to be of similar shape in order for this scaling to be valid. The backscattering cross section by a given target in the geometric scattering region (i.e., $k_1 a_{\mathrm{esr}} \gtrsim 1$) can be estimated by the formula

$$\langle \sigma_{\mathrm{bs}} \rangle \simeq (\langle \sigma_{\mathrm{bs}}^{(m)} \rangle / \pi a_{\mathrm{esr}}^{(m)^2}) \pi a_{\mathrm{esr}}^2 \quad (k_1 a_{\mathrm{esr}} \gtrsim 1), \tag{1}$$

where the quantity in the parentheses represents the average of the backscattering cross section, $\sigma_{\mathrm{bs}}^{(m)}$, measured or modeled in the $k_1 a_{\mathrm{esr}}^{(m)} \gtrsim 1$ region, normalized by the measured/modeled geometric cross section of the body, $\pi a_{\mathrm{esr}}^{(m)^2}$ (this quantity in parentheses corresponds to the measured/modeled reduced target strength defined later). This quantity is an empirical or modeled scaling factor that relates $\langle \sigma_{\mathrm{bs}} \rangle$ to a_{esr}. The brackets $\langle \cdots \rangle$ denote an average over orientation and/or a narrow range of size or frequency so as to remove the dependence of $\sigma_{\mathrm{bs}}^{(m)} / \pi a_{\mathrm{esr}}^{(m)^2}$ upon $k_1 a_{\mathrm{esr}}^{(m)}$ that is related to various (narrow) resonances and directivity: although $\sigma_{\mathrm{bs}}^{(m)} / \pi a_{\mathrm{esr}}^{(m)^2}$ varies rapidly versus $k_1 a_{\mathrm{esr}}^{(m)}$ (and angle of orientation for nonspherical objects) in the geometric scattering region (i.e., $k_1 a_{\mathrm{esr}}^{(m)} \gtrsim 1$), $\langle \sigma_{\mathrm{bs}}^{(m)} \rangle / \pi a_{\mathrm{esr}}^{(m)^2}$ is relatively constant in this region which makes the below estimates convenient.

For a dense solid sphere, the independence of $\langle \sigma_{\mathrm{bs}}^{(m)} \rangle / \pi a_{\mathrm{esr}}^{(m)^2}$ with respect to $k_1 a_{\mathrm{esr}}^{(m)}$ for $k_1 a_{\mathrm{esr}}^{(m)} \gtrsim 1$ is apparent by examining the component of $\sigma_{\mathrm{bs}}^{(m)}$ due to the front interface. This component makes up a significant fraction of the echo for $k_1 a_{\mathrm{esr}}^{(m)} \gtrsim 1$ and is proportional to $a_{\mathrm{esr}}^{(m)^2}$ (and does not depend upon k_1). The total echo will vary with respect to $k_1 a_{\mathrm{esr}}^{(m)}$ due to interferences between different ''partial'' waves (e.g., circumferential waves). Once averaged over $k_1 a_{\mathrm{esr}}^{(m)}$, the structure due to the interferences is smoothed out and a relatively smooth curve remains that is proportional to $a_{\mathrm{esr}}^{(m)^2}$. This phenomenon has been demonstrated empirically with scattering by a large range of sizes of irregular scatterers and over a large range of frequencies. In a study by Thorne *et al.* (1995), the scattering of irregular solid elastic objects ranging in size (radii) from 50 μ to 2.5 cm were analyzed (the objects included sand grains and rocks) over a frequency range of 40 kHz to 5 MHz. The average echoes, based on an average over orientations, were plotted on the same figure (Fig. 9 of that paper) on a normalized scale. Plotted was form function (on a logarithmic scale) versus $k_1 a_{\mathrm{esr}}^{(m)}$, which is equivalent (to within some constants) to reduced target strength ($\mathrm{RTS}^{(m)}$ defined below) versus $k_1 a_{\mathrm{esr}}^{(m)}$. Also plotted was the exact modal series solution to a

smooth sphere, averaged over size. There was little structure in the data and modal-series-based (averaged) solution for $k_1 a_{\mathrm{esr}}^{(m)} \gtrsim 1$. In fact, for $k_1 a_{\mathrm{esr}}^{(m)} \gtrsim 2$, the data and theory (normalized by $a_{\mathrm{esr}}^{(m)}$) were essentially independent of $k_1 a_{\mathrm{esr}}^{(m)}$. A similar independence (but of k_1 only) was observed in Stanton *et al.* (2000) involving elastic shelled animals, but involving only a single animal. Finally, a study has been published in which the echoes from randomly oriented shrimp were averaged (Stanton *et al.*, 1993). These scatterers are very elongated and possessed a strong directional scattering pattern. However, once averaged over orientation, the scattering was nearly independent of $k_1 a_{\mathrm{ecr}}^{(m)}$ for above about 3 ($a_{\mathrm{ecr}}^{(m)}$ is the equivalent *cylindrical* radius). Although data only involved a narrow range of sizes, the theory showed an independence of average reduced target strength upon $k_1 a_{\mathrm{ecr}}^{(m)}$ for high $k_1 a_{\mathrm{ecr}}^{(m)}$.

For an array of N similarly sized, random-phase scatterers on a section of the seafloor of area A, the ensemble average echo energy is proportional to

$$s_N = N\langle \sigma_{\mathrm{bs}} \rangle. \tag{2}$$

The quantity N is related to the area and packing factor $\mathcal{F}$ as

$$N = (A/\pi a_{\mathrm{esr}}^2)\mathcal{F}, \tag{3}$$

where $\mathcal{F}$, which is equal to the fraction of seafloor covered by the objects, is less than unity and the shape of the body is assumed not to deviate significantly from a sphere (i.e., not to be needlelike so that the equivalent spherical radius can be used here). The area scattering coefficient, which is proportional to the average scattered energy per unit area, is equal to

$$s_A = s_N / A. \tag{4}$$

Inserting the above expressions into Eq. (4) gives

$$s_A = (\langle \sigma_{\mathrm{bs}}^{(m)} \rangle / \pi a_{\mathrm{esr}}^{(m)^2})\mathcal{F}. \tag{5}$$

Expressing this in terms of logarithms for the sonar equation, the area scattering strength on a decibel scale is equal to

$$S_A = 10 \log s_A \tag{6}$$

Applying this to Eq. (5) gives

$$S_A = \langle \mathrm{RTS}^{(m)} \rangle + 10 \log \mathcal{F}, \tag{7}$$

where the average reduced target strength $\langle \mathrm{RTS}^{(m)} \rangle$ is defined by

$$\langle \mathrm{RTS}^{(m)} \rangle \equiv \langle \mathrm{TS}^{(m)} \rangle - 10 \log \pi a_{\mathrm{esr}}^{(m)^2} \tag{8}$$

and the average (free-field) target strength $\langle \mathrm{TS}^{(m)} \rangle$ is defined by

$$\langle \mathrm{TS}^{(m)} \rangle = 10 \log \langle \sigma_{\mathrm{bs}}^{(m)} \rangle. \tag{9}$$

Note that both $\langle \mathrm{TS}^{(m)} \rangle$ and $\langle \mathrm{RTS}^{(m)} \rangle$ here are based on averages of $\sigma_{\mathrm{bs}}^{(m)}$ over orientation and/or a narrow range of size or frequency which makes these averaged forms of $\mathrm{TS}^{(m)}$ and $\mathrm{RTS}^{(m)}$.

Equation (7) is a very interesting result, as it shows that for high $k_1 a_{\mathrm{esr}}$, the area scattering strength from a bed of discrete scatterers can be related to the sum of the average reduced target strength of one scatterer and the packing factor on a decibel scale. Note that it is similar to the first-order scattering term of the predictions by arrays of bosses published in Twersky (1957). Because of the scaling properties of the average reduced target strength, the term $\langle \mathrm{RTS}^{(m)} \rangle$ in Eq. (7) can be determined from a benthic animal of a different size than that in the bed provided that the scattering is in the geometric scattering region and that the animal is morphologically similar.

For a very simple case of scattering by a dense solid sphere (smooth round rock), the average backscattering cross section can be approximated in the high $k_1 a_{\mathrm{esr}}^{(m)}$ region as

$$\langle \sigma_{\mathrm{bs}}^{(m)} \rangle = \tfrac{1}{4} a_{\mathrm{esr}}^{(m)^2} \mathcal{R}_{12}^2, \tag{10}$$

where $\mathcal{R}_{12}$ is the reflection coefficient [$\mathcal{R}_{12} = (gh-1)/(gh+1)$, where g and h are the mass density and sound speed of the object, respectively, normalized by the corresponding quantities for the surrounding water]. This term represents the echo from the front interface which makes up much of the total echo [see, for example, Eq. (16) of Marston, 1988].

Inserting Eq. (10) into Eq. (7) gives

$$S_A = 10 \log(\mathcal{R}_{12}^2/4\pi) + 10 \log \mathcal{F}, \tag{11}$$

where now the expression for scattering strength has been reduced to depending only on the material properties and packing factor.

These simple equations, Eqs. (7) and (11), show that the acoustic scattering by arrays of random phase scatterers, such as on the seafloor, can be reduced to being related to the reflective properties of the scattering, $\mathcal{R}$ (or more generally, normalized cross section) and the packing factor. This formula is made simple, in part, because of the fact that the cross-sectional area dependence of the scattering was cancelled out in the calculation of number per unit area. Of course, the above formula is a very crude approximation and, at best, only applies near normal incidence. For shallow (near horizontal) grazing angles, shadowing effects will become important. Also, for all angles, scattering by the surrounding substrate seafloor material plays a role. Nonetheless, the above formulas can be useful for certain estimates.

In order to extend the results to other angles, the seafloor scattering is assumed to obey Lambert's law. In this approximate approach, the area scattering strength is expressed as

$$S_A(\theta_g) = 10 \log \mu + 10 \log \sin^2 \theta_g, \tag{12}$$

where θ_g is the grazing angle ($\theta_g = 90$ degrees is normal incidence) and $10 \log \mu$ is the scattering strength at normal incidence (Urick, 1983). This formula has proven to be useful in studies of scattering by the seafloor. For example, Stanic *et al.* (1989) showed that the scattering has followed this angular dependence for 5 degrees $\leqslant \theta_g \leqslant$ 30 degrees. For accurate predictions over a wider range of angles and conditions, other approaches are required (see, for example, Jackson *et al.*, 1986a; Gensane, 1989; Ivakin, 1998). Equating the expressions for S_A in Eqs. (7) and (11) to the term $10 \log \mu$, Eq. (12) becomes

$$S_A(\theta_g) = \langle \mathrm{RTS}^{(m)} \rangle + 10 \log \mathcal{F} + 10 \log \sin^2 \theta_g , \quad (13)$$

$$S_A(\theta_g) = 10 \log(\mathcal{R}_{12}^2/4\pi) + 10 \log \mathcal{F} + 10 \log \sin^2 \theta_g , \quad (14)$$

respectively.

These formulas estimate the scattering as a function of grazing angle, in terms of the average reduced target strength of an individual target [Eq. (13)] or material properties [Eq. (14)] and the packing factor of the targets.

III. COMPARISON WITH SEAFLOOR SCATTERING DATA

There have been very few controlled experiments involving acoustic backscattering by the seafloor in regions where there is a significant presence of shells. Two such studies were published by Jackson *et al.* (1986a, b) and Stanic *et al.* (1989). In the Jackson *et al.* (1986a, b) studies, the acoustic scattering by the seafloor was measured as a function of grazing angle, acoustic frequency, and seafloor type. One of the seafloor types involved a bottom material that consisted of very fine sand with a dense covering of live shellfish. The scattering by the bed that contained the shellfish was elevated relative to the section of seafloor that contained sandy silt and no shellfish, indicating that the shellfish played a significant role in the scattering. In the studies by Stanic *et al.* (1989), the studies were focused entirely on a region where the seafloor was covered with shells and the acoustic scattering was measured as a function of grazing angle and acoustic frequency. Characterization of the shells was made possible through the use of samples collected at the site.

The above formulas are now directly applied to the above-mentioned seafloor scattering data. Although the size distribution of the shells was not presented in the Jackson *et al.* (1986a, b) papers, it is assumed for the purpose of this analysis that the scattering by the shells is in the geometric scattering region. For the frequencies of 20 to 50 kHz used in that study, the sizes of the shells would need to be at least about 1 cm long in order to be in the geometric scattering region for the lowest frequency. For the 4-mm- (mean) diam shells observed by Stanic *et al.* (1989), the frequencies need to be about 60 kHz or higher. Also, the shapes of the shells were not documented in either paper. Any differences between the shapes of the shells in the seafloor studies and those used as a basis of modeling the scattering is a potential source of error.

With the assumption that the shells are in the geometric scattering region, the expression given in Eq. (7) for area scattering strength can be used without detailed knowledge of the shell size. As discussed above, since the average reduced target strength in the expression based on an averaged backscattering cross section is relatively independent of size and frequency in the geometric scattering region, it is very convenient for use in this type of application. It is employed simply by using a typical value of the $\langle \mathrm{RTS}^{(m)} \rangle$ from the measurements from Stanton *et al.* (2000) of target strength of the periwinkles in the geometric scattering region. The average value of target strength at high $k_1 a_{\mathrm{esr}}^{(\mathrm{m})}$ of animal No. 97-1 was approximately -55 dB (Fig. 11 of that paper). Using an equivalent spherical radius of the animal to be 2.28 mm, the average reduced target strength of that animal at high $k_1 a_{\mathrm{esr}}^{(m)}$ is approximately -7 dB. Using a packing factor of about 0.8 for closed-packed circles, the estimated area scattering strength for near-normal incidence is -8 dB. This value should be considered an upper bound to the estimate of the scattering by the shells. The presence of shells that have sizes in the Rayleigh scattering region as well as any shadowing effects due to the dense packing of the shells will tend to reduce the estimated value of scattering. Nonetheless, when compared with the values of area scattering strength reported in Figs. 6 and 7 in Jackson *et al.* (1986b) for near-normal incidence, the estimated value of -8 dB lies within the range of observed values which range from about -10 dB to about -2 dB. Figure 21 of Stanic *et al.* (1989) consists of values of $10 \log \mu$ plotted versus frequency that were derived from best fits to data for 5 degrees$\leqslant \theta_g \leqslant$30 degrees. For frequencies above 60 kHz, their values of $10 \log \mu$ range from -22 to -10 dB. Thus, the estimated value of -8 dB using the simplistic discrete-target-based approach overestimates their maximum value of scattering by 2 dB.

IV. DISCUSSION AND CONCLUSION

As a result of recent extensive measurements and modeling of acoustic scattering by shelled animals, an estimate of the contribution of shelled animals to the scattering by a shell-covered seafloor was made possible. The estimate involved a simple formula that was derived herein which related the area scattering strength to the average reduced target strength or simply material properties and packing factor of the objects. The discrete-target-based estimate of scattering due to the presence of a dense covering of shells was close to or consistent with backscattering data from two different shell-covered seafloors. Furthermore, the discrete-target-based formula used in the estimates illustrated that for sufficiently high acoustic frequencies (i.e., in the geometric scattering region), the area scattering strength (at least near normal incidence) is generally independent of size and acoustic frequency and only depends upon material properties and packing factor. This set of dependencies, or lack thereof, is broadly consistent with much of the backscattering data involving the seafloor (shell-covered and otherwise as well as other angles of incidence) which generally show a weak dependence of scattering upon frequency and size of features.

While the measurements and modeling of the scattering by individual shelled animals provided a high-quality basis for the estimates of scattering by a shell-covered seafloor, the estimates were still far from rigorous. Clearly, a rigorous analysis would need to take into account, for example, multiple scattering of the shells, size and shape distribution of the shells, and scattering contributions due to the seafloor substrate. The results of these estimates show promise for incorporating discrete-target-based information into a more general model.

ACKNOWLEDGMENTS

The author gratefully acknowledges useful communications with Mike Richardson of Marine Geosciences Division, Naval Research Laboratory, Stennis Space Center, MS, Darrel Jackson of the Applied Physics Laboratory, University of Washington, Seattle, WA, and Ralph Goodman of The Pennsylvania State University, State College, PA. The author also wishes to thank Shirley Barkley and Sheila Hurst for preparation of the manuscript to this paper. This research was supported by the U.S. Office of Naval Research Grant No. N00014-95-1-0287. This is Contribution No. 9767 for the Woods Hole Oceanographic Institution.

Gensane, M. (**1989**). ''A statistical study of acoustic signals backscattered from the sea bottom,'' IEEE J. Ocean. Eng. **14**, 84–93.

Ivakin, A. N. (**1998**). ''A unified approach to volume and roughness scattering,'' J. Acoust. Soc. Am. **103**, 827–837.

Jackson, D. R., and Briggs, K. B. (**1992**). ''High-frequency bottom backscattering: roughness vs. sediment volume scattering,'' J. Acoust. Soc. Am. **92**, 962–977.

Jackson, D. R., and Ivakin, A. N. (**1998**). ''Scattering from elastic sea beds: First-order theory,'' J. Acoust. Soc. Am. **103**, 336–345.

Jackson, D. R., Winebrenner, D. P., and Ishimaru, A. (**1986a**). ''Application of the composite roughness model to high-frequency bottom backscattering,'' J. Acoust. Soc. Am. **79**, 1410–1422.

Jackson, D. R., Baird, A. M., Crisp, J. J., and Thomson, P. A. G. (**1986b**). ''High-frequency bottom backscatter measurements in shallow water,'' J. Acoust. Soc. Am. **80**, 1188–1199.

Marston, P. L. (**1988**). ''GTD for backscattering from elastic spheres and cylinders in water and the coupling of surface elastic waves with the acoustic field,'' J. Acoust. Soc. Am. **83**, 25–37.

Medwin, H., and Clay, C. S. (**1998**). *Fundamentals of Acoustical Oceanography* (Academic, Boston).

Ogilvy, J. A. (**1991**). *Theory of Wave Scattering from Random Rough Surfaces* (Hilger, New York).

Richardson, M. D., and Briggs, K. B. (**1996**). ''*In situ* and laboratory geoacoustic measurements in soft mud and hard-packed sand sediments: Implications for high-frequency acoustic propagation and scattering,'' GeoMarine Letters **16**, 196–203.

Stanic, S., Briggs, K. B., Fleischer, P., Sawyer, W. B., and Ray, R. I. (**1989**). ''High-frequency acoustic backscattering from a coarse shell ocean bottom,'' J. Acoust. Soc. Am. **85**, 125–136.

Stanton, T. K., Chu, D., Wiebe, P. H., and Clay, C. S. (**1993**). ''Average echoes from randomly oriented random-length finite cylinders: Zooplankton models,'' J. Acoust. Soc. Am. **94**, 3463–3472.

Stanton, T. K., Chu, D., Wiebe, P. H., Eastwood, R. L., and Warren, J. D. (**2000**). ''Acoustic scattering by benthic and planktonic shelled animals,'' J. Acoust. Soc. Am. **108**, 535–550.

Tang, D. J., Frisk, G. V., Sellers, C. J., and Li, D. (**1995**). ''Low-frequency acoustic backscattering by volumetric inhomogeneities in deep-ocean sediments,'' J. Acoust. Soc. Am. **98**, 508–516.

Tang, D. J., Jin, G., Jackson, D. R., and Williams, K. L. (**1994**). ''Analysis of high-frequency bottom and subbottom backscattering for two distinct shallow water environments,'' J. Acoust. Soc. Am. **96**, 2930–2936.

Thorne, P. D., Waters, K. R., and Brudner, T. J. (**1995**). ''Acoustic measurements of scattering by objects of irregular shape,'' J. Acoust. Soc. Am. **97**, 242–251.

Twersky, V. (**1957**). ''On scattering and reflection of sound by rough surfaces,'' J. Acoust. Soc. Am. **29**, 209–225.

Urick, R. J. (**1983**). *Principles of Underwater Sound* (McGraw–Hill, New York).

Zhang, L. (**1996**). ''High Frequency Seafloor Acoustic Backscattering with the Presence of Turbulence in Seawater,'' Ph.D. dissertation, University of Southern Mississippi.

On the applicability of Kramers–Krönig relations for ultrasonic attenuation obeying a frequency power law

Kendall R. Waters
Laboratory for Ultrasonics, Department of Physics, Washington University, St. Louis, Missouri 63130-4899

Michael S. Hughes
Mallinckrodt, Inc., 675 McDonnell Boulevard, Hazelwood, Missouri 63042

Joel Mobley
Laboratory for Ultrasonics, Department of Physics, Washington University, St. Louis, Missouri 63130-4899

Gary H. Brandenburger
Mallinckrodt, Inc., 675 McDonnell Boulevard, Hazelwood, Missouri 63042

James G. Miller
Laboratory for Ultrasonics, Department of Physics, Washington University, St. Louis, Missouri 63130-4899

(Received 3 May 1999; accepted for publication 4 April 2000)

In the recent literature concern has been raised regarding the validity of Kramers–Krönig relations for media with ultrasonic attenuation obeying a frequency power law. It is demonstrated, however, that the Kramers–Krönig dispersion relations for application to these types of media are available. The developed dispersion relations are compared with measurements on several liquids, and agreement is found to better than 1 m/s over the experimentally available bandwidth. A discussion regarding the validity of these dispersion relations, in particular how the dispersion relations relate to the so-called Paley–Wiener conditions, forms the conclusion. © *2000 Acoustical Society of America.* [S0001-4966(00)03107-6]

PACS numbers: 43.35.Bf [HEB]

INTRODUCTION

Recently there have been suggestions in the acoustics literature[1–3] that Kramers–Krönig relations[4,5] are not applicable to media whose attenuation apparently obeys a frequency power law. In particular, experimental evidence[2,3] has shown that a *nearly local approximation* to the Kramers–Krönig relations[6,7] is adequate only for a restricted class of media, namely those media whose attenuation is roughly linear with frequency.

The development of ultrasonic Kramers–Krönig relations is well documented in the literature.[7–9] A technique commonly employed in deriving the Kramers–Krönig dispersion relations is the method of subtractions.[9,10] Through use of this technique we show that valid Kramers–Krönig dispersion relations can be formed for media whose attenuation obeys a frequency power law. To this end we investigate four liquids with attenuation behaviors of the desired frequency power law.

In Sec. I we state the assumed attenuation model and the appropriate forms of the Kramers–Krönig relations necessary to investigate the attenuation model. We then obtain the phase velocity in terms of the attenuation coefficient from these Kramers–Krönig relations. In Sec. II we describe the liquid specimens and the experimental techniques. Experimental measurements are presented in Sec. III, and compared with predictions based on the Kramers–Krönig dispersion relations. We follow in Sec. IV with a discussion of how the Kramers–Krönig dispersion relations for the particular ultrasonic attenuation model under investigation compare with another proposed dispersion theory, and address concerns of applicability of these relations. Section V presents our summary.

I. THEORY

In a variety of media (e.g., liquids and tissue) over a finite bandwidth, the attenuation of ultrasonic waves appears to be adequately modeled by a power-law dependence on frequency,[2,3,11]

$$\alpha(\omega)=\alpha_0|\omega|^y, \tag{1}$$

where we assume α_0 and y are real constants, with $0<y\leqslant 2$ typically. Throughout this paper we assume the attenuation model of Eq. (1), from which we will determine the form of the dispersion. We also assume for purposes of this paper that the power-law form of the attenuation exists for all frequencies, which for most media is an idealized condition. The frequency response of a medium of thickness d can be characterized by its transfer function

$$H(\omega)=e^{iK(\omega)d}=e^{i(\omega/c(\omega)+i\alpha(\omega))d}=A(\omega)e^{i(\omega/c(\omega))d}, \tag{2}$$

where $K(\omega)=\omega/c(\omega)+i\alpha(\omega)$ is the conventional complex wave number and $A(\omega)\equiv e^{-\alpha(\omega)d}$. With the goal of determining the relationships between attenuation coefficient $[\alpha(\omega)]$ and phase velocity $[c(\omega)]$, we consider $\ln H(\omega)/d=iK(\omega)$ instead of $H(\omega)$. The ultrasonic parameters then become more readily accessible by considering the logarithm of the transfer function. The approach of using the logarithm of the transfer function from which to form Kramers–Krönig relations has been discussed in the literature.[12,13]

We begin by stating one of a pair of Kramers–Krönig relations for the modified, ultrasonic complex wave number $[iK(\omega)]$. The dispersion relation, available from several sources in similar forms,[8,9,13–15] is given by

$$\frac{1}{c(\omega)}=\frac{2}{\pi}\int_0^\infty [\alpha(\omega')-\alpha(\omega)]\,\frac{d\omega'}{\omega'^2-\omega^2}, \tag{3}$$

where ω represents angular frequency. It is evident that Eq. (3) is an improper integral for the case $y\geqslant 1$, as has been observed by several authors.[1–3,16] In addition, it has been noted that the Paley–Wiener condition[17,18] regarding casual transforms does not appear to be satisfied for these media $(y\geqslant 1)$. This apparent violation of the Paley–Wiener condition has led some to conclude Kramers–Krönig relations cannot be formed for these types of media (i.e., $y\geqslant 1$) due to the apparent causality violation. We address this issue now as we consider the method of subtractions.

In general, for a medium with an attenuation behavior that goes as $\alpha(\omega)\sim\omega^{n-1}$ with n considered a real positive number, one can form dispersion relations *with n subtractions* to address the aforementioned concern of a divergent integral. (Further discussion of the method of subtractions can be found in Sec. IV.) That is, if the divergence of the attenuation is no worse than a power of the frequency, valid Kramers–Krönig relations can be formed.[10,19] For reasons that will be addressed below, we consider the function $iK(\omega)$ for the case $y\neq 1$, while we consider $iK(\omega)/\omega$ for the case $y=1$. In both cases ($y\neq 1$ and $y=1$) a new subtracted function in terms of either $iK(\omega)$ or $iK(\omega)/\omega$ can be created. The mathematical validity of this technique has been developed for the particular case of the ultrasonic complex wave number,[9,13] as well as other fields of study.[10,14,18–24] We now investigate the Kramers–Krönig dispersion relations for different ranges of the exponent y. For each case considered we determine the phase velocity in terms of the attenuation coefficient.

A. Case: $0<y<1$ (one subtraction)

For media with attenuation that obeys a frequency power law [Eq. (1)] for $0<y<1$, the Kramers–Krönig relations with one subtraction are valid. The phase velocity is related to the attenuation coefficient as given by Eq. (3). Using Eqs. (1) and (3) for the case $0<y<1$, we can write

$$\frac{1}{c(\omega)}=\frac{2}{\pi}\alpha_0\int_0^\infty\frac{\omega'^y-\omega^y}{\omega'^2-\omega^2}d\omega', \tag{4}$$

which can be written (using a change of variables) as

$$\frac{1}{c(\omega)}=\frac{2}{\pi}\alpha_0\omega^{y-1}\int_0^\infty\frac{u^y-1}{u^2-1}du. \tag{5}$$

Solving the Mellin transform of u^{s-1} $(s=1)$ for the frequencies ω and ω_0, one finds[25,26]

$$\frac{1}{c(\omega)}-\frac{1}{c(\omega_0)}=\alpha_0\tan\left(\frac{y\pi}{2}\right)(\omega^{y-1}-\omega_0^{y-1}). \tag{6}$$

B. Case: $1<y<3$ (two subtractions)

The Kramers–Krönig relation with two subtractions, valid for the case $1<y<3$, can be written as

$$\frac{1}{c(\omega)}-\frac{1}{c(0)}=\frac{2}{\pi}\int_0^\infty\left[\frac{\alpha(\omega')-\alpha(0)}{\omega'^2}-\frac{\alpha(\omega)-\alpha(0)}{\omega^2}\right]\frac{d\omega'}{\omega'^2-\omega^2}. \tag{7}$$

Using Eqs. (7) and (1), we write

$$\frac{1}{c(\omega)}-\frac{1}{c(0)}=\frac{2\omega^2}{\pi}\alpha_0\int_0^\infty\left[\frac{\omega'^y}{\omega'^2}-\frac{\omega^y}{\omega^2}\right]\frac{d\omega'}{\omega'^2-\omega^2}, \tag{8}$$

which can be written (using a change of variables) as

$$\frac{1}{c(\omega)}-\frac{1}{c(0)}=\frac{2}{\pi}\alpha_0\omega^{y-1}\int_0^\infty\frac{u^{y-2}-1}{u^2-1}du. \tag{9}$$

Solving this Mellin transform of u^{s-1} $(s=1)$ for the frequencies ω and ω_0, one finds[25,26]

$$\frac{1}{c(\omega)}-\frac{1}{c(\omega_0)}=\alpha_0\tan\left(\frac{y\pi}{2}\right)(\omega^{y-1}-\omega_0^{y-1}). \tag{10}$$

It is now evident the cases $y\neq 1$ considered here have the same dispersion relations. This is in general agreement with previously published Kramers–Krönig dispersion relations,[14] as well as a time-causal theory.[1]

C. Case: $y=1$ (one subtraction)

As mentioned previously, we consider $iK(\omega)/\omega$ instead of $iK(\omega)$ when the attenuation coefficient goes as the first power of frequency [i.e., $\alpha(\omega)\sim\omega$]. If we were to consider $iK(\omega)$ for this case, two subtractions would be necessary as for the case $1<y<3$. [However, the condition that $K'(\omega)$ exists at the origin fails because $\mathrm{Im}\,K(\omega)\sim|\omega|$, hence $K'(\omega)$ is discontinuous at $\omega=0$; see Eq. (17) for further details regarding the method of subtractions. We have placed the defining equations for that method near the end of the paper so that the reader may more quickly get to our main experimental results.] We can avoid any problem this may cause by developing the Kramers–Krönig relations with one subtraction for $iK(\omega)/\omega$. The phase velocity can then be written in terms of the attenuation coefficient as

$$\frac{1}{c(\omega)}-\frac{1}{c(\omega_0)}=\frac{2(\omega-\omega_0)}{\pi}\times P\int_0^\infty\frac{\alpha(\omega')(\omega+\omega_0)-\dfrac{\alpha(\omega_0)}{\omega_0}(\omega'^2+\omega\omega_0)}{(\omega'^2-\omega^2)(\omega'^2-\omega_0^2)}d\omega', \tag{11}$$

where P denotes the Cauchy Principle Value. The use of the specific attenuation model [Eq. (1)] will permit us to change the Cauchy Principle Value into a proper integral. Using Eqs. (11) and (1), we find

$$\frac{1}{c(\omega)}-\frac{1}{c(\omega_0)}=-\frac{2(\omega-\omega_0)}{\pi}\alpha_0 \times P\int_0^\infty \frac{(\omega'-\omega)(\omega'-\omega_0)}{(\omega'^2-\omega^2)(\omega'^2-\omega_0^2)}d\omega' =-\frac{2(\omega-\omega_0)}{\pi}\alpha_0\int_0^\infty \frac{d\omega'}{(\omega'+\omega)(\omega'+\omega_0)}. \tag{12}$$

This Stieltjes transform[25] of $(\omega'+\omega)^{-1}$ provides the dispersion relation

$$\frac{1}{c(\omega)}-\frac{1}{c(\omega_0)}=-\frac{2}{\pi}\alpha_0\ln\left(\frac{\omega}{\omega_0}\right). \tag{13}$$

Note that this result is in agreement with Kramers–Krönig dispersion relations previously published.[14,22,27]

While we develop Kramers–Krönig dispersion relations for the previously mentioned cases, we experimentally investigate only the case of $1<y\leqslant 2$. For the case $y<1$ there is no question regarding the causal validity of the Kramers–Krönig relations. In addition, many investigations available in the literature have shown agreement between experiment and the Kramers–Krönig relations for the case $y\approx 1$.[2,7,28]

II. MATERIALS AND METHODS

A. Liquid specimens and specimen chamber

Four liquids were investigated during this study. Three of the liquids, typically utilized as mechanical and diffusion pump fluids, were: DC 705 silicone fluid (Dow Corning Corp., Midland, MI), Permavisc 10 synthetic hydrocarbon oil (Kurt J. Lesker Company, Clairton, PA), and TKO-1170 high-viscosity hydrocarbon oil (Kurt J. Lesker Company, Clairton, PA). The fourth liquid was castor oil. These liquids were chosen for their ultrasonic attenuation behaviors so as to provide a range of frequency power laws to study.

The liquids were contained in a specimen chamber composed of two circular Lexan® face plates (approximately 6-in. diameter, 0.5-in. thick) bolted to each side of a circular Lexan® middle section (approximately 6-in. diameter, 1-in. thick). Both the face plates and the middle section had a 2-in.-diameter hole passing through their centers. Each face plate served to clamp O-rings, a washer, and a Saran Wrap® sheet against its side of the middle section, so that the completed chamber had a hollow, circular section with parallel faces of Saran Wrap® serving as acoustic windows. Each liquid was contained within this part of the chamber during measurements. The distance between the Saran Wrap® walls could be varied by using washers of various thickness; for this experiment, the wall separation distance was maintained at approximately 0.45 cm. Further details of the specimen chamber are available in the literature.[29]

B. Experimental setup and data acquisition

The measurements were performed in an immersion through-transmission setup at the laboratories of Mallinckrodt, Inc. using a pair of broadband immersion polyvinylidene fluoride (PVDF) transducers (0.4-in. diameter, 3.25-in. focal length, model ZF3002-SU, Panametrics, Waltham, MA). Active control of the temperature of the water bath (37.0 ± 0.1 °C) was maintained, in addition to control of the insonifying acoustic pressures. The peak positive and negative pressure were 30 and -30 kPa, respectively. Further details of the experimental arrangement and data acquisition protocol can be found in a previous publication.[29]

The through-transmitted signal was time-averaged in sets of 1000 traces (8-bit single shot, 250-MHz sampling rate) in order to enhance dynamic range and reduce effects from random noise. Five time-averaged sample traces (4.092 μsec long, 1024-point record length) were acquired in this manner for each liquid investigated. Total acquisition time was approximately 5 sec. Immediately after each data run, the specimen path length was determined ultrasonically by measuring the time interval between the through-transmitted signal and the signal from the first round-trip reverberation. The thickness of the specimen chamber was approximately 0.45 cm for the measurements in this study. Five reference traces were acquired in the same fashion as for the specimen traces, using the specimen chamber filled only with water.

C. Data reduction and analysis

Both the frequency-dependent attenuation coefficient and phase velocity were determined from Fourier analysis of the through-transmitted waveforms. The attenuation coefficient is determined from a log-spectral subtraction technique,

$$\alpha_s(\omega)=\frac{1}{d}\left(\ln\frac{|\tilde{F}_w(\omega)|}{|\tilde{F}_s(\omega)|}-\ln\frac{|\tilde{T}_w(\omega)|}{|\tilde{T}_s(\omega)|}\right), \tag{14}$$

where $\tilde{F}_s(\omega)$ is the Fourier transform of the through-specimen trace, $\tilde{F}_w(\omega)$ is the Fourier transform of the reference trace, $\tilde{T}_w(\omega)$ is the transmission coefficient through the Saran Wrap® windows with a water-filled chamber, $\tilde{T}_s(\omega)$ is the transmission coefficient through the Saran Wrap® windows with a specimen-filled chamber, and d is the thickness of the specimen. Attenuation losses in the water compared to the attenuation losses of the specimen are negligible. The differences in the ultrasonic transmission coefficients of the specimen-filled chamber versus the water-filled chamber are also small compared to the attenuation of the specimens. [For the specimens considered here, the difference in transmission coefficients accounted for a few tenths of a (Np/cm).] In addition, diffraction effects are negligible for this particular substitution arrangement. In order for diffraction effects to have an impact in the context of the substitution technique, the specimen has to significantly alter the diffraction pattern through the combination of the velocity contrast with the host medium and a sufficiently thick sample. For our system, the velocity mismatches are not significantly different, and the sample chambers are too thin for significant alteration of the diffraction patterns to occur.[30]

By use of an ultrasonic broadband spectroscopy technique,[31] the phase velocity of the specimen can be determined by

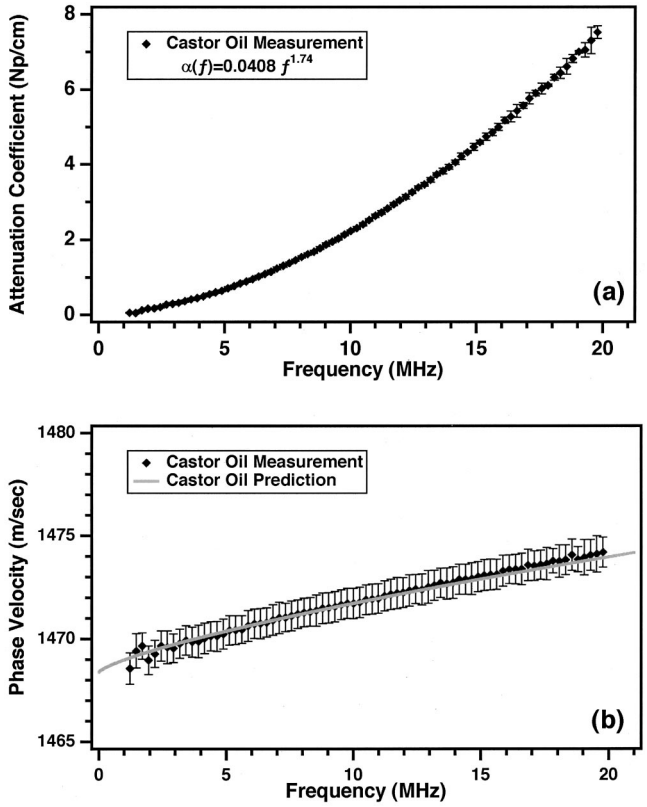


FIG. 1. (a) Attenuation coefficient and (b) phase velocity measurements with standard deviation bars for castor oil. The Kramers–Krönig predicted phase velocity [using Eq. (10)] is also shown. These measurements were conducted at 37.0 °C with an insonifying acoustic peak positive pressure of 30 kPa and a peak negative pressure of −30 kPa.

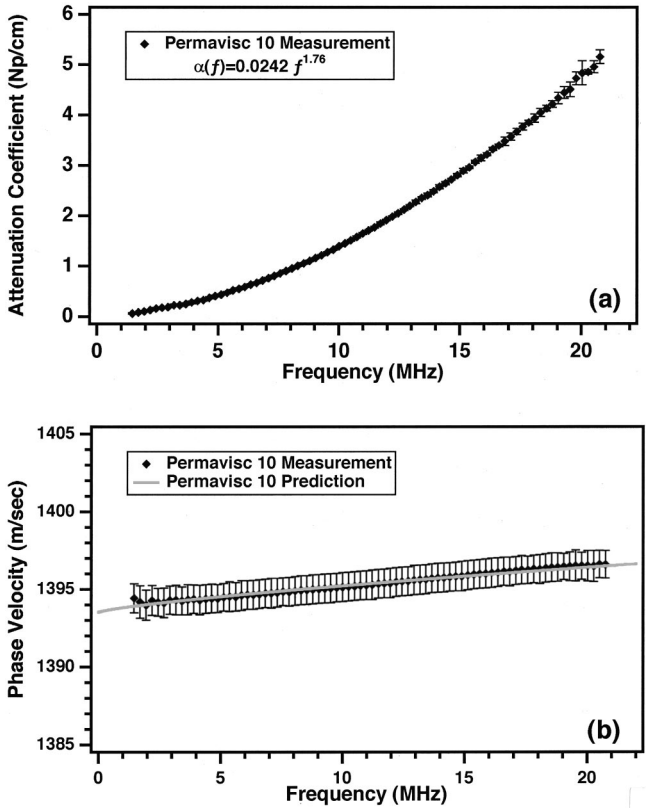


FIG. 2. (a) Attenuation coefficient and (b) phase velocity measurements with standard deviation bars for Permavisc 10 (synthetic hydrocarbon oil). The Kramers–Krönig predicted phase velocity [using Eq. (10)] is also shown. These measurements were conducted at 37.0 °C with an insonifying acoustic peak positive pressure of 30 kPa and a peak negative pressure of −30 kPa.

$$c_s(\omega) = c_w \frac{1}{1 - c_w \dfrac{\Delta\phi(\omega)}{\omega d}}, \tag{15}$$

where c_w is the temperature-dependent speed of sound in the reference medium,[32] $\Delta\phi(\omega) \equiv \phi_s(\omega) - \phi_w(\omega)$ is the difference in the phase spectra of the through-specimen and reference traces, and d is the specimen thickness. Multiple reverberations in the specimen chamber (≈0.45-cm thick) itself are not an issue because the ultrasonic signal is gated to include only the first received pulse and no echoes.

Prior to determining the magnitude and phase spectra, the acquired radio frequency (rf) waveforms were baseline-subtracted and symmetrized. These steps remove the uninteresting zero frequency [or direct current (dc)] term and the steep linear ramp of the phase, respectively, as has been discussed in Ref. 15. For each specimen investigated, the magnitude and (unwrapped) phase of the five time-averaged reference and specimen waveforms were determined via the fast Fourier transform (FFT) algorithm. The attenuation coefficient and phase velocity for the five specimen waveforms were determined using Eqs. (14) and (15), respectively. The results were averaged, and the standard deviations determined. This procedure was repeated for each liquid investigated.

We fit the attenuation coefficient of each specimen to the frequency power-law model [Eq. (1)]. The best fit was determined through a χ^2-minimization technique, while the standard error for the model parameter y (i.e., the power law) was determined by evaluation of the variation of χ^2 with respect to y.[33] The attenuation model fit was then used in conjunction with the Kramers–Krönig dispersion relations to predict the dispersion. Because the developed dispersion relations predict only changes in the velocity of sound with frequency, and not the absolute level of the phase velocity, we chose to match the predicted phase velocity to the measured phase velocity at 7.5 MHz. Analyses were implemented in both Igor Pro (Wavemetrics, Inc., Lake Oswego, OR) at Washington University and in Mathematica (Wolfram Research, Inc., Champaign, IL) using a C MathLink module at Mallinckrodt, Inc. The independent development of both sets of analysis routines provides additional confirmation of the results presented in this paper.

III. RESULTS

Figures 1–4 show the experimentally measured attenuation coefficient and phase velocity with their respective standard deviations for each of the four liquids. In addition, the Kramers–Krönig predicted phase velocity is compared to the measured phase velocity. Note that the vertical scales may differ among the four liquids.

The experimentally observed attenuation coefficients obey a frequency power law over the usable bandwidth for each of the four liquids. The fits to the attenuation coeffi-

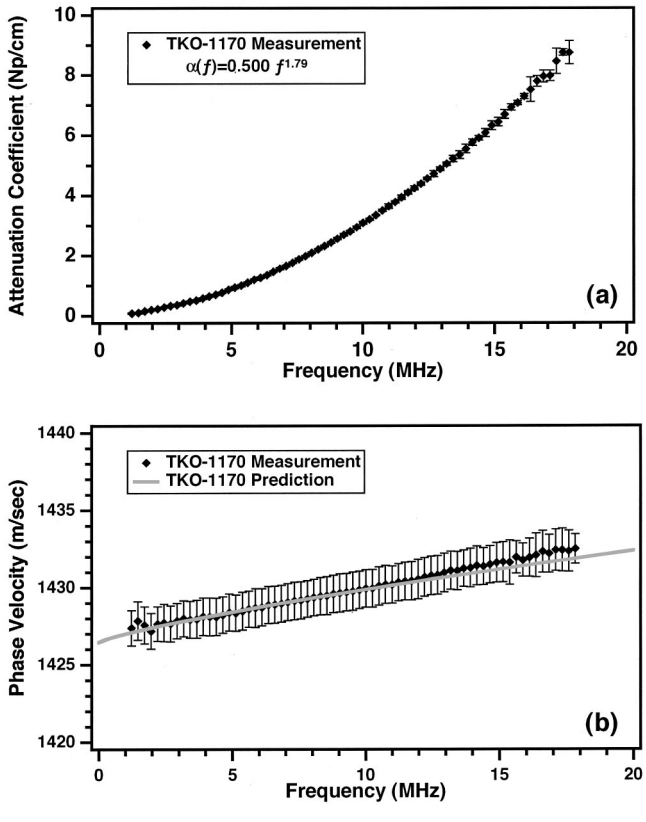


FIG. 3. (a) Attenuation coefficient and (b) phase velocity measurements with standard deviation bars for TKO-1170 (high-viscosity hydrocarbon oil). The Kramers–Krönig predicted phase velocity [using Eq. (10)] is also shown. These measurements were conducted at 37.0 °C with an insonifying acoustic peak positive pressure of 30 kPa and a peak negative pressure of −30 kPa.

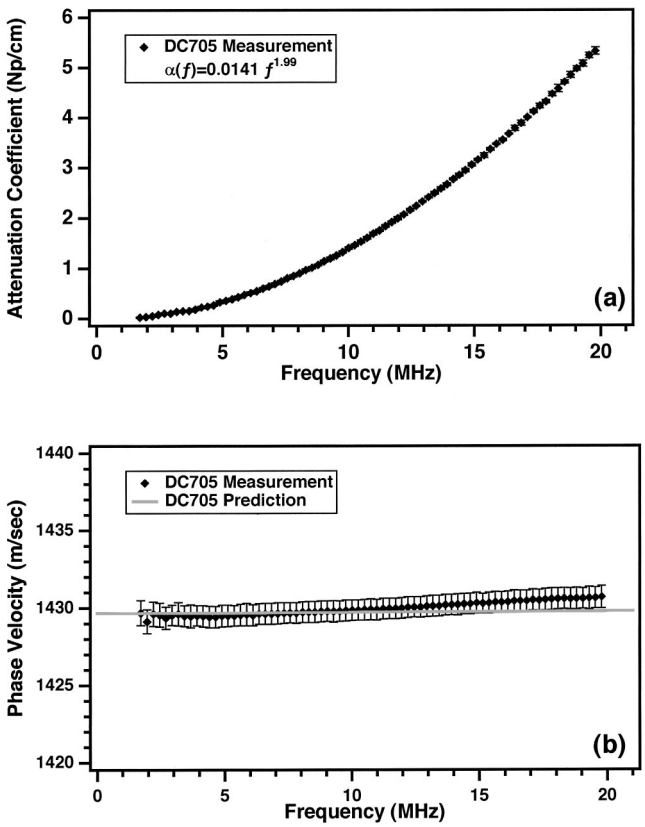


FIG. 4. (a) Attenuation coefficient and (b) phase velocity measurements with standard deviation bars for DC 705 (silicone fluid). The Kramers–Krönig predicted phase velocity [using Eq. (10)] is also shown. These measurements were conducted at 37.0 °C with an insonifying acoustic peak positive pressure of 30 kPa and a peak negative pressure of −30 kPa.

cients using the frequency power-law model [Eq. (1)] are listed in Table I in addition to relevant experimental conditions. As it is the power law that is of most interest in this study, we also report the standard error of the power-law fit parameter (y) for each liquid. In all cases, the standard error is small (± 0.01) and has a negligible impact on the Kramers–Krönig dispersion predictions.

In general, we find excellent agreement between the measured and predicted phase velocities for the liquids. In all cases we see agreement to better than 1 m/s over the experimentally available bandwidth. Of particular note is the case of DC 705 silicone fluid, which exhibits an attenuation with a frequency power law *near* 2 ($y=1.99$). As expected from the Kramers–Krönig dispersion relation [Eq. (10)], we observe very little dispersion. It is for this case that we find our largest discrepancy (0.9 m/s) between the measured and predicted dispersions over the experimental bandwidth. This may be due in part to the measurement being only over a modest bandwidth (~2 to 20 MHz). A previously reported measurement available in the literature for a similar liquid (DC 710 silicone fluid) over a larger bandwidth has shown that the absorption of the silicone fluid is more appropriately modeled as a single relaxation process rather than a frequency power law.[34] Hence, we might expect some discrepancy between the measured and predicted dispersions.

IV. DISCUSSION

A. Comparison to the time-causal theory

A recently published theory has taken a time-domain-based approach to model the behavior of media with attenuation obeying a frequency power law.[1,2] One impetus for the development of the time-causal theory was the aforementioned *apparent* lack of validity for Kramers–Krönig relations for media with $y \geqslant 1$, in accordance with the Paley–Wiener condition.[17,18] In the limit of generalized functions, the time-causal model is able to predict dispersion relations given the attenuation obeys a frequency power law.

The author comments that the *nearly local approximation* to the Kramers–Krönig relation overestimates the dispersion for cases $y \neq 1$. (We note that the previously published nearly local approximation to the Kramers–Krönig relation might also be called the nearly local approximation to the Kramers–Krönig relation with one subtraction.[6,7]) One highly notable discrepancy between the nearly local approximation and time-causal dispersion predictions is for the value of $y=2$, where the nearly local approximation to the Kramers–Krönig relation with one subtraction predicts a dispersion increasing linearly with frequency,[6] whereas the time-causal model predicts zero dispersion. While relative discrepancies for dispersion predictions exist between the nearly local approximation and time-causal theory for all values of $y \neq 1$, absolute discrepancies have been difficult to

TABLE I. Power-law fits to the attenuation coefficient of each liquid. The temperature, peak positive acoustic pressure, and peak negative acoustic pressure are reported as well. Note the fit parameter α_0 is in units of linear frequency. The uncertainty in the power law (y) represents the standard error.

Liquid	Temperature (°C)	Peak positive/negative pressure (kPa)	α_0 [Np/(cm·MHz^y)]	y
Castor oil	37.0	30/−30	0.0408	1.74±0.01
Permavisc 10	37.0	30/−30	0.0242	1.76±0.01
TKO-1170	37.0	30/−30	0.0500	1.79±0.01
DC 705	37.0	30/−30	0.0141	1.99±0.01

measure for $y \lesssim 1.5$. The discrepancy between the nearly local approximation and time-causal dispersion predictions decreases as y approaches one. Furthermore, the time-causal dispersion prediction has been shown to be a Taylor-series expansion of the nearly local approximation dispersion prediction for values of y near one.[2] As suggested by the author, the nearly local approximation of the Kramers–Krönig relation with one subtraction appears to be valid only for a restricted range of media, that is media with $\alpha(\omega) \sim \omega^y$, where $y \approx 1$. Based on this evidence concerning the *nearly local approximation* to the Kramers–Krönig relation with one subtraction, the author appears to suggest that the full (not approximate) Kramers–Krönig relations are invalid for these types of media due to a causality violation (as discussed further below).

In contrast, we have shown that the full Kramers–Krönig dispersion relations with subtractions presented here predict dispersion relations identical to those predicted by the time-causal theory, which by definition is a causal model. It is also worth restating that the Kramers–Krönig relations require no assumption on the attenuation form (for example, that the attenuation exhibit a power-law dependence on frequency). Instead, the Kramers–Krönig relations serve only to measure the consistency of a given model.[6,7]

B. Paley–Wiener condition

According to Titchmarsh's theorem,[10,35] the general conditions necessary for the Kramers–Krönig relations to exist for a linear, passive system are square integrability and causality. A condition on the form of the transfer function [Eq. (2)] suggested by Paley and Wiener might seem to place further restrictions on the form of the transfer function. It states that a necessary and sufficient condition for a nonnegative, square-integrable function $A(\omega)$ to be the Fourier transform of a causal function is that the magnitude of $A(\omega)$ satisfy[17,18]

$$\int_{-\infty}^{\infty} \frac{|\ln A(\omega)|}{1+\omega^2} d\omega < \infty. \tag{16}$$

Here, we consider $|\ln A(\omega)| = |-\alpha(\omega) d|$ as defined in Eq. (2). The Paley–Wiener condition states that a function must approach zero at high frequencies but cannot be too close to zero too often. (See p. 1767 of Ref. 19.)

If one assumes an attenuation model that obeys a frequency power law [Eq. (1)], as we have for this paper, and then attempts to employ the Kramers–Krönig dispersion relation with one subtraction [Eq. (3)], convergence problems occur for the case $y \geqslant 1$ as has been noted above. However, as demonstrated in Sec. I, for cases $y \geqslant 1$ we can develop dispersion relations with subtractions which, *prima facie*, are convergent. Furthermore, the development of the dispersion relations in Sec. I made no mention of the Paley–Wiener condition. In spite of this omission, there is no real paradox since we did not develop Kramers–Krönig dispersion relations using $iK(\omega)$, but instead used a subtracted function in terms of $iK(\omega)$ to derive a relationship between the attenuation coefficient and phase velocity [see Eqs. (3), (7), and (11)]. As we will show shortly [Eqs. (19)–(22)], these subtracted functions patently obey the Paley–Wiener condition.

The approach taken in Sec. I [Eqs. (3), (7), and (11)] may be described in a unified, and more general, fashion using the subtracted function defined by[10]

$$D_n(\omega,\omega_0) = \frac{iK(\omega) - iK(\omega_0) - (\omega-\omega_0)\dfrac{d}{d\omega} iK(\omega)\Big|_{\omega=\omega_0} - \ldots - \dfrac{(\omega-\omega_0)^{n-1}}{(n-1)!}\dfrac{d^{n-1}}{d\omega^{n-1}} iK(\omega)\Big|_{\omega=\omega_0}}{(\omega-\omega_0)^n}, \tag{17}$$

where n represents the subtraction order. We assume that $K(\omega)$ is differentiable up to order n at $\omega = \omega_0$. It is from this subtracted function that we derive the Kramers–Krönig relations. Inspection of Eq. (17) reveals the method of subtractions to be equivalent to forming a Taylor polynomial in terms of the complex wave number. It is clear from Eq. (17) that the subtracted function, $D_n(\omega,\omega_0)$, is related to the nth-order derivative of the complex wave number.

We can express the subtracted function in terms of modified forms of the attenuation coefficient and wave number,

$$D_n(\omega,\omega_0)=\mathrm{Re}[D_n(\omega,\omega_0)]+i\,\mathrm{Im}[D_n(\omega,\omega_0)]$$
$$\equiv-\bar{\alpha}_n(\omega,\omega_0)+i\bar{k}_n(\omega,\omega_0). \tag{18}$$

(Here, the overscore is simply used to denote a subtracted form of either the attenuation coefficient or wave number.)

Consider the case for $1<y<3$. We developed the dispersion relation with two subtractions ($n=2$) using a subtracted form of the complex wave number,

$$D_2(\omega,\omega_0)=\frac{iK(\omega)-iK(\omega_0)-(\omega-\omega_0)\dfrac{d}{d\omega}iK(\omega)\Big|_{\omega=\omega_0}}{(\omega-\omega_0)^2}. \tag{19}$$

We choose $\omega_0=0$ without loss of generality, and use Eq. (1) for the attenuation model [$\alpha(0)=\alpha'(0)=0$ for $1<y<3$]. We find

$$\bar{\alpha}_2(\omega,0)=\frac{\alpha(\omega)}{\omega^2}, \tag{20}$$

$$\bar{k}_2(\omega,0)=\frac{\dfrac{1}{c(\omega)}-\dfrac{1}{c(0)}}{\omega}. \tag{21}$$

Applying the Paley–Wiener condition [Eq. (16)] to the twice-subtracted form of the attenuation coefficient [Eq. (20)] yields

$$\int_0^\infty\frac{|\bar{\alpha}_2(\omega)|}{(1+\omega^2)}d\omega=\int_0^\infty\frac{|\alpha(\omega)|}{\omega^2(1+\omega^2)}d\omega, \tag{22}$$

which is finite for $1<y<3$. Hence, the Paley–Wiener condition is satisfied for the twice-subtracted function [Eq. (19)] in terms of the complex wave number $K(\omega)$. In turn, the twice-subtracted function is a causal transform.

Finally, it should be noted that the Paley–Wiener condition can be shown to be a consequence of Titchmarsh's theorem, rather than an additional restriction. This has been reported elsewhere for the dispersion relations with one subtraction,[9] and more generally for the case of the dispersion relations with n subtractions.[12] Recalling the typical derivation of the Kramers–Krönig relations, one eventually arrives at an equation of the form [Ref. 10, Eq. (1.6.6) or (1.9.7)],

$$\mathrm{Re}[iK(\omega)]\propto\oint_C\frac{\mathrm{Im}[iK(\omega')]}{\omega'^2-\omega^2}d\omega', \tag{23}$$

where $iK(\omega')/(\omega'^2-\omega^2)$ is required to be analytic in the interior of the contour C (the upper half-plane in Ref. 10). With the choice $\omega=i$, Eq. (23) immediately reduces to Eq. (16) because $\mathrm{Re}[iK(i)]$ is finite [since $iK(\omega')/(\omega'^2-\omega^2)$ is analytic for $\omega>0$]. We note that this argument comprises the ''easy'' part of the Paley–Wiener theorem which actually states the equivalence of Eq. (16) with the fact that $iK(\omega)$ is a causal transform.

In the language of distribution theory, the subtracted form of the attenuation coefficient, $\bar{\alpha}_n(\omega,\omega_0)$, for some $n\geqslant 0$ is said to be a distribution of slow growth at infinity (i.e., the distribution is tempered).[36] It is then permissible to relax the condition of square integrability to the much weaker condition of square temperedness. In doing so, a generalized Paley–Wiener theorem for tempered distributions is satisfied.[37]

V. SUMMARY

We have presented ultrasonic Kramers–Krönig dispersion relations developed using the method of subtractions to predict dispersion in media with attenuation obeying a frequency power law. We have shown the Paley–Wiener condition is satisfied for these dispersion relations. Furthermore, the dispersion relations developed here have been shown to be equivalent to those developed using a time-causal approach. The time-causal theory[1,2] appears to face the same dilemma regarding the Paley–Wiener condition, yet convergence of the Paley–Wiener integral was not directly addressed in that theory. We suspect that the time-causal approach and the method of subtractions taken in this paper are actually equivalent, and may be related as Fourier transforms of each other. This possible relationship will be the subject of further research.

[1] T. L. Szabo, ''Time domain wave equations for lossy media obeying a frequency power law,'' J. Acoust. Soc. Am. **96**, 491–500 (1994).
[2] T. L. Szabo, ''Causal theories and data for acoustic attenuation obeying a frequency power law,'' J. Acoust. Soc. Am. **97**, 14–24 (1995).
[3] P. He, ''Simulation of ultrasound pulse propagation in lossy media obeying a frequency power law,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 114–125 (1998).
[4] R. d. L. Kronig, ''On the theory of dispersion of x-rays,'' J. Opt. Soc. Am. **12**, 547–557 (1926).
[5] H. A. Kramers, ''La diffusion de la lumière par les atomes,'' Atti Congr. Intern. Fis., Como **2**, 545–557 (1927), available in H. A. Kramers, *Collected Scientific Papers* (North-Holland, Amsterdam, 1956), pp. 333–345.
[6] M. O'Donnell, E. T. Jaynes, and J. G. Miller, ''General relationships between ultrasonic attenuation and dispersion,'' J. Acoust. Soc. Am. **63**, 1935–1937 (1978).
[7] M. O'Donnell, E. T. Jaynes, and J. G. Miller, ''Kramers–Krönig relationship between ultrasonic attenuation and phase velocity,'' J. Acoust. Soc. Am. **69**, 696–701 (1981).
[8] V. L. Ginzberg, ''Concerning the general relationship between absorption and dispersion of sound waves,'' Sov. Phys. Acoust. **1**, 32–41 (1955).
[9] R. L. Weaver and Y.-H. Pao, ''Dispersion relations for linear wave propagation in homogeneous and inhomogeneous media,'' J. Math. Phys. **22**, 1909–1918 (1981).
[10] H. M. Nussenzveig, *Causality and Dispersion Relations*, 1st ed. (Academic, New York, 1972), pp. 3–53.
[11] N. Akashi, J. Kushibiki, and F. Dunn, ''Acoustic properties of egg yolk and albumen in the frequency range 20–400 MHz,'' J. Acoust. Soc. Am. **102**, 3774–3778 (1997).
[12] R. E. Burge, M. A. Fiddy, A. H. Greenaway, and G. Ross, ''The phase problem,'' Proc. R. Soc. London, Ser. A **350**, 191–212 (1976).
[13] J. Mobley, ''Ultrasonic dispersion in suspensions and solids: A study of fundamental dynamics and the Kramers–Kronig relations,'' Ph.D. thesis, Washington University, 1998, pp. 189–207.
[14] C. W. Horton, ''Dispersion relations in sediments and sea water,'' J. Acoust. Soc. Am. **55**, 547–549 (1974).
[15] M. S. Hughes, ''Applications of local approximations to the Kramers–Kronig relations in ultrasonics,'' Ph.D. thesis, Washington University, 1987, pp. 117–120.
[16] K. V. Gurumurthy and R. M. Arthur, ''A dispersive model for the propagation of ultrasound in soft tissue,'' Ultrason. Imaging **4**, 355–377 (1982).
[17] R. E. A. C. Paley and N. Wiener, *Fourier Transforms in the Complex*

Domain, 1st ed. (American Mathematical Society, New York, 1934), pp. 16–19.
[18] A. Papoulis, *The Fourier Integral and its Applications*, reissue edition (McGraw-Hill, New York, 1987), pp. 213–217.
[19] J. S. Toll, ''Causality and the dispersion relations: Logical foundations,'' Phys. Rev. **104**, 1760–1770 (1956).
[20] K. Aki and P. G. Richards, *Quantitative Seismology: Theory and Methods* (W. H. Freeman, San Francisco, 1980), pp. 173–175.
[21] H. W. Bode, ''Relations between attenuation and phase in feedback amplifier design,'' Bell Syst. Tech. J. **19**, 421–454 (1940).
[22] W. I. Futterman, ''Dispersive body waves,'' J. Geophys. Res. **67**, 5279–5291 (1962).
[23] K. W. Winkler and T. J. Plona, ''Technique for measuring ultrasonic velocity and attenuation spectra in rocks under pressure,'' J. Geophys. Res. B **87**, 10776–10780 (1982).
[24] H. C. Booij and G. P. J. M. Thoone, ''Generalization of Kramers–Kronig transforms and some approximations of relations between viscoelastic quantities,'' Rheol. Acta **21**, 15–24 (1982).
[25] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954), Vol. I, p. 311, Eq. 33, and Vol. II, p. 216, Eq. 2.
[26] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 4th ed. (Academic, New York, 1980), pp. 293, Eq. 3.246.
[27] C. W. Horton, ''Comment on 'Kramers–Kronig relationship between ultrasonic attenuation and phase velocity,' '' J. Acoust. Soc. Am. **70**, 1182 (1981).
[28] B. Hosten, ''Heterogeneous structure of modes and Kramers–Kronig relationship in anisotropic viscoelastic materials,'' J. Acoust. Soc. Am. **104**, 1382–1388 (1998).
[29] J. N. Marsh, C. S. Hall, M. S. Hughes, J. Mobley, J. G. Miller, and G. H. Brandenburger, ''Broadband through-transmission signal loss measurements of Albunex® suspensions at concentrations approaching *in vivo* doses,'' J. Acoust. Soc. Am. **101**, 1155–1161 (1997).
[30] J. Mobley, J. N. Marsh, C. S. Hall, M. S. Hughes, G. H. Brandenburger, and J. G. Miller, ''Broadband measurements of phase velocity in Albunex® suspensions,'' J. Acoust. Soc. Am. **103**, 2145–2153 (1998).
[31] W. Sachse and Y.-H. Pao, ''On the determination of phase and group velocities of dispersive waves in solids,'' J. Appl. Phys. **49**, 4320–4327 (1978).
[32] V. A. Del Grosso and C. W. Mader, ''Speed of sound in pure water,'' J. Acoust. Soc. Am. **52**, 1442–1446 (1972).
[33] P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, 2nd ed. (McGraw-Hill, New York, 1992), pp. 141–147.
[34] F. Dunn and J. E. Breyer, ''Generation and detection of ultra-high-frequency sound in liquids,'' J. Acoust. Soc. Am. **34**, 775–778 (1962).
[35] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, 2nd ed. (Clarendon, Oxford, 1948), pp. 119–131.
[36] E. J. Beltrami and M. R. Wohlers, *Distributions and the Boundary Values of Analytic Functions* (Academic, New York, 1966), pp. 34–44.
[37] E. Pfaffelhuber, ''Generalized impulse response and causality,'' IEEE Trans. Circuit Theory **CT-18**, 218–23 (1971).

Diffraction effects on bulk-wave ultrasonic velocity and attenuation measurements

Jun-ichi Kushibiki and Mototaka Arakawa
Department of Electrical Engineering, Tohoku University, Sendai 980-8579, Japan

(Received 1 September 1999; accepted for publication 3 May 2000)

The loss and phase advance due to diffraction are experimentally observed by measuring the amplitude and phase of radio frequency (rf) tone burst signals in the VHF range, in an ultrasonic transmission line consisting of a buffer rod with an ultrasonic transducer on one end, a couplant of water, and a solid specimen of synthetic silica glass. The measured results agree well with the calculated results from the exact integral expression of diffraction. The diffraction effects on the velocity and attenuation measured in this frequency range and their corrections are investigated to realize more accurate measurements. It is shown that attenuation measurements are influenced by diffraction losses and can be corrected by numerical calculations, and that velocity measurements are affected by the phase advance caused by diffraction. This investigation demonstrates that, in complex-mode velocity measurements, in which the velocity is determined from the measured phase of the signals, the true velocity at each frequency can be obtained by correction using the numerical calculation of diffraction. Based on this result, a new correction method in amplitude-mode velocity measurements is also proposed. In this new method, the velocity is determined from the intervals of interference output obtained by sweeping the ultrasonic frequency for the superposed signals generated by the double-pulse method. Velocity may be measured accurately at frequencies in the Fresnel region, and diffraction correction is essential to obtain highly accurate values with five significant figures or more. [S0001-4966(00)02408-5]

PACS numbers: 43.35.Cg, 43.35.Yb [HEB]

INTRODUCTION

In ultrasonic metrology, sound velocity and attenuation, which reflect elastic properties of materials, have been measured using a variety of techniques.[1] In principle, it is necessary to accurately measure the propagation distance and propagation time for precise velocity measurements, and the propagation distance and signal amplitude for precise attenuation measurements. In measuring velocity and attenuation, the following factors generally introduce measurement error, producing amplitude and phase variations: diffraction; bonding layer between the transducer and specimen; parallelism of the specimen surfaces; velocity dispersion; media inhomogeneity; and temperature gradients in experiments.[1] These problems have been investigated both theoretically and experimentally.

The losses and phase advances caused by diffraction of ultrasonic waves have been analyzed and calculated.[2–6] Also, attempts have been made to explain those effects through experiments.[6–9] However, very little data have been collected so far on the phase advance due to diffraction in the VHF and UHF ranges.

The authors have developed an ultrasonic spectroscopy system that operates in the VHF and UHF ranges,[10] and have measured the acoustic properties, viz., sound velocity, attenuation, acoustic impedance, and density, of solid, liquid, and biological tissue specimens in ultrasonic transmission line models using radio frequency (rf) pulses (for example, see Fig. 1). This technique has been realized with the help of feedback from a previously developed technique for an acoustic measurement system[11] and the movement, alignment, and signal processing techniques developed for the line-focus-beam (LFB) acoustic microscopy system.[12]

Rf tone burst pulses are used to measure acoustic properties. Sound velocity can be precisely measured by the bulk-wave ultrasonic pulse interference method.[1,10,13] The measurement method is broadly subdivided into the frequency sweeping (*f*-interference) method and the variable path length (*z*-interference) method. In the former, the velocity is determined from the frequency interval Δf in the interference waveform obtained by sweeping the ultrasonic frequency. In the latter, the velocity is determined from the distance interval Δz in the interference waveform obtained by changing the propagation distance. The *f*-interference method is suitable for measuring solids and biological tissues whose propagation length cannot be changed, or for determining the gap length in the measurements of acoustic properties of liquid specimens. However, when this method is applied to measure dispersive specimens, only the average value over the frequency range employed for analysis can be obtained. In contrast, in the *z*-interference method, the velocity for liquid specimens is measured at every frequency. The attenuation is measured with the pulse echo method.[1,10]

By applying this method to velocity measurements of solid specimens, elastic constants have been determined and standard specimens for LFB acoustic microscopy have been prepared.[13] Consequently, it is important to fully understand the effects of diffraction in this method for more accurate measurements and for examining the significant figures in the obtained values.

This paper investigates the effects of diffraction in the VHF range with experiments and numerical calculations on an isotropic solid specimen of synthetic silica (SiO_2) glass,

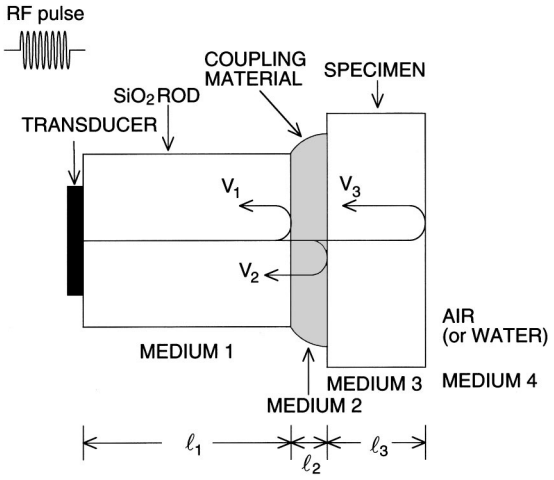


FIG. 1. Experimental arrangement for bulk velocity and attenuation coefficient measurements of solid specimens using bulk ultrasonic rf pulses.

in which attenuation is negligibly small in the VHF range and velocity is nondispersive. The diffraction effect on the velocity measured with the *f*-interference method and its correction method are also discussed.

I. MEASUREMENT PRINCIPLE

Figure 1 shows the experimental arrangement employed for measuring acoustic properties of a solid specimen using bulk-wave rf tone burst pulse signals. The ultrasonic device is made of an SiO_2 glass buffer rod with a transducer fabricated on one end. Here, we describe measuring properties of a specimen of SiO_2 glass (N-ES, Nippon Silica Glass Co., 5001.7 μm in thickness) with a water couplant using longitudinal waves. The specimen will be described in detail in Sec. III. $\mathbf{V}_1$, $\mathbf{V}_2$, and $\mathbf{V}_3$ are the reflected signals from the buffer rod end and from the front and back surfaces of the specimen, respectively. Pure water is used for the couplant, and $\mathbf{V}_1$ can be separated from $\mathbf{V}_2$ in the time domain by properly setting the rf pulse width and the propagation length of the couplant, l_2. Based on the ultrasonic transmission line model, the signals $\mathbf{V}_2$ and $\mathbf{V}_3$ are expressed as

$$\mathbf{V}_2 = A \cdot ATT_2 \cdot T_{12} \cdot R_{23} \cdot T_{21} \cdot \exp(-2\gamma_1 l_1 - 2\gamma_2 l_2), \quad (1)$$

$$\mathbf{V}_3 = A \cdot ATT_3 \cdot T_{12} \cdot T_{23} \cdot R_{34} \cdot T_{32} \cdot T_{21} \cdot \exp(-2\gamma_1 l_1 - 2\gamma_2 l_2 - 2\gamma_3 l_3), \quad (2)$$

where A is the amplitude coefficient, ATT_i is the effect of diffraction during the propagation of each signal, T_{ij} and R_{ij} are the transmission and reflection coefficients when the ultrasonic wave is incident from medium i to medium j, and γ_i is the propagation constant of medium i. R_{34} is -1 for air at the back surface of the specimen, and is equal to $-R_{23}$ for water. γ_i, T_{ij}, and R_{ij} are defined as

$$\gamma_i = \alpha_i + j\beta_i, \quad (3)$$

$$T_{ij} = \frac{2Z_j}{Z_j + Z_i}, \quad (4)$$

$$R_{ij} = \frac{Z_j - Z_i}{Z_j + Z_i}, \quad (5)$$

$$\beta_i = \frac{\omega}{V_i}, \quad (6)$$

$$Z_i = \rho_i V_i, \quad (7)$$

where α_i is the attenuation constant; β_i, the phase constant; Z_i, the acoustic impedance; ω, the angular frequency; V_i, the velocity; and ρ_i, the density of medium i. The acoustic impedance is determined on the assumption that it is lossless. The effect of the diffraction is represented as

$$ATT_i = |ATT_i| \cdot \exp(j\theta_i), \quad (8)$$

and the amplitude and phase are examined separately.

When the velocity is obtained by just measuring the amplitude of signals (the amplitude mode), $\mathbf{V}_2$ and $\mathbf{V}_3$ are superposed in the time domain by the double-pulse method,[10,13–15] and the interference output $|\mathbf{V}_2 + \mathbf{V}_3|$ is measured. $|\mathbf{V}_2 + \mathbf{V}_3|$ is given as follows:

$$|\mathbf{V}_2 + \mathbf{V}_3| = |A| \cdot |ATT_2| \cdot T_{12} \cdot R_{23} \cdot T_{21} \cdot \exp(-2\alpha_1 l_1 - 2\alpha_2 l_2) \left| 1 + \frac{|ATT_3|}{|ATT_2|} \cdot \left| \frac{T_{23} \cdot R_{34} \cdot T_{32}}{R_{23}} \right| \cdot \exp(-2\alpha_3 l_3) \cdot \exp\{j(-2\beta_3 l_3 + \pi + \Delta\theta)\} \right|, \quad (9)$$

where $\Delta\theta = \theta_3 - \theta_2$. While $|\mathbf{V}_2 + \mathbf{V}_3|$ obtained from Eq. (9) exhibits an interference waveform which reflects the phase rotation in the specimen $(-2\beta_3 l_3 + \pi + \Delta\theta)$, it is impossible to extract the phase rotation in the specimen only, since the result, as seen from the above equation, really includes the device performance or the frequency characteristics of the conversion efficiency; the influence of the diffraction loss; and the influence of the attenuation in the buffer rod, water couplant, and specimen. To solve this problem, the amplitudes of each signal, $|\mathbf{V}_2|$ and $|\mathbf{V}_3|$, are measured as well, and the operation is performed using the following equation to extract the phase rotation in the specimen alone:

$$\frac{|\mathbf{V}_2 + \mathbf{V}_3|^2 - |\mathbf{V}_2|^2 - |\mathbf{V}_3|^2}{2 \cdot |\mathbf{V}_2| \cdot |\mathbf{V}_3|} = \cos(-2\beta_3 l_3 + \pi + \Delta\theta). \quad (10)$$

The difference in the phase advance due to diffraction, $\Delta\theta$, is included in the phase term and affects the velocity measurements.

When the velocity is obtained by measuring the amplitude and phase of the signals (the complex mode), the phase rotation of the specimen is extracted by normalizing $\mathbf{V}_3$ by $\mathbf{V}_2$ as follows:

$$\frac{\mathbf{V}_3}{\mathbf{V}_2}=\frac{|ATT_3|}{|ATT_2|}\cdot\left|\frac{T_{23}\cdot R_{34}\cdot T_{32}}{R_{23}}\right|\cdot\exp(-2\alpha_3 l_3)$$
$$\cdot\exp\{j(-2\beta_3 l_3+\pi+\Delta\theta)\}. \tag{11}$$

The difference $\Delta\theta$ is included in the phase term of the equation above, in the same way as in the equation for the amplitude mode. In either case, it is necessary to consider the difference in the phase advance for more accurate velocity measurements. If the diffraction effect is ignored, the rotation period during which the phase rotates 2π, i.e., the interference interval Δf, is given by

$$\Delta f=\frac{V_3}{2l_3}. \tag{12}$$

In this case, the velocity of the specimen V_3 is obtained from Eq. (12) using Δf and the measured specimen thickness l_3.

In the complex mode, when the diffraction effect is ignored and the phase of $\mathbf{V}_3/\mathbf{V}_2$ is ϕ, the velocity of specimen V_3 is given by

$$V_3=-\frac{2\omega l_3}{\phi-\pi}. \tag{13}$$

In both the amplitude and complex modes, the attenuation coefficient is obtained from the following equation:

$$\alpha_3=\frac{1}{2\cdot l_3}\ln\left\{\left|\frac{\mathbf{V}_2}{\mathbf{V}_3}\right|\cdot\left|\frac{T_{23}\cdot R_{34}\cdot T_{32}}{R_{23}}\right|\right\}-\frac{1}{2\cdot l_3}\ln\left|\frac{ATT_2}{ATT_3}\right|. \tag{14}$$

In the measurements, measured values are influenced by the ratio of the diffraction loss, $|ATT_2/ATT_3|$.

II. MEASUREMENT METHOD AND SYSTEM

Measurements were performed with the ultrasonic spectroscopy system. This system is capable of very accurately measuring bulk acoustic properties, viz., velocity, attenuation, acoustic impedance and density of solid, liquid, and biological tissue specimens in the VHF and UHF ranges. The system is described in detail in the literature.[10] We also have developed a system that can measure the amplitude and phase of rf pulses in the VHF range. This system has been used for analyzing the performance of a scanning reflection acoustic microscope system with a concave transducer[16] and the moving characteristics of the precision mechanical-translation stage using ultrasonic plane waves.[17] With the accumulated knowledge from these experiences, we have developed a new system to enhance stability in measurements. Figure 2(a) shows a block diagram of the system; the main part of the pulse mode measurement system is shown in Fig. 2(b).[11] The network/spectrum analyzer (Model HP-4396A with the option 1D5 and a custom-made option T01, Hewlett-Packard Japan, Ltd.) is used to transmit and receive electrical signals with its special option of measuring the amplitude and phase of rf tone burst pulse signals in the zero-span mode. Rf switches of a GaAs Monolithic Microwave Integrated Circuit (MMIC) switch (Model CDS0671, DAICO Industries, Inc.), with an on-off ratio of 80 dB or higher in the range 10–2000 MHz were employed for an rf modulator to generate rf tone burst pulse signals and an rf gate circuit in the receiving unit.

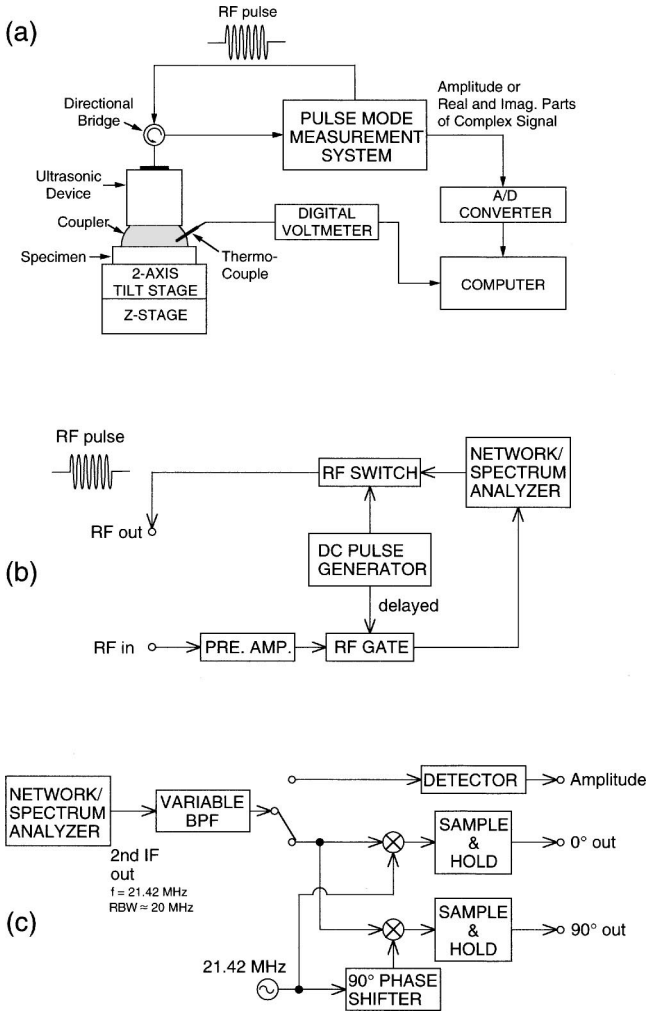


FIG. 2. Block diagram of the ultrasonic spectroscopy system. (a) Whole system. (b) Pulse-mode measurement system. (c) Special option unit of the network/spectrum analyzer.

Measurements are carried out employing the spectrum analyzer mode of the network/spectrum analyzer. Continuous signals at the frequency synchronized to the network/spectrum analyzer receiver are output from its internal signal source, converted into rf tone burst pulse signals by the rf switch, then applied to the ultrasonic transducer through the rf SWR bridge (Model 60NF50-1, WILTRON Co.). The signals are converted into ultrasonic pulses by the ultrasonic transducer. The ultrasonic waves propagate through the SiO_2 buffer rod, reflect or transmit at each interface in the ultrasonic transmission line, and form a train of rf pulse echoes. The repetition frequency is set to around 1 kHz, considering reverberation in the SiO_2 buffer rod. The rf gate extracts only the pulse of interest from the rf pulse echoes, which is then detected with the spectrum analyzer.

The second IF signal (21.42 MHz) from the network/spectrum analyzer is input to the optional unit at a maximum IF frequency bandwidth of 20 MHz [Fig. 2(c)]. The detected amplitude output, or the real and imaginary parts of complex signals, which are quadrature detected and held at their

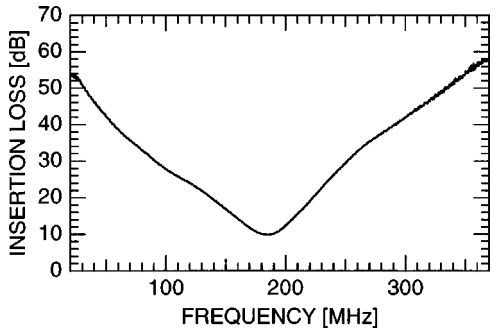


FIG. 3. Typical frequency characteristics of insertion loss of a plane-wave ultrasonic device consisting of a ZnO film transducer fabricated on a 10-mm-long SiO_2 buffer rod. ZnO film thickness=14.2 μm, transducer diameter=2.5 mm.

peaks, is output. The video output from the network/spectrum analyzer is converted into a video output through a peak-holding circuit or a sampling and holding circuit, A/D-converted by the 16-bit A/D converter, and recorded by the computer.

A ZnO piezoelectric film transducer with a film thickness of 14.2 μm was fabricated on one end of a 10-mm-long, 8.4-mm-diameter SiO_2 rod. The typical frequency characteristics of insertion loss of the ultrasonic device are shown in Fig. 3. The operating center frequency is around 185 MHz.

Stabilization of the measurement temperature environment plays an important role in measuring the phase. The entire system was installed in a temperature-controlled room where the temperature was maintained around 23 ± 0.1 °C. To keep the measurement temperature constant, the specimen, couplant, and ultrasonic device were contained in a chamber whose temperature was controlled within ±0.01 °C by water circulation. The temperature was monitored with a copper-constantan thermocouple (JIS T-Model, Class 1, CHINO Co.) which is calibrated within ±0.01 °C by a platinum resistance thermometer (Model R800-2, CHINO Co.).

Unlike the measurements conducted in the transmission mode, a number of spurious signals appear in the measurements in the reflection mode. Figure 4 shows the oscilloscope traces of reflected signals observed for an SiO_2 glass specimen (thickness 5001.7 μm) with a propagation length of 824.3 μm in a water couplant and an rf pulse width of 500 ns. Figure 4(a) shows the pulse echoes observed at an rf pulse carrier frequency of 180 MHz, and Fig. 4(b), those at 80 MHz. Signal $\mathbf{V}_2$ is delayed in time from signal $\mathbf{V}_1$ by 1.11 μs, and signal $\mathbf{V}_3$ is delayed from signal $\mathbf{V}_2$ by 1.69 μs. Internal multiple-reflection signals in the SiO_2 buffer rod appear every 3.35 μs after signal $\mathbf{V}_1$. The signals are reflected multiple times in the water couplant and in the specimen and appear every 1.11 μs (1.69 μs) after signal $\mathbf{V}_2$ ($\mathbf{V}_3$). Although the ZnO piezoelectric film transducer excites longitudinal waves efficiently, it also excites some shear waves at lower frequencies. Consequently, the shear wave signal $\mathbf{V}_s$, which travels back and forth within the buffer rod, appears as a spurious signal approximately 1.99 μs after signal $\mathbf{V}_1$. This signal increases around 80 MHz. In the figure, the shear pulse waveform is deformed in the latter half part due to interference with a longitudinal wave component which is received by the transducer after going and returning twice in

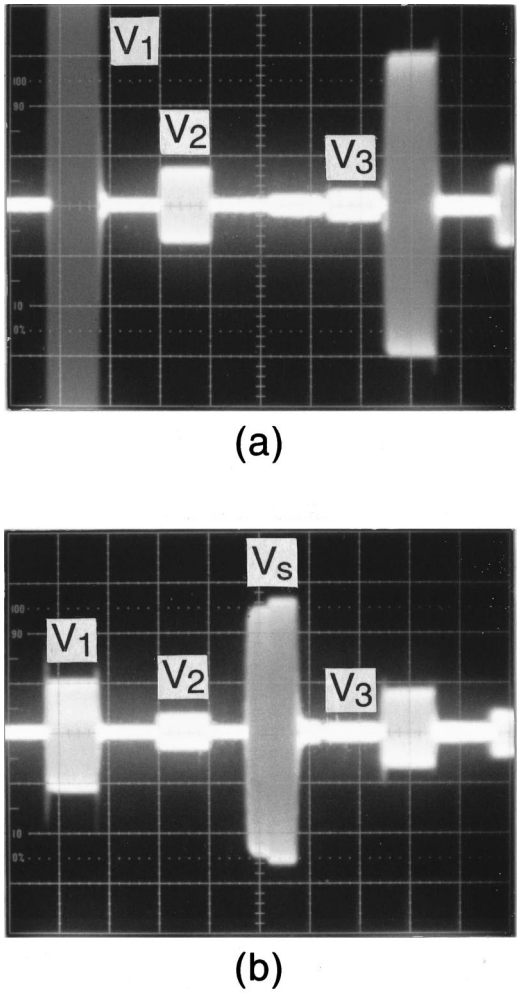


FIG. 4. Rf pulse echoes observed for an SiO_2 glass specimen through water couplant. The specimen thickness is 5001.7 μm, and the couplant length is 824.3 μm. (a) $f=180$ MHz (input attenuator 25 dB). (b) $f=80$ MHz (input attenuator 20 dB). Scales: vertical, 0.1 V/div; horizontal, 0.5 μs/div.

the water couplant. Therefore, we have to set the rf pulse width and the propagation length in the water couplant with a great care when making measurements in order to prevent signals $\mathbf{V}_2$ and $\mathbf{V}_3$ from overlapping with those spurious signals.

In the amplitude mode, $\mathbf{V}_2$ and $\mathbf{V}_3$ have to be superposed in the time domain to obtain an interference output by the double-pulse method, as shown in Fig. 5. Figure 5(a) shows the pulse echoes obtained when rf pulse signals were applied to the ultrasonic transducer; Fig. 5(b) shows the pulse echoes obtained when rf pulse signals were generated with a time delay equal to the time difference between $\mathbf{V}_2$ and $\mathbf{V}_3$. Figure 5(c) shows the transducer output when the transducer was excited with the double rf pulses to superpose $\mathbf{V}_2$ and $\mathbf{V}_3$. Only the superposed signals of $\mathbf{V}_2$ and $\mathbf{V}_3$ are gated out and input into the spectrum analyzer, as shown in Fig. 5(d). $\mathbf{V}_2$ and $\mathbf{V}_3$, shown in Fig. 5(a), are also gated out individually, and their amplitudes, $|\mathbf{V}_2|$ and $|\mathbf{V}_3|$, are measured. In the complex mode, $\mathbf{V}_2$ and $\mathbf{V}_3$ are also gated out individually and measured.

A typical waveform obtained by the bulk ultrasonic pulse interference method and its analysis procedure are given in Fig. 6. In addition to the interference signal

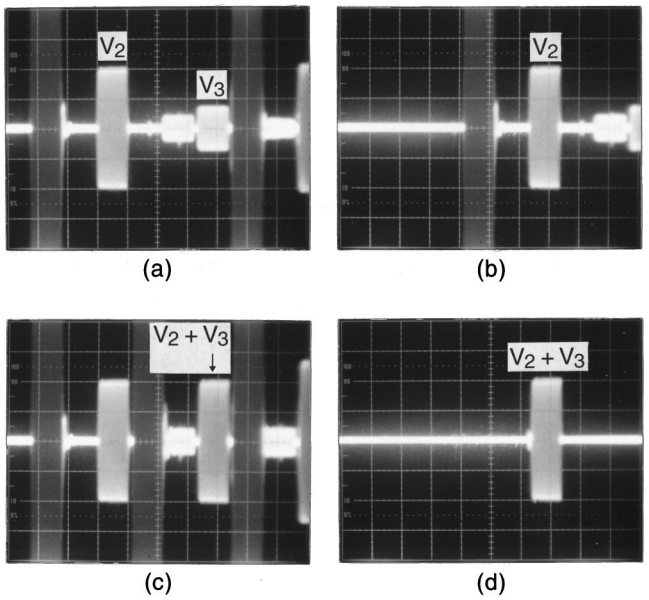


FIG. 5. Interference output by the double-pulse *f*-interference method (reflection mode). $f=180$ MHz (input attenuator 16 dB). (a) Rf pulse echoes by the first rf pulse signal. (b) Rf pulse echoes by the second rf pulse signal. (c) Superposition of (a) and (b). (d) Interference output gated out. The delayed time between two rf pulses of signals $\mathbf{V}_2$ and $\mathbf{V}_3$ is 1.69 μs. Scales: vertical, 0.1 V/div; horizontal, 0.5 μs/div.

$|\mathbf{V}_2+\mathbf{V}_3|$ superposed with the double-pulse method, the amplitude of each signal, $|\mathbf{V}_2|$ and $|\mathbf{V}_3|$, is measured [Fig. 6(a)], and only the interference component is extracted by applying Eq. (10) [Fig. 6(b)]. The frequency interval Δf is obtained through FFT analysis of this waveform [Fig. 6(c)].

In complex-mode measurements, the real (V_{Re}) and imaginary (V_{Im}) parts of the complex signals are measured as shown in Fig. 7(a). V_{Re} and V_{Im} are expressed as follows:

$$V_{\mathrm{Re}}=B\cdot\cos\,\phi, \tag{15}$$

$$V_{\mathrm{Im}}=B\cdot\sin\,\phi, \tag{16}$$

where B is the amplitude and ϕ is the phase. The amplitude and phase are given by the following equations:

$$B=\sqrt{V_{\mathrm{Re}}^2+V_{\mathrm{Im}}^2}, \tag{17}$$

$$\phi=\tan^{-1}\left(\frac{V_{\mathrm{Im}}}{V_{\mathrm{Re}}}\right). \tag{18}$$

The amplitudes and phases of signals $\mathbf{V}_2$ and $\mathbf{V}_3$ are measured, and then the velocity and attenuation are obtained by extracting the amplitude variation and phase rotation in the specimen using Eq. (11). Before a complex waveform can be analyzed by FFT, its real (cosine waveform) and imaginary (sine waveform) parts have to be obtained by calculations.

To increase the certainty of the frequency employed, the spectrum analyzer is set in the zero-span mode, and measurements are made by repeatedly changing its set frequency. The rf pulse width was selected to be approximately 500 ns and the IF bandwidth of the spectrum analyzer was set to 3 MHz for measurements both in the amplitude mode and in the complex mode. When signal $\mathbf{V}_1$ was measured at a fixed frequency in the complex mode, the stability of the phase

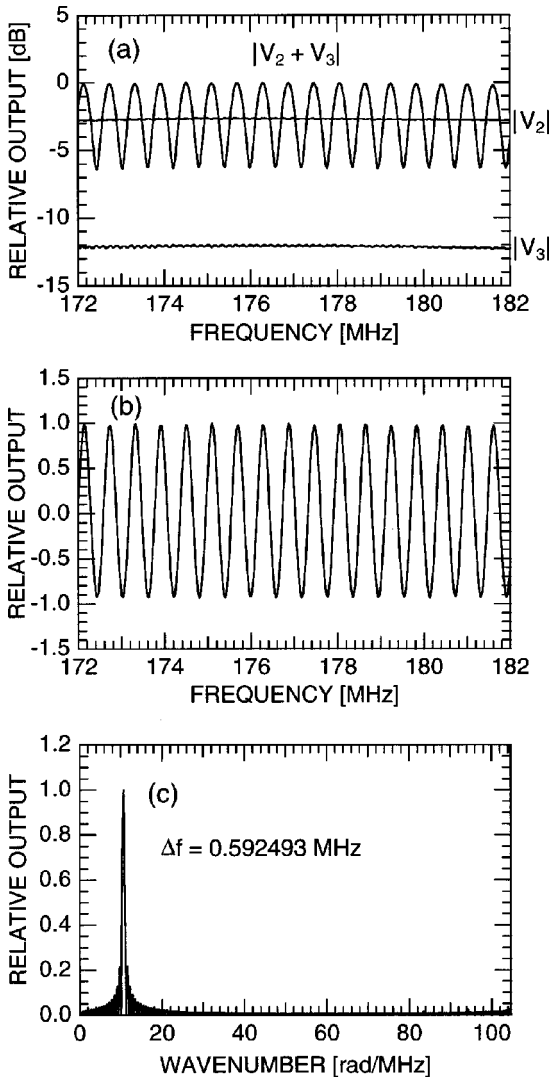


FIG. 6. An example of analysis of velocity measurement by the bulk ultrasonic pulse *f*-interference method. (a) Measured frequency responses. (b) Extracted interference output. (c) Spectral distribution analyzed by an FFT of the interference output shown in (b).

was found to be within $\pm0.12°$ for $\pm2\sigma$ (σ: the standard deviation). When the amplitude and phase of a complex signal are calculated from its real and imaginary parts, measurement errors occur due to the zero offset of each channel, the gain ratio of the two channels, and the quadrature of the detector. Although those effects probably do not exceed $\pm0.15°$ in the present experiment, it is possible to reduce the errors with further improvement.

III. EXPERIMENTS AND DISCUSSIONS

An isotropic solid of SiO_2 glass (N-ES, Nippon Silica Glass Co.) is used for the specimen. The specimen is 5001.7-μm-thick and 25 mm in diameter with both sides optically polished. The thickness was measured by a digital length gauging system (CERTO, DR. JOHANNES HEIDENHAIN GmbH, Traunreut, Germany) with an accuracy of ±0.1 μm. The parallelism of the specimen is within 14 seconds. The influence of the parallelism is negligible in the velocity measurements in the frequency range used here.

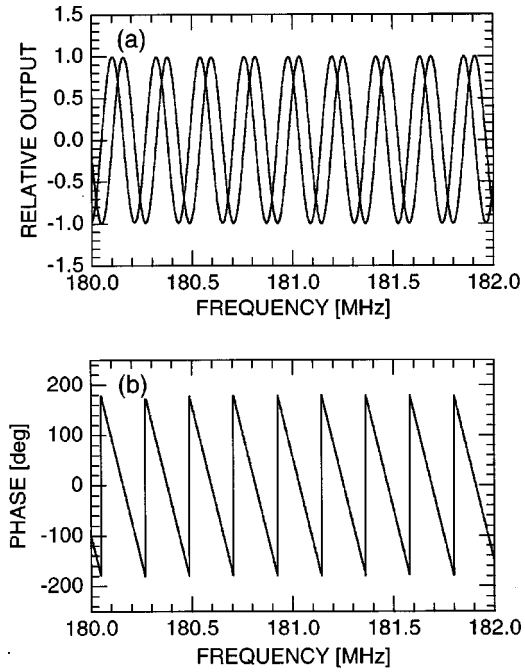


FIG. 7. An example of frequency response measured in complex mode. (a) Complex waveforms for $\mathbf{V}_2$ signals. (b) Phase for $\mathbf{V}_2$ signals.

A. Phase variation due to diffraction

Signals $\mathbf{V}_2$ and $\mathbf{V}_3$ were measured in 0.05-MHz steps from 20 to 250 MHz in the complex mode, and the phase of $\mathbf{V}_3/\mathbf{V}_2$ was obtained. The couplant was pure water, and the distance between the buffer rod and the specimen was set to 824.3 μm. The back surface of the specimen here is also pure water, since the specimen is immersed in a container filled with pure water to stabilize the temperature of the entire system composed of the specimen, couplant, and buffer rod. The phase of $\mathbf{V}_3/\mathbf{V}_2$ changes linearly and exhibits a waveform in which the phase is delayed approximately 608° with every 1-MHz frequency increase. It is extremely difficult to detect the slight phase variations caused by diffraction from this waveform. Therefore, an approximated straight line (ϕ_a) is obtained with the least squares method for the data of the phases obtained over the frequencies 80 to 200 MHz, where the S/N of the signals is good. This line is then extrapolated to the frequency range used for the measurements. The differences ($\phi_m - \phi_a$) were then obtained by subtracting the obtained values (ϕ_a) from the measured values (ϕ_m), and moving-averaged with ± 5 MHz (± 100 points). The result is given in Fig. 8(a), in which some significant variations were observed with sufficient S/N.

1. Numerical calculations

To confirm that the phase variations observed in Fig. 8(a) are due to diffraction, the diffraction was calculated numerically. The exact integral expression of diffraction by Williams[2] is used for calculations. This expression is used to calculate diffraction in an isotropic, homogeneous medium during the transmission and reception of ultrasonic waves made by two circular transducers with an axially concentric alignment and the same size. It depends upon $S\ (=l/l_F)$ (the propagation distance normalized by the Fresnel length l_F which is equal to a^2/λ, where λ is the wavelength and a is the transducer radius), and a product of ka where k is the wave number.

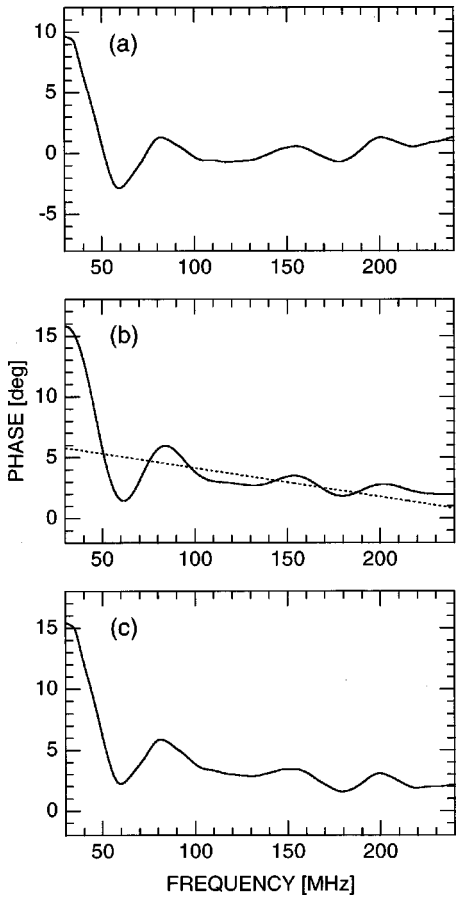


FIG. 8. Phase variations of $\mathbf{V}_3/\mathbf{V}_2$ by diffraction for the SiO_2 specimen as a function of frequency. (a) Measured. (b) Calculated solid curve and approximated dotted straight line obtained by the least squares method. (c) Experimental results obtained by adding the approximated dotted straight line shown in (b) to the experimental data in (a).

In this measurement method, ultrasonic waves propagate through different media, i.e., buffer rod, water couplant, and specimen. Therefore, the total propagation length is obtained using the Papadakis method[18] by adding the propagation length of each medium normalized by the Fresnel length in each medium, so that the diffraction effect is calculated using the expression. ka is calculated using the wave number in the buffer rod. This is primarily because the wave number in the rod is equal to that of the specimen, and S in the water couplant accounts for approximately 1.4% of the total S since the Fresnel length in water is large and therefore has very little influence on the calculated values. The parameters used for the calculations are given in Table I. The velocity of longitudinal waves in water was obtained from the literature,[19] referring to the water temperature in the measurements. The difference between the phase advances of signals $\mathbf{V}_2$ and $\mathbf{V}_3$ caused by diffraction, $\Delta\theta(=\theta_3-\theta_2)$, is calculated. The results are given in Fig. 8(b).

2. Comparison of measured and calculated values

Both the measured [Fig. 8(a)] and the calculated [Fig. 8(b)] differences in phase advance due to diffraction exhibit

TABLE I. Longitudinal velocities in propagation media and their lengths in calculating diffraction effects at 23 °C.

	Velocity (m/s)	Length (μm)
SiO_2 rod[a]	5954.1	9983.5
Water	1491.24	824.3
Specimen[b]	5927.38	5001.7

[a]Synthetic Silica glass (T-4040, Toshiba Ceramics Co.).
[b]Synthetic Silica glass (N-ES, Nippon Silica Glass Co.).

similar variations with minima and maxima at nearly equal frequencies. However, the waveform of the measured values are slightly inclined as a whole with a positive slope with respect to frequency. This may have resulted from the subtraction using the approximated straight line of the phase data over the frequency range 80–200 MHz, in order to obtain the tiny variations in phase shown in Fig. 8(a). Similarly, the approximated straight line for the phase data of the calculated values over the same frequencies (80–200 MHz) is represented by the dotted line in Fig. 8(b), and its values are added to the measured values in Fig. 8(a). The result is shown in Fig. 8(c) and is considered to be the real, experimental result of the phase difference $\Delta\theta$ under the actual experimental conditions. The result in Fig. 8(c) is in excellent agreement with that in Fig. 8(b). It is thus seen that the phase variation due to diffraction in the VHF range is successfully extracted in this measurement. The variations in $\Delta\theta$ are remarkable at lower frequencies and slight at frequencies above 116 MHz at which the normalized propagation length S is equal to unity for the signal reflected from the back surface, $\mathbf{V}_3$.

B. Diffraction effect on attenuation measurements

Figure 9 shows the attenuation coefficient obtained from the amplitudes of $\mathbf{V}_2$ and $\mathbf{V}_3$ measured in the complex mode using Eq. (14). The broken curve represents the measured values of the attenuation coefficient obtained from the first term of Eq. (14), not considering any diffraction effects. The density of the specimen, which was used in calculating the reflection and transmission coefficients, was $\rho=2200.3$ kg/m^3, measured using the Archimedes method.[20] The measurement error of the density was within an accuracy of ± 0.1 kg/m^3. The density of water needed in the density measurement was obtained from the literature,[21] referring to the water temperature. The dash-dotted curve shows the calculated values of amplitude variations due to diffraction, obtained from the calculation for the second term of Eq. (14). The calculated values exhibit similar variations with maxima and minima resembling those for the measured values.

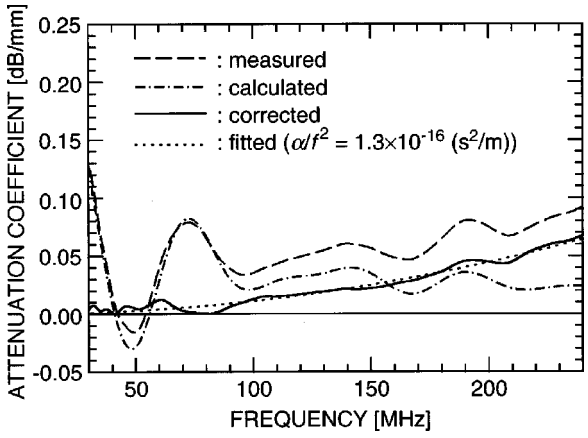


FIG. 9. Diffraction loss for the SiO_2 specimen.

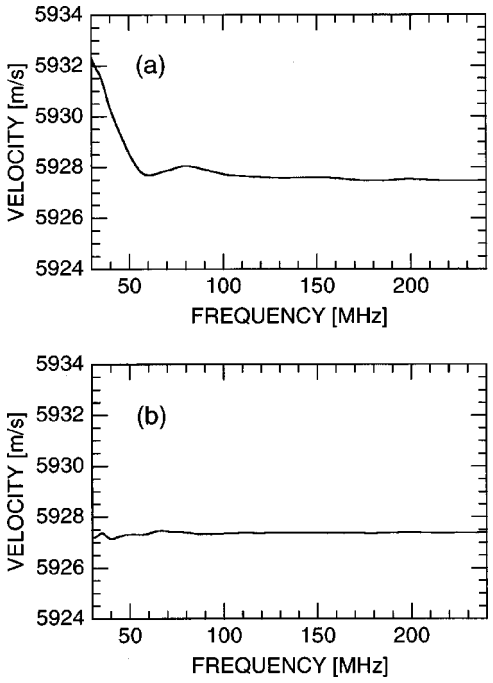


FIG. 10. Velocities of the SiO_2 specimen calculated from the phases measured in complex mode. (a) Measured. (b) Diffraction corrected.

From Eq. (14), it is seen that the diffraction effect in attenuation measurements can be corrected by subtracting the calculated values from the measured values. The attenuation coefficient corrected for the diffraction effect is plotted by the solid curve. The diffraction loss exceeds the propagation loss at frequencies below 150 MHz. Assuming that the attenuation coefficient is proportional to the square of frequency, the relation $\alpha/f^2=1.3\times10^{-16}$ (s^2/m) is obtained by the dotted curve. While the results for SiO_2 glass have large variations in this frequency range because of its low attenuation, the attenuation coefficients can be accurately determined for acoustically lossy specimens with higher attenuation, such as of borosilicate glasses and metal materials.

IV. DIFFRACTION EFFECT ON VELOCITY MEASUREMENTS

In the previous section, the phase advance due to diffraction in the VHF range was experimentally detected. In this section, we discuss how the phase advance due to diffraction affects the velocity measurements in the complex and amplitude modes, focusing on velocity, which is considered to be the most important of the acoustic properties.

A. Velocity measurements in complex mode

Figure 10(a) shows the calculated result of velocity obtained by substituting the phase of $\mathbf{V}_3/\mathbf{V}_2$ obtained in the complex mode measurement into Eq. (13). Deviations of the measured values were observed, particularly at lower fre-

quencies. As a whole, the velocity decreases as the frequency increases. If the diffraction effect is considered, the velocity measured in the complex mode is expressed as follows:

$$V_3 = -\frac{2\omega l_3}{\phi - \pi - \Delta\theta}. \tag{19}$$

From Eq. (19), it is seen that $\Delta\theta$ affects the velocity in the complex mode. Therefore, the velocity was obtained using $l_3 = 5001.7$ μm after the calculated values of $\Delta\theta$ in Fig. 8(b) were subtracted from the measured phase values. The result is given in Fig. 10(b). From the diffraction correction, nearly constant values were obtained independent of frequency, i.e., no velocity dispersion, and the true velocity was determined to be 5927.38 ($\pm$0.03) m/s at 100–240 MHz. Measurement errors of the corrected velocities were slightly larger at the lower frequencies than 70 MHz because of small S/N. Since $\Delta\theta > 0$ at all frequencies used in this measurement as shown in Fig. 8(b), the measured velocity is larger at all frequencies due to diffraction effect. For example, the measured velocity at 30 MHz is about 5.1 m/s larger. As described in this section, the diffraction effect in the complex mode can be corrected at each frequency, and the correct velocity value can be obtained.

B. Velocity measurements in amplitude mode

The amplitude mode using the pulse interference mode is usually adopted for precise velocity measurements,[1,10,13–15] so the effect of phase advance due to diffraction on the interference frequency intervals and its correction are discussed. When the frequency interval is affected by diffraction, the apparent frequency interval $\Delta f'$ [MHz] is expressed as follows:

$$\Delta f' = \left| \frac{360^\circ}{(-2\beta_3 l_3/f) + [d(\Delta\theta)/df]} \right| = \frac{\Delta f}{1 - (\Delta f/360^\circ)\cdot[d(\Delta\theta)/df]}, \tag{20}$$

where Δf [MHz] is the true frequency interval and $d(\Delta\theta)/df$ [deg/MHz] is the differentiation of the difference in phase variation due to diffraction, $\Delta\theta (= \theta_3 - \theta_2)$, in the analysis region with respect to f. From Eq. (20), it is clearly seen that $d(\Delta\theta)/df$ affects the velocity in the amplitude mode measurements. The apparent frequency interval $\Delta f'$ is larger than the true frequency interval Δf when $d(\Delta\theta)/df$ is positive, and smaller when $d(\Delta\theta)/df$ is negative. As shown in Fig. 8, $\Delta\theta$ varies considerably in the lower frequency range due to the diffraction effect, so that the frequency intervals also vary markedly at some frequencies. The variations in frequency interval are small in the higher frequency range ($f > 116$ MHz), where S for $\mathbf{V}_3$ is less than unity.

In the amplitude mode measurements, the phase variation caused by diffraction cannot be corrected for the interference waveform obtained. Instead, the correction is made for the frequency intervals obtained through analysis.

By transforming Eq. (20), the true frequency interval Δf is given by

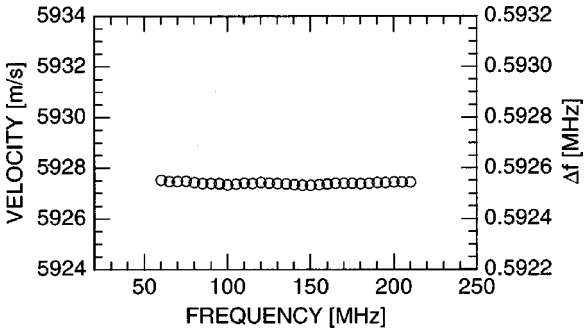


FIG. 11. Measured velocities and frequency intervals after diffraction correction in amplitude mode. The frequency span in FFT analysis is 80 MHz.

$$\Delta f = \frac{\Delta f'}{1 + (\Delta f'/360^\circ)\cdot[d(\Delta\theta)/df]} = C\cdot\Delta f', \tag{21}$$

where the correction coefficient C is defined as

$$C = \frac{1}{1 + (\Delta f'/360^\circ)\cdot[d(\Delta\theta)/df]} = 1 - \frac{\Delta f}{360^\circ}\cdot\frac{d(\Delta\theta)}{df}. \tag{22}$$

C is obtained with $\Delta f'$ or Δf and $d(\Delta\theta)/df$. C can be obtained from the first equation with $\Delta f'$ known to correct measured values. Similarly, as shown in the next section, C can be obtained from the second equation with Δf known to estimate the correction coefficient by numerical calculations. C is less than unity when $d(\Delta\theta)/df$ is positive, and is larger than unity when $d(\Delta\theta)/df$ is negative. In the amplitude mode, analyzing an interference waveform obtained with many oscillations is a very effective way to reduce the measurement errors of the interference intervals in the analysis.[22] Here, a frequency span of 80 MHz was employed for data processing by changing the center frequency of the analysis at intervals of 5 MHz.

The frequency intervals measured in the amplitude mode were corrected for diffraction using Eq. (21) with the calculated values in phase given in Fig. 8(b); the result is plotted with circles in Fig. 11. For example, $\Delta f'$ at 180 MHz is 0.592 527 MHz before diffraction correction, and is corrected to $\Delta f = 0.592\,539$ MHz with $C = 1.000\,019$. By correcting the diffraction effect, nearly constant values were obtained for the amplitude mode, even for the range where $S > 1$. In the amplitude mode measurements, 0.592 537 $\pm$0.000 006 MHz was obtained in the frequency range of 85–190 MHz. From these values, the longitudinal velocity is determined to be 5927.39 m/s, and the error in velocity corresponding to $\pm$0.000 006 MHz is estimated to be $\pm$0.06 m/s, ignoring the measurement error due to the specimen thickness.

The velocity in the amplitude mode is in excellent agreement with that in the complex mode, so that velocities obtained in the amplitude mode measurements can be corrected using Eq. (21).

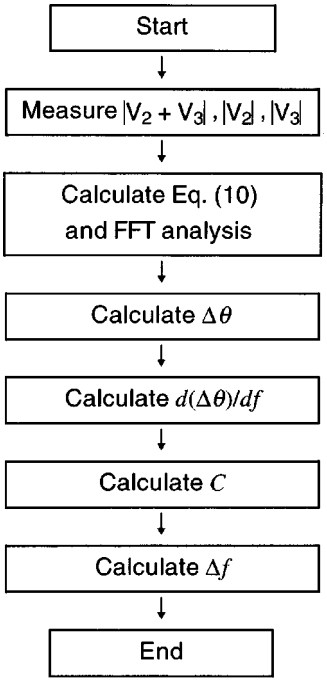


FIG. 12. A measurement procedure to obtain true frequency intervals by the bulk ultrasonic pulse *f*-interference method in amplitude mode.

C. Summarized correction procedure in amplitude mode

A flow chart showing the procedure for measuring longitudinal velocity precisely in the amplitude mode is given in Fig. 12.

For example, when velocities for the same SiO_2 glass specimens ($V_3=5927.38$ m/s) with the different thicknesses of 3, 4, and 5 mm are measured, the correction coefficients at 180 MHz with an FFT-analysis span of 80 MHz are 1.000 011, 1.000 019, and 1.000 019, respectively. The respective S values for the $\mathbf{V}_3$ signals are 0.561, 0.599, and 0.642. All measurements can be conducted in the Fresnel region, but it is necessary to correct diffraction to determine the velocities with five significant figures or more.

We will now numerically calculate the correction coefficient for the frequency interval in the longitudinal velocity measurements for other solid specimens with different longitudinal velocities, using the SiO_2 buffer rod ($l_1=10$ mm) and the water couplant. The C values are obtained for a 5-mm-thick specimen in the velocity range of 4000 m/s–15 000 m/s. The result of the calculation of Eq. (22) using Δf obtained from Eq. (12) with velocity V_3 and the specimen thickness l_3 is given in Fig. 13. In measuring the velocity for solid specimens in the VHF range, the appropriate rod length is 8–12 mm,[22] as selected in this measurement. A transducer radius of 1.25 mm was adopted for the calculation, and the propagation length of the couplant was set so that $\mathbf{V}_2$ and $\mathbf{V}_3$ would not overlap with spurious signals, as shown in Fig. 4, and so that a sufficient pulse width is obtained. $\Delta\theta$ was obtained over 20–250 MHz, and $d(\Delta\theta)/df$ is obtained for $\Delta\theta$ with an analysis span of 80 MHz. C is shown as a function of S for $\mathbf{V}_3$ at the center frequency for analysis. S decreases as the frequency increases. In this calculation, the C values generally exceed unity, but are sometimes less than unity. When S is less than 1.7, the percentage of the correction in the frequency interval value is small: -1.8×10^{-3} to $+25.1\times10^{-3}$%. When S is less than 0.9, the percentage decreases further: -1.8×10^{-3} to $+5.9\times10^{-3}$%. To ensure accurate measurements, velocity measurements are usually conducted in the Fresnel region, where diffraction effects are small.[1] This practice also applies to this method. Since the diffraction effects in the composite ultrasonic transmission line model used in this method depend upon not only the velocities and propagation lengths of the specimen but also those of the SiO_2 buffer rod and water couplant, and the analyzing frequency spans, it is impossible to make a universal correction chart with S only. However, if diffraction effects are corrected by following the procedure chart shown in Fig. 12, the velocity values close to six significant digits can be obtained.

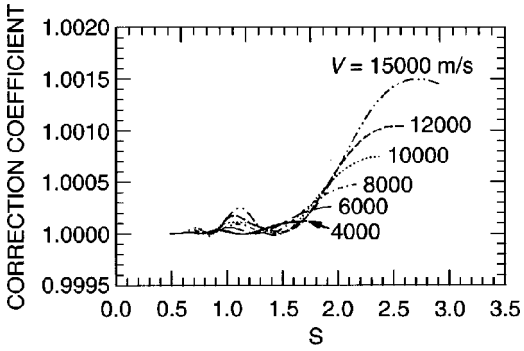


FIG. 13. Numerical calculations of correction coefficients in amplitude mode. The 10-mm-long SiO_2 rod, water couplant, and a 1.25-mm-radius transducer are used for calculations for 5-mm-thick specimens in the velocity range 4000–15 000 m/s. FFT analysis frequency span: 80 MHz.

Since the acoustic impedance of water differs from that of the solid specimen, some of the ultrasonic waves incident on the interface are transmitted and the remainder are reflected. When the specimen is SiO_2 glass and its back medium is water, the ratio of amplitudes of $\mathbf{V}_2$ and $\mathbf{V}_3$ is 8.7 dB. This ratio increases as the acoustic impedance of the specimen increases, and the amplitude of the interference observed in the interference waveform in Fig. 3(a) decreases. One material with a very high impedance is (111) diamond, whose impedance is 63.9×10^6 (Ns/m^3),[23] and the ratio of the amplitudes of $\mathbf{V}_2$ and $\mathbf{V}_3$ is approximately 21.0 dB with the back medium of water. Even in such a case, measurements can be successfully made with this method. While in this paper the correction for diffraction effects has been discussed for an isotropic solid, it must be noted that the anisotropy parameter for anisotropic solids must be considered before the diffraction effect is corrected.[6]

V. CONCLUSIONS

In this paper, we investigated the effects of diffraction on measuring velocity and attenuation of a solid specimen in a composite ultrasonic transmission line model using rf tone burst bulk ultrasonic pulses. The amplitude and phase were measured for an isotropic solid of SiO_2 glass in the VHF range, and both amplitude variations and phase variations

due to diffraction were successfully detected. Also, to calculate the diffraction effect in the composite ultrasonic transmission line model, numerical calculations based on the exact integral expression of diffraction were carried out using the total normalized propagation length. The measured values were in excellent agreement with the calculated values, so that the diffraction effects in the VHF range were experimentally explained. It was also confirmed that the phase variations due to diffraction were less in the frequency range where $S<1$. Furthermore, it was demonstrated that, in attenuation measurements, the true frequency dependence of attenuation can be obtained after correcting diffraction loss by numerical calculations.

The diffraction effect in velocity measurements was then discussed. It was demonstrated that, in complex-mode measurements, the true velocity at each frequency can be obtained by numerically calculating and correcting for diffraction. Generally, velocity measurements were conducted in the amplitude mode, in which the amplitude of the interference output of the signals superposed with the double-pulse method is measured by sweeping the frequency, and the velocity is obtained from the frequency intervals. By comparing and investigating the data obtained from both the amplitude-mode and complex-mode measurements, we found that the frequency intervals in the amplitude-mode measurements were frequency-dependent because of the diffraction effect. Also, a new method for correcting phase advance due to diffraction was proposed and demonstrated to obtain true velocity values. In any case, to measure velocity with high accuracy without correction, measurements have to be made in the Fresnel zone in the frequency range where $S<0.9$. With the application of this newly proposed method for correcting diffraction effects, it is possible to obtain velocity to nearly six significant figures.

This paper discussed solid specimens, but these approaches can be extended to accurately measure acoustic parameters for liquid and biological tissue specimens. Studies are to be continued on the diffraction effects of liquids in measurements using the variable path length method, and on the diffraction effects and velocity measurements for solid and liquid specimens which exhibit velocity dispersion.

ACKNOWLEDGMENTS

The authors are very grateful to T. Sannomiya, N. Akashi, and Y. Hashimoto for their invaluable cooperation in constructing the measurement system used here, and to K. Yagi, Hewlett Packard Japan, Ltd., for producing the network/spectrum analyzer with a custom-made option for us. This work was supported in part by a Research Grant-in-Aid from the Ministry of Education, Science and Culture of Japan.

[1] For example, M. A. Breazeale, J. H. Cantrell, Jr., and J. S. Heyman, ''Ultrasonic wave velocity and attenuation measurements,'' in *Methods of Experimental Physics*, Vol. 19, Ultrasonics, edited by P. D. Edmonds (Academic, New York, 1981), Chap. 2, pp. 67–135; J. H. Cantrell and W. T. Yost, ''Ultrasonic velocity,'' in *Encyclopedia of Acoustics*, Vol. 2, edited by M. J. Crocker (Wiley, New York, 1997), Chap. 55, pp. 629–639.

[2] A. O. Williams, Jr., ''The piston source at high frequencies,'' J. Acoust. Soc. Am. **23**, 1–6 (1951).

[3] 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).

[4] A. S. Khimunin, ''Numerical calculation of the diffraction corrections for the precise measurement of ultrasound absorption,'' Acustica **27**, 173–181 (1972).

[5] G. C. Benson and O. Kiyohara, ''Tabulation of some integral functions describing diffraction effects in the ultrasonic field of a circular piston source,'' J. Acoust. Soc. Am. **55**, 184–185 (1974).

[6] E. P. Papadakis, ''Ultrasonic diffraction loss and phase change in anisotropic materials,'' J. Acoust. Soc. Am. **40**, 863–876 (1966).

[7] H. J. McSkimin, ''Empirical study of the effect of diffraction on velocity of propagation of high-frequency ultrasonic waves,'' J. Acoust. Soc. Am. **32**, 1401–1404 (1960).

[8] A. S. Khimunin, ''Ultrasonic propagation parameter measurements incorporating exact diffraction corrections,'' Acustica **39**, 87–95 (1978).

[9] K. Negishi, K. Takagi, and H. Ozawa, ''Observation of phase advance due to diffraction of ultrasonic pulse,'' J. Acoust. Soc. Jpn. (E) **1**, 11–15 (1980).

[10] J. Kushibiki, N. Akashi, T. Sannomiya, N. Chubachi, and F. Dunn, ''VHF/UHF range bioultrasonic spectroscopy system and method,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 1028–1039 (1995).

[11] J. Kushibiki, T. Sannomiya, and N. Chubachi, ''A useful acoustic measurement system for pulse mode in VHF and UHF ranges,'' IEEE Trans. Sonics Ultrason. **SU-29**, 338–342 (1982).

[12] J. Kushibiki and N. Chubachi, ''Material characterization by line-focus-beam acoustic microscope,'' IEEE Trans. Sonics Ultrason. **SU-32**, 189–212 (1985).

[13] J. Kushibiki and M. Arakawa, ''A method for calibrating the line-focus-beam acoustic microscopy system,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 421–430 (1998).

[14] J. Williams and J. Lamb, ''On the measurement of ultrasonic velocity in solids,'' J. Acoust. Soc. Am. **30**, 308–313 (1958).

[15] C. Wong and P. Garikepati, ''Ultrasonic velocity measurement by the pulse interference method using a single transducer,'' J. Acoust. Soc. Am. **70**, 1185 (1981).

[16] Y. Sugawara, J. Kushibiki, and N. Chubachi, ''Performance of concave transducers in acoustic microscopy,'' Proceedings of the IEEE Ultrasonics Symposium 751–756 (1988).

[17] Y. Ono, J. Kushibiki, and N. Chubachi, ''A measurement method of moving characteristics of precision mechanical-translation stages using ultrasonic plane waves and its application to a line-focus-beam acoustic microscopy system,'' Trans. IEICE Japan **J78-A**, 279–286 (1995).

[18] E. P. Papadakis, ''Ultrasonic attenuation in thin specimens driven through buffer rods,'' J. Acoust. Soc. Am. **44**, 724–734 (1968).

[19] W. Kroebel and K.-H. Mahrt, ''Recent results of absolute sound velocity measurements in pure water and sea water at atmospheric pressure,'' Acustica **35**, 154–164 (1976).

[20] H. A. Bowman and R. M. Schoonover, ''Procedure for high precision density determinations by hydrostatic weighing,'' J. Res. Natl. Bur. Stand., Sect. C **71**, 179–198 (1967).

[21] G. S. Kell, ''Density, thermal expansivity, and compressibility of liquid water from 0° to 150 °C: Correlations and tables for atmospheric pressure and saturation reviewed and expressed on 1968 temperature scale,'' J. Chem. Eng. Data **20**, 97–105 (1975).

[22] M. Arakawa and J. Kushibiki (unpublished).

[23] H. J. McSkimin and W. L. Bond, ''Elastic moduli of diamond,'' Phys. Rev. **105**, 116–121 (1957).

Elastic constants of an aluminum–alumina unidirectional composite

J. E. Vuorinen[a] and R. B. Schwarz
Materials Science and Technology Division, MST-8, Los Alamos National Laboratory, MS G755, Los Alamos, New Mexico 87545

C. McCullough
Metal Matrix Composite Program, 3M Company, St. Paul, Minnesota 55144

(Received 10 June 1999; accepted for publication 1 May 2000)

The five independent second-order elastic constants of a transversely isotropic aluminum/alumina fiber composite have been measured for the first time using a resonant ultrasound spectroscopy technique. These data were used to deduce the elastic constants and engineering moduli for off-axis loading conditions. © *2000 Acoustical Society of America.* [S0001-4966(00)01608-8]

PACS numbers: 43.35.Cg [HEB]

INTRODUCTION

Continuous-fiber (CF) reinforced composites can be tailored to achieve higher strength, stiffness, and creep resistance than those found in monolithic materials and particulate composites. In recent years, there has been extensive work in developing CF composites based on metallic and polymeric matrices. Metallic matrices have the advantage that they are nonflammable, do not outgas in vacuum, and are minimally attacked by organic fluids such as fuels and solvents.[1] Furthermore, since the elastic properties of available reinforcing fibers better match those of metals than polymers, CF metal-matrix composites have smaller anisotropy than polymer-based composites.

Most of the commercial work on CF-MMCs has been based on using aluminum and its alloys as the metal matrix, reinforced with fibers such as alumina, boron, or graphite. Among them, alumina fibers offer several advantages including high strength and stiffness, good bonding to the aluminum matrix, and lack of galvanic corrosion between matrix and fibers.[2] Much emphasis has been given to correlating strength to damage mechanisms in aluminum/alumina CF-MMCs.[3,4]

In the past, the elastic properties of MMCs have mainly been determined from tensile and compressive tests.[5,6] In a few instances, the elastic properties have been derived from measurements of the second-order elastic constants using ultrasonic techniques.[7,8]

Microscopically, the elastic properties of any composite are heterogeneous. However, if the scale of interest is large enough, the material may exhibit statistical homogeneity. A volume element, L^3, can be considered statistically homogeneous if the distribution of fibers within this volume is homogeneous and the separation between the fibers is much smaller than L.

Predicting the deformation of a MMC under arbitrary loading conditions requires a detailed knowledge of its average elastic properties. The simplest CF-MMC is one for which the fibers are straight and homogeneously distributed, all aligned in one direction. The elastic properties of this solid are homogeneous on a plane normal to the fiber direction. Such a *transversely isotropic* solid is characterized by only five independent elastic constants. With the [100] axis oriented along the fiber direction, the average or effective constitutive stress and strain components are related by (in reduced two-index matrix notation):

$$\begin{Bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{31} \\ \sigma_{12} \end{Bmatrix} = \begin{Bmatrix} C_{11} & C_{12} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{22} & C_{23} & 0 & 0 & 0 \\ C_{12} & C_{23} & C_{22} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{C_{22}-C_{23}}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{66} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{66} \end{Bmatrix} \times \begin{Bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{33} \\ 2\varepsilon_{23} \\ 2\varepsilon_{31} \\ 2\varepsilon_{12} \end{Bmatrix}. \quad (1)$$

These five moduli describe the elastic properties of a homogeneous solid that is elastically equivalent to the heterogeneous composite.[9] Although this elastic constant matrix gives a complete description of the stiffness of the composite, such description has not been generally adopted. The main difficulty has been measuring the five independent elastic constants. Engineers have thus opted for a pragmatic, often incomplete, description of the elastic properties of composites in terms of *engineering* moduli, which are amenable to determination via static tests. Engineering moduli are also preferred because they are more directly related to applications.

Advances in resonant ultrasound spectroscopy (RUS) techniques have made it possible to deduce the complete elastic-constant tensor from a single frequency scan of the

[a]Permanent address: Institute of Materials Science, Tampere University of Technology, P.O. Box 589, FIN-33101 Tampere, Finland.

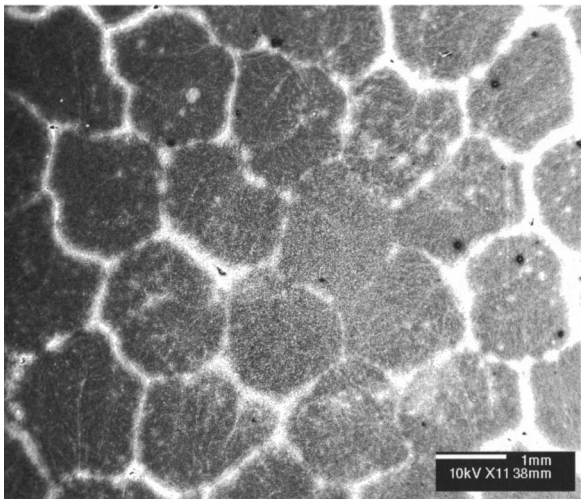


FIG. 1. Low magnification scanning electron micrograph of the surface normal to the fiber direction.

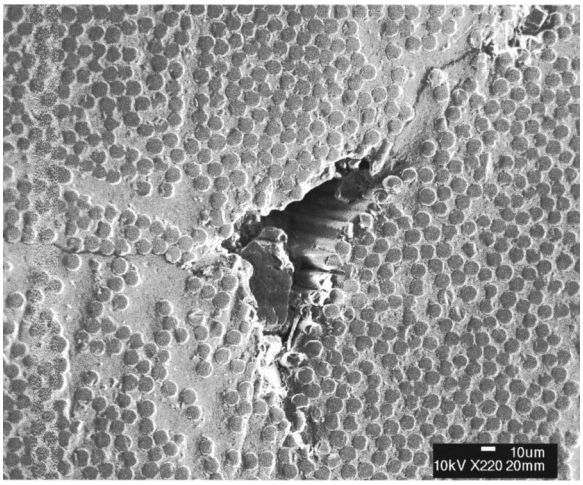


FIG. 2. High magnification scanning electron micrograph of surface normal to the fiber direction. A micro crack at the triple junction of pultruded Al/Al_2O_3 wires is clearly visible.

mechanical resonances of a sample.[10,11] Currently, the technique is limited to solids having orthorhombic elastic symmetry, or higher. Orthorhombic solids are characterized by nine independent elastic constants. A composite having an arbitrary distribution of reinforcing fibers would most likely have symmetry lower than orthorhombic, and would thus require a large number of independent elastic constants. Numerical codes to analyze the mechanical resonant spectrum of such low-symmetry solids are not yet available.

The present work addresses the elastic properties of a transversely isotropic aluminum-alumina CF-MMC. We used the RUS method to measure the five independent elastic constants of the composite. These constants were then used to calculate engineering moduli, which were then compared to engineering moduli determined by averaging the known elastic properties of the matrix and fibers.

I. EXPERIMENT

The measurements were done on a 12×12 mm cross section rod of CF $Al-Al_2O_3$ composite material provided by the 3M Company (3M Center, St. Paul, MN 55144-1000). This material was experimental in nature and contained approximately 50 vol% of Nextel 610™ alumina fibers in a matrix of pure aluminum. The initial starting material is made by infiltrating approximately 35 tows of fiber with molten aluminum, to produce a wire precursor. This precursor is then used in an experimental second process to pultrude the rod. This latter process involves simultaneously drawing many wires through a bath of molten aluminum and combining them through a shaping orifice to form the final product. Figure 1 shows a scanning electron microscope image of the CF-AMC surface normal to the fiber direction. The micrograph shows the wires (dark regions), approximately 2 mm in diameter, separated by regions rich in aluminum (lighter regions). Figure 2 is a higher magnification micrograph showing one of the black features of Fig. 1 in a region joining three pultruded wires. Here, individual Al_2O_3 fibers, approximately 12 μm in diameter, are clearly seen. Small cracks running parallel to the fiber direction were detected at some of the triple joints. Such cracks arise from insufficient consolidation of the wires as they pass through the exit orifice during the processing of the composite bar. In addition, a thin oxide layer can be occasionally found at the boundary between consolidated wires.

Two parallelepiped-shaped samples, each having one axis parallel to the fiber direction, were cut from the original rod using a low-speed diamond-impregnated wire saw. The sample dimensions, mass, and densities are given in Table I.

The resonant ultrasound measurements were done using a commercial ultrasound system from Dynamic Resonant Systems (Powell, Wyoming 82435). The as-cut samples were lightly held at a pair of opposite faces between a pair of pinducers (Valpey Fisher, Hopkinton, MA 01748). Because obtaining accurate elastic constants requires measuring a large number of resonances, the excitation frequency was scanned from 100 kHz to 700 kHz, measuring approximately 63 resonance frequencies. The first 43 resonances were used to deduce the elastic constants. The following 20 resonance frequencies were used to verify the calculations, by comparing the measured frequencies with those predicted by the deduced elastic constants.

There is no direct method to calculate the elastic constant tensor from the measured set of mechanical resonances, f_n^{mea} $(n=1,2,...)$. We thus use an indirect method. We start by calculating an approximate resonance spectrum, $f_n^{\mathrm{cal}}(n=1,2,...)$, using as input the known sample dimensions and mass, and a set of *guessed* elastic constants. The difference between the calculated and measured resonance frequency spectrum is quantified by a figure-of-merit function,

TABLE I. Dimensions, mass, and density of samples 1 and 2.

Sample	d_1 (mm)	d_2 (mm)	d_3 (mm)	Mass (g)	ρ (g/cm^3)
1	9.116	9.661	11.553	3.329	3.272
2	6.441	7.771	8.752	1.428	3.260

$$F=\sum_n w_n(f_n^{\mathrm{cal}}-f_n^{\mathrm{mea}})^2, \tag{2}$$

where w_n $(n=1,2,...)$ are weight coefficients reflecting the confidence on individual resonance measurements. We then seek a minimization of the function F by regressing the values of all the elastic constants, using computer software developed for the process.[12] Certainly, the number of mechanical resonances used to deduce the five elastic constants of our composite is largely redundant. However, a large number of resonances is needed to ensure that the minimization has reached the global minimum in the function F and not a local minimum. The extra 20 measured resonances, not used in the minimization process, are then carefully compared to the calculated resonances, paying special attention to the overall spacing and grouping of these extra resonances. This procedure makes it unnecessary to identify each of the resonant modes.[10]

II. CALCULATED ELASTIC CONSTANTS OF THE COMPOSITE

The minimization of F in Eq. (2) requires we start from a reasonable good set of *guessed* C_{ij} values, since starting from an arbitrary set of C_{ij} values may also lead to a local, rather than global, minimum. We did not find in the literature the required C_{ij} values, nor any set of engineering moduli. Researchers have used various micromechanics models to calculate the elastic properties of composites. Herakovich[13] has compared the elastic moduli of transversely isotropic composites deduced from the predictions of several micromechanical models, including the simpler methods of Reuss and Voigt. These two models are based on the assumption that the stress or the strain (respectively) are uniform throughout the composite, and give extreme values for the elastic moduli. From Herakovich's comparison it can be seen that the Reuss or Voigt predictions give, depending on the particular modulus, results that are quite close to those predicted by the more elaborate micromechanical models. We thus evaluated the starting C_{ij} values using the following formulas:

$$E_{11}=V_fE_f+V_mE_m \quad \text{(Voigt average)}, \tag{3}$$

$$\frac{1}{E_{22}}=\frac{V_f}{E_f}+\frac{V_m}{E_m} \quad \text{(Reuss average)}, \tag{4}$$

$$\frac{1}{G_{12}}=\frac{V_f}{G_f}+\frac{V_m}{G_m} \quad \text{(Reuss average)}, \tag{5}$$

$$\nu_{12}=V_f\nu_f+V_m\nu_m \quad \text{(Voigt average)}, \tag{6}$$

$$\nu_{23}=\frac{E_mV_f\nu_f+E_fV_m\nu_m}{E_mV_f+E_fV_m} \quad \text{(Reuss average)}. \tag{7}$$

These formulas are basically rules of mixtures that predict the engineering moduli of transversely isotropic MMCs in terms of the isotropic elastic properties of the matrix and fiber. Here E, G, ν, and V denote the Young modulus, shear modulus, Poison ratio, and volume fraction, respectively. The sub-indices m and f refer to the matrix and fiber. In calculating the Young's modulus and Poisson's ratio for a tensile elongation along the fibers, strain is considered to be uniform, leading to Voigth-type averaging. To calculate the Young's modulus and Poisson's ratio for deformation normal to the fibers, stress is considered uniform, leading Reuss-type averaging. These results are identical to those proposed by Brody and Ward[14] in the limit of isotropic fibers. In this approximation, $G_{23}=G_{12}$.

TABLE II. Estimated and measured engineering moduli.

Engineering modulus	Estimated from Eqs. (3) to (7)	Calculated using measured C_{ij}'s of sample 1	Calculated using measured C_{ij}'s of sample 2
E_{11}(GPa)	221.0	231.0	230.7
E_{22}(GPa)	116.5	136.6	136.5
G_{12}(GPa)	44.2	56.9	57.0
G_{23}(GPa)	44.2	51.6	52.7
ν_{12}	0.290	0.268	0.275
ν_{23}	0.318	0.322	0.295

Knowing the engineering moduli, one can easily calculate the elastic compliance tensor $[S_{ij}]$, given by:

$$[S_{ij}]=\begin{bmatrix} \frac{1}{E_{11}} & \frac{-\nu_{12}}{E_{11}} & \frac{-\nu_{12}}{E_{11}} & 0 & 0 & 0 \\ \frac{-\nu_{12}}{E_{11}} & \frac{1}{E_{22}} & \frac{-\nu_{23}}{E_{22}} & 0 & 0 & 0 \\ \frac{-\nu_{12}}{E_{11}} & \frac{-\nu_{23}}{E_{22}} & \frac{1}{E_{22}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2(1+\nu_{23})}{E_{22}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{G_{12}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{G_{12}} \end{bmatrix}, \tag{8}$$

where the Poisson's ratio is defined as $\nu_{ij}=-(\epsilon_{jj}/\epsilon_{ii})$. A numerical inversion of $[S_{ij}]$ gives the $[C_{ij}]$ matrix.

Pure polycrystalline aluminum has Young's modulus of 69.0 GPa, shear modulus of 25.9 GPa, and Poisson's ratio of 0.330. Alumina fibers have Young's modulus of 373 GPa, shear modulus of 149 GPa, and Poisson's ratio of 0.250.[13,15] We assume the present alumina fibers to be elastically isotropic. The presence of anisotropy would only affect the cal-

TABLE III. Estimated and measured elastic constants.

	Estimated from Eqs. (3) to (8) (GPa)	Measured, sample 1 (Gpa)	Measured, sample 2 (Gpa)	Difference between samples 1 and 2 (%)
C_{11}	254.0	264.1	264.3	0.1
C_{12}	55.9	61.5	61.0	0.8
C_{22}	142.3	166.5	163.6	1.8
C_{23}	54.0	63.3	58.2	8.1
C_{44}	44.2	51.6	52.7	2.1
C_{66}	44.2	56.9	57.0	0.2

TABLE IV. First ten calculated and experimental resonance frequencies for samples 1 and 2, and their differences.

Sample 1			Sample 2		
f_{ex}(kHz)	f_{calc} (kHz)	Difference (%)	f_{ex} (kHz)	f_{calc} (kHz)	Difference (%)
164.081	164.032	−0.03	208.201	208.729	0.25
210.213	208.806	−0.67	288.542	289.520	0.34
249.221	248.420	−0.32	316.017	315.215	−0.25
251.219	252.083	0.34	321.620	318.101	−1.09
256.914	256.610	−0.12	349.183	345.772	−0.98
260.159	259.665	−0.19	349.899	350.646	0.21
264.311	264.186	−0.05	352.160	352.215	0.02
271.233	270.058	−0.43	352.580	352.919	0.10
278.297	281.014	0.98	358.270	355.663	−0.73
298.216	298.279	0.02	392.670	388.286	−1.12

culated input values for elastic constants of the composite and not the measured elastic moduli of the composite. The assumption of elastically isotropic fibers is less justified for carbon or SiC fibers.[16]

The fiber volume fraction was calculated using the formula:

$$V_f=\frac{\rho_c-\rho_m}{\rho_f-\rho_m}, \tag{9}$$

where the sub-index c denotes the composite. The density of the samples was measured by the Archimedes' principle. The density of both samples was 3.289 g/cm^3. Using Eq. (9) and density values of 2.698 g/cm^3 for pure Al and 3.88 g/cm^3 for alumina fibers, we calculated the volume fraction of fibers to be 0.50.[15,17] We also determined V_f by a linear-intercept metallographic method applied to the SEM micrograph shown in Fig. 2. This method gave a fiber volume fraction in the 0.44–0.60 range, which is consistent with the value determined from the measured sample density. The metallographic method also shows the range in local volume fractions in the material. Using a value of $V_f=0.5$, and Eqs. (3)–(8), we calculated the engineering moduli and elastic constants of the MMC. The results are given in Tables II and III.

III. MEASURED ELASTIC CONSTANTS

A minimization of the error function F, as explained above, yielded the true elastic constants of the Al–Al_2O_3 composite, which are given in Table III. The first ten resonant frequencies calculated from the true elastic constants (and the known sample dimensions and mass) are given in Table IV and their agreement with the corresponding measured resonance frequencies is good. The total rms error between measured and calculated resonance frequencies for first 43 resonances is 0.48% for sample 1, and 0.64% for sample 2. The agreement between the following 20 calculated and measured resonance frequencies is equally good. Adding the 20 resonant frequencies to the minimization of F does not improve in the accuracy of the fit nor changes the values of the fitted elastic constants.

TABLE V. Estimated and measured elastic compliances.

	Estimate using Eqs. (3)–(8) (GPa^{-1})	Measured, sample 1 (GPa^{-1})	Measured, sample 2 (GPa^{-1})
S_{11}	4.50×10^{-3}	4.33×10^{-3}	4.34×10^{-3}
S_{12}	1.28×10^{-3}	1.16×10^{-3}	1.19×10^{-3}
S_{22}	8.57×10^{-3}	7.33×10^{-3}	7.33×10^{-3}
S_{23}	2.75×10^{-3}	2.36×10^{-3}	2.16×10^{-3}
S_{44}	22.6×10^{-3}	19.4×10^{-3}	19.0×10^{-3}
S_{66}	22.6×10^{-3}	17.6×10^{-3}	17.6×10^{-3}

We used the measured elastic constants (Table III) to deduce experimental values for the engineering moduli (Table II). To do this, we first calculated in Table V the compliance $[S_{ij}]$ tensor by inverting the measured $[C_{ij}]$ tensor. The engineering moduli follow from:

$$E_{11}=\frac{1}{S_{11}};\quad E_{22}=\frac{1}{S_{22}};\quad G_{12}=\frac{1}{S_{66}};$$
$$G_{23}=\frac{1}{S_{44}};\quad \nu_{12}=\frac{-S_{12}}{S_{11}};\ \text{and}\ \ \nu_{23}=\frac{-S_{23}}{S_{22}}. \tag{10}$$

The engineering moduli are given in Table II. The anisotropy in the Young's modulus is $E_{11}/E_{22}=1.68$. The anisotropy in the shear modulus is lower, $G_{12}/G_{23}\approx1.1$.

IV. ENGINEERING MODULI FOR OFF-AXIS LOADING CONDITIONS

Engineering applications of MMC often require knowing the elastic properties of the material at off-axis loading conditions. The moduli in Tables II and III are given with respect to a coordinate system whose [100] axis is oriented along the fibers. The engineering moduli for a coordinate system rotated by an angle β about the 3-axis (transformed [100] axis makes an angle β with the original [100] axis) can be easily calculated by expressing the elastic compliances in the fourth-order tensor notation, and transforming this tensor according to:

$$S'_{ijkl}=a_{im}a_{jn}a_{ko}a_{1p}S_{mnop}, \tag{11}$$

where $a_{11}=\cos(\beta)$, $a_{12}=0$, $a_{13}=\sin(\beta)$, $a_{21}=-\sin(\beta)$, $a_{22}=0$, $a_{23}=\cos(\beta)$, $a_{31}=0$, $a_{32}=1$, and $a_{33}=0$. In converting the elastic compliances from 2-index notation to the tensorial 4-index notation, one has to carefully consider

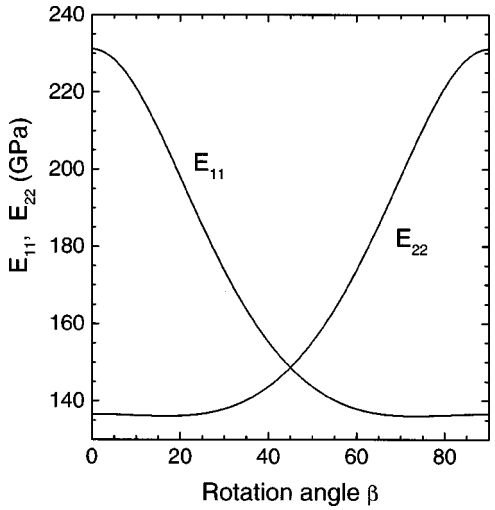


FIG. 3. Young's moduli E_{11} and E_{22} as a function of rotation angle β.

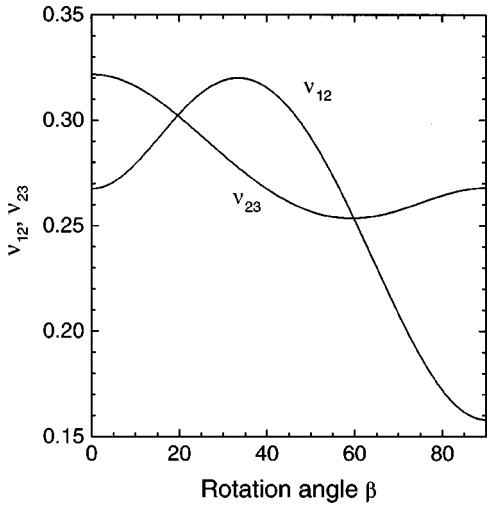


FIG. 5. Poisson's ratios ν_{12} and ν_{23} as functions of rotation angle β.

factors of 2 and 4 arising from the definitions of strains.[18] Knowing the elastic compliance tensor in the rotated coordinate system, the engineering moduli follow Eq. (10).

Figure 3 shows the engineering moduli E_{11} and E_{22} as a function of rotation angle β. For increasing values of β, the Young's modulus E_{11} decreases rapidly from 230 GPa to about 137 GPa, reflecting the typical behavior of a soft matrix reinforced with stiff unidirectional continuous fibers. It is remarkable that the stiffening effect of the fibers has been almost completely lost for $\beta=45°$. As expected, the angular dependence of the E_{22} is mirror image of E_{11} about $\beta=45°$. The dependence of the shear moduli G_{12} and G_{23} on angle β are shown in Fig. 4. G_{12} shows a maximum at 45°, which is typical of unidirectional CF composites.[19] As expected for this fiber geometry, $G_{12}(\beta=0)=G_{12}(\beta=90)=G_{23}(\beta=90)$. The Poisson ratios ν_{12} and ν_{23} are given in Fig. 5. The ν_{12} shows the typical increased lateral contractions of CF composites caused by shear strains.[19] The change in the value of ν_{23} is less pronounced.

V. DISCUSSION

There is a clear difference between the elastic moduli calculated using micromechanical averaging formulas and moduli measured by ultrasonic spectroscopy. Because *all* the calculated elastic moduli are smaller than the measured ones, the difference could simply be attributed to an error in the volume fraction of fiber used in the calculations. Another possible reason for the discrepancy is that the averaging formulas underestimate the moduli of the composite. This is especially true for the transverse Young's modulus E_{22} and for the shear moduli, all of which were determined using Reuss-type averages.[13] This is not true for the longitudinal Young's modulus E_{11}, which was determined by a Voigt-type average that overestimated the modulus.[13] A third possible reason for the discrepancy is that the values for the fiber and aluminum matrix elastic moduli used on the calculations were too low. Finally, the presence of micro cracks in the aluminum matrix could also affect the elastic moduli, but this would lower the measured elastic moduli and thus would not explain the observed differences between the calculated and measured values.

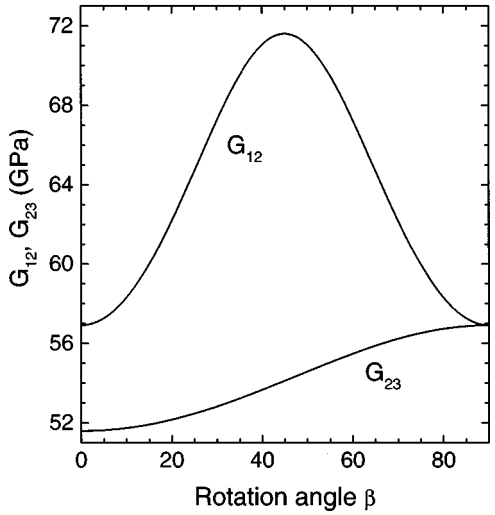


FIG. 4. Shear moduli G_{12} and G_{23} as a function of rotation angle β.

As shown in Table III, the difference between the elastic constant measured in samples 1 and 2 ranges from 0.08% for C_{11}, to 8.5% for C_{23}. Differences less than 1% could be attributed to the accuracy of the resonance ultrasound method itself since, after minimization of F, the rms of the differences between the calculated and measured resonance frequencies was about 0.5% for sample 1, and 0.7% for sample 2. The large difference between the C_{23} values measured in samples 1 and 2 must therefore be attributed to differences between the samples. Figure 1 shows that the Al_2O_3 fibers are distributed in bundles, approximately 1 mm in diameter. Thus each sample contains about 100 bundles (see sample dimensions in Table I). But, depending on how the sample is cut, this number can fluctuate, changing the overall volume fraction of fibers. However, the equality in the sample densities suggests that the fiber volume fractions are also similar. Furthermore, a significant difference in fiber volume fraction should change the values of the other elastic constants as well, which is contrary to observation. A more plausible reason for the discrepancy in the C_{23} values is the presence of micro cracks in the aluminum matrix, as seen near the triple points between fiber bundles in Fig. 2. Be-

cause C_{23} depends strongly on the elastic properties of the matrix, a slightly larger number of micro cracks in sample 2 could explain its lower C_{23} value.

VI. CONCLUSIONS

Resonance ultrasound spectroscopy can be used to measure the five independent second-order elastic constants of transversely isotropic matrix composites.

The calculation of elastic moduli from mechanical resonance data is greatly simplified by starting from an approximate set of elastic moduli. For unidirectional fibers (transversally isotropic solid), a simple rule of mixture give reasonable estimates for the elastic moduli of the composite in terms of the elastic moduli of the matrix and fibers.

ACKNOWLEDGMENTS

The work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences. J. E. V. acknowledges partial support from the Tampere University of Technology, Institute of Materials Science.

[1] J. V. Foltz, Adv. Mater. Processes **12**, 19 (1998).
[2] H. E. Deve and C. McCullough, J. Minerals, Metals and Materials Society **47**, 33 (1995).
[3] U. Ramamurty, F. W. Wok, F. A. Leckie, and H. E. Deve, Acta Mater. **45**, 4603 (1997).
[4] J. He, I. J. Beyerlein, and D. R. Clarke, J. Mech. Phys. Solids **47**, 465 (1999).
[5] C. McCullough, H. E. Deve, and T. E. Channel, Mater. Sci. Eng., A **189**, 147 (1994).
[6] N. Tsangarakis, G. P. Pelletier, and M. S. Pepi, Exp. Mech. **32**, 49 (1992).
[7] H. Ledbetter, C. Fortunko, and P. Heylinger, J. Appl. Phys. **78**, 1542 (1995).
[8] H. Ogi, K. Takashima, H. Ledbetter, M. L. Dunn, G. Shimoike, M. Hirao, and P. Bowen, Acta Mater. **47**, 2787 (1999).
[9] R. M. Christensen, *Mechanics of Composite Materials* (Wiley, New York, 1979).
[10] *Resonant Ultrasound Spectroscopy*, edited by A. Migliori and J. L. Sarrao (Wiley-Interscience, New York, 1997).
[11] V-T. Kuokkala and R. B. Schwarz, Rev. Sci. Instrum. **63**, 3136 (1992).
[12] W. M. Visscher, A. Migliori, T. M. Bell, and R. A. Reinert, J. Acoust. Soc. Am. **90**, 2154 (1991).
[13] C. T. Herakovic, *Mechanics of Fibrous Composites* (Wiley, New York, 1998).
[14] H. Brody and I. M. Ward, Polym. Eng. Sci. **11**, 129 (1971).
[15] 3M Nextel Ceramic Fiber Technical Notebook, 3M Ceramic Fiber Products, St. Paul, MN, USA, 1997.
[16] D. S. Li and M. R. Wisnom, J. Compos. Mater. **30**, 561 (1996).
[17] J. Emsley, *The Elements*, 2nd ed. (Clarendon, Oxford, 1991).
[18] E. Schreiber, O. L. Anderson, and N. Soga, *Elastic Constants and Their Measurement* (McGraw-Hill, New York, 1974).
[19] D. Hull and T. W. Clyne, *An Introduction to Composite Materials*, 2nd ed. (Cambridge University Press, Cambridge, 1996).

Scattering-induced attenuation of an ultrasonic beam in austenitic steel

Thomas Seldis and Claudio Pecorari[a)]
Institute for Advanced Materials, European Commission, PO Box 2, 1755 ZG Petten, The Netherlands

(Received 6 May 1999; accepted for publication 18 April 2000)

The scattering-induced attenuation coefficient of a beam of longitudinal waves propagating through an austenitic steel plate is measured as a function of the texture angle. The experimental data were obtained by mapping the incident and the transmitted ultrasonic field, and by evaluating the energy loss experienced by each plane wave component of the beam. Contrary to the behavior of data obtained by means of conventional techniques, that of the data reported in this work agrees qualitatively with the theoretical predictions. The reasons for the disagreement between theory and data obtained by means of conventional techniques are also discussed. © *2000 Acoustical Society of America.* [S0001-4966(00)00408-2]

PACS numbers: 43.35.Cg [HEB]

INTRODUCTION

A reliable assessment of the integrity of an austenitic steel weld requires a good understanding of the inspecting beam's behavior within the weld. Among the physical parameters characterizing the interaction of the beam with its supporting medium, ultrasonic attenuation is very important because it limits the volume of the system that can be inspected, and may affect the quantitative assessment of the properties of a defect the inspecting wave may detect. Several theoretical models have been developed to predict the dependence of plane wave attenuation on the ultrasonic frequency and on the structural properties of the polycrystalline medium in which the wave propagates. Both materials without[1] and with texture[2–5] have been considered in these studies. Measurements of longitudinal wave attenuation in textured austenitic steel as a function of the propagation direction have also been reported in the literature.[6,7] A comparison between theoretical predictions and experimental results shows a surprising disagreement between the two: while all the models predict the attenuation coefficient reaching a minimum in the texture direction, the experimental results display a relative maximum around that direction.

Neumann[7] conjectures that two mechanisms may explain the apparent increase of the attenuation coefficient around the texture direction. The first mechanism would explain the relative maximum in terms of the longitudinal-into-shear wave mode conversion occurring at the grain boundaries as a longitudinal wave propagates along the main axis of the grains. The second mechanism would be the destructive interference among the plane wave components of the inspecting beam at the receiving transducer's surface.

Panetta, Margetan *et al.*[8,9] investigated the behavior of an ultrasonic beam crossing a plate of titanium alloy having a textured structure similar to that of the austenitic steel plates considered in this work. They measured the attenuation of longitudinal waves in the texture direction and along two directions normal to it by employing both a conventional pulse–echo method and a novel approach by which the energy of the transmitted beam is used to evaluate the attenuation. The results obtained by the conventional approach showed the attenuation along the texture direction considerably larger than that in directions normal to it at all the ultrasonic frequencies considered in the investigation. By using the energy approach, on the other hand, they obtained results that were comparable in all the directions examined. They explained this finding by showing that the wavefront of the beam crossing the plate is distorted more when the beam travels along the texture direction than along directions normal to it. In other words, Panetta, Margetan *et al.*[8,9] recognized that the surprisingly large attenuation along the texture direction measured by the pulse–echo method is the result of destructive interference between the beam's components. However, in their analysis Panetta, Margetan *et al.* left open the question whether the phase distortion is caused mainly by the beam interaction with the material microstructure, or results from the interaction of the beam with the anisotropic macrostructure of the sample. Furthermore, they estimated the energy transmitted through the sample by considering only the transmission coefficient at normal incidence, and neglecting the fact that the beam components propagating in directions other than the normal may be affected differently by the interaction with the water–sample interface.

The aims of this paper are twofold. First, measurements of the scattering-induced attenuation coefficient of a beam of longitudinal waves propagating in textured austenitic steel plates are presented, together with a description of the new experimental approach used in this work. These results agree qualitatively with the theoretical predictions. Second, the causes of the disagreement between theoretical predictions and measurements of the longitudinal wave attenuation coefficient in textured polycrystalline materials by means of a conventional technique are discussed.

I. DESCRIPTION OF THE SAMPLES

Austenitic, cast samples containing 5% carbon, 10% nickel, and 18% chromium were utilized for the longitudinal wave attenuation measurements. The average lateral dimen-

a)Electronic mail: pecorari@jnr.nl

TABLE I. Elastic constants of the austenitic steel plates used in this work.

C_{11} (10^{11} Nm^{-2})	C_{13} (10^{11} Nm^{-2})	C_{33} (10^{11} Nm^{-2})	C_{44} (10^{11} Nm^{-2})	C_{66} (10^{11} Nm^{-2})
2.575	1.505	2.201	1.187	0.768

sion of the grains was estimated to be of the order of a few millimeters. The mass density of this material is $\rho = 7880$ kg/m^3, and its structure is characterized by hexagonal symmetry. The average stiffness constants are given in Table I. Figure 1 reports the values of the longitudinal phase velocity as a function of the texture angle. A reasonable agreement between the experimental values and those obtained theoretically by using the elastic constants in Table I is observed. Figure 2 illustrates the sample-slicing plan employed to prepare the textured, 8-mm-thick plates. The texture direction coincides with the direction labeled ''z'' in Fig. 2. The raw material was cut in the $Y-Z$-plane into samples from 0 to 90 deg at 10 deg increments.

II. EXPERIMENTAL SETUP AND DATA ACQUISITION

A schematic representation of the experimental setup used in this work is shown in Fig. 3. Similar experimental setups have been used to characterize the ultrasonic fields generated by immersion transducers.[10–12] In this setup a flat, $\frac{1}{2}$-in.-diameter transducer operating at a nominal frequency of 2.25 MHz is used to generate an ultrasonic beam in water. A hydrophone of a needle-type design with a 0.635-mm diameter is used as a receiver. The position of the hydrophone can be changed in a plane that is parallel to the transmitter face and in the direction normal to it by using a three-axial motorized positioning stage. Each axis of the stage can cover a maximum distance of 200 mm, by steps of 1 μm, and with repeatability of 1 μm. The stage is connected to a computer-controlled motor controller.

The experimental approach used in this work maps the ultrasonic field of the flat emitter twice by recording the incident pressure field first at the location where the front face

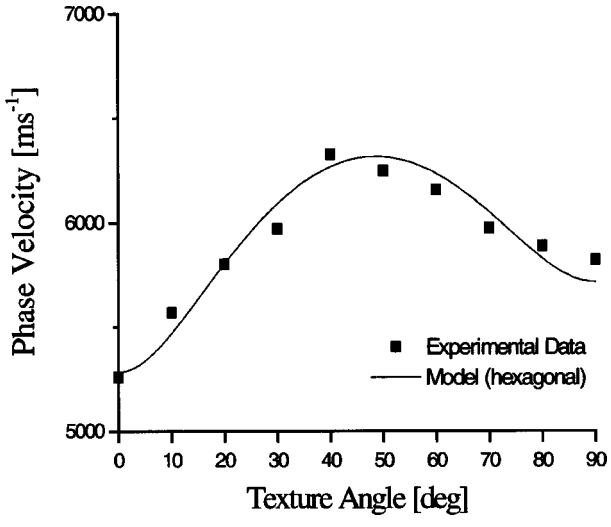


FIG. 1. Comparison between longitudinal phase velocity calculated by using the elastic constants of Table I and experimental data.

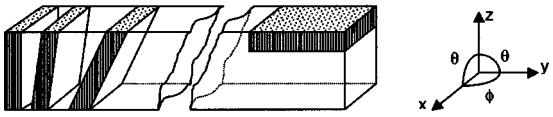


FIG. 2. Sample slicing plan.

of the sample is placed, and then the transmitted pressure field in the proximity of the sample's back face. Each map of the beam consists of a matrix of 77×77 observation points. At each point the *A*-scan is averaged 100 times, sampled at 25 MS/s, and digitized with at most 16 bits using a Tektronix TDS 524 digital oscilloscope. Each digitized signal consists of 512 time samples. First neighboring points are separated by a distance equal to the linear dimension of the hydrophone, which is 0.635 mm. The scanned area is approximately 18 times larger than the transducer's surface area. By placing the sample at a distance of about 1.2 near-field lengths calculated at a frequency of 2.25 MHz, this configuration allows the energy carried by the beam to be recovered almost entirely. This feature of the scanning technique makes corrections for beam diffraction and beam steering unnecessary. In this experimental setup, a second transducer is simultaneously used to monitor the water temperature variations during the measurement. Velocity variations caused by changes of the water temperature are incorporated into the signal-processing software.

III. BEAM ATTENUATION

Two distinct data-processing procedures are applied to the recorded maps of the pressure field. The first approach simulates a conventional technique used to measure ultrasonic attenuation. The second procedure is developed to account only for the energy loss due to ultrasonic scattering occurring within the samples. Given the lateral dimension of the hydrophone used in the second method, which is of the order of the wavelength of the longitudinal wave in water, this approach is expected to be sensitive only to variations of the beam's phase occurring over a distance larger than the hydrophone diameter.

A. Simulated conventional approach

This procedure simulates a receiving transducer having the same area and the same axis as the transmitter. To this end, all the *A*-scans recorded by the hydrophone within the

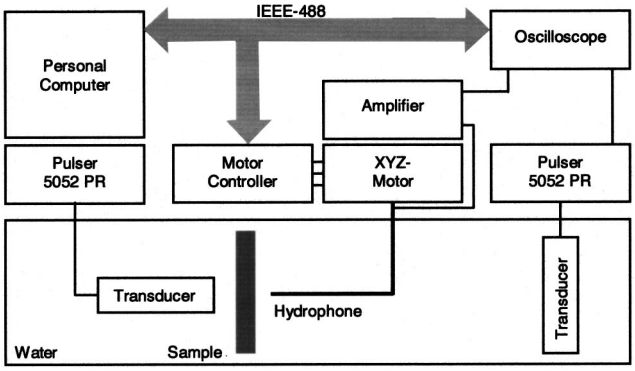


FIG. 3. Experimental setup for the scanning technique.

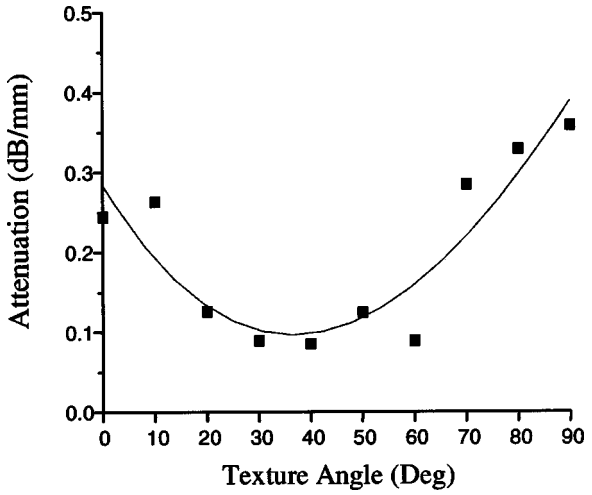


FIG. 4. Longitudinal wave attenuation coefficient vs texture angle at 1.7 MHz. These data have been obtained by means of a conventional method.

surface of the simulated receiver are integrated over the transducer's surface. The raw data are corrected for beam diffraction following a well-known procedure based on the analysis of Rogers and Van Buren.[13] Rogers and Van Buren's analysis is valid only for isotropic solids, and therefore is equivalent to the procedures that were adopted to obtain the results published in the literature[6] to the extent that it disregards additional effects due to the anisotropy of the medium. A second correction to the raw data accounts for the energy loss due to the beam transmission through the water–solid interface. For this purpose the transmission coefficient for a plane wave traveling along the direction normal to the interface is used. Energy losses due to mode conversion experienced by the beam components insonifying the sample at oblique incidence are, therefore, not taken into account. Finally, the longitudinal wave attenuation in dB/(unit length) can be obtained using the following equation;

$$\alpha_{\mathrm{conv}}(\omega,\theta)=\frac{20}{d}\left(\lg\left(\frac{\mathrm{cor}_{\mathrm{tr}}(\omega)}{\mathrm{cor}_{\mathrm{in}}(\omega)}\right)+\lg(T_{\mathrm{sw}}(\theta)T_{\mathrm{ws}}(\theta))+\lg\left(\frac{|S_{\mathrm{in}}(\omega)|}{|S_{\mathrm{tr}}(\omega,\theta)|}\right)\right), \tag{1}$$

where ω is the angular frequency, θ is the texture angle, that is the angle between the texture direction and the normal to the sample surface, and d is the sample thickness. The quantity ''cor'' is the diffraction correction; $T_{\mathrm{sw}}(\theta)$ and $T_{\mathrm{ws}}(\theta)$ are the normal incidence transmission coefficients for the particle velocity at the solid–water and water–solid interfaces, respectively. Finally, S is the signal amplitude that is proportional to the modulus of the particle velocity. The subscripts ''in'' and ''tr'' indicate the incident and transmitted signals, respectively.

Figure 4 illustrates the behavior of the estimated longitudinal wave attenuation coefficient at 1.7 MHz versus the texture angle, θ. In an attempt to compensate partially for beam steering caused by the anisotropy of the samples, the thickness of the samples was corrected by the factor $1/\cos(\delta)$, where δ is the skewing angle, the angle between the group velocity and the wave vector. Such a correction is not greater than 5 percent, that is, about half of the scattering of the experimental results. The data acquired using the cast samples exhibit a relative maximum near the grain growth direction and an absolute minimum at 40 deg of texture angle. These results, therefore, confirm the findings already published in the literature.

B. Novel approach

According to the second data-processing procedure, each *A*-scan $U(x,y,z,t)$ acquired by the hydrophone is Fourier-transformed into the angular frequency domain to obtain the spectrum of the incident and of the transmitted signals at each observation point

$$U_{\mathrm{in}}(x,y,z=0/\omega)=\int U_{\mathrm{in}}(x,y,z=0,t)e^{-j\omega t}\,dt, \tag{2a}$$

$$U_{\mathrm{tr}}(x,y,z=d/\omega,\theta)=\int U_{\mathrm{tr}}(x,y,z=d,t/\theta)e^{-j\omega t}\,dt, \tag{2b}$$

where the coordinates $z=0$ and $z=d$ indicate the locations of the planes containing the front and back faces of the sample, respectively. The symbol d represents the thickness of the sample. In this work, the field $U(x,y,z)$ is the modulus of the particle velocity that is proportional to the acoustic pressure. The transformation yields a frequency-dependent, complex field. Figures 5 and 6 illustrate examples of the amplitude and phase distribution of the incident and transmitted recorded signal, respectively, at a frequency of 1.7 MHz, while Figs. 7(a) and 7(b) present examples of the amplitude of the corresponding spectra. Note that the angular half-width of the main lobe of the amplitude spectra measures about 5 degs.

The transformed fields, $U_{\mathrm{in}}(x,y,z=0/\omega)$ and $U_{\mathrm{tr}}(x,y,z=d/\omega)$, can be further represented in terms of plane waves to obtain the corresponding two-dimensional spectra in the *k*-space domain

$$U_{\mathrm{in}}(\xi,\zeta/\omega)=\int\int U_{\mathrm{in}}(x,y,z=0/\omega)e^{j(\xi x+\zeta y)}dx\,dy, \tag{3a}$$

$$U_{\mathrm{tr}}(\xi,\zeta/\omega,\theta)=\int\int U_{\mathrm{tr}}(x,y,z=d/\omega,\theta)e^{j(\xi x+\zeta y)}dx\,dy, \tag{3b}$$

The quantities $U_{\mathrm{in}}(\xi,\zeta,\omega)$ and $U_{\mathrm{tr}}(\xi,\zeta,/\omega\theta)$ are the complex amplitudes of the incident and transmitted plane waves, respectively, having a wave vector $\mathbf{k}$, with components (ξ,ζ,κ), where $\kappa=((\omega/C_L)^2-\xi^2-\zeta^2)^{1/2}$, and C_L is the phase velocity of the longitudinal plane wave in the sample. The phase velocity, $C_L=C_L(\theta)$, is a function of the wave vector's direction, θ, if the medium is anisotropic.

A simple analysis of the propagation of a plane wave through an attenuating plate of thickness d leads to the following relationship between $U_{\mathrm{in}}(\xi,\zeta,/\omega)$ and $U_{\mathrm{tr}}(\xi,\zeta,/\omega\theta)$:

$$U_{\mathrm{tr}}(\xi,\zeta/\omega,\theta)=U_{\mathrm{in}}(\xi,\zeta/\omega)T_{\mathrm{ws}}(\xi,\zeta,\theta)T_{\mathrm{sw}}(\xi,\zeta,\theta)e^{-\alpha d}e^{-j\kappa d}. \tag{4}$$

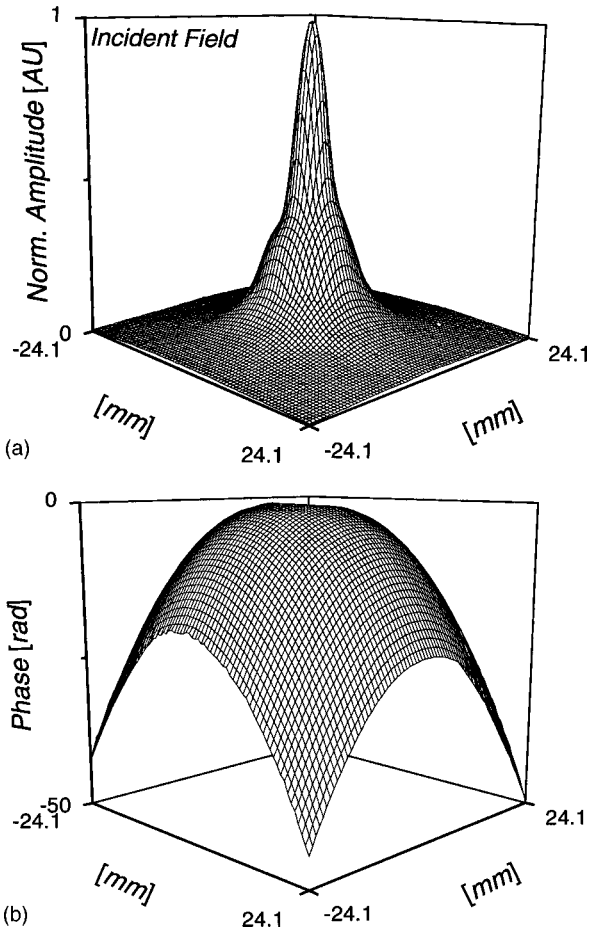


FIG. 5. Distribution of the pressure amplitude (a) and phase (b) at 1.7 MHz of the incident beam.

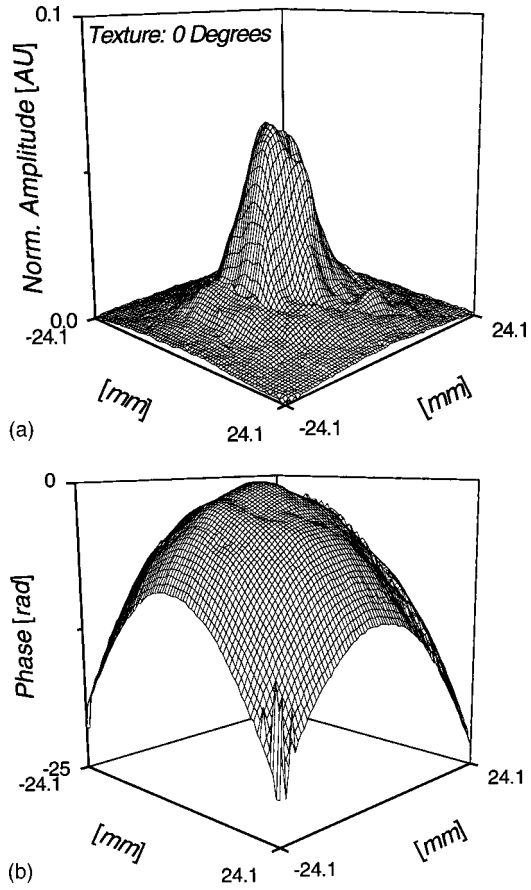


FIG. 6. Distributions of the pressure amplitude (a) and phase (b) at 1.7 MHz of the beam transmitted through a sample with texture direction of 0 deg.

In Eq. (4), $T_{ws}(\xi,\zeta,\theta)$ and $T_{sw}(\xi,\zeta,\theta)$ are the particle velocity transmission coefficients at the water–solid and solid–water interfaces, respectively. They account for the anisotropy of the solid medium and for mode conversion, and depend on the crystallographic orientation of the sample, θ. Since only propagating waves are considered here, such coefficients are real quantities. The coefficient $T_{ws}(\xi,\zeta\theta)$ is purely imaginary for those values of the components ξ,ζ corresponding to angles of incidence larger than the critical angle for longitudinal waves. Like the phase velocity, C_L, the plane wave attenuation coefficient α depends also on the direction of the wave vector, **k**, i.e., on the texture angle, θ. In addition, α is a function of the angular frequency, ω. The plane wave attenuation coefficient can be easily recovered from Eq. (4), and, in particular, at normal incidence, that is, for $\xi=0$, $\zeta=0$, the attenuation coefficient is given by

$$\alpha(\omega,\theta)=\frac{20}{d}\left(\lg\left(\frac{|U_{in}(0,0/\omega)|}{|U_{tr}(0,0/\omega,\theta)|}\right)+\lg(T_{ws}(0,0,\theta)T_{sw}(0,0,\theta))\right). \tag{5}$$

This data-processing procedure was tested on computer-simulated data obtained by sampling a simulated beam over the same area and at the same points used in the actual measurements. In the range between 1 and 6 near-field lengths at 2.25 MHz, fluctuations of a few percent were observed in the transmitted amplitude. Such amplitude fluctuations that may be attributed to the limited extension of the scanned area determine relative variations of the estimated attenuation coefficient larger than 10 percent. It was found, on the other hand, that the integral over the k space of the square of the modulus of the beam plane wave components is a very stable quantity with respect to the distance between the emitter and the observation plane. Such a quantity, proportional to the energy carried by the beam and insensitive to the phase relationships among the beam components, can be used to estimate the amount of energy lost by the beam while crossing the attenuating slab. Therefore, if E_{in} and E_{tr} are defined by

$$E_{in}(\omega)=\int\int |U_{in}(\xi,\zeta/\omega)|^2 d\xi\, d\zeta, \tag{6a}$$

$$E_{tr}(\omega,\theta)=\int\int |U_{tr}(\xi,\zeta/\omega,\theta)|^2 d\xi\, d\zeta, \tag{6b}$$

the beam attenuation coefficient can be introduced as follows:

$$\alpha_{beam}(\omega,\theta)=\frac{10}{d}\lg\left(\frac{E'_{in}(\omega)}{E_{tr}(\omega,\theta)}\right), \tag{7}$$

where E'_{in} is defined by

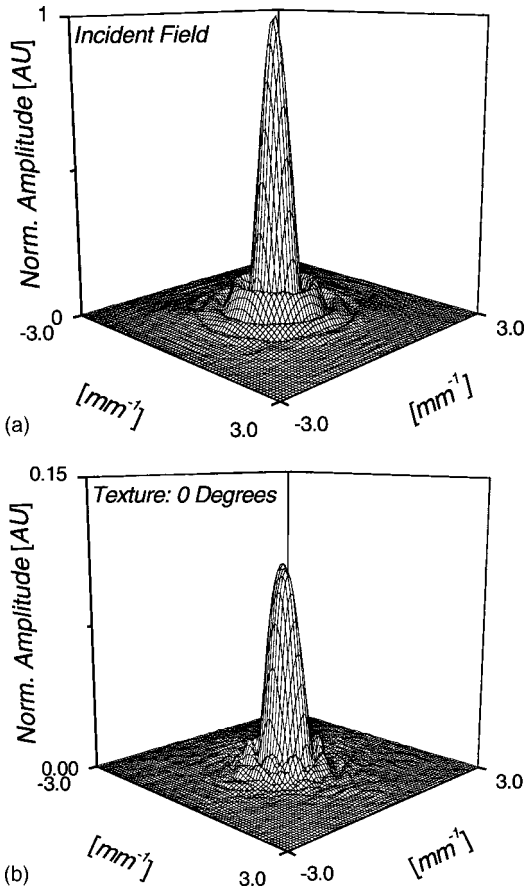


FIG. 7. Distribution of the spectrum amplitude at 1.7 MHz of the incident beam (a), and of the beam transmitted through a sample with texture direction of 0 deg (b).

$$E'_{\text{in}}(\omega,\theta) = \int\int |U_{\text{in}}(\xi,\zeta/\omega)|^2 T_{\text{ws}}^2(\xi,\zeta,\theta) T_{\text{sw}}^2(\xi,\zeta,\theta)\, d\xi\, d\zeta. \quad (8)$$

In Eq. (8), the energy carried by each component of the incident field is corrected for the losses occurring at the two interfaces by multiplying it by the product of the two velocity transmission coefficients. In this way, the beam attenuation evaluated according to Eq. (7) accounts for the energy losses of the coherent beam within the sample. Note that quantities $E_{\text{in}}(\omega)$ and $E_{\text{tr}}(\omega\theta)$ in Eqs. (6a) and (6b) may also be calculated by integrating the square of the amplitude of the space-domain signal over the scanned area. However, if such an approach were adopted, the possibility to correct the data for the energy loss due to the transmission at the water–sample interface in a way that accounts for the anisotropy of the sample would be lost. Since the analysis of the beam propagation is carried out in terms of plane wave components, the use of the transmission coefficients of each plane wave seems to be the natural way to account for the energy loss caused by the impedance mismatch at the water–solid and solid–water interfaces. Finally, the integration is carried out over the spatial frequency range within which $T_{\text{ws}}(\xi,\zeta,\theta)$ is real.

The definition of the attenuation coefficient of a plane wave [Eq. (5)] can be recovered from that of a beam [Eq. (7)] if the plane wave spectrum of the beam is reduced to a delta function. Like the plane wave attenuation coefficient, α_{beam} is a function of the texture angle (see Fig. 8). Note that the relative maximum in the neighborhood of the texture direction is no longer present, in agreement with the predictions of the theoretical models. It is also important to stress that α_{beam} is the result of a weighted-average process over the attenuation coefficients of the plane wave components of the beam. Since the angular half-width of the main lobe of the beam does not exceed 10 deg in steel, the weighting factor favors the component at normal incidence. Therefore, α_{beam} is expected to be rather close to the attenuation coefficient of a plane wave.

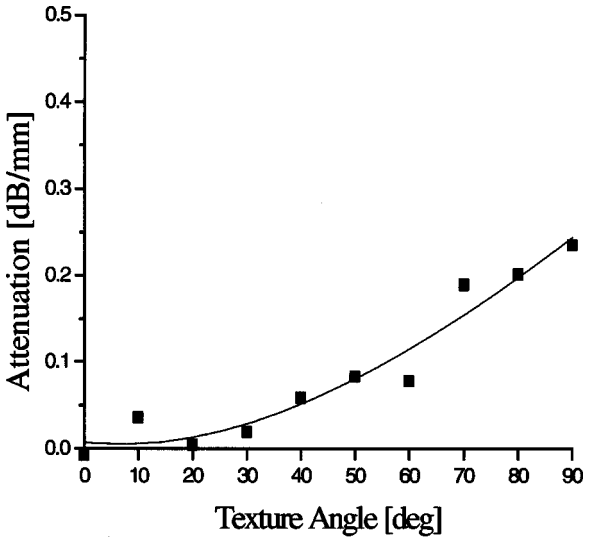


FIG. 8. Longitudinal beam attenuation coefficient versus texture angle.

A final remark is in order to point out that the results obtained by using the approach just described may be affected by the incoherent field that is scattered within the spatial frequency range over which the integration in Eq. (8) is carried out. In fact, the modeling of the beam propagation through the sample presented here does not account for the incoherent waves that may be generated by the interaction of the coherent beam with the sample microstructure. However, as it will be shown in the next section, such a phenomenon seems to be negligible in the case of the samples examined in this work.

IV. DISCUSSION

An understanding of the difference between the theoretical attenuation coefficient and the data presented in Fig. 4 can be achieved by analyzing the conventional measurement process. Let $U_{\text{in}}(x,y,z=0,t)$ be the incident field at the front face of the sample. If a transducer with a finite radius a is used as a receiver, the spectrum of the recorded signal is given by

$$U_{\text{in}}^{\text{rec}}(\omega) = C\int\int U_{\text{in}}(\xi,\zeta/\omega) D(\xi,\zeta,a)\, d\xi\, d\zeta, \quad (9)$$

where C is a constant that depends on the area of the receiving transducer, of which $D(\xi\zeta a)$ is the directivity function. Similarly, the spectrum of the signal received by the same transducer at the back face of the sample is

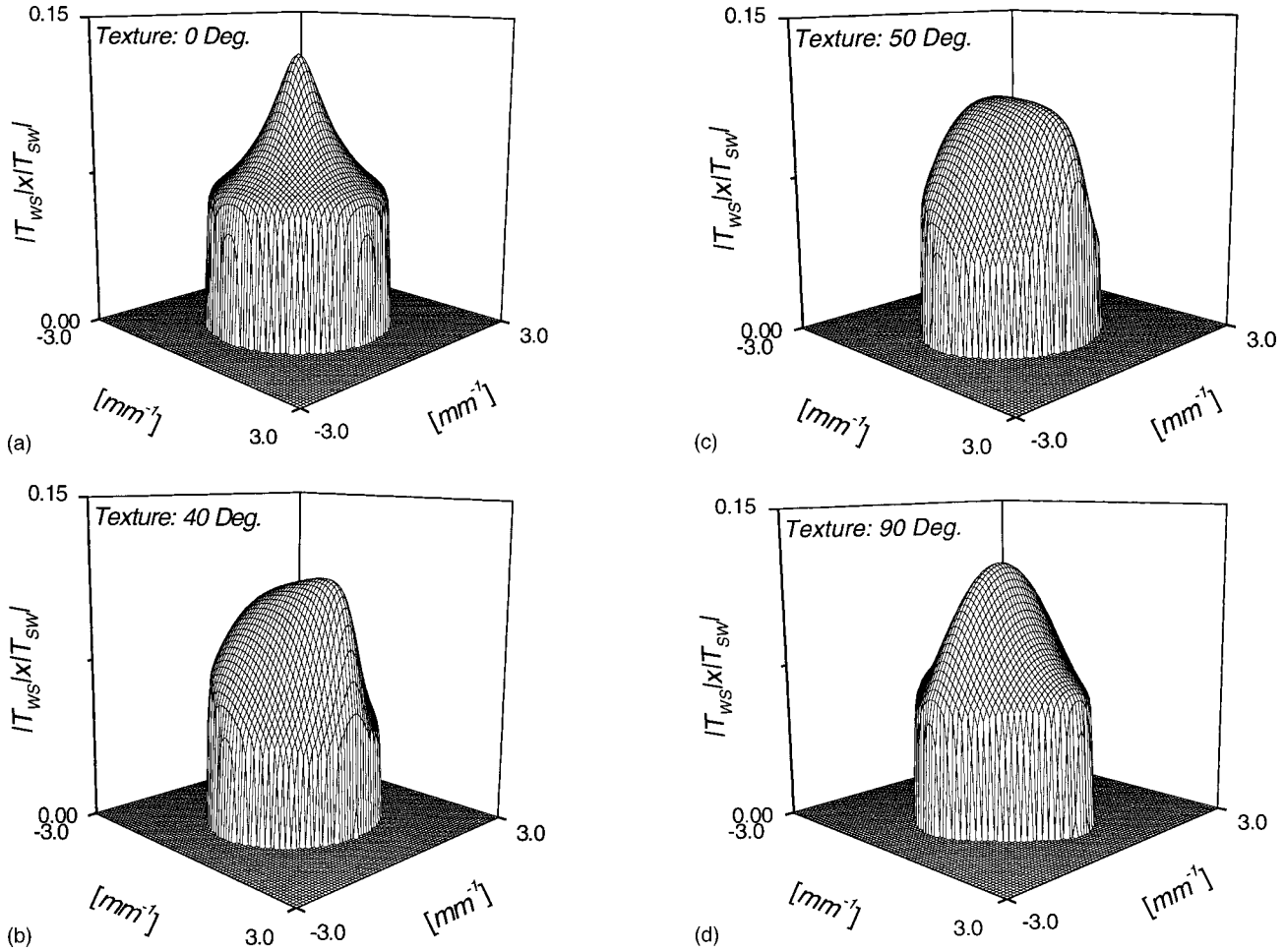


FIG. 9. Product of the transmission coefficients, $T_{ws}(\xi,\zeta,\theta)T_{sw}(\xi,\zeta,\theta)$, vs the spatial frequency for four values of the texture angle: (a) 0 deg, (b) 40 deg, (c) 50 deg, (d) 90 deg.

$$U_{tr}^{rec}(\omega,\theta)=C\int\int U_{in}(\xi,\zeta/\omega)T_{ws}(\xi,\zeta,\theta)T_{sw}(\xi,\zeta,\theta)\times e^{-j\kappa d}e^{-\alpha d}D(\xi,\zeta,a)\,d\xi\,d\zeta. \quad (10)$$

If the spectrum of the beam could be approximated by a two-dimensional delta function, $\delta(\xi)\delta(\zeta)$, then Eqs. (9) and (10) would become

$$U_{in}^{rec}(\omega)\approx C' U_{in}(0,0/\omega), \quad (11a)$$

$$U_{tr}^{rec}(\omega,\theta)\approx C' U_{in}(0,0/\omega)T_{ws}(0,0,\theta)T_{sw}(0,0,\theta)e^{j\kappa_o d}e^{-\alpha_o d}, \quad (11b)$$

respectively, where C' is a new constant. In Eq. (11b), κ_o and α_o are the wave number and attenuation coefficient, respectively, of the longitudinal wave with a wave vector normal to the sample's surface. By using the modulus of the recorded spectra given by Eqs. (11a) and (11b), Eq. (5) provides the attenuation coefficient of a longitudinal plane wave.

When dealing with samples of rather limited extension, the measurements must be carried out at a distance such that the beam does not interact with the edges of the sample. In the case of the samples used in this work, such a distance was chosen to be 1.2 near-field lengths. In this region the inspecting beam can no longer be represented by a plane wave, and Eq. (5) becomes inadequate for two reasons.

The first reason concerns the inappropriate correction for the transmission at the sample–water interfaces. In fact, by using the transmission coefficients at normal incidence in Eq. (8) as a correction factor for all the spatial frequencies, the

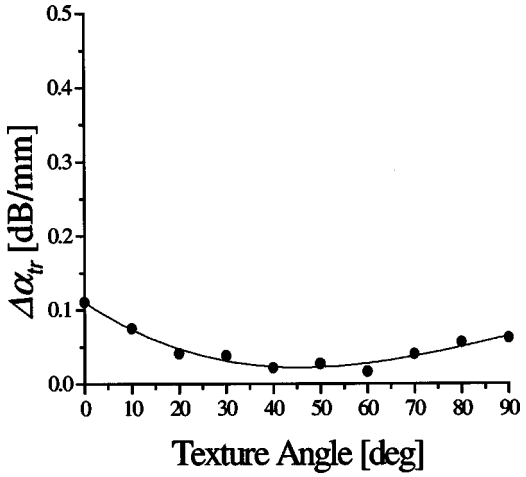


FIG. 10. Contribution to the ultrasonic attenuation excess due to the use in Eq. (8) of the product of the normal transmission coefficients, $T_{ws}(0,0,\theta)$ and $T_{sw}(0,0,\theta)$, in place of the correct quantities.

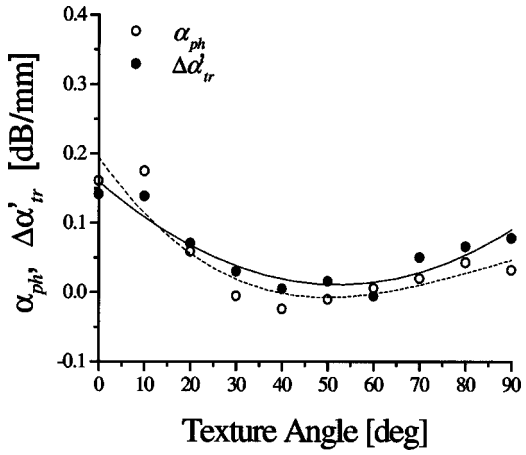


FIG. 11. Comparison between the contribution to the attenuation excess caused by the anisotropy of the phase velocity, α_{ph}, and the difference between the data of Figs. 4 and 10, $\Delta\alpha'_{\mathrm{tr}}$.

energy loss due to wave transmission across the water–solid interfaces is overestimated. Figure 9(a)–9(d) illustrates examples of the product $T_{\mathrm{ws}}(\xi,\zeta,\theta)T_{\mathrm{sw}}(\xi,\zeta,\theta)$. These plots should be compared with that of the flat-top weighting function, defined as

$$F(\xi,\zeta,\theta)=\begin{cases} T_{\mathrm{ws}}(0,0,\theta)T_{\mathrm{sw}}(0,0,\theta) & (\xi^2+\zeta^2)^{1/2}\leqslant\kappa_{L,\mathrm{cr}} \\ 0 \quad (\xi^2+\zeta^2)^{1/2}>\kappa_{L,\mathrm{cr}} & \end{cases}, \tag{12}$$

where $\kappa_{L,\mathrm{cr}}$ is the critical value of the parallel component of the wave vector above which the transmitted wave becomes evanescent. Figure 9 shows that the actual correction factor resembles a flat-top function better around a texture angle of 45 deg than at 0 deg and 90 deg from the texture direction. In fact, as ξ and/or ζ increases, the product $T_{\mathrm{ws}}(\xi,\zeta,\theta)T_{\mathrm{sw}}(\xi,\zeta,\theta)$ decreases more rapidly at 0 deg and 90 deg than in other cases. Therefore, the use of Eq. (12) in Eq. (8) leads to an overestimate of the energy loss, $\Delta\alpha_{\mathrm{tr}}$, which is minimal at texture angles about 45 deg, and reaches its maximum along the texture direction and in any direction normal to it. Such an overestimation is illustrated in Fig. 10, where the difference between the attenuation values obtained using Eq. (12) in Eq. (8) and the data of Fig. 8 are presented.

The second cause for the discrepancy between theory and experimental data obtained by means of conventional methods resides in the corrections for beam diffraction that now must include the effects of anisotropy. In fact, as illustrated in Eq. (10), the phase of each plane wave component depends, through the term κd, on the phase velocity, C_L, which, as pointed out earlier, is a function of the direction of the wave vector, **k**. Therefore, when the medium is anisotropic, the phase relationship among the plane wave components of the beam may differ from that which would occur if the medium were isotropic. Furthermore, if the emitting and receiving transducers are placed coaxially, beam steering due to the nonparallelism between the wave and the Poynting vectors, which generally occurs in anisotropic media, should be accounted for. *Ad hoc* diffraction corrections are, thus, required. In his review of the subject, Papadakis[14] presents correction curves for energy loss due to diffraction of a beam propagating in anisotropic media. Such curves are valid for a beam propagating along axes of symmetry, and, therefore, cannot be used in more general situations like those considered here. The effect of neglecting these additional wave phenomena occurring to the inspecting beam because of material anisotropy is assessed as follows. Two new signals are considered. The first one is the superposition of plane waves propagating with phase velocity determined by the properties of the anisotropic medium of interest as given in Eq. (10), while the second signal is formed by plane waves that propagate with the phase velocity of the central component of the beam in a fictitious isotropic medium. The attenuation of the second signal with respect to the first one, α_{ph}, is illustrated in Fig. 11, and is caused by the extra phase variation introduced by the material anisotropy. The same figure also reports the difference, $\Delta\alpha'_{\mathrm{tr}}$, between the data of Fig. 4 and those obtained by using Eqs. (7) and (8), in which only the transmission coefficients at normal incidence have been used to correct for the loss caused by the impedance mismatch. The substantial similarity of the two series of data proves that the remaining excess of attenuation, $\Delta\alpha'_{\mathrm{tr}}$, measured by conventional methods is caused by the additional destructive interference caused by the anisotropic structure of the medium. Furthermore, it shows that the attenuation induced by the additional phase modulation of the wavefront is caused almost entirely by the sample texture, rather than by the material microstructure, at least for the samples considered in this work.

V. SUMMARY

The scattering-induced attenuation coefficient of a beam of longitudinal plane waves traveling through an austenitic steel plate has been measured as a function of the texture direction of the inspected material. The experimental data have been obtained by evaluating the energy loss due to grain scattering and wave transmission through the water–solid interfaces experienced by each plane wave component of the beam. These results are in qualitative agreement with the theoretical predictions of several models. The analysis presented here also shows that the phase distortion caused by the interaction of the beam with the texture of the material contributes considerably to the attenuation of the wave when measured by conventional methods. Finally, although the attenuation coefficient of the beam is a weighted average over the attenuation coefficients of the beam components, its values and behavior closely approximate those of the attenuation coefficient of a plane wave provided that the beam is generated by a large aperture ($ka\gg1$) transducer.

ACKNOWLEDGMENTS

This work was carried out as part of the SMT project ''Effect of ultrasonic scattering on inspection of austenitic welds,'' Contract Number SMT4-CT95-2012, sponsored by the DG XII of the Commission of the European Communities, and constitutes the core of T.S.'s doctoral thesis to be presented at the Technische Universität zu Berlin, Germany.

T. S. would like to express his gratitude to the Joint Research Center of the European Commission for the fellowship awarded to him. The authors would also like to thank Dr. E. Neumann for providing the samples used in this work, Professor H. Wüstenberg for the useful discussions, and all the partners in this project for their continuous support.

[1] F. E. Stanke and G. S. Kino, ''A unified theory for elastic wave propagation in polycrystalline materials,'' J. Acoust. Soc. Am. **75**, 664–681 (1983).
[2] S. Hirsekorn, ''Directional dependence of ultrasonic propagation in textured polycrystals,'' J. Acoust. Soc. Am. **79**, 1269–1279 (1986).
[3] S. Ahmed and R. B. Thompson, ''Propagation of elastic waves in equiaxed stainless-steel polycrystals with aligned [001] axes,'' J. Acoust. Soc. Am. **99**, 2086–2096 (1996).
[4] V. K. Munikoti and E. Neumann, ''Extension of Ahmed & Thompson theory to general elastic plane quasi-wave propagation in textured polycrystalline material,'' in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1997), Vol. 17, pp. 1657–1664.
[5] E. Neumann, M. Bieth, P. Krarup, F. Hardie, and E.-B. Pers-Anderson, ''Effect of ultrasonic scattering on inspection of austenitic welds,'' in *Proceedings of the First International Conference on NDE in Relation to Structural Integrity for Nuclear and Pressurized Components*, edited by M. Bieth and J. L. Monjaret (Woodhead, Cambridge, 1999), Vol. 2, pp. 758–768.
[6] B. L. Baikie, A. R. Wagg, M. J. Whittle, and D. Yapp, ''Ultrasonic inspection of austenitic welds,'' J. Br. Nucl. Energy Soc. **15**(3), 257–261 (1976).
[7] E. Neumann, *Ultraschallprüfung von austenitischen Plattierungen, Mischnähten und austenitischen Schweißnähten* (Expert, Renningen-Malmshein, 1995).
[8] P. D. Panetta, F. J. Margetan, I. Yalda, and R. B. Thompson, ''Ultrasonic attenuation measurements in jet-engine titanium alloys,'' in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1996), Vol. 15B, pp. 1525–1532.
[9] P. D. Panetta, F. J. Margetan, I. Yalda, and R. B. Thompson, ''Observation and interpretation of microstructurally induced fluctuations of back-surface signals and ultrasonic attenuation in titanium alloys,'' in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1997), Vol. 16B, pp. 1545–1554.
[10] F. P. Higgins, S. J. Norton, and M. Linzer, ''Optical interferometric visualization and computerized reconstruction of ultrasonic fields,'' J. Acoust. Soc. Am. **68**, 1169–1176 (1980).
[11] M. E. Schafer and P. A. Lewin, ''A computerized system for measuring the acoustic output from diagnostic ultrasonic equipment,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**(2), 102–109 (1988).
[12] M. E. Schafer and P. A. Lewin, ''Transducer characterization using the angular spectrum method,'' J. Acoust. Soc. Am. **85**, 2202–2214 (1989).
[13] P. H. Rogers and A. L. Van Buren, ''An exact expression for the Lommel diffraction correction integral,'' J. Acoust. Soc. Am. **55**, 724–728 (1974).
[14] E. P. Papadakis, ''Ultrasonic diffraction from single apertures with application to pulse measurements and crystal physics,'' in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1975), Vol. XI, pp. 151–211.

The Huygens entrainment phenomenon and thermoacoustic engines

P. S. Spoor[a] and G. W. Swift
Condensed Matter and Thermal Physics Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 14 September 1999; accepted for publication 1 May 2000)

The earliest known reference to the mode-locking, or entrainment, of two maintained oscillators is Christiaan Huygens' description of two pendulum clocks ''falling into synchrony'' when hung on the same wall. We describe an analogous phenomenon in acoustics—the mode-locking of two thermoacoustic engines which have their cases rigidly welded together, but which are otherwise uncoupled. This ''mass-coupling'' might compete with acoustic coupling when the latter is used to enforce antiphase mode-locking in such engines, for vibration cancellation. A simple theory relating the phase difference between the engines in the locked state to the corresponding ratio of their pressure amplitudes is in excellent agreement with theory and numerical simulations. The theory's prediction relating the phase difference to the engines' natural frequency difference is qualitatively confirmed by experiment, despite larger experimental uncertainties. The mass coupling is relatively weak compared to the aforementioned acoustic coupling, and in general occurs in antiphase, so we conclude that mass coupling will not interfere with vibration cancellation by acoustic coupling in most circumstances. © *2000 Acoustical Society of America.* [S0001-4966(00)01708-2]

PACS numbers: 43.35.Ud, 43.40.Vn [HEB]

INTRODUCTION

Entrainment, or mode-locking, whereby two or more self-maintained oscillators lock together in frequency and phase, exists in a fascinating array of physical and biological systems. The earliest known reference to this phenomenon in the annals of science was made by Christiaan Huygens, who wrote to his father in a letter dated 26 February 1665:[1]

> . . . I have noted an impressive effect which no one has yet been able to explain. This is that two clocks, hanging side by side and separated by one or two feet, keep between them a consonance so exact that the two pendula always strike together, never varying.

In this work we describe the equivalent of the Huygens phenomenon for acoustic resonators. Two cylindrical-duct resonators are welded together side by side, the oscillating gas inside the resonators acting the part of the clock pendula and the duct cases acting the part of the wall upon which the clocks were hung. The resonators exhibit mode-locking in the same way that Huygens' clocks synchronized their motion. The oscillations in the ducts are maintained thermoacoustically, although to observe the Huygens phenomenon, any method of producing spontaneous oscillations will do.

A thermoacoustic engine of a size and power density suitable for heavy industrial applications (such as large-scale natural gas liquefaction[2]) experiences tons of axial dynamic force on its structure from the pressure oscillations of the gas inside it, potentially inducing intense, damaging vibrations in this structure. In a separate paper[3] we show how these vibrations can be canceled by rigidly attaching two engines together, side by side, and acoustically coupling two neighboring ends through a narrow tube a half-wavelength long, thus encouraging the oscillations in the two engines to be in antiphase. If the engines are attached rigidly enough to move as a unit, however, they are of necessity coupled through their mutual structure as well as through the acoustic tube. Hence, our particular goal in the present work was to discover how efforts to cancel external case vibrations in two rigidly attached thermoacoustic engines, by coupling them acoustically, might be helped or hindered by the Huygens-style structural coupling. We refer to this coupling as ''mass coupling,'' since the coupling occurs through forces on a shared mass (the case) and is stronger if the oscillating gas is heavy and the case is light. Therefore, to increase the mass coupling we used a pressurized, heavy gas—xenon—in the lightly constructed engines used for studies of acoustic coupling.[3] Without acoustic coupling, we found that the engines would tend to mass-lock in antiphase, as desired for vibration cancellation. For the range of mass-locked states we could achieve, we measured the dependence of the amplitude ratio of the two engines and the phase difference between them as a function of their natural frequency difference. Even with 3 bars of xenon gas in our engines, the mass-coupling was relatively weak; locked states existed only over a very narrow ''capture bandwidth'' of frequency differences (fractional differences of order 0.001). This article provides a quantitative explanation of these measurements, through both theory and numerical simulations.

I. ESSENTIAL DYNAMICS OF MODE-LOCKING

Two self-maintained oscillators, such as pendulum clocks, organ pipes, or thermoacoustic engines, may alter each other's frequencies enough to lock in frequency and phase if they are sufficiently coupled. When the ''detuning'' (the difference in natural uncoupled frequencies) is large, or

[a]Present address: Clever Fellows Innovation Consortium, 302 10th St., Troy, NY 12180.

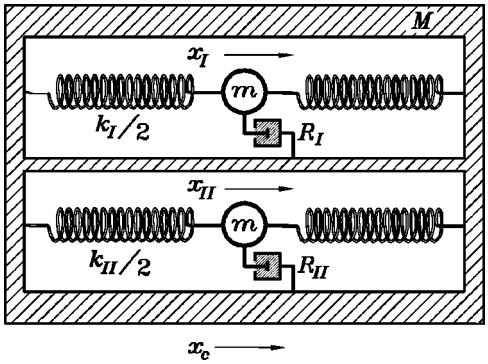


FIG. 1. Mass–spring model of mass-coupled oscillators.

the coupling is weak, the oscillators ''beat'' at the difference frequency, just like coupled linear passive oscillators. As the detuning is decreased, or as the coupling is increased, the beating slows down and eventually stops, at which point the oscillators are locked. The change in frequency of one oscillator depends on its phase relative to the other, so the locked state at a particular detuning is characterized by a corresponding phase difference between the oscillators. In general, the amplitudes in the oscillators also depend on this phase difference, and there may be a net flow of energy from one oscillator to the other if the amplitudes are not equal. In an intermediate state, where the coupling is not quite strong enough to lock the oscillators, they beat with a nonsinusoidal envelope—the beat period elongates when the frequencies have a close approach.

We use the following notation for two oscillators that are mode-locked or nearly mode-locked:

$$\psi_{\mathrm{I}}(t)=\Re\{\Psi_{\mathrm{I}}(t)e^{i[\omega t+\phi_{\mathrm{I}}(t)]}\}, \tag{1}$$

$$\psi_{\mathrm{II}}(t)=\Re\{\Psi_{\mathrm{II}}(t)e^{i[\omega t+\phi_{\mathrm{II}}(t)]}\}, \tag{2}$$

where $\Re\{\}$ indicates taking the real part. Here $\psi_{\mathrm{I,II}}$ could be displacements, pressures, velocities, or any other similar variables of interest that depend on time t. We assume that the amplitudes $\Psi_{\mathrm{I,II}}$ and the phases $\phi_{\mathrm{I,II}}$ are real and slowly varying compared to the angular frequency ω, and $\Psi_{\mathrm{I,II}}\geqslant 0$. The locked state is characterized by a relative phase between the oscillators and a ratio of their amplitudes, which are

$$\phi=\phi_{\mathrm{II}}-\phi_{\mathrm{I}}, \quad \zeta=\frac{\Psi_{\mathrm{II}}}{\Psi_{\mathrm{I}}}. \tag{3}$$

We assume that oscillator I has natural frequency ω_{I} and oscillator II has ω_{II} when I and II are uncoupled, so that the difference in natural frequencies, or detuning, is

$$\Delta\omega=\omega_{\mathrm{II}}-\omega_{\mathrm{I}}. \tag{4}$$

We let the compromise frequency ω that the oscillators possess when they are coupled be a constant in time; the apparent *instantaneous* difference in frequency between resonators II and I is simply $\Delta\omega_{\mathrm{inst.}}=\partial\phi/\partial t$.

The obvious mechanism for frequency shift in our system is the moving case. If a simple harmonic oscillator of mass m and stiffness k is anchored to an unfettered mass M, the natural frequency is shifted according to

$$\omega_0^2=\frac{k}{m}\left(1+\frac{m}{M}\right). \tag{5}$$

For a straight cylindrical half-wave acoustic resonator whose case of mass M is free to move, the corresponding expression is [see Sec. II B, especially Eqs. (32) and (36)]

$$\omega_0^2\simeq\frac{\pi^2a^2}{l^2}\left[1+\left(\frac{8}{\pi^2}\right)\frac{m_g}{M}\right], \quad m_g\ll M, \tag{6}$$

where $\pi a/l$ is the natural angular frequency of a half-wave resonator with sound speed a and length l, and m_g is the total mass of gas in the resonator. Thus one might conclude that the effective mass of the oscillating gas in a half-wave resonator is $m=(8/\pi^2)m_g\simeq 0.8m_g$ in the present context.

II. THEORY

A. Mass-and-spring model

Before discussing coupled acoustic engines, we will consider a simple system of masses and springs, arranged to resemble two mass-coupled acoustic resonators. Figure 1 shows a pair of mass–spring oscillators that attach to the same moving mass. The oscillators are identical except for a slight difference in frequency; they have masses m and springs k_{I} and k_{II}, hence also natural frequencies $\omega_{\mathrm{I}}=\sqrt{k_{\mathrm{I}}/m}$ and $\omega_{\mathrm{II}}=\sqrt{k_{\mathrm{II}}/m}$. Following Pippard[4] and Van der Pol,[5] the feedback mechanism that maintains the oscillations is modeled by a resistance that can be either positive or negative, depending on the amplitude. We let

$$R_{\mathrm{I}}=\frac{m\omega_{\mathrm{I}}}{Q}\left(1-\frac{X_0^n}{X_{\mathrm{I}}^n}\right), \tag{7}$$

where X_{I} is the amplitude in oscillator I and X_0 is the steady-state amplitude to which the oscillator relaxes. Recall from Sec. I that the amplitude X_{I} is assumed to vary slowly with time compared to ω. The equation of motion for oscillator I in the absence of coupling is then

$$\ddot{x}_{\mathrm{I}}+\omega_{\mathrm{I}}^2x_{\mathrm{I}}+\frac{\omega_{\mathrm{I}}}{Q}\left(1-\frac{X_0^n}{X_{\mathrm{I}}^n}\right)\dot{x}_{\mathrm{I}}=0, \tag{8}$$

where $x_{\mathrm{I}}(t)$ is the displacement of the mass m from equilibrium, and the dot indicates a time derivative. We use this particular form because if X_0 is set to zero, one obtains the equation for an ordinary damped oscillator, with Q akin to the quality factor of the oscillator.

The exponent n depends on the mechanism maintaining the oscillations. Here, we argue that a thermoacoustic engine is governed by $n=2$ when comparing steady-state amplitudes at different operating points. In the mass-and-spring model, the power delivered or absorbed by the oscillator (the ''load'') is $\dot{W}_L=-(m\omega/Q)(1-X_0^n/X^n)|\dot{x}|^2$. In a thermoacoustic engine, such things as resonator dissipation and stack power are also proportional to amplitude squared,[6] so the power produced by the stack is $\dot{W}_{\mathrm{stack}}=\beta P_0^2=\beta P^2+\dot{W}_L$, where P_0 is the unloaded pressure amplitude and P is the loaded amplitude. Then we have $\dot{W}_L=\beta P_0^2-\beta P^2$, or $\dot{W}_L=-\beta(1-P_0^2/P^2)P^2$.

In practice, we find that only the ratio Q/n, and not Q or n separately, can be measured experimentally (and therefore, for our purposes, Q/n is all that matters). In principle, one should be able to determine both Q and n from transient measurements; however, the thermal mass of a thermoacoustic engine causes the growth and decay time constants to be artificially long. The ratio Q/n that governs the steady state can be measured by comparing states at various loads, but if we consider only small loads (i.e., $P \approx P_0$, or $X \approx X_0$) such as the engines experience when they are coupled, it is difficult to distinguish Q or n separately.

To fully analyze the dynamics of the system, we could write down the equations of motion with full time dependence, carry out the time derivatives in these equations, with $x(t)$ having the form suggested by Eqs. (1) and (2), and find expressions for $\partial\phi/\partial t$ and $\partial\zeta/\partial t$. Pippard[7] treats a problem very similar to ours in this manner. However, we are interested mostly in locked states, where everything is in the steady state. We seek expressions for the state variables ϕ and ζ in terms of known or measurable quantities like m, M, Q, and $\Delta\omega$, and an expression for the ''sensitivity'' of the system, $\partial\phi/\partial\Delta\omega$. This last quantity is of particular interest in the vibration-cancellation context, because it tells how fast a system tends to move away from some ideal phase (e.g., $\phi=\pi$) as it is detuned; when considering a mode-locked pair of acoustic engines, a movement of phase away from $\phi=\pi$ means more case vibration.

One may readily show, as has Pippard,[7] that a reactively coupled pair of oscillators like ours has two possible locked states, one nominally near $\phi=0$ and one nominally near $\phi=\pi$. Which one is actually selected is of some interest; a tendency for our coupled engines to mass-lock at $\phi=0$, where both ''masses'' are moving in the same direction, could impair vibration cancellation. Fortunately, we may take some comfort in the words of Huygens[1] describing his mode-locked pendulum clocks:

> . . . the motion of the pendula, while they are in consonance, are not in phase, but they approach and fall back with contrary motion . . .

Due to external friction on our moving case, we expect the mode with the least case motion to be favored. In a passive, linear system, the modes that experience more damping tend to decay faster, and we expect similar behavior in our system, although it is neither passive nor strictly linear. In any event, like Huygens, we nearly always observe that our engines mass-lock in antiphase rather than in-phase. [If one extracts the mechanisms from two metronomes and mounts them on a very light balsa-wood or perforated-aluminum frame, and suspends this frame so it is free to swing, the Huygens phenomenon is readily observable. One of us (PS) has experimented with such a system, observing both in-phase and antiphase locking of the metronomes, the former especially if the two metronome pendula are initially started out with in-phase motion.]

To proceed with our analysis, then, we will let $d/dt = i\omega$, and assume amplitudes and phases are constant in time. We obtain

$$(\omega_{\mathrm{I}}^2-\omega^2)X_{\mathrm{I}}+\frac{i\omega\omega_{\mathrm{I}}}{Q}\left(1-\frac{X_0^2}{X_{\mathrm{I}}^2}\right)X_{\mathrm{I}}=\omega_{\mathrm{I}}^2 e^{i(\phi_c-\phi_{\mathrm{I}})}X_c\,, \tag{9}$$

$$(\omega_{\mathrm{II}}^2-\omega^2)X_{\mathrm{II}}+\frac{i\omega\omega_{\mathrm{II}}}{Q}\left(1-\frac{X_0^2}{X_{\mathrm{II}}^2}\right)X_{\mathrm{II}}=\omega_{\mathrm{II}}^2 e^{i(\phi_c-\phi_{\mathrm{II}})}X_c\,, \tag{10}$$

$$-\omega^2 X_c=\mu\omega_{\mathrm{I}}^2(e^{i(\phi_{\mathrm{I}}-\phi_c)}X_{\mathrm{I}}-X_c)+\mu\omega_{\mathrm{II}}^2(e^{i(\phi_{\mathrm{II}}-\phi_c)}X_{\mathrm{II}}-X_c), \tag{11}$$

where $\mu=m/M$. Referring to Fig. 1, note that X_c is the motion amplitude of the common mass, and hence the right sides of Eqs. (9) and (10) are coupling terms. Equation (11) can be solved for X_c; substituting this result in Eqs. (9) and (10) yields

$$(\omega_{\mathrm{I}}^2-\omega^2)+\frac{i\omega\omega_{\mathrm{I}}}{Q}\left(1-\frac{X_0^2}{X_{\mathrm{I}}^2}\right)=-\mu(\omega_{\mathrm{I}}^2+\omega_{\mathrm{II}}^2\,\zeta\, e^{i\phi})\,, \tag{12}$$

$$(\omega_{\mathrm{II}}^2-\omega^2)+\frac{i\omega\omega_{\mathrm{II}}}{Q}\left(1-\frac{X_0^2}{X_{\mathrm{II}}^2}\right)=-\mu(\omega_{\mathrm{II}}^2+\omega_{\mathrm{I}}^2 e^{-i\phi}/\zeta)\,. \tag{13}$$

We assume that $\mu \ll 1$ and $|1-X_0^2/X_{\mathrm{I,\,II}}^2| \ll 1$, and that $\omega_{\mathrm{II}} \sim \omega_{\mathrm{I}} \sim \omega$. Taking the real part of Eq. (13) minus Eq. (12), keeping terms to first order, and simplifying gives

$$\frac{\Delta\omega}{\omega}=\frac{\mu}{2}(\zeta-1/\zeta)\cos\phi\,. \tag{14}$$

To find a second equation, we follow Pippard[7] and assume that in the steady state, energy is conserved in the system, i.e.,

$$\frac{m\omega_{\mathrm{I}}}{Q}\left(\frac{X_0^2}{X_{\mathrm{I}}^2}-1\right)|\dot{x}_{\mathrm{I}}|^2+\frac{m\omega_{\mathrm{II}}}{Q}\left(\frac{X_0^2}{X_{\mathrm{II}}^2}-1\right)|\dot{x}_{\mathrm{II}}|^2=0\,. \tag{15}$$

[Equation (15) may also be obtained from the ratio of the imaginary parts of Eqs. (12) and (13).] After some algebra, this leads to the identity

$$\frac{X_0^2}{X_{\mathrm{I}}^2}-\frac{X_0^2}{X_{\mathrm{II}}^2}=\frac{1}{2}\left(\zeta^2-\frac{1}{\zeta^2}\right)\,. \tag{16}$$

Using this identity, with the imaginary part of Eq. (12) minus Eq. (13), yields

$$(\zeta-1/\zeta)=2\mu Q\sin\phi\,. \tag{17}$$

Equations (14) and (17) can be combined to eliminate ζ, to obtain

$$\frac{\Delta\omega}{\omega}=\mu^2 Q\sin\phi\cos\phi\,; \tag{18}$$

in the limit $\phi\to\pi$, this leads to

$$\frac{\omega}{\pi}\frac{\partial\phi}{\partial\Delta\omega}=\frac{1}{\pi\mu^2 Q}\,. \tag{19}$$

This last equation is particularly useful for comparing the present results with those for acoustically coupled engines.[3]

The results in this section are simple and roughly true for any mass-coupled pair of oscillators. In the next section, to obtain results specific to mass-coupled acoustic engines,

we will derive differential equations isomorphic with Eqs. (9)–(11) for the acoustic pressure in the engines and the forces on the case. We will derive equations valid for engines with resonators of arbitrary cross section, and then modify these for specific examples of interest.

B. Acoustic model

Consider two nearly identical acoustic engines, called engines I and II, which are rigidly attached side-by-side and have a combined case mass M. Each engine consists of a prime mover section, where heat is applied and the acoustic power to sustain the oscillations is generated, and a resonator section, which comprises the bulk of the volume and is mostly responsible for determining the frequency at which the engine runs. The engine cross section may vary with position, but the radius is always much smaller than the length l. The engines are also assumed to be supporting what is more or less a half-wavelength mode. The long axis is labeled x, and is positive in the same direction as positive displacements and positive forces. The prime movers are located near the $x=0$ end, and we assume the acoustic pressure at the $x=l$ end is known.

We assume that the acoustic pressure throughout each engine as a function of x and t can be written as a product of the pressure $p(t)$ at the $x=l$ end and a spatial weighting function $w(x)$, that is,

$$\text{pressure}_{\text{I}}(x,t)=w(x)p_{\text{I}}(t)\,, \tag{20}$$

$$\text{pressure}_{\text{II}}(x,t)=w(x)p_{\text{II}}(t)\,. \tag{21}$$

Furthermore, in the manner of Eqs. (1) and (2), we assume

$$p_{\text{I}}(t)=\Re\{P_{\text{I}}e^{i(\omega t+\phi_{\text{I}})}\}\,, \tag{22}$$

$$p_{\text{II}}(t)=\Re\{P_{\text{II}}e^{i(\omega t+\phi_{\text{II}})}\}. \tag{23}$$

Following Morse,[8] and guided by the results in Sec. II A, we write the inhomogeneous wave equation for engine I as

$$p_{\text{I}}\nabla^2 w-w\ddot{p}_{\text{I}}/a^2-\frac{\omega_{\text{I}}}{Q}\left(1-\frac{P_0^2}{P_{\text{I}}^2}\right)w\dot{p}_{\text{I}}/a^2=-\rho\dot{s}\,, \tag{24}$$

where a is the sound speed, ρ is the density, and the source density $s(x,t)=(1/V)\partial V/\partial t$ is the volume velocity per unit volume imposed by the moving boundaries. Multiplying both sides by $-wa^2$ and integrating over the volume leads to

$$\ddot{p}_{\text{I}}-\frac{\int_V a^2 w\nabla^2 w\,dV}{\int_V w^2\,dV}p_{\text{I}}+\frac{\omega_{\text{I}}}{Q}\left(1-\frac{P_0^2}{P_{\text{I}}^2}\right)\dot{p}_{\text{I}}=\rho a^2\frac{\int_V \dot{s}w\,dV}{\int_V w^2\,dV}. \tag{25}$$

(Note that ρa^2 is a constant, so it appears outside the integral.) This last equation has exactly the same form as Eq. (8), if the right-hand side is zero (no coupling), with the coefficient of p_{I} identified as ω_{I}^2. [In the limit that a is constant, and ∇w is zero at the walls, the vector identity $\nabla\cdot(w\,\nabla w)=(\nabla w)^2+w\,\nabla^2 w$ and Gauss' theorem (for turning the volume integral of a divergence into a surface integral) yield

$$-\frac{\int_V a^2 w\nabla^2 w\,dV}{\int_V w^2\,dV}=\frac{a^2\int_V(\nabla w)^2 dV}{\int_V w^2\,dV}=\omega_{\text{I}}^2, \tag{26}$$

in accordance with Rayleigh's principle relating the frequency of a normal mode oscillation to the ratio of the maximum potential and kinetic energies.[9,10]] Hence, an appropriate equation for p_{I} is

$$\ddot{p}_{\text{I}}+\omega_{\text{I}}^2 p_{\text{I}}+\frac{\omega_{\text{I}}}{Q}\left(1-\frac{P_0^2}{P_{\text{I}}^2}\right)\dot{p}_{\text{I}}=\rho a^2\frac{\int_V \dot{s}w\,dV}{\int_V w^2\,dV}. \tag{27}$$

To make this equation specific to mass coupling, consider how the moving case of an acoustic resonator couples to the gas inside. If the motion of the resonator is essentially rectilinear, then (in the low-amplitude limit) changes in cross section act like local volume-velocity sources (or sinks). In a cylindrical resonator of varying cross-sectional area $A(x)$, one may write

$$s(x,t)=\frac{1}{A(x)}\frac{dA(x)}{dx}u_c(t), \tag{28}$$

where u_c is the case velocity in the x direction, which is the same for all x (if the case is assumed rigid). We can also find u_c in terms of the forces on the case; considering the forces exerted by *both* engines, assuming they are identical in geometry, and neglecting viscous forces, we can write

$$\dot{u}_c=-\frac{p_{\text{I}}+p_{\text{II}}}{M}\int_0^l\frac{dA(x)}{dx}w\,dx, \tag{29}$$

where M is the total case mass. Note that for each *abrupt* change in cross section, say between $x_j-\varepsilon$ and $x_j+\varepsilon$,

$$\int_{x_j-\varepsilon}^{x_j+\varepsilon}\frac{dA(x)}{dx}w\,dx=[A(x_j+\varepsilon)-A(x_j-\varepsilon)]w(x_j)\,. \tag{30}$$

For example, Eq. (30) would apply to the endcaps of a typical resonator. Combining Eqs. (27)–(29), and using $\int_V\{\}dV=\int_0^l\{\}A\,dx$, we obtain

$$\ddot{p}_{\text{I}}+\omega_{\text{I}}^2 p_{\text{I}}+\frac{\omega_{\text{I}}}{Q}\left(1-\frac{P_0^2}{P_{\text{I}}^2}\right)\dot{p}_{\text{I}}=-\frac{\rho a^2}{M}\frac{[\int_0^l(dA/dx)w\,dx]^2}{\int_0^l w^2 A\,dx}p_{\text{I}}(1+\zeta e^{i\phi}). \tag{31}$$

We may now make use of the results previously obtained for the mass–spring system. Starting with Eq. (31) for p_{I} and a similar equation for p_{II}, letting $d/dt=i\omega$, and then dividing the first equation by $p_{\text{I}}(t)$ and the second by $p_{\text{II}}(t)$ results in a pair of equations exactly like Eqs. (12) and (13), except that the constant $\mu\omega^2$ has been replaced by a more complicated expression. We can therefore identify the effective mass ratio μ for the acoustic system as

$$\mu=\frac{\rho a^2}{\omega^2 M}\frac{[\int_0^l(dA/dx)w\,dx]^2}{\int_0^l w^2 A\,dx}, \tag{32}$$

and then all the results obtained for the mass–spring system should apply to the acoustic system as well.

1. Obtaining Q of engines from experiment

We measure Q from the ''load response'' of each engine. Suppose that instead of a moving case, the only volume-velocity source is a small orifice at $x=l$ (e.g., a

valve) leading to a small volume, forming an acoustic RC network which serves as a load. In the steady state, the differential equation for the loaded engine leads to

$$(\omega_0^2-\omega^2)\tilde{p}+i\omega\frac{\omega}{Q}\left(1-\frac{P_0^2}{P^2}\right)\tilde{p}=-\rho a^2\frac{i\omega\tilde{U}_e}{\int_0^l w^2 A\,dx}, \tag{33}$$

where $\tilde{p}$ is the complex pressure amplitude at the loaded end of the resonator, P is its magnitude, and $\tilde{U}_e$ is the volume velocity exiting that end. Multiplying both sides by $\tilde{p}^*/i\omega$, where $\tilde{p}^*$ is the complex conjugate of $\tilde{p}$, and taking the real part leads to

$$\frac{\omega}{Q}(P^2-P_0^2)=-\left[\frac{2\rho a^2}{\int_0^l w^2 A\,dx}\right]\Re\{\tilde{p}^*\tilde{U}_e\}/2. \tag{34}$$

Since $\Re\{\tilde{p}^*\tilde{U}_e\}/2=\dot{W}_L$, the time-averaged load on the engine, it follows that

$$Q=-\left[\int_0^l w^2 A\,dx\right]\frac{\omega}{2\rho a^2}\left(\frac{\partial P^2}{\partial \dot{W}_L}\right). \tag{35}$$

(In this context, power flowing out of the engine is positive; for constant heat input to the engine, pressure amplitude drops as the load increases, so $\partial P^2/\partial\dot{W}_L$ is negative.)

Strictly speaking, the left side of Eq. (35) should be $2Q/n$, but this is moot unless one can find n independently, say by measuring the curvature of P^2 vs $\dot{W}_L$. We find that for the engines in the present work, P^2 vs $\dot{W}_L$ is quite linear, and the data have sufficient scatter to discourage higher-order fitting. Since $n=2$ simplifies some of the math, it is a useful assumption. The interested reader may verify that carrying out our derivations with arbitrary n leads to the same results as ours, but for the appearance of $2Q/n$ instead of Q (as long as $P_{\mathrm{I,II}}\approx P_0$).

2. Uniform-diameter engine

To enable ready comparison of theory to numerical simulations in a later section, we consider the simple example of an engine with a single cross-sectional area A_0 throughout. If we further assume that the pressure distribution in such an engine is simply $w(x)=-\cos(\omega/a)x$, then $\int_0^l (dA/dx)w\,dx=-2A_0$ and $\int_0^l w^2 A\,dx=A_0 l/2$. Our formulas for μ and Q become

$$\mu=\frac{8\rho a^2 A_0}{\omega^2 lM}=\frac{8}{\pi^2}\frac{m_g}{M} \tag{36}$$

and

$$Q=\frac{\omega A_0 l}{4\rho a^2}\left(\frac{\partial P^2}{\partial\dot{W}_L}\right)=\frac{\omega}{2\rho^2 a^2}\frac{m_g}{2}\left(\frac{\partial P^2}{\partial\dot{W}_L}\right), \tag{37}$$

where $m_g=\rho A_0 l$ is the mass of gas in the engine and ρ and a are assumed constant.

3. Application to acoustic coupling

A useful check on part of our theory is to compare its predictions for *acoustic* coupling with those published in Ref. 3. The sensitivity to frequency difference of two engines coupled by a narrow half-wavelength tube of area A_c is[3]

$$\left(\frac{\omega}{\pi}\right)\left.\frac{\partial\phi}{\partial\Delta\omega}\right|_{\text{acoust.}}=F_{\mathrm{H}}\frac{\pi}{4}\left(\frac{A_\tau}{A_c}\right)\frac{\delta}{r_c}, \tag{38}$$

where r_c is the radius of the coupler and δ is the effective thermoviscous boundary layer thickness $\delta=\delta_\nu+(\gamma-1)\delta_\kappa$; δ_ν is the viscous penetration depth, δ_κ is the thermal penetration depth, and γ is the ratio of isobaric to isochoric specific heats. F_{H} is a quantity called the ''Helmholtz factor,'' which is proportional to $1/A_\tau$ but depends only weakly on r_c and δ. A_τ is arbitrary and normally refers to a typical cross-sectional area of the engines; if the engines have a constant cross section A_0, then choosing $A_\tau=A_0$ makes $F_{\mathrm{H}}=1$. The Helmholtz factor represents the effect on sensitivity of nonuniform cross section, e.g., of using engine geometries that are more like Helmholtz resonators and less like straight ducts. Combining our present general theory with that of Ref. 3, we may find F_{H} analytically. A full expression for F_{H} that is precisely accurate over a wide range of coupler diameters has a very complicated dependence on r_c, δ_κ, and δ_ν, but reasonable accuracy is obtained with

$$F_{\mathrm{H}}=\frac{2}{A_\tau l_c}\int_0^l w^2 A(x)\,dx, \tag{39}$$

which is exact in the limit $r_c\gg\delta$, $l_c\to\lambda/2$, where l_c is the resonant coupler length and λ is the wavelength. In this limit, Eq. (39) predicts $F_{\mathrm{H}}=2.83$ for the engines in the present work, letting A_τ equal the endcap area at $x=l$ and performing the integral numerically using a DELTAE[11] simulation of a single engine (with air at 80 kPa as the working gas). For comparison, our DELTAE simulation of two acoustically coupled engines predicts $F_{\mathrm{H}}=[(\omega/\pi)\partial\phi/\partial\Delta\omega]/[(\pi/4)A_\tau/A_c(\delta/r_c)]=2.80$, a difference of only 1%. We may also compare our present theory with the experimental results in Ref. 3; in that work, two sets of sensitivity data were obtained, one on the engines used in this work (which have a variable cross section—see Fig. 4) and another on engines that were nearly uniform in cross section. The geometry of the engines was the only thing that varied between the two experiments, so if A_τ is chosen to be the same for both types of engine, the ratio of the measured sensitivities should be

$$\frac{\text{sensitivity (var. } A)}{\text{sensitivity (const. } A)}=\frac{F_{\mathrm{H}}(\text{var. } A)}{F_{\mathrm{H}}(\text{const. } A)}=\frac{\int_0^l w^2 A(x)\,dx\ (\text{var. } A)}{A\int_0^l w^2\,dx\ (\text{const. } A)}. \tag{40}$$

This formula predicts a ratio of 3.09, whereas the data and simulations in Ref. 3 both indicate a ratio of 3.08, a negligible difference.

III. COMPARING THEORY WITH NUMERICAL SIMULATION

To compare our theory with numerical simulations, we used DELTAE[11] to model a mass-coupled pair of engines with uniform cross section. DELTAE (Design Environment

for Low-amplitude ThermoAcoustic Engines) integrates the one-dimensional wave equation in the small-amplitude (''acoustic'') approximation, including thermal and viscous effects, according to the thermoacoustic theory of Rott.[12] To simulate a mass-coupled pair of engines, we force all four endcaps in the model to move with the complex case velocity, and require that the case acceleration be the complex sum of all the end pressures times the cross-sectional area, divided by the case mass (we ignore forces on the heat exchangers and stack). We then vary ω_{II} relative to ω_{I} by varying the ambient gas temperature inside simulated engine II, and note the effect on the phase difference ϕ between the engines and on their amplitude ratio ζ. The results of this simulation may be compared to the predictions of Eqs. (17)–(19), using Eqs. (36) and (37) for μ and Q. In order to obtain Q, we simulate a single engine with no moving endcaps, and a large real impedance instead of a perfectly rigid endcap at $x=l$. As the end impedance is lowered, the power flowing out of the end increases and the pressure drops; the derivative $\partial P^2/\partial \dot{W}_L$ is therefore obtained.

While these simulated engines have a simple shape, we strive to make them similar in other respects to the real engines we studied. The simulated engines are filled with 3 bar of xenon, with an ambient temperature of 300 K; they run at 56 Hz and have $Q\approx 200$ and $m_g\approx 0.06$ kg; they are 1.6 m long and have $A_0\approx 2.3\times 10^{-3}$ m^2. They are also set to run with cold-end acoustic pressure of about 16 kPa, since this is close to where our experimental engines were operated.

First, we will compare the Q that best describes the simulated mode-locked states according to Eq. (17) with the Q determined by observing the simulated load-response of one engine as suggested by Eq. (37). By combining the two equations, as well as Eq. (36) for μ, one may obtain the putative identity

$$\frac{\zeta-1/\zeta}{\sin\phi}=\frac{4A_0^2}{\omega M}\left|\frac{\partial P^2}{\partial \dot{W}_L}\right| . \tag{41}$$

The quantities on the left-hand side define the locked state, and the quantities on the right are well-defined constants that are independent of the locked state, if we restrict our attention to the limit $\phi\to\pi$, $\dot{W}_L\to 0$. In our simulations, everything is held constant but Q; for each point, the exact pore geometry of the stack is altered slightly to change the performance of the engine (i.e., change $\partial P^2/\partial \dot{W}_L$). We find that over the range $70\leqslant Q\leqslant 230$, there is virtually no difference between the left and right sides of Eq. (41). These Q values are similar to those of our actual devices, and typical of the passive Q's of acoustic resonators with geometry and gas properties similar to our engines.

Next, we may inquire whether the sensitivity $\partial\phi/\partial\Delta\omega$ of our simulated system matches theory. We use Eq. (19), along with Eqs. (36) and (37) for μ and Q, and compare that result to the value of $(\omega/\pi)\partial\phi/\partial\Delta\omega$ suggested by the simulations in the limit $\phi\to\pi$. In order to make a meaningful comparison, we must address one subtlety, that of calculating $\Delta\omega$, the difference in natural frequencies of two acoustic engines. This question has direct bearing on our experiments as well. One can imagine defining the ''natural frequency'' of an engine as the natural resonance of a column of gas with the engine's geometry and temperature distribution, or as the frequency with which an engine with that geometry and temperature distribution will run. When the frequency difference between two engines is largely determined by the temperature of the gas in the resonators, the distinction between these definitions is unimportant; but when two engines are weakly coupled, even a very small change in frequency can mean a relatively large change in phase, which can mean a significant change in the acoustic power flow between the engines. This in turn can cause significant differences in the engines' hot-end temperatures, and these differences can ultimately dominate the frequency difference. In this circumstance, the difference in the running frequencies of the two engines may vary considerably from the difference in their gas-column resonance frequencies.

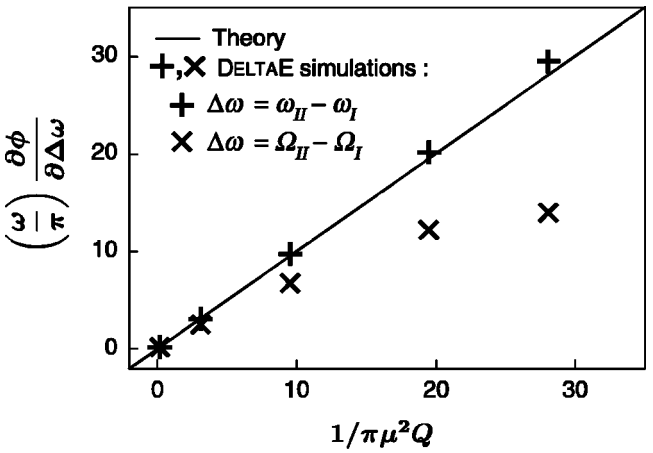


FIG. 2. The sensitivity $\partial\phi/\partial\Delta\omega$ of a pair of mass-locked thermoacoustic engines with uniform cross section, as predicted by theory (line) and DELTAE simulation (symbols).

This distinction is of little practical importance in the big picture, since one is generally more interested in how sensitive the locked states are to objective external changes (one engine sitting in the sun, the other in the shade; one engine driving a bigger load, etc.) and not to changes in the ''natural frequencies,'' particularly if those frequencies are dominated by the engines' internal responses to external changes. However, the natural frequency difference is a very convenient analytic tool, and useful for expressing general results about mode-locked systems—e.g., the phase difference ϕ between the engines is determined by $\Delta\omega$, however that $\Delta\omega$ comes about.

We compare our theory with simulation by calculating the frequency difference two ways; the first, $\Delta\omega=\omega_{\mathrm{II}}-\omega_{\mathrm{I}}$, is the difference in the gas-column resonances; the second, $\Delta\omega=\Omega_{\mathrm{II}}-\Omega_{\mathrm{I}}$, is the difference between the frequencies at which the engines will run, uncoupled, if they have the same temperature distribution as when coupled.

Figure 2 shows the comparison of simulations and theory; moving along the horizontal axis from left to right, the coupling is weakened by increasing the mass M of the combined engine cases, thereby lowering μ, but leaving Q unchanged. As the coupling gets weaker, $\Omega_{\mathrm{II}}-\Omega_{\mathrm{I}}$ overestimates the value of $\Delta\omega$. Therefore we might expect that in an experiment on a weakly coupled system, the values of ζ and

ϕ for a given locked state might agree well with theory, but the experimentally determined $\partial\phi/\partial\Delta\omega$ will be smaller than its ''true'' value.

Acoustically coupling[3] two engines so that they lock near $\phi=\pi$, and keeping $|\phi-\pi|$ small enough so that 95% of the case vibration is canceled when their frequency difference is as much as 1%, leads to $(\omega/\pi)\partial\phi/\partial\Delta\omega\approx 2$. Typically, an engine might have $Q\sim 100$ and $\mu\sim 10^{-3}$, implying that the mass-coupled sensitivity is $(\omega/\pi)\partial\phi/\partial\Delta\omega = 1/\pi\mu^2 Q\approx 10^4$. This implies that the mass coupling is really very weak, and can be ignored in the vibration-cancellation application. Absent heroic efforts, we are confined to weakly coupled engines in the laboratory as well. The engines we use for our experiments are charged with 3 bar of xenon (as in the simulations) and lightly constructed; still, we find that $1/\pi\mu^2 Q\approx 10^3$.

A. Possible complications

In many presentations on mode-locking, few quantitative results are given because it is usually so difficult to estimate the coupling strength. In Huygens' workshop, his clocks could have been coupling by uniform motion of the wall, elastic waves in the wall, or even motion of the nearby air. (Huygens himself suggested that it was the ''imperceptible agitation of the air caused by the motion of the pendula'' that enabled the clocks to ''communicate.''[1]) In addition, one would have to characterize the ''load response'' of the clock mechanism to obtain the coupling strength.

By contrast, we can successfully characterize the coupling of our thermoacoustic engines, given the nature of the analytic results and our experience at making acoustic and thermodynamic measurements. To illustrate how difficult the problem can become, however, and why it often is not worth the effort to make quantitative analyses of mode-locked systems, we will consider two of many possible subtle complications: external dissipation on the case and heat leak between the engines' hot ends.

1. External dissipation

So far we have treated the coupling between the engines as purely reactive, but we know the actual coupling impedance must be richer in detail, and in particular it must contain some resistance. One possible consequence of external resistance (friction in the engine mounts, etc.) is that it may discourage the $\phi\simeq 0$ mode. In an extreme limit, imagine that the case mass becomes negligible and the only impedance is a dashpot connecting the case to a rigid wall. Obviously the mode that encourages maximum motion of the dashpot will suffer more damping than the mode that encourages minimum motion. This may help explain why the in-phase, $\phi\simeq 0$ mode was rarely observed by us (or Huygens). There are other, more subtle consequences of adding resistance to the coupling, however.

Consider the mass-and-spring model of Fig. 1. Imagine connecting one side of a dashpot with mechanical resistance R_c to the case M, and the other side to a rigid wall, so that the coupling impedance $Z_c=R_c+i\omega M=\omega M(i+\alpha)$, where $\alpha=R_c/\omega M$. The additional dissipation alters the analysis presented in Sec. II A. A new term $i\omega^2\alpha X_c$ is added to the left side of Eq. (11), with consequences in Eqs. (12)–(14). The energy conservation equation, Eq. (15), becomes

$$(X_0^2-X_{\rm I}^2)+(X_0^2-X_{\rm II}^2)=\frac{\alpha Q}{\mu}X_c^2, \tag{42}$$

with consequences in Eq. (16). Equation (17) becomes

$$\zeta-1/\zeta=\frac{2\mu Q\sin\phi}{1+\alpha^2+\alpha\mu Q}. \tag{43}$$

Finally, in the limit $\phi\to\pi$, and with the approximation $\zeta+1/\zeta\approx 2$ (which is true for our experimental system, but is *not* generally true for all systems far from the center of the ''capture band''), we find

$$\frac{\omega}{\pi}\frac{\partial\phi}{\partial\Delta\omega}=\frac{1}{\pi\mu^2 Q}\left[\frac{1+\alpha^2+\alpha\mu Q}{1+\alpha/\mu Q}\right]. \tag{44}$$

Recall from Eq. (19) that without external resistance, the sensitivity $(\omega/\pi)\partial\phi/\partial\Delta\omega=1/\pi\mu^2 Q$, so we may view everything in the square brackets on the right side of Eq. (44) as the effect of external resistance on sensitivity. Figure 3 shows this multiplying factor as a function of α, for a number of different values of μQ.

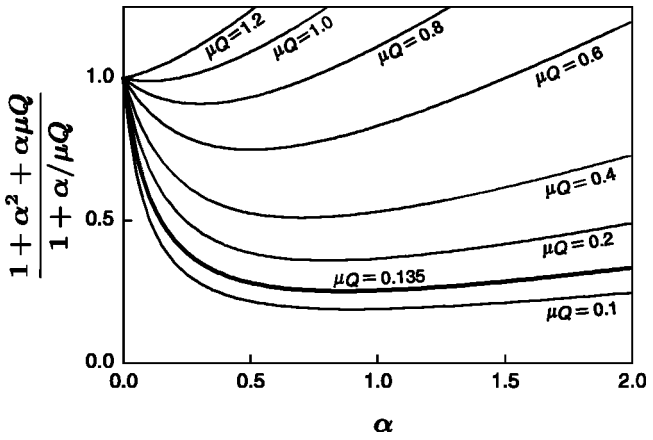


FIG. 3. Normalized sensitivity of a mass-locked system with external resistance R_c as a function of $\alpha=R_c/\omega M$, for various strengths of mass coupling. The darker line corresponds to $\mu Q=0.135$, the value we estimate for our experimental system of coupled thermoacoustic engines.

The behavior of these curves is not immediately intuitive. For strong coupling (large μQ) the addition of dissipation increases the sensitivity, making the system more weakly coupled; but for systems that are weakly coupled already, the addition of small amounts of external damping actually *enhances* the coupling, quite dramatically. (For our acoustic engines, this suggests that adding a small real component to the coupling impedance results in a phasing, between the pressure at the endcaps of the resonator and the case motion, that is more favorable to energy transfer between the engines.) Each curve has a minimum at $\alpha=1-\mu Q$, if $\mu Q\leqslant 1$. Only after the resistance becomes fairly large does the coupling strength begin to diminish again. This could have quite a noticeable effect on any ϕ vs $\Delta\omega$ data.

In contrast, Eq. (43) is only a small perturbation away from Eq. (17), for small external resistance. In addition, the

perturbation appears only as a multiplicative constant, so it is virtually impossible to tell from ζ vs ϕ data whether there is external dissipation, as opposed to whether one's values for μ and/or Q are inaccurate.

We have not pursued this much further, or tried to devise a reliable method for measuring R_c. In principle, one can determine it from measurements of case motion; that is, the phase between the motion of the case and the pressures inside may indicate whether the pressure forces are driving a reactance or a reactance and a resistance. Based on such measurements, we estimate that for our system, $\alpha = R_c/\omega M \sim 0.1$, a plausible value.

2. Heat leaks between engines

We now briefly discuss another complication: heat flowing between neighboring prime movers due to a temperature difference that develops as a result of mass coupling. Imagine that in a given locked state, engine I becomes ''loaded,'' feeding power into engine II, which is then ''driven.'' The amplitude P_{I} in engine I will fall and its hot temperature $T_{\mathrm{H,I}}$ will rise; conversely, P_{II} will rise and $T_{\mathrm{H,II}}$ will fall. If the two hot ends have some thermal contact, a heat leak $\dot{Q}_{\mathrm{cross}} = UA(T_{\mathrm{H,I}} - T_{\mathrm{H,II}})$ will flow between them, where UA is some effective heat-transfer coefficient. This has the effect of draining more power from engine I and feeding more into engine II, making the amplitude imbalance between the engines that much greater. Hence, it makes $\partial P^2/\partial \dot{W}_L$ greater than it would otherwise be, making Q larger and enhancing the coupling. This additional coupling will not be apparent when making a load measurement on a single engine, unless great care is taken to duplicate the conditions of the coupled state (e.g., pump the gas out of the other engine so it cannot start, and manipulate its heat input so it always has the T_{H} that it would have in the locked state corresponding to the first engine's P).

This also points to a larger issue: the effect of different modes of operation on engine coupling. Engines may be run with constant power, constant hot-end temperature, or something in between (such as constant pressure amplitude, or constant power delivered to a load, etc.) The general principles developed so far should still apply; that is, if one can find the ''Q'' that characterizes the engines in their typical mode of operation, along with the effective moving gas mass and the case mass, one should be able to estimate the coupling strength. Care must be taken, of course, to always reexamine the energy balance whenever the mode of operation is changed. For instance, consider two mass-locked engines whose hot-end and cold-end stack temperatures are held constant. This example is of interest because material properties usually limit how hot an engine may be run, and the need for maximum power density and efficiency may require that the engine be run at this upper limit. If one such engine experiences an extra load as a result of being mass-coupled to another engine, its hot temperature will tend to rise and its oscillating pressure amplitude will fall; the temperature control will decrease the power flowing to that engine to maintain its temperature at the upper limit, lowering the pressure still further. Thus one might expect that such engines would have very high $\partial P^2/\partial \dot{W}_L$, and hence high Q, implying that they would allow very little phase shift for a given difference in natural frequency [see Eq. (19)]; this behavior is indeed observed in simulations. However, the engines allow large differences in pressure amplitude for small frequency differences.

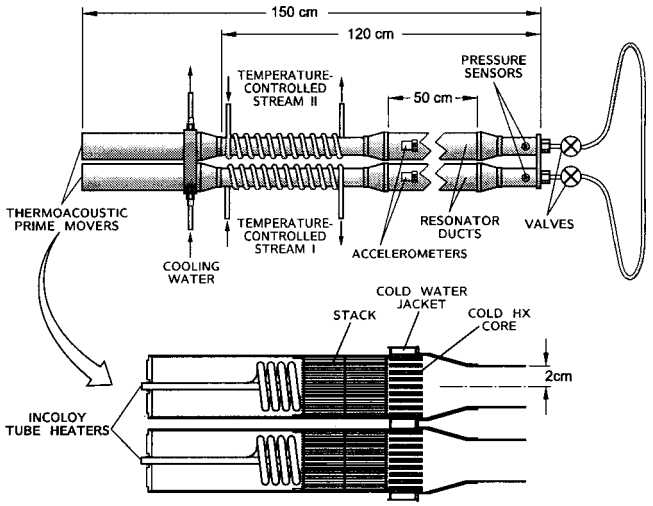


FIG. 4. Experimental setup for exploring the behavior of coupled thermoacoustic engines. The insulation surrounding the prime movers and the temperature-control coils is not pictured.

Having given the reader a taste of the challenges involved in making a good mass-locking measurement, we will proceed to describe our experiments with small, mass-coupled thermoacoustic engines.

IV. APPARATUS AND EXPERIMENTS

To explore mass-locking of acoustic engines experimentally, we use two nearly identical thermoacoustic engines which are rigidly attached side by side (these are the same engines that were used for a portion of the acoustic coupling work described in Ref. 3). Two neighboring ends of the engines are connected by a coupling duct which can be inserted or removed from the system by means of valves, enabling the system to be acoustically coupled, or not, as desired. Each engine consists of a section of duct of varying cross section (the resonator) connected to a thermoacoustic prime mover, by which we mean an electric heater, a water-cooled heat exchanger, and a stack of parallel plates between them; these sustain the oscillations in the resonator when the engine is running.

Figure 4 shows the essentials of the experimental setup, and some details of the thermoacoustic prime mover hardware. Note that the engines share a water jacket that cools their cold heat exchangers, and both are welded to a common plate at the other end. The engines are also joined by two narrow, 50-cm-long plates that are welded between the 50-cm-long ''fat'' sections of the resonators. A section of each resonator is wrapped with copper refrigeration tubing, which circulates water from a temperature-controlled bath. Thus the gas temperature in each engine can be varied independently, enabling control of $\Delta\omega$. A pressure sensor is mounted at the end of each resonator where it joins the coupling duct. We

take the signal from the sensor in resonator I (pressure sensor I) to represent $p_1(t)$, and that from pressure sensor II to represent $p_{II}(t)$. Using p_1 as a reference to a lock-in amplifier and using p_{II} as the input signal allows a direct measurement of ϕ. Independent monitoring of pressure sensors I and II gives $P_1(t)$ and $P_{II}(t)$. Accelerometers on the cases measure case vibration in the axial (x) direction, and numerous thermocouples (not pictured) provide temperature data on the gas near the hot end of the stack, the cooling water, the water in the coils wound around the resonators, and the gas inside the resonators. Multimeters monitor the input voltages and currents to the tube heaters that power the engines. The accelerometers are used to verify that the two engines, welded together, act like a rigid structure, and that vibrations are indeed canceled when the phase between the engines approaches π—we find that cancellation of case vibration is at least 99.5% complete when $\phi=\pi\pm0.002$. The engines are filled with xenon, with the mean pressure $p_m=303$ kPa; they run at 50 Hz, in a room temperature environment (rejecting heat at 300 K). The combined case mass of the engines, including water in the heat exchangers and such, is $M\simeq11$ kg.

To assist us in interpreting our data, we simulate the system using DELTAE. We use a single simulated engine, without moving endcaps, to help us estimate μ and Q for the system; we use a pair of simulated engines, whose combined case mass moves in response to the oscillating pressure, to compare simulated mode-locked states with measured ones. The simulated mode-locked system involves a few approximations; for instance, the forces on the heat exchangers and stacks are neglected, and the changes in cross-section are all treated as abrupt (including the truncated cones shown in Fig. 4). Nevertheless, we expect that the agreement between experiment and simulation should be reasonable.

The single simulated engine is used to estimate $w(x)$, the pressure waveform in the engines. The only pressure reading taken in each experimental engine is at the end of the resonator (the ''cold end''), as shown in Fig. 4, so we trust DELTAE's linear acoustics to estimate what the pressure distribution in a given geometry should be. Combined with our knowledge of $A(x)$, this allows us to numerically carry out the integrals in Eqs. (32) and (35).

To obtain Q, we also need $\partial P^2/\partial\dot{W}_L$. To measure this quantity experimentally, we use a precision leak valve in series with a small cylindrical ''tank'' of volume $V=150$ cm^3 to form an acoustic RC load, and attach this load to the end of one engine resonator at a time, observing changes in the end pressure of that engine as the valve is gradually opened. The other engine is pumped out for this measurement, so it cannot start; some heat is still applied to it, however, to maintain it near the temperature it would have *if* it were running. The actual power dissipated at each valve setting is estimated by using a pressure transducer within the tank, along with the one at the end of the engine, to measure the pressure difference across the valve. This pressure difference and the tank volume V are used to calculate the dissipated power according to[13] $\dot{W}_L=\omega V|p||p_T|\sin\theta/2\gamma p_m$, where p is the oscillating pressure at the end of the engine, p_T is the pressure in the tank, and θ is the phase by which p leads p_T.

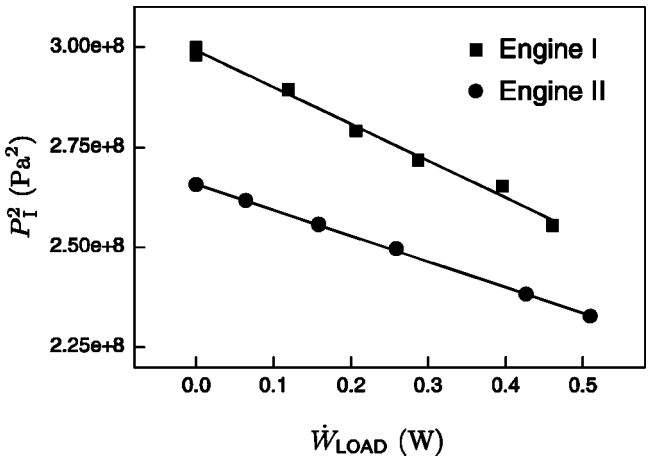


FIG. 5. Load responses of the engines used in the mass-locking experiment. The squares are engine I data and the circles, engine II; the lines are linear fits. Each data set on a given engine was taken with the neighboring engine hot, but not running, to approximate the heat leak that exists when both engines are running.

In these measurements, we not only had to contend with heat leaks between the engines, as discussed earlier, but also ordinary heat leaks from the engines to the ambient. Rather than attempt to subtract the ambient heat leaks from the electric power going into the engines and manipulate the electric power to maintain constant heat input, we chose to simply maintain constant electric power, and characterize the load performance of the engines as they were, leaks and all. Figure 5 shows the results; evidently $\partial P^2/\partial\dot{W}_L$ is fairly linear in the region studied, but the data have sufficient scatter that a higher-order fit is probably not warranted. The data for each engine were taken over the range of pressures which that engine experienced when the data on locked states were taken. Of interest is the difference in slopes between I and II; apparently the two engines do not perform identically, despite being constructed ''identically.'' One difficulty in predicting the performance of these particular engines, and in modeling them, is the use of tubular heaters in the prime movers (see detail in Fig. 4). These heaters are robust and easy to bend into a desired shape, but are not as well understood thermodynamically as standard shell-and-tube or parallel-plate heat exchangers.

We find that $Q_1=73$ and $Q_{II}=50$, with $\overline{Q}$, the average, being 62; by simulating this measurement in DELTAE, we find $Q=61$. Because of the difficulties in precisely modeling the hot ends of these engines, and the large difference between Q_1 and Q_{II}, this striking agreement must be a bit coincidental. The difference in measured Q between the engines adds a slight offset of order $[(Q_{II}-Q_1)/\overline{Q}]^2$ to Eq. (17). We do not treat the question of the offset rigorously, but merely measure it and subtract it from our simulated and theoretical results.

We may also use an experimental observation to obtain an approximate value for μ, the effective mass ratio. Observing the engines when they are coupled but not mode-locked, and measuring the frequency and phase of both engines as a

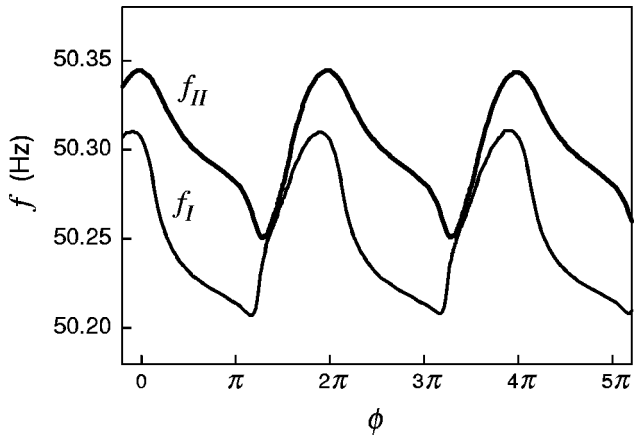


FIG. 6. Frequency as a function of the phase between two mass-coupled engines that are on the verge of locking.

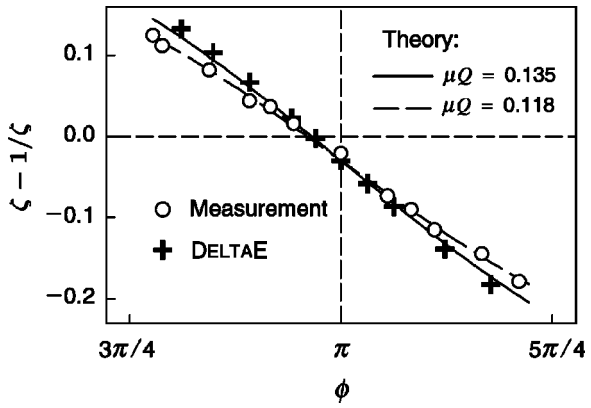


FIG. 7. Amplitude imbalance $\zeta - 1/\zeta$, where $\zeta = P_{II}/P_I$, versus phase difference ϕ of mass-locked state. Circles are data points, crosses are DELTAE simulations, and the broken line is a fit to $\zeta - 1/\zeta = 2\mu Q \sin\phi + C$, with μQ and C free parameters. The solid line is the same equation, but using our estimated μ and Q. The difference in their slopes at $\phi = \pi$ is about 15%. The DELTAE points have had the offset C subtracted from them as well.

function of time, one may obtain a plot of frequency versus phase for the engines as they beat. Such a plot is shown in Fig. 6; note that the frequencies are their lowest near $\phi = \pi, 3\pi, 5\pi, \ldots$ and their highest near $\phi = 0, 2\pi, 4\pi, \ldots$, consistent with the idea that a moving case shifts frequencies up. By comparing the high and low values, we find that the frequency shift $|\Delta f|/f \approx 0.002$; if we assume $f^2 = f_0^2(1+\mu)$, then $\mu \approx 0.004$. Since this includes the moving gas in two engines, we may guess that μ for one engine, that is, the $\mu = m/M$ to use in Eqs. (17) and (18), is μ (expt.) ≈ 0.002. Using our theory, together with our DELTAE model to find $w(x)$, we find that the effective mass $m = 0.024$ kg for our engines; with $M = 11$ kg, μ (theo.) ≈ 0.0022.

We are now poised to predict $\zeta - 1/\zeta$ versus ϕ for our engines; we choose $\mu Q = 0.0022 \times 62 = 0.135$ as our ''theoretical'' coefficient in Eq. (17). This equation can be compared to actual phase differences and amplitude ratios at various locked states for our engines. We found that the engines would readily mode-lock, and the phase could be varied easily by changing the temperature of one of the cooling coils; however, we found that the states were not particularly stable, but very sensitive to small changes in temperature. Only by very carefully ''teasing'' the system could it be made to settle at a particular state (largely because the hot temperatures need to equilibrate to establish a stable state). Figure 7 shows a comparison between theory and experiment. Considering all the approximations that have been made to get to this point, the agreement between the three methods is remarkably good.

We also try to compare theory and experiment for Eq. (19), but here we must grapple with the difficulty in measuring $\Delta\omega$. The most straightforward way to measure the difference in frequency between two coupled engines is to decouple them and measure the beat frequency; however, we know from Sec. III that this frequency difference is not necessarily the ''natural frequency difference'' required by theory. In addition, when the engines are decoupled, the hot-end temperatures immediately begin to shift, making the measurement of a long beat period somewhat difficult. We have chosen an indirect, but simpler, method to find $\Delta\omega \approx \Omega_{II} - \Omega_I$. We record phase versus $\Omega_{II} - \Omega_I$ for the engines when they are *acoustically* coupled, using the coupler that is already in place. An extra 130 kg of mass is clamped to the engines during the acoustic measurement, to remove all effects of mass-coupling. Subsequently, during the mass-locking measurement, after the engines have established a stable mass-locked state we momentarily open the acoustic coupler valves and record the phase, which can then be related to a frequency difference. Figure 8 shows a plot of ϕ_A, the phase at which the system is mode-locked due to acoustic coupling, versus the normalized frequency difference $\Delta\omega/\omega \approx (\Omega_{II} - \Omega_I)/\omega$. Near $\phi_A = \pi$, the data are linear, and we use the slope in this region to obtain the relationship between ϕ and $\Omega_{II} - \Omega_I$. We have chosen a relatively narrow-diameter coupler for this part of the work, so that the acoustic coupling is fairly weak (although not as weak as the mass coupling). Thus the phase of the acoustically locked state is a sensitive measure of frequency difference. We allow the engines to lock through mass coupling (i.e., with the acoustic coupler's valves closed), record the steady-state amplitudes and phases, and then throw the valves open momentarily and record the phase ϕ_A, thus obtaining $\Omega_{II} - \Omega_I$. As

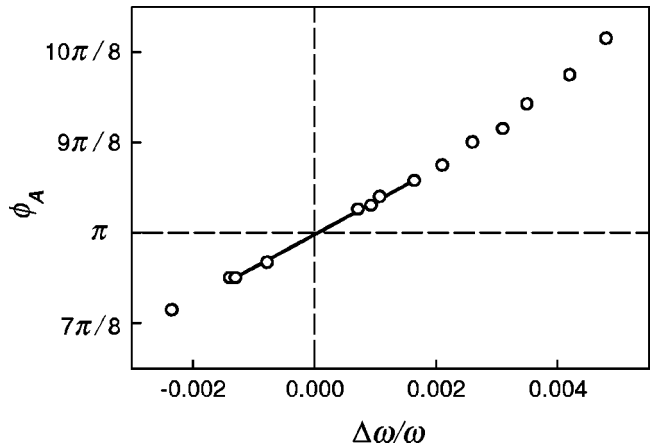


FIG. 8. The phase of the locked state when the engines are acoustically coupled, achieved when the coupler valves in Fig. 4 are open, versus the frequency difference estimated by measuring the beat period when the valves are closed. In this measurement, an extra 130 kg of mass is clamped to the engines, to discourage any mass-coupling.

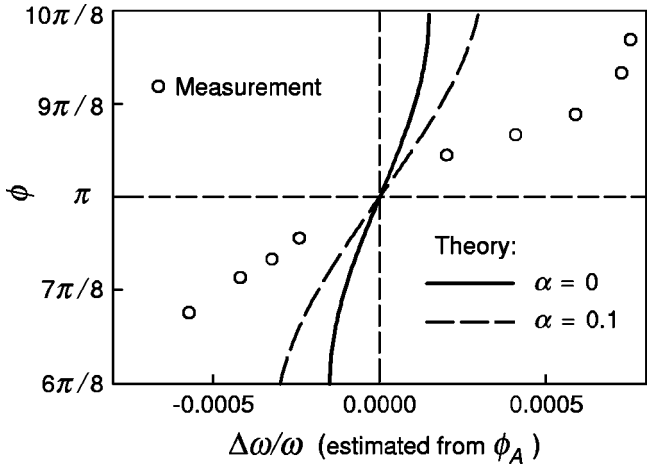


FIG. 9. The phase difference of the mass-locked state versus the frequency difference estimated from the corresponding acoustically-locked state. The quantity α is defined in Sec. III A 1, and represents the degree of external dissipation in the system, such as that caused by friction in the supports.

expected, we find that ϕ_A drifts, since the hot temperatures drift as soon as the valves are closed, and thus $\Omega_{\mathrm{II}}-\Omega_{\mathrm{I}}$ is a moving target. We find that measuring the phase about 5 s after the valves are opened gives a repeatable, if not wholly accurate, value for ϕ_A.

Figure 9 shows our mass-locked ϕ vs $\Delta\omega/\omega$ data, compared with two theoretical predictions—one assuming no external dissipation, and another assuming moderate external dissipation. We were able to achieve mass-locking only between $3\pi/4<\phi<5\pi/4$; this is in contrast to the acoustically coupled engines studied in Ref. 3, which would lock between $\pi/2<\phi<3\pi/2$. We may use our theory to predict the capture bandwidth; starting with Eq. (18), we ask what angle ϕ makes $\Delta\omega$ a maximum. This is equivalent to seeking ϕ such that $d(\sin\phi\cos\phi)/d\phi=0$. This occurs when $\phi=(2n+1)\pi/4$, $n=0,1,2,\ldots$, and hence when $\sin\phi\cos\phi=\frac{1}{2}$. For a system whose band center is nominally at $\phi=\pi$, then, we predict

$$\pm\left(\frac{\Delta\omega}{\omega}\right)_{\max}=\pm\frac{\mu^2 Q}{2}\quad\text{at}\quad\phi=\pi\pm\frac{\pi}{4}. \tag{45}$$

The experimental phase limits agree well with this prediction, but the predicted frequency bandwidth is much narrower.

The shapes of the experimental and theoretical curves have some qualitative similarities, but as expected (recall Fig. 2), the slope of the experimental curve is much less than that predicted by theory. We found that measuring ϕ_A ''immediately'' (within 0.5 s) after the valves were opened appeared to approximately double the slope; however, the valves were operated by hand and took several tenths of a second to open, so these data are of dubious value. While the simplest theory (no dissipation) predicted that the mass-coupling in our system should have been about 25 times weaker than the acoustic coupling, the data suggest it is only about 10 times weaker. External dissipation brings these numbers into closer agreement, as shown by the dashed line in Fig. 9, but it would require three times as much dissipation as we estimate for our system to make the experimental and theoretical curves coincide. We believe inability to measure the ''true'' value of $\Delta\omega/\omega$ is the biggest source of error.

Since $\Delta\omega/\omega$ is of mostly theoretical interest anyway, we consider what effect objective changes might have on our system. The DELTAE model, which agrees well with theory and experiment as demonstrated by Fig. 7, suggests that a 0.03-K change in the resonator gas temperature of one engine, within the section wrapped in copper tubing, would cause a change in phase angle $\Delta\phi=\pi/16$. The measurement of gas temperature in our experiment was too coarse to verify this, but if we assume it is correct, it demonstrates again how very sensitive mass-coupled engines are, compared to the acoustic coupling necessary to cancel vibration. For comparison, consider an acoustically coupled system of two engines, designed so that 95% of the case vibration is canceled when $\Delta\omega/\omega=0.01$. This implies $|\Delta\phi|=|\phi-\pi|\leqslant 0.05$ when the difference in resonator gas temperatures is $\Delta T/T_m=0.02$, or $\Delta T=6$ K if the mean temperature T_m is about 300 K. For such a system, $(T_m/\pi)\Delta\phi/\Delta T\approx 1$; for our mass-locked system, $(T_m/\pi)\Delta\phi/\Delta T\approx 600$.

In conclusion, we have shown that the Huygens effect, or mass-coupling, is sufficiently weak in thermoacoustic engines that it may be ignored in many circumstances, although a precise measure of its weakness will often be difficult to obtain. It is also apparent that the natural mode for the engines to mass-lock has the correct phase relationship to minimize motion of the common mass. This is further encouragement that mass-locking will not interfere with the canceling of engine vibrations by acoustic coupling.

There are some circumstances, of course, where the mass-coupling could be significant. Engines that use a liquid as the working fluid (such as a liquid-sodium thermoacoustic engine[6]) could have $\mu\sim 1$. This could also be true of acoustic engines whose resonators include moving parts, such as magnets for generating electricity. This strong coupling might in fact be an advantage, if the proper suspension or supports for the engines can encourage them to lock in antiphase, thus lessening or eliminating the need for any acoustic coupling.

ACKNOWLEDGMENTS

The authors wish to thank Bob Keolian and Bob Ecke for teaching us about mode-locking; Bill Ward, Scott Backhaus, and Bob Hiller for useful discussions; and David Gardner for assistance with the experiments. We are also grateful to Andrew Piacsek (of Central Washington University) for providing an English translation of Huygens' letter. This work has been supported by the Office of Basic Energy Sciences in the U.S. Department of Energy.

[1] C. Huygens, in *Ouevres Completes de Christian Huyghens*, edited by M. Nijhoff (Societe Hollandaise des Sciences, The Hague, The Netherlands, 1893), Vol. 5, p. 243 (a letter to his father, dated 26 Feb. 1665).
[2] G. W. Swift, ''Thermoacoustic natural gas liquefier,'' in Proceedings of the DOE Natural Gas Conference, Houston, TX, March 1997.
[3] P. S. Spoor and G. W. Swift, ''Mode locking of acoustic resonators and its application to vibration cancellation in acoustic heat engines,'' J. Acoust. Soc. Am. **106**, 1353–1362 (1999).

[4] A. B. Pippard, *The Physics of Vibration (Omnibus Edition)* (Cambridge U. P., Cambridge, 1989), p. 392.
[5] B. Van der Pol, ''The nonlinear theory of electric oscillations,'' Proc. Inst. Radio Engineers **22**, 1051 (1934).
[6] G. W. Swift, ''Thermoacoustic engines,'' J. Acoust. Soc. Am. **84**, 1145–1180 (1988).
[7] A. B. Pippard, *The Physics of Vibration (Omnibus Edition)* (Cambridge U.P., Cambridge, 1989), Chap. 12.
[8] P. M. Morse, *Vibration and Sound* (American Institute of Physics, New York, 1986), pp. 104, 313.
[9] J. W. Strutt and Lord Rayleigh, *Theory of Sound* (Dover, New York, 1945), Vol. 1, Sec. 88.
[10] A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, Woodbury, New York, 1989), p. 53.
[11] W. C. Ward and G. W. Swift, ''Design environment for low amplitude thermoacoustic engines (DeltaE),'' J. Acoust. Soc. Am. **95**, 3671–3672 (1994). Fully tested software and user's guide available from Energy Science and Technology Software Center, US Department of Energy, Oak Ridge, TN. To review DeltaE's capabilities, visit the Los Alamos thermoacoustics web site at `http://www.lanl.gov/thermoacoustics`. For a beta-test version, contact ww@lanl.gov (Bill Ward) via Internet.
[12] N. Rott, ''Thermoacoustics,'' Adv. Appl. Mech. **20**, 135–175 (1980).
[13] A. M. Fusco, W. C. Ward, and G. W. Swift, ''Two-sensor power measurements in lossy ducts,'' J. Acoust. Soc. Am. **91**, 2229–2235 (1992).

A new method for the absolute measurement of piezoelectric coefficients on thin polymer films

François M. Guillot and Jacek Jarzynski
School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

(Received 2 October 1999; revised 6 April 2000; accepted 10 April 2000)

A new quasistatic method to measure piezoelectric coefficients on thin polymer films is presented. This method is based on a combined experimental/analytical approach, where small polymer samples (6 mm×3 mm×110 μm) are encapsulated in a soft silicone rubber and an electric field is applied across their thickness (3-direction). Strains are measured optically along three perpendicular directions using a laser Doppler vibrometer, and the experimental measurements are used in a Rayleigh–Ritz energy minimization procedure implemented symbolically in MATHCAD, which yields the absolute piezoelectric coefficients d_{3ii}. These measured coefficients are material properties of the polymer and do not depend on the specific boundary conditions of the problem. The validity of the method is established using the ATILA finite element code. Experimental values of d_{311}, d_{322}, and d_{333} obtained for polyvinylidene fluoride (PVDF) at room temperature, in the frequency range 500–2000 Hz, are presented and compared with existing data; excellent agreement is found. The extension of the method to the determination of electrostrictive coefficients on soft polyurethane materials is introduced.
[S0001-4966(00)03607-9]

PACS numbers: 43.38.Ar, 43.38.Fx, 43.35.Sx, 43.28.Tc [SLE]

INTRODUCTION

Piezoelectric properties of materials can be characterized in two ways: (1) by measuring the electrical response to a mechanical excitation (referred to as the direct effect) or (2) by measuring the mechanical response of the materials due to an electrical excitation (converse effect). A major issue in measuring the converse effect on polymer films consists in applying the electrical excitation with minimal disturbances on the motion of the film itself. For example, the presence of metal electrodes deposited on the film surface stiffen the sample and considerably reduce the effective d_{311} and d_{322} piezoelectric coefficients, rendering their measurement quite difficult.[1] The experimental apparatus itself can also influence the sample motion, either because of the way the film is configured for measurements (sample holders often impose nontrivial boundary conditions on the film)[2] or because of the technique used to measure the mechanical response (for instance, if an accelerometer is used to measure the thickness response, its presence on the sample surface loads the film).[3] Because of these inherent difficulties, most measuring techniques cannot yield both in-plane and out-of-plane coefficient data simultaneously, or if they do, the data correspond to the apparent coefficients rather than the true ones. This article presents a new experimental method that provides converse effect measurements of the true d_{311}, d_{322}, and d_{333} coefficients of piezoelectric polymer films. This method can also be readily extended to the measurement of electrostrictive coefficients, and it can therefore be applied to other types of materials, such as electrostrictive polyurethanes.

Measurements of electric field-induced strains on thin polymer films present two main challenges: (1) the measuring technique should not interfere with the film motion, and (2) the mechanical boundary conditions should be very well defined, because they have a great effect on the motion of such soft materials. To address the first point, optical techniques such as laser Doppler interferometry are ideal, because they offer a noncontact way of measuring small displacements (greater than or equal to 5×10^{-10} m). In a previous study,[4] the authors used such a technique to measure the piezoelectric coefficient d_{333} of polyvinylidene fluoride (PVDF) on freely suspended and clamped electroded samples. Although the results were fairly accurate, it was found that a significant error was introduced by the presence of bending waves in the sample. In the new method, a similar laser interferometer is used, but the sample is prepared differently: using a soft silicone rubber, it is attached to a rigid metal backing on one side, and to an aluminum foil on the other side. The rigid backing effectively suppresses the bending motion, and the sample is excited in a pure thickness mode when an electric field is applied in the 3-direction. The use of the silicone rubber is also the key factor to address the boundary condition issue. In previous studies, the samples were mostly constrained by the mechanical holder used to position them in the measuring apparatus, and by the metal electrodes deposited on their surface to apply an electrical voltage. In the new configuration used in this study, the driving voltage is applied between the metal backing and the aluminum foil, thereby eliminating the need for deposited electrodes. Furthermore, the only mechanical constraint on the film is that imposed by the silicone rubber, which is a soft material compared to the polymer. Therefore, the constraint on the polymer film is small and the boundary conditions can be precisely evaluated from the knowledge of the silicone rubber elastic properties.

The optical measurements on polymer samples thus prepared are used in conjunction with a Rayleigh–Ritz variational method that yields electromechanical constants. An-

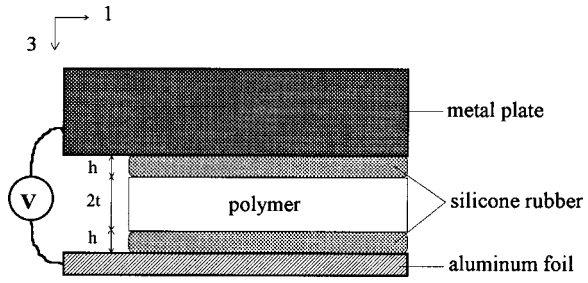


FIG. 1. Top view of the polymer/rubber composite stack and its electrical connections. ''1'' is the stretched direction and ''3'' is the poled direction.

other advantage of the method is that the experimental configuration allows the determination of three independent piezoelectric coefficients. It should be also noted that the method provides an absolute measurement of these coefficients, in the sense that the mechanical boundary conditions are essentially ''factored out,'' and that the resulting piezoelectric constants are the true material properties of the polymer. In other words, the experimentally measured coefficients are independent of the specific boundary conditions and represent the intrinsic electromechanical behavior of the material. Finally, it should be pointed out that the method presented in this work relies on a quasistatic assumption: it is assumed that the driving frequency is well below the fundamental thickness resonance frequency of the polymer films, which is on the order of several MHz for the sample sizes considered in this study.

This article is divided into five sections. In Sec. I we first explain the method, detailing the sample preparation, the optical measurements, and the variational approach. In Sec. II we then deal with the validation of the method, obtained through the use of finite elements. In Sec. III we further validate the technique by comparing experimental results on PVDF to published data. In Sec. IV we discuss the extension of the present method to the study of electrostrictive materials. Finally, conclusions are drawn in Sec. V.

I. PRINCIPLE OF THE METHOD

The method described in this article can be divided into two distinct steps. The first step is experimental and consists in measuring displacements optically on a rubber-encapsulated polymer film. The second step is analytical and uses the experimental data in an energy minimization procedure that yields the d_{3ii} piezoelectric coefficients ($i=1,2,3$). Details of these steps are presented in the following subsections.

A. Sample preparation

Figure 1 depicts how the polymer samples are prepared for the experimental measurements. A two-component, low-viscosity silicone compound (Master Sil Polymer 151, from Master Bond Inc.) is used to sandwich a thin rectangular polymer film (thickness $2t=110\,\mu$m) between a metal plate and a piece of aluminum foil (thickness $d=17\,\mu$m). The compound cures to a soft rubber and forms two layers, each of uniform thickness h (h varies between 3 and 12 μm, depending on the pressure applied to the ''stack'' during curing). An electrical voltage V is applied between the metal plate and the aluminum foil, creating an electric field E_p inside the polymer and an electric field E_r inside the two rubber layers; both electric fields are along the 3-direction. E_p causes the polymer film to undergo piezoelectric-induced strains. The basic idea of the method is to measure these strains in all three directions and to use these measured values to compute the piezoelectric tensile coefficients d_{311}, d_{322}, and d_{333}. This procedure requires an accurate assessment as to how the two rubber layers affect the strain response of the piezoelectric film when a voltage V is applied to the rubber-polymer film-rubber composite stack. The layers essentially influence the film response in two ways: electrically and mechanically. Electrically, the field E_p inside the polymer depends on t and h, as well as ε_p and ε_r, the dielectric constants of the polymer and of the silicone rubber, respectively; it is given by $E_p=V/(2t+2h\varepsilon_p/\varepsilon_r)$. Mechanically, the rubber imposes certain boundary conditions to the polymer film that can be estimated if the elastic constants of both materials are known. The polymer used in this study is a commercially available piezoelectric PVDF material for which dielectric and elastic constants are available. For the silicone rubber, its manufacturer specifies a value of $\varepsilon_r=2.7$ at 1 MHz; capacitance measurements revealed that this value remained the same in the kHz range. The Young's modulus of the rubber was also measured using a resonant rod technique[5] and was found to be equal to $Y_r=5.31$ MPa at 100 Hz. Finally, Poisson's ratio was computed using wave speed measurements at 1 MHz (longitudinal wave speed $c_L=1085.5$ m/s), and was found to be equal to $\nu_r=0.4993$. In Sec. I C we will detail how Y_r and ν_r are used to estimate the boundary conditions at the rubber/polymer interface.

B. Optical measurements

The strains associated with the piezoelectric response of the samples were optically measured using a laser Doppler vibrometer configured for out-of-plane (normal) displacement detection. This optical system is a one-probe version of a dual-beam interferometer previously described,[6] and it is shown in Fig. 2. The source is a 15-mW helium–neon laser whose beam is split into two beams (the ''reference'' and ''object'' beams) by an acousto-optic Bragg cell operating at 40 MHz. Each beam is coupled into single-mode optical fibers using X20 microscope objectives. The other end of the fibers is terminated by quarter-pitch gradient index (GRIN) lenses. The lens in the reference arm is used to collimate the light which is sent to one side of a beamsplitter cube; the lens in the object arm is used to focus the light onto the sample surface. The spot diameter of the incident beam on the sample is 80 μm. Depending on the nature of the surface, the incident light is either reflected or scattered back into a multimode optical fiber that sends the object beam through the cube so that it interferes with the reference beam. The interference produces a 40-MHz optical heterodyne signal. The out-of-plane motion of the sample creates a Doppler shift (phase modulation) in the 40-MHz signal. This phase shift is directly proportional to the displacement of the surface. This signal is then converted into an ac voltage using a Hewlett-Packard PIN photodiode. It is then amplified and

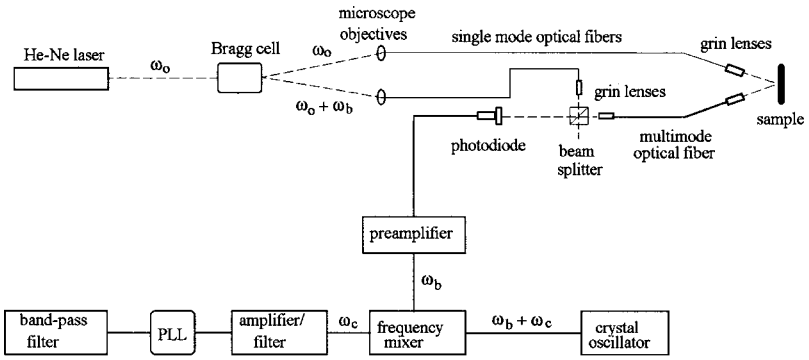


FIG. 2. Laser Doppler interferometer for out-of-plane displacement detection. ω_0 is the optical frequency; ω_b is the Bragg cell frequency (40 MHz); and ω_c is the carrier frequency of the signal to be demodulated (100 kHz).

combined (via a frequency mixer) with a 40.1-MHz sine wave from a crystal oscillator, in order to get a down-shifted 100-kHz heterodyne signal, which is then amplified and filtered. FM demodulation is used to extract the displacement information contained in the phase of the heterodyne voltage. This is done using a phase-locked loop (PLL) electronic circuit, which produces an output proportional to the surface velocity. After some band-pass filtering, the output of the PLL is displayed on a Tektronix digital oscilloscope (TDS 420) equipped with a GPIB board for data transfer to a microcomputer.

Figure 3 illustrates how the vibrometer is used in this work. When the sample (of dimensions $2L\times 2w\times 2t$) is excited with an ac voltage of amplitude V and frequency f, the optical probe is positioned to successively measure the resulting displacements $u_1(L)$, $u_2(w)$, and $u_3(t)$, along the 1-, 2-, and 3-directions, respectively. To measure $u_3(t)$, the probe is positioned along the 3-axis and it is focused on the aluminum foil, where the incident light is reflected into the receiving lens. For the other two measurements of $u_1(L)$ and $u_2(w)$, the probe is focused on the side of the polymer film where a small amount of white ink has been deposited, in order to scatter the incident light, as shown in Fig. 3 (the ink is deposited before the film is encapsulated in the silicone rubber).

C. Rayleigh–Ritz minimization procedure

This section describes how the experimental values of the displacements $u_1(L)$, $u_2(w)$, and $u_3(t)$ are used to compute d_{311}, d_{322}, and d_{333}. The chosen approach uses a Rayleigh–Ritz variational method. The first step is to assume admissible displacement fields $u_1(x_1,x_2,x_3)$, $u_2(x_1,x_2,x_3)$, and $u_3(x_1,x_2,x_3)$ for the polymer film, expressed in terms of the experimentally measured displacements $u_1(L)$, $u_2(w)$, and $u_3(t)$. The second step consists in calculating the total energy of the polymer sample undergoing first-order (piezoelectric) electromechanical deformations. The total energy functional is then minimized with respect to $u_1(L)$, $u_2(w)$, and $u_3(t)$. This is implemented in Mathcad, whose symbolic calculation capabilities allow us to specify the piezoelectric coefficients d_{3ii} as generic parameters (without any assigned numerical value). The minimization procedure results in expressions of $u_1(L)$, $u_2(w)$, and $u_3(t)$ as a function of d_{311}, d_{322}, and d_{333}. The experimentally measured values of $u_1(L)$, $u_2(w)$, and $u_3(t)$ are finally used in Mathcad to solve the system of three equations for the three unknowns d_{311}, d_{322}, and d_{333}. These steps are outlined next.

1. Admissible displacement fields

The Rayleigh–Ritz method requires the choice of trial functions for the displacement fields of the polymer film in the three directions. These functions must be compatible with the boundary conditions imposed to the film. Referring to Fig. 3, which shows the orientation of the coordinate system with its origin at the center of the sample, it is clear that the sides of the film perpendicular to the 1- and 2-axes are stress-free, whereas the surfaces of the film perpendicular to the 3-axis are subjected to a shear stress due to the presence of the silicone rubber. The aluminum foil is treated as a moving rigid surface: it does not expand laterally along the 1- or 2-direction, but it is free to move in the 3-direction. It is assumed that the inertial loading on the sample from the motion of the Al foil is negligible. It is further assumed that, away from the edges of the film, (1) u_1 is uniform across the film thickness $(2t)$ and along the film width $(2w)$; (2) u_2 is uniform across the film thickness $(2t)$ and along the film length $(2L)$; and (3) u_3 is uniform along the film length and the film width (these assumptions were verified using a finite element simulation, as will be discussed in Sec. II). Therefore, denoting the total elongations of the film along the 1- and 2-directions as ΔL and Δw, respectively, we have

$$u_1(x_1)=\frac{\Delta L}{2L}x_1, \tag{1}$$

$$u_2(x_2)=\frac{\Delta w}{2w}x_2. \tag{2}$$

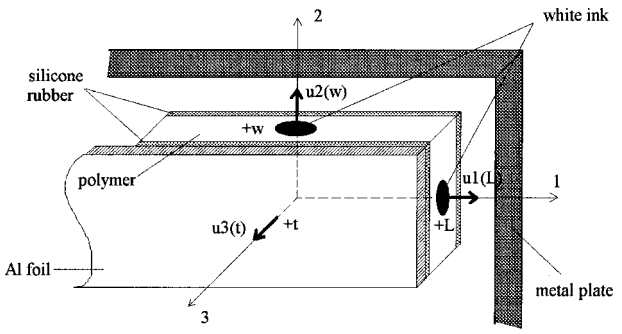


FIG. 3. Sample geometry showing axis orientation, measurement locations, and measured displacements.

Furthermore, if the total thickness change is denoted as Δt, we have

$$u_3(x_3)=\frac{\Delta t}{2}\left(1+\frac{x_3}{t}\right). \tag{3}$$

Thus, the experimentally measured displacements are

$$u_1(L)=\frac{\Delta L}{2}, \tag{4}$$

$$u_2(w)=\frac{\Delta w}{2}, \tag{5}$$

$$u_3(t)=\Delta t. \tag{6}$$

2. Energy functional

In this section, an expression for the energy of the sample is derived. Here the term ''sample'' refers to the whole system whose energy is calculated, and it includes the polymer film and the rubber layers. The samples are subjected to a constant electric field, as well as a constant temperature. Furthermore, the Rayleigh–Ritz method employed in this work requires the energy minimization to be performed with respect to the displacements. Given these assumptions, the most convenient choice of independent variables is therefore the strains S_{kl}, the electric fields E_m (all indices assume the values 1, 2, and 3), and the temperature T; corresponding to this choice, the thermodynamic potential function to consider in order to compute the energy density is the electric Gibbs free energy function G_2, defined by[7,8]

$$G_2=U-E_nD_n-TS, \tag{7}$$

where U is the internal energy density, the D_n are the electric displacements, and where S is the entropy. (One should note that the use of G_2 is purely dictated by the choice of the strains and the electric fields as the independent variables; had the stresses and electric fields been chosen instead, the thermodynamic potential function to use would have been the Gibbs free energy function G, similarly defined by Mason[7] as $G=U-S_{ij}X_{ij}-E_nD_n-TS$.) The differential of the internal energy density is given by

$$dU=X_{ij}\,dS_{ij}+E_n\,dD_n+T\,dS, \tag{8}$$

where the X_{ij} are the stresses. Therefore,

$$dG_2=X_{ij}\,dS_{ij}-D_n\,dE_n-S\,dT. \tag{9}$$

Using thermodynamic considerations, one can show that, under the assumptions of constant temperature and electric field, the energy G_2 is a minimum for the considered system. Furthermore, it is assumed that the thermal effects have a negligible impact on the deformations of the system, and the energy density of interest for the minimization procedure is therefore[9] [from Eq. (9)]

$$G_2=\tfrac{1}{2}X_{ij}S_{ij}-\tfrac{1}{2}D_nE_n. \tag{10}$$

This energy density now needs to be evaluated for the polymer film and for the rubber layers. Starting with the polymer, its complete electromechanical behavior is described by the following constitutive laws, written using the strains and the electric fields as independent variables:

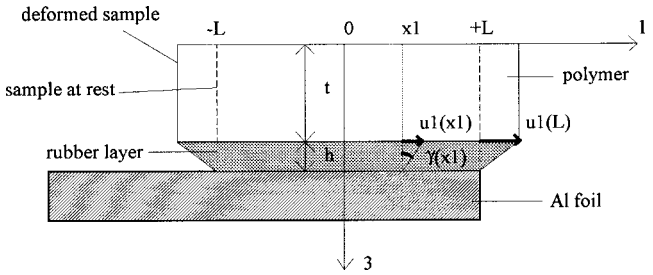


FIG. 4. Deformation at the polymer/rubber interface along the length of the sample. $\gamma(x_1)$ is the shear strain S_{13} inside the rubber layer.

$$X_{ij}=c^E_{ijkl}S_{kl}-e_{mij}E_m, \quad D_n=e_{nkl}S_{kl}+\varepsilon^S_{nm}E_m. \tag{11}$$

In these equations, the c^E_{ijkl} are the stiffness coefficients (at constant electric field), the ε^S_{nm} are the permittivity coefficients (at constant strain), and the e_{mij} are the piezoelectric coefficients relating electric displacement to strain. However, the d_{mij} (piezoelectric coefficients relating electric displacement to stress) are more widely used quantities, and they are related to the e_{mij} by

$$e_{mij}=d_{mpq}c^E_{pqij}. \tag{12}$$

Combining Eqs. (10)–(12) yields

$$G_2^{\text{poly}}=\tfrac{1}{2}S_{ij}c^E_{ijkl}S_{kl}-S_{ij}d_{mpq}c^E_{pqij}E_m-\tfrac{1}{2}E_n\varepsilon^S_{nm}E_m, \tag{13}$$

or, in matrix form,

$$G_2^{\text{poly}}=\tfrac{1}{2}(S)_t[c^E](S)-(E)_t[d][c^E](S)-\tfrac{1}{2}(E)_t[\varepsilon^S](E). \tag{14}$$

In Eq. (14), the first term is the elastic energy density, the second term is the mutual energy density (representing first-order electromechanical coupling), and the last term is the dielectric energy density. Since the last term does not depend on strain, the value of its derivative with respect to displacements is zero. Therefore, the dielectric energy does not affect the results in the Rayleigh–Ritz method and is omitted in what follows. The elastic and mutual energies of the piezoelectric film are found by integrating the corresponding energy densities over the film volume, and they are respectively given by

$$G^{\text{poly}}_{2,\text{elas}}=\frac{1}{2}\int_{-t}^{t}\int_{-w}^{w}\int_{-L}^{L}(S)_t[c^E](S)\,dx_1\,dx_2\,dx_3, \tag{15}$$

$$G^{\text{poly}}_{2,\text{mut}}=-\int_{-t}^{t}\int_{-w}^{w}\int_{-L}^{L}(E)_t[d][c^E](S)\,dx_1\,dx_2\,dx_3. \tag{16}$$

Next, the energy contribution of the rubber layers is examined. Since the silicone rubber used to prepare the samples is not a piezoelectric material, its Gibbs free energy density possesses only two terms: the elastic term and the dielectric term. The latter is irrelevant to this work because, like in the polymer case, it does not depend on strain. This leaves the elastic energy density as the only contribution to the total energy from the rubber layers. Figure 4 details the deformation in the 1-direction of one of the rubber layers. For small deformations, the shear strain S_{13} in the rubber layer is given by $S_{13}=\gamma(x_1)\cong\tan\gamma(x_1)=u_1(x_1)/h=(\Delta L/2hL)x_1$. The corresponding shear stress is therefore given by

$$X_{13}=\mu_r S_{13}=\mu_r \frac{\Delta L}{2hL}x_1, \tag{17}$$

where μ_r is the silicone rubber shear modulus; $\mu_r=Y_r/2(1+\nu_r)$. Similarly, the shear stress in the rubber layer corresponding to the deformation of the film along the 2-direction is

$$X_{23}=\mu_r S_{23}=\mu_r \frac{\Delta w}{2hw}x_2. \tag{18}$$

Since the rubber layer is in a state of pure shear, its elastic energy density is given by

$$G_{2,\text{elas}}^{\text{rubber}}=\tfrac{1}{2}(X)_t(S), \tag{19}$$

where $(X)_t=(0\ 0\ 0\ X_{23}\ X_{13}\ 0)$ is the rubber stress vector and $(S)_t=(0\ 0\ 0\ S_{23}\ S_{13}\ 0)$ is the rubber strain vector (the subscript t denotes the transpose of the vectors). The total elastic energy from both rubber layers is therefore given by

$$G_{2,\text{elas}}^{\text{rubber}}=8h\int_0^w\int_0^L \frac{1}{2}(X)_t(S)\,dx_1\,dx_2. \tag{20}$$

Equation (20) can be easily evaluated and yields

$$G_{2,\text{elas}}^{\text{rubber}}=\frac{\mu_r}{3}\frac{Lw}{h}[(\Delta L)^2+(\Delta w)^2]. \tag{21}$$

Because of the quasistatic assumption of the model, the kinetic energy of the system is neglected. The total energy of the sample relevant to the Rayleigh–Ritz minimization procedure is therefore given by the sum of the quantities defined in Eqs. (15), (16), and (21):

$$G_2=G_{2,\text{elas}}^{\text{poly}}+G_{2,\text{mut}}^{\text{poly}}+G_{2,\text{elas}}^{\text{rubber}}. \tag{22}$$

The strain vector (S) inside the polymer is defined in terms of ΔL, Δw, and Δt, through the displacement fields [Eqs. (1)–(3)]:

$$\begin{aligned}(S)_t&=\left(\frac{du_1}{dx_1}\quad\frac{du_2}{dx_2}\quad\frac{du_3}{dx_3}\quad 0\quad 0\quad 0\right)\\&=\left(\frac{\Delta L}{2L}\quad\frac{\Delta w}{2w}\quad\frac{\Delta t}{2t}\quad 0\quad 0\quad 0\right).\end{aligned} \tag{23}$$

The electric field vector is given by

$$(E)_t=(0\quad 0\quad E_p)=\left(0\quad 0\quad \frac{V}{2t+2h\varepsilon_p/\varepsilon_r}\right). \tag{24}$$

The stiffness tensor is defined by the elastic properties of the polymer material. Finally, for poled, uniaxially stretched piezoelectric polymers such as PVDF, the piezoelectric tensor is given by

$$[d]=\begin{bmatrix}0 & 0 & 0 & 0 & d_{113} & 0\\ 0 & 0 & 0 & d_{223} & 0 & 0\\ d_{311} & d_{322} & d_{333} & 0 & 0 & 0\end{bmatrix}. \tag{25}$$

However, for an electric field vector in which only $E_3\neq 0$, d_{113} and d_{223} do not play any role in the deformation of the film. Therefore, G_2 is a function of ΔL, Δw, Δt, d_{311}, d_{322}, and d_{333} only. This functional can be computed in Mathcad, defining the d_{3ii} coefficients as symbolic parameters. G_2 [as given by Eq. (22)] is then minimized with respect to the displacements ΔL, Δw, and Δt:

$$\frac{dG_2}{d\Delta L}=0,\quad \frac{dG_2}{d\Delta w}=0,\quad \frac{dG_2}{d\Delta t}=0. \tag{26}$$

The above system of equations is solved symbolically for ΔL, Δw, and Δt in Mathcad, yielding expressions for the displacements as a function of the d_{3ii} coefficients:

$$\begin{aligned}\Delta L&=A_1d_{311}+A_2d_{322}+A_3d_{333},\\ \Delta w&=B_1d_{311}+B_2d_{322}+B_3d_{333},\\ \Delta t&=C_1d_{311}+C_2d_{322}+C_3d_{333},\end{aligned} \tag{27}$$

where the A's, B's, and C's are constants depending on the experimental parameters (sample dimensions, electrical excitation,...). Finally, system (27) is solved for the d_{3ii} coefficients, using the experimentally measured values of ΔL, Δw, and Δt. The solution to this system constitutes an absolute measurement of the three piezoelectric coefficients of the polymer material: since the mechanical boundary conditions are automatically taken into account in the Rayleigh–Ritz minimization procedure, the new method provides a way to directly evaluate the piezoelectric coefficients that represent the intrinsic electromechanical behavior of the polymer.

II. VALIDATION OF THE METHOD BY FINITE ELEMENTS

Finite elements were used in order to assess the reliability of the method. The finite element code ATILA[10] was utilized to compute the response of a piezoelectric polymer film modeled in the experimental configuration described in this article. All the relevant parameters were assigned specific numerical values and the output of the code provided the approximate displacements undergone by the film. These displacements were then used in the Rayleigh–Ritz analytical model to compute the corresponding piezoelectric coefficients. Finally, the resulting d_{3ii} values obtained from Mathcad were compared to the values input into the ATILA code.

A. Two-dimensional finite element model

A finite element program was written in ATILA to model the behavior of a piezoelectric film sandwiched between two layers of silicone rubber. Due to memory space limitation on the particular machine on which the program was implemented, the model had to be restricted to a two-dimensional, plain strain analysis. Figure 5 shows the geometry of the model. The polymer sample is a 110×10^{-6}-m (4.33 mil) thick and 500×10^{-6}-m-long film sandwiched between two 11.43×10^{-6}-m (0.45 mil) thick layers of silicone rubber. The thickness of the aluminum layer is 25.4×10^{-6}-m (1 mil). The rubber and aluminum layers are divided into 40 elements lengthwise; the polymer layer is divided into 40 elements lengthwise and 4 elements thicknesswise. The prescribed mechanical boundary conditions are of zero displacement in all directions along the back plane of the upper rubber layer; this corresponds to the metal plate to which the

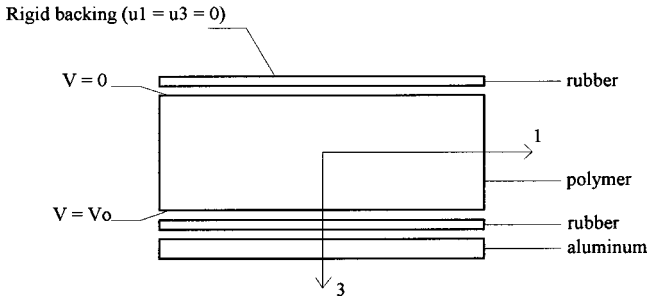


FIG. 5. Geometry of the ATILA finite element model.

sample is attached. Since the silicone rubber is only defined as an elastic material in ATILA (i.e., its dielectric properties are not taken into account by the code), the driving voltage Vo is directly applied between the two planes corresponding to the rubber/polymer interface ($Vo=42$ V). The driving frequency f is 2000 Hz. The polymer material is defined in the code using typical experimental values for the elastic constants of stretched PVDF, as well as the following set of piezoelectric constants:[11]

$$d_{311}=23\times10^{-12}\ \text{m/V},$$
$$d_{322}=1.5\times10^{-12}\ \text{m/V}, \tag{28}$$
$$d_{333}=-33\times10^{-12}\ \text{m/V}.$$

Figure 6 shows the results of the finite element simulation. Figures 6(a) and 6(b) confirm the validity of the choice of linear functions for the admissible displacement fields [Eqs. (1) and (3)]. Figures 6(c) and 6(d) validate the assumptions that the thickness change is mostly uniform along the film length and that the elongation is uniform across the film thickness. The ATILA results indicate that $\Delta t=-1.337\times10^{-9}$ m at the center of the film [cf. Fig. 6(c), $\Delta t=u_3$ (Al foil)] and that $\Delta L=4.372\times10^{-9}$ m [cf. Fig. 6(d), $\Delta L=2u_1(L/2)$]. These results remained unchanged when the same code was run without the aluminum layer, indicating that the inertial loading from the foil is negligible.

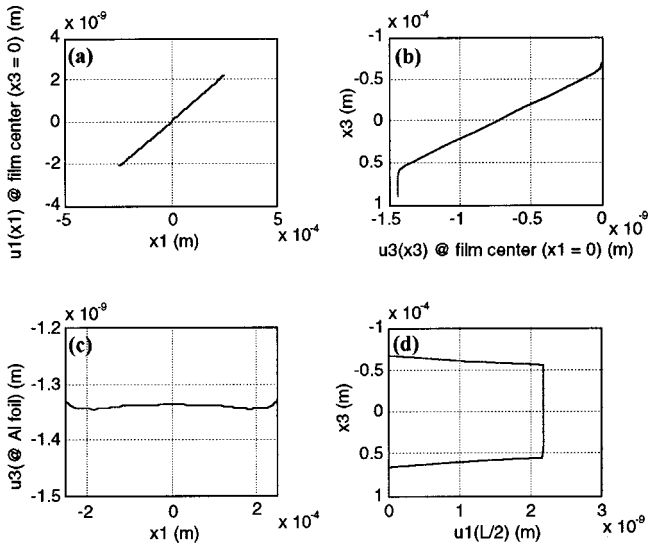


FIG. 6. Finite element simulation results: (a) displacement u_1 at mid-thickness; (b) displacement u_3 at mid-length; (c) thickness change of the film; and (d) displacement at one end of the film.

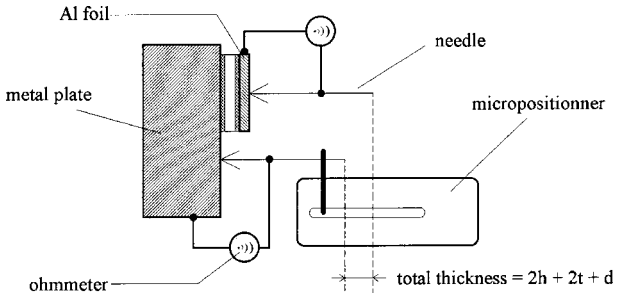


FIG. 7. Nominal sample thickness measurement. $2t$ is the polymer thickness (110 μm), d is the aluminum foil thickness (17 μm), and h is the thickness of each rubber layer.

B. Two-dimensional analytical model

In order to use the ATILA results, a two-dimensional, plain strain version of the procedure described in Sec. I C was developed. In the corresponding Mathcad program, only the displacement fields $u_1(x_1)$ and $u_3(x_3)$ are defined, and the procedure therefore yields ΔL and Δt. These two quantities are used to compute d_{311} and d_{333} (d_{322} is defined numerically). The elastic energy of the rubber layers depends only on ΔL, and is given by

$$G_{2,\text{elas}}^{\text{rubber}}=\frac{\mu_r}{6}\frac{L}{h}(\Delta L)^2. \tag{29}$$

The rest of the program is similar to the three-dimensional case. In the plain strain case, the Mathcad output yields the following system:

$$\Delta L=A_1'd_{311}+A_3'd_{333}+A_2',\quad \Delta t=C_1'd_{311}+C_3'd_{333}+C_2', \tag{30}$$

where the A's and C's depend on the experimental parameters, and where A_2' and C_2' also depend on d_{322}. After setting the different parameters to the values corresponding to the ATILA model, the system of equations (30) was solved using the values of ΔL and Δt provided by the finite element code, giving the following results:

$$d_{311}=23.02\times10^{-12}\ \text{m/V},\quad d_{333}=-32.93\times10^{-12}\ \text{m/V}. \tag{31}$$

The values in Eq. (31) are very close to the values input into ATILA [Eq. (28)]. It is therefore concluded that the analytical approach based on the Rayleigh–Ritz method can be successfully implemented to experimentally measure piezoelectric coefficients. The next section presents the results of such measurements.

III. EXPERIMENTAL RESULTS ON PVDF SAMPLES

The piezoelectric material used in this study is a stretched (draw ratio: 5) PVDF polymer, commercially available from AMP Inc. (Piezo Film Sensors Division). Rectangular samples (approximately 3 mm by 6 mm) were cut with a razor blade from 110×10^{-6}-m-thick sheets. The rubber-polymer–rubber-aluminum composite was assembled using a soft elastic ball that compressed the stack with a uniform pressure; the assembly was allowed to rest for 24 h in order to let the silicone rubber cure. Next, a needle attached to a micropositionner (moving in steps of 10^{-6} m) was utilized to

TABLE I. Typical displacements measured by the optical system.

Driving voltage (V)	Measured ΔL (10^{-9} m)	Measured Δw (10^{-9} m)	Measured Δt (10^{-9} m)
60	7.6	3.4	−1.0

measure the thickness h of the rubber layers, and to make sure that h was uniform over the sample surface. This arrangement is depicted in Fig. 7. Here h varied between 5×10^{-6} m and 12×10^{-6} m from one sample to another, depending on the pressure applied with the ball. The samples were driven with an ac voltage at frequencies ranging from 500 to 2000 Hz, and displacements in the three directions were measured with the optical probe. Typical experimental values are shown in Table I. Table II shows the 2000-Hz experimental results for the piezoelectric coefficients at room temperature: the values obtained with the new method (averaged from results on six samples), along with their standard deviations, are compared against the values available from the manufacturer of the material and from published values for similarly stretched PVDF. Data for d_{322} is not available from the manufacturer, who specifies a margin of error of about 20% for the other two piezoelectric constant values. It is unclear whether these manufacturer values result from experimental measurements or from a theoretical model. The last row of Table II gives an indication of the spread in the published data. Wang[12] remarked that experimental evidence has shown that the piezoelectric coefficients vary with polymer processing methods, even for samples with the same composition and studied under the same experimental conditions. It is therefore concluded that the experimental data presented in Table II represent an accurate assessment of the piezoelectric activity of PVDF. Results obtained over the range 500–2000 Hz were identical to those presented in Table II. Repeatability was established by performing the same measurements on different samples (cf. standard deviation values in Table II) and by measuring displacements at various locations on a given sample: these latter measurements also yielded results identical to those from Table II, but with a negligible standard deviation. Therefore, the small discrepancies that can be observed from one set of data to the other are mainly due to imperfections in the sample preparation.

Next, the sensitivity of the measured d_{3ii} coefficients to the displacements and to the elastic constants input into Mathcad was studied. Results of this study are presented in Tables III and IV, respectively. Table III shows how an increase of 5% in the measured displacements (which corresponds to the largest estimate of the optical system overall uncertainty) affects the values of the computed piezoelectric coefficients. Table IV indicates which elastic constants affect the d_{3ii}. To get these results, values of the stiffness coefficients were successively modified in Mathcad, corresponding to an increase by 10% in one of the Young's moduli Y_i or Poisson's ratios ν_{ij}. This 10% variation of the elastic constants is considered typical of the spread in their published values. In both tables, the ''/'' symbols indicate a change of less than 1%. The d_{322} coefficient is the most sensitive one to both displacement and elastic constant values, which is consistent with the fact that it is one order of magnitude smaller that the other two piezoelectric coefficients and that it is therefore the most difficult one to measure with accuracy. However, the variations displayed in Tables III and IV are small (especially for d_{333}), and it is concluded that the piezoelectric coefficients are relatively insensitive to reasonable variations in the values of the measured displacements and of the elastic constants. Therefore, the uncertainties associated with the measuring apparatus and the spread in the published data of elastic constants do not significantly affect the accuracy of the method.

TABLE II. Comparison of experimentally measured piezoelectric coefficients at 2000 Hz (obtained from six samples) with published data.

	d_{311} (10^{-12} m/V)	d_{322} (10^{-12} m/V)	d_{333} (10^{-12} m/V)
Experimental	19	3	−37
Standard deviation	4	2	5
Manufacturer data	23±5	/	−33±7
Published data (Ref. 12)	12–28	1–5	−15–−33

TABLE III. Results sensitivity to measured displacements.

Increase of 5% in measured displacements	% change in d_{311}	% change in d_{322}	% change in $\lvert d_{333}\rvert$
ΔL	+5	−10	+2
Δw	/	+14	/
$\lvert\Delta t\rvert$	/	/	+3

Finally, the effects of varying the value of the rubber layer thickness h in Mathcad was studied. h is a critical parameter because it affects both the mechanical boundary conditions and the electrical excitation. The estimated error on the h measurement is 1 μm, which corresponds to a percent error comprised between 10% and 20%, depending on the measured value of the rubber layer. Considering the least favorable case of an increase by 20% in the value of h used in Mathcad, the results change as follows: d_{311} decreases by 10%, d_{322} increases by 9% and $\lvert d_{333}\rvert$ decreases by 2%. Therefore, the error associated with the measurement of h does not compromise the accuracy of the method.

These results show that the new method presented in this paper is a valid and useful approach to piezoelectric coefficient measurements on thin polymer films.

IV. EXTENSION OF THE METHOD TO ELECTROSTRICTIVE MATERIALS

The topic of electromechanical activity in polymers has received renewed attention since large electrostrictive strains have been reported for several polyurethane materials.[3] Dif-

TABLE IV. Results sensitivity to PVDF elastic constants.

Increase of 10% in elastic constants	% change in d_{311}	% change in d_{322}	% change in $\lvert d_{333}\rvert$
Y_1	−9	/	/
Y_3	/	/	−4
ν_{12}	/	−18	/
ν_{13}	/	/	+4

ferent mechanisms have been proposed to explain electrostriction in polymers,[13,14] but the origins of the mechanism are not yet completely understood, and the data available in the literature are conflicting.[3,15] An in-depth investigation of polyurethane electrostriction requires a dependable method for accurately measuring electrostrictive coefficients, which the technique presented in this paper can provide. Below is an outline of how the method is to be modified in order to be used in the electrostrictive case.

Polyurethane materials differ from piezoelectric polymers such as PVDF in two essential ways: (1) they are significantly softer than PVDF (their Young's modulus is lower by about two orders of magnitude) and (2) they are electrostrictive. The first fact stated implies that the silicone rubber has a greater mechanical influence on the polyurethane, effectively modifying the admissible displacement fields [Eqs. (1)–(3)]. In the polyurethane case, a quadratic dependence on x_3 must be assumed for u_1 and u_2, as well as a cubic dependence on x_3 for u_3. For electrostriction, using the same formalism as before, the complete constitutive laws [Eqs. (11)] become

$$\begin{aligned} X_{ij} &= c^E_{ijkl} S_{kl} - e_{mij} E_m - Q_{moij} E_m E_o , \\ D_n &= e_{nkl} S_{kl} + \varepsilon^S_{nm} E_m + O^S_{nmo} E_m E_o + 2 Q_{nmkl} E_m S_{kl} , \end{aligned} \tag{32}$$

where the electrostrictive coefficients Q_{moij} represent the variations of the dielectric coefficients with respect to the strains and where the O^S_{nmo} are the second-order equivalent to the ε^S_{nm} coefficients. Furthermore, the electrostrictive coefficients M_{mokl} analogous to the d_{mkl} coefficients [Eq. (12)] are given by

$$Q_{moij} = M_{mokl} c^E_{klij} . \tag{33}$$

One should note that polyurethanes do not exhibit any first-order piezoelectric behavior, and that the e_{mij} coefficients are therefore zero, leaving the M_{mokl} as the only unknowns. The total energy of the sample has the same form as Eq. (22), except that the mutual energy contribution is a second-order term involving the electrostrictive tensor $[M]$, and that another second-order term, corresponding to the Maxwell stress effect, must be added. When the sample is excited with both an ac voltage at frequency f and a large dc bias voltage, it undergoes electrostrictive deformations at frequencies f and $2f$. Optical measurement of the displacements of the sample at any of these two frequencies yields the necessary information to compute M_{3311}, M_{3322}, and M_{3333} from the Rayleigh–Ritz procedure. As in the piezoelectric case, the values of the M_{33ii} coefficients are material parameters, since both the mechanical boundary conditions and the effects of the Maxwell stress-induced deformations are taken into account in the energy approach.

This method has already been extended by the authors to the case of measurements on soft electrostrictive polyurethanes: the Rayleigh–Ritz approach has been modified to take into account the second-order electromechanical coupling, and finite element tests have been successfully performed. Preliminary experimental data has shown that measurable displacements can be obtained on the polyurethanes, and electrostriction measurements are currently in progress. A subsequent paper will detail the second-order version of the method introduced here and will present electrostrictive coefficient data measured on various types of polyurethanes.

V. CONCLUSIONS

Finite element analysis and experimental measurements on PVDF samples established the validity of the new method presented in this paper. It was shown that the method is a useful tool for the quasistatic evaluation of piezoelectric coefficients on thin films, because the knowledge of the exact mechanical boundary conditions allows for the direct determination of a complete set of three piezoelectric material constants. The technique can also be extended to perform similar measurements of electrostrictive coefficients on soft polyurethane materials.

ACKNOWLEDGMENTS

This work was partially supported by the Office of Naval Research, Code 332. The authors thank Dr. Philip B. Abraham for his support and his interest in this project.

[1] M. B. Moffett, D. Ricketts, and J. L. Butler, ''The effect of electrode stiffness on the piezoelectric and elastic constants of a piezoelectric bar,'' J. Acoust. Soc. Am. **83**, 805–811 (1988).
[2] T. Furukawa and T. T. Wang, ''Measurements and Properties of Ferroelectric Polymers,'' in *The Applications of Ferroelectric Polymers*, edited by T. T. Wang, J. M. Herbert, and A. M. Glass (Chapman and Hall, New York, 1988), Chap. 5, p. 95.
[3] M. Zhenyi, J. I. Scheinbeim, J. W. Lee, and B. A. Newman, ''High Field Electrostrictive Response of Polymers,'' J. Polym. Sci., Part B: Polym. Phys. **32**, 2721–2731 (1994).
[4] F. M. Guillot, J. Jarzynski, and E. Balizer, ''Electromechanical response of polymer films by laser Doppler vibrometry,'' J. Acoust. Soc. Am. **103**, 1421–1427 (1998).
[5] W. M. Madigosky and G. F. Lee, ''Improved resonance technique for materials characterization,'' J. Acoust. Soc. Am. **73**, 1374–1377 (1983).
[6] F. M. Guillot and J. Jarzynski, ''A New Method for Measuring the Bulk Modulus of Compliant Acoustic Materials,'' J. Sound Vib., in press.
[7] W. P. Mason, *Piezoelectric Crystals and Their Applications to Ultrasonics* (Van Nostrand, New York, 1950), p. 34.
[8] C. S. Lynch and R. M. McMeeking, ''Finite Strain Ferroelectric Constitutive Laws,'' Ferroelectrics **160**, 177–184 (1994).
[9] O. B. Wilson, *Introduction to Theory and Design of Sonar Transducers* (Peninsula, Los Altos, 1988), pp. 59–60.
[10] *ATILA: Finite-Element Code for Piezoelectric and Magnetostrictive Transducer Modeling*, User's Manual (Institut Supérieur d'Electronique du Nord, Lille, 1993).
[11] H. Schewe, ''Piezoelectricity of Uniaxially Oriented Polyvinylidene Fluoride,'' IEEE Ultrason. Symp. Proc. **1**, 519–524 (1982).
[12] H. Wang, ''Electromechanical Effects in Polymeric Materials,'' Ph.D. Dissertation, Pennsylvania State University, 1994, p. 25.
[13] Y. M. Shkel and D. J. Klingenberg, ''Material Parameters for Electrostriction,'' J. Appl. Phys. **80**(8), 4566–4572 (1996).
[14] J. Su, Q. M. Zhang, and R. Y. Ting, ''Space-Charge-Enhanced Electromechanical Response in Thin-Film Polyurethane Elastomers,'' Appl. Phys. Lett. **71**(3), 386–388 (1997).
[15] H. Wang, ''Electromechanical Effects in Polymeric Materials,'' Ph.D. Dissertation, Pennsylvania State University, 1994, pp. 146–159.

Integrated photoelectric device made of a piezoelectric ceramic exhibiting pyroelectricity and an internal photoeffect

Li Quanlu
Institute of Applied Acoustics, Shaanxi Normal University, Xi'an, 710062, People's Republic of China

(Received 16 February 1999; revised 18 September 1999; revised 28 December 1999; accepted 3 May 2000)

This paper reports on the preparation of a piezoelectric ceramic, which also exhibits pyroelectricity and an internal photoeffect. It then describes the design and trial production of a new integrated photoelectric device using this ceramic with a piezoelectric transformer, a solar cell, and an inverter. Important applications of this device are given. © *2000 Acoustical Society of America.* [S0001-4966(00)02108-1]

PACS numbers: 43.38.Fx, 43.38.Ar [SLE]

INTRODUCTION

Rapid development of modern electronics, photoelectric techniques, solar energy applications, and other technologies places an urgent demand for new multifunctional materials. The objective is to be able to apply their intersected coupling effects among the basic energy types, such as electricity, acoustics, heat, and optics. A piezoelectric ceramic, which also exhibits pyroelectricity and an internal photoelectric effect, is the example chosen for this study. Starting with basic theory, we carry the development of this new type of material through to the design and experiment of a new integrated photoelectric device. This new device includes a piezoelectric ceramic transformer, an inverter, and a solar cell, and it applies the intersecting couplings of piezoelectricity, pyroelectricity, and an internal photoelectricity in the multifunctional piezoelectric ceramic. In this novel device, when light illuminates the solar cell, an output voltage (dc photocurrent) is sent to the inverter, where it is converted into an ac output that is fed to the ceramic transformer which steps up the voltage. The output voltage may be used to correct for power failures, to provide high voltage for civil electronic apparatus, equipment, etc.

I. FUNDAMENTAL CONCEPTS

This paper is mainly concerned with piezoelectricity, pyroelectricity, and internal photoeffect, which are briefly described below.

A. Piezoelectricity

Piezoelectricity is the phenomenon whereby electric dipoles are generated in certain anisotropic materials when subjected to a mechanical stress. This property is called the direct piezoelectric effect, and it was discovered in 1880 by Jacques and Pierre Curie. The same materials exhibit the inverse effect, in that they display a dimensional change under the influence of an electric field. The piezoelectric materials observed initially were those that occurred in nature, and they include quartz, tourmaline, and Rochelle salt. Man has been able to create newer piezoelectric materials, some of which are also producible as single crystal (e.g., $NaKC_4H_4O_6 \cdot H_2O$, Li_2SO_4, $LiNbO_3$, and $NaBa_2Nb_5O_{15}$) and others which become piezoelectric upon applying a biasing electric field, such as piezoelectric ceramics [e.g., PZT, P(MN)ZT, and $BaTiO_3$], piezopolymers (e.g., piezoelectric rubbers, piezoelectric plastics, piezoelectric resin such as polyvinylidene fluoride), piezoelectric composites (e.g., ceramic and polymer), piezoelectric liquid crystals, piezoelectric thin film, piezoelectric biomaterials, and piezoelectric biological film. All of these are used as reversible electroacoustical or electromechanical transducing materials in applications as transducers, sensors, actuators, etc.

B. Pyroelectricity

The polarization observed in piezoelectric materials arises from an unbalancing of the permanent dipole moments by the applied stress. Relative to piezoelectricity, in certain materials, a polar axis exists in the absence of an applied stress. When such materials are heated or cooled, thermal dilation changes the magnitude of the polarization, and consequently a change appears across the polar axis. This phenomenon is called pyroelectricity. Wurtzite (hexagonal ZnS), for example, has a polar axis and, as a result, is both pyro- and piezoelectric. In contrast, Sphalerite (cubic ZnS) does not have a polar axis in the absence of an applied stress, therefore, it is only piezoelectric. Among existing materials which exhibit pyro- and piezoelectricity there are modified PZT, PL_xZT, PSZT, $BaTiO_3$ ceramics, GASH, $LiNbO_3$, $Li_2SO_4 \cdot H_2O$, $LiTaO_3$, $NaNO_2$, $Sr_xBa_{1-x}Nb_2O_6$, TGS, PVDF thin film, and others. They are used as thermal detectors, piezoelectric pyrometers, pyroelectric detectors, and optical or infrared detectors.

C. Photoelectricity

The photoelectric effect is the phenomenon whereby the substances exhibit a change of some of their electrical characteristics when they are subjected to light irradiation. Photoelectric effects may be further classified as the photoconductive effect, the photovoltaic effect, and the photoemissive effect. The first two are generated in the interior of the body and are referred to as the internal photoeffect. The last, for

which a bound electron is ejected from the surface of a solid or liquid due to light irradiation, is called the external photoeffect. The internal photoeffect is mostly generated in the interior of semiconductors, while the external photoeffect is mostly generated in metals.

1. Photoconductive effect

Many substances exhibit a marked increase in electric conductivity and temperature of their carrier when illuminated by light; this phenomenon is called photoconductive effect. Photoresistors are an application of this effect.

2. Photovoltaic effect

When some substances are irradiated by light, the resultant electromotive force (EMF) that is produced generates a boundary of two substances. This effect is called the photovoltaic effect. Photocells are an application of this effect.

3. Photoemissive effect

The phenomenon of emission of electrons by certain materials upon exposure to light radiation, when the electrons have absorbed enough light energy, is called the photoemissive effect. Photoemissive cells, phototubes, and photomultipliers are an application of this effect.

II. EXPERIMENTAL PROCEDURES

The present research is on piezoelectric ceramics, which also exhibit pyroelectricity and an internal photoelectric effect;[1–3] these materials are selected and formulated, their properties are analyzed, and an experimental basis is established for their application. Multifunctional integrated devices (multifunctional modules), and research and development of integrated devices using solar energy[4] provide a theoretical basis.

A. Preparation of materials

In the present study, we have applied the theory and practice of molecular design to determine the correct formulation using dry and wet methods, such as sandwich forming by dry pressing, sintering, and other methods. The main procedures in the development of formulation included: (a) designing the formulation; (b) selecting and mixing the raw materials; (c) utilizing the chemical coprecipitation method and classified calcination, such as sintering, grinding, granulation, and sandwich forming. The final steps included optical treatment, side-face polarizing, and testing to determine suitability of the final product as a piezoelectric, pyroelectric, and photoelectric material deemed an acceptable multifunctional test sample.

B. Properties of materials

Through experimentation we have sought the proper formulation and the appropriate technological process to achieve a piezoelectric ceramic which would also exhibit pyroelectricity and the internal photoeffect. If successful, this ceramic material would be of a new type not presently available. At the present time several well known piezoelectrics are available such as PZT, modified $BaTiO_3$, etc., as PZT series, PL_xZT, $Pb(Sn_{1/2}Sb_{1/2})ZrTiO_3$ series, $BaTiO_3$, $PbTiO_3$, $Sr_xBa_{1-x}Nb_2O_6$, $LiTaO_3$, $NaBa_2Nb_5O_{15}$, $KSr_2Nb_5O_{15}$, $LiNbO_3$, $KNbO_3$, $KTa_xNb_{1-x}O_3$, and others. Although some of these piezoelectric ceramic or piezoelectric single-crystal materials may possess, in varying degrees, the pyroelectric and/or the photoelectric effect, they have been found unable to exhibit all three effects on a single ceramic with acceptable performance. However, in the present work, we have developed and screened some useful materials which exhibit the three parameters being sought. Indeed, through experimentation, we have achieved a fundamental formulation for a multifunctional piezoelectric ceramic, which can be used for pyroelectric and photoelectric applications. This formulation can be written as follows:

$$m[\mathrm{Pb}_x\mathrm{Sr}_y\mathrm{Ba}_z(\mathrm{Zr}_u\mathrm{Ti}_v)\mathrm{O}_3]+n\mathrm{BiFeO}_3$$

$$(m+n=1,\ \ m=0.995-0.998;\ \ x+y+z=1,$$

$$x=0.93-0.96,\ \ y=0.02-0.035;\ \ u+v=1,$$

$$u=0.54-0.57).$$

Added to the total weight are 0.1%–0.3% La_2O_3 and 0.1%–0.3% Nb_2O_5, and a small amount of Nd_2O_3 and In_2O_3.

The abovementioned formulation is only an experimental one. Further alterations are to be done and the technology of processing is to be improved so that the finished multifunctional piezoelectric ceramic can be qualified for use in piezoelectricity, pyroelectricity, and internal photoelectricity.

C. Applications

The development of a piezoelectric ceramic transformer, which utilizes the multifunctional material described previously, provides a device capable of a substantial voltage gain and high voltage capabilities. In this paper this device generates an increase of the output voltage from a solar cell illuminated by sunlight.

1. The piezoelectric ceramic transformer

A horizontal–vertical-type ceramic transformer was constructed with the multifunctional piezoelectric ceramic described in Sec. II B. Figure 1 illustrates the device construction. The ceramic slab is divided into two parts, namely, a left side and a right side, as shown. Each side of the transformer consists of a length, *l*, a thickness, *t*, and a width, *w*. The upper and lower surfaces of the left side are metallized using sinter-permeate silver and the piezoelectric material is polarized in the thickness direction. The right side is polarized in the length, *l*, direction, while the area of the slab thickness of the left side is also metallized with sinter-permeate silver along the length direction. In addition, the upper surface of the left side is designed to be a solar cell. If an ac voltage is applied to the input end of the transformer, vibrational motion will be generated in the length direction of the ceramic due to the electrostrictive effect. At same time, the resultant strain on the left side of the transformer will generate a voltage output due to the direct piezoelectric effect. Due to the operational characteristics of the device, we refer to the left side as the ''driver'' and the right side as the ''electrical generator.'' At resonance the strain in the left

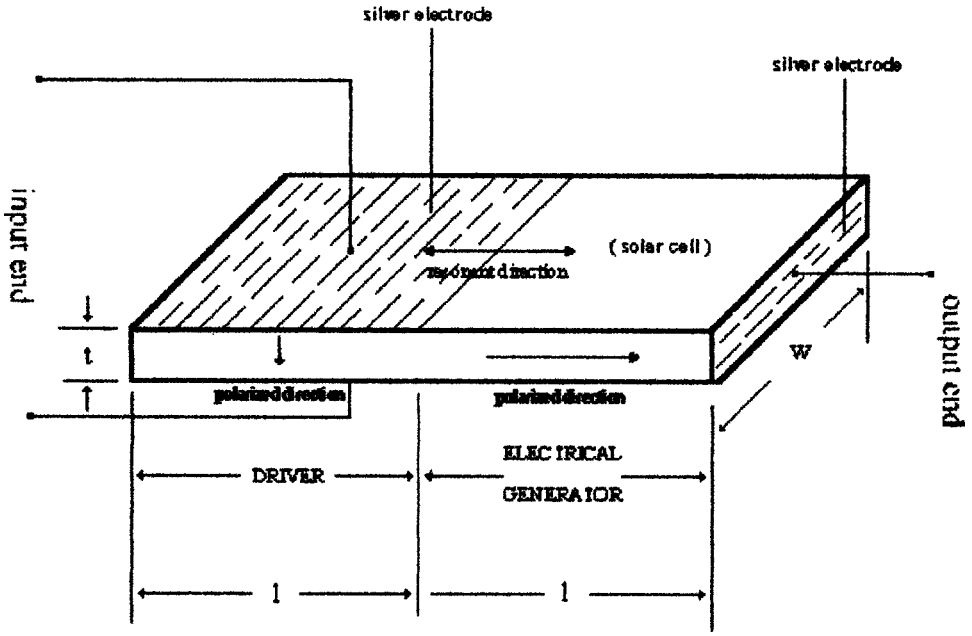


FIG. 1. Schematic diagram of horizontal–vertical-type ceramic transformer.

side of the transformer can be considerable, thus generating a fairly high voltage at the output. Moreover, since the entire length, $2l$, of the ceramic transformer is greater than the width, w, the output voltage becomes greater than the input. The operating frequency of the ceramic transformer depends on the sound velocity, v, in the material and the total length, $2l$, of the piezoelectric slab. For example, at half-wave resonance, the center frequency, f_0, of the ceramic transformer is[5]

$$f_0 = v/4l. \tag{1}$$

Based on the electromechanical equivalent circuit of the ceramic transformer, the ratio of the boosted voltage under no load conditions can be calculated and is given by

$$\frac{V_2}{V_1} = \frac{4}{\pi^2} \cdot Q_m k_{31} k_{33} \cdot \frac{l}{t} \cdot \frac{C_0}{C_0 + C}, \tag{2}$$

where V_1 and V_2 are the input voltage and output voltage of the piezoelectric ceramic transformer, respectively, Q_m is the mechanical quality factor of piezoelectric ceramic material, k_{31} and k_{33} are the electromechanical coupling factors of the piezoelectric ceramic material in the horizontal and vertical directions, respectively, l is the length of each half of the ceramic piece, t is the thickness of the ceramic piece, and C_0 and C are the equivalent distributed capacitances connected to the load. The piezoelectric ceramic transformer can operate at either the half-wave or full-wave modes. In addition, the ceramic transformer works under high voltage–low current conditions, and, in contrast with iron and winding transformers, has the advantage of small volume, light weight, and simple structure and manufacturing techniques. In particular, it cannot produce jamming magnetic fields nor burns in working conditions.

2. The solar cell

Solar cells are used to convert radiant energy directly into electrical energy. The cell consists of a thin layer of *p*-type silicon on an *n*-type silicon base. The *n*-type and *p*-type regions are heavily doped so that the resistance of the cell is small. There is an optimum thickness of the *p*-region so that as much as possible of the light is absorbed near the junction. When a photon is absorbed in the *p*- or *n*-region, it can create a hole–electron pair. Usually the hole and electron quickly recombine, but in a solar cell the internal electric field in the depletion layer at the *p*-*n* junction (directed from the *n*-side to the *p*-side) can separate the hole and electron before they have a chance to recombine. Electrons produced in the *p*-side that wander near the junction are pulled into the *n*-side, and holes produced near the junction in the *n*-side are pulled into the *p*-side, resulting in an electrical current across the junction. In flowing through an external load, the charge carriers do work, which has come directly from the energy of the incident photon. As shown in Fig. 1, the solar cell is located on the right side of the multifunctional piezoelectric ceramic. One may produce a solar cell with silicon by means of the chemical vapor deposition method. The design of the solar cell is shown in Fig. 2.

3. The inverter

Since the efficiency of a solar cell is quite low and the surface area of the cell in our integrated design is quite small,

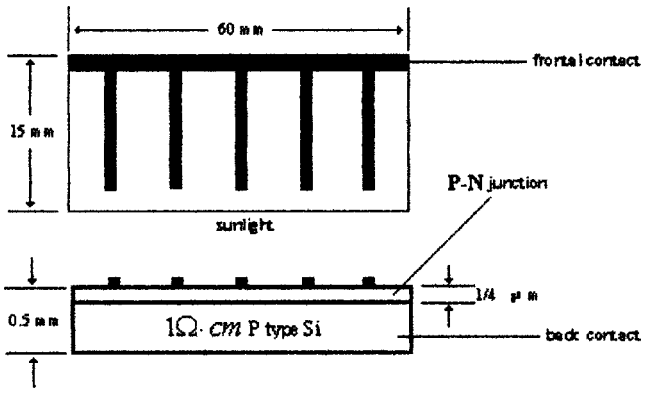


FIG. 2. Schematic diagram of principle of a solar cell.

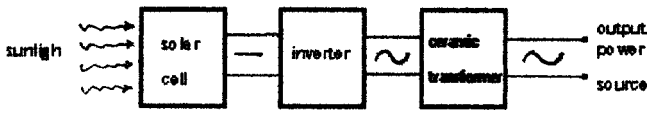


FIG. 3. Schematic diagram of integrating principle of solar cell, inverter, and ceramic transformer.

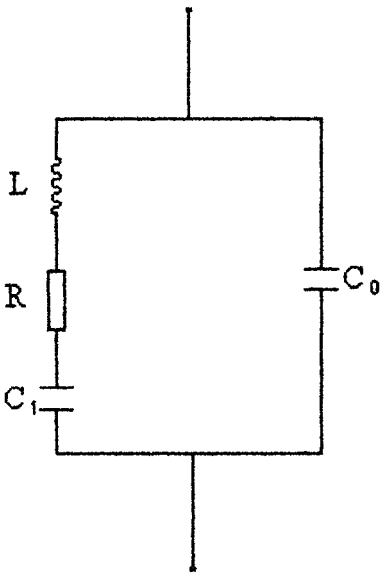
we found it necessary to boost the voltage output utilizing an inverter. This electronic device converts the dc voltage generated by the solar cell into an ac voltage, acting as an additional element connecting the circuits between the solar cell, the ceramic transformer, and the outside of the ceramic body. Their relationship is illustrated in Fig. 3.

4. Development of integrated device

In the present study, from preparations of material to the analyses of their properties, from independent elements (as unitary ceramic transformer, inverter, solar cell) to the development of integrated devices, we have made good progress, and received some qualitative results.

① The principal properties of multifunctional piezoelectric ceramic in the present study based on the Van Dyke equivalent circuit (Fig. 4) are as follows:

(a) pyroelectric coefficient $p=2.1\times10^{-8}$ C/cm^2 K
(b) Curie temperature $T_c=310\,^\circ$C
(c) mechanical quality factor $Q_m>500$



Notes :

L — series inductance in resonator equivalent circuit;
R — series resistance in resonator equivalent circuit;
C_1 — series capacitor in resonator equivalent circuit;
C_0 — shunt capacitor in resonator equivalent circuit.

FIG. 4. The electromechanical equivalent circuit of the multifunctional ceramic.

(d) relative dielectric constant $\epsilon_{33}^T/\epsilon_0>3000$
(e) piezoelectric strain constant $d_{33}=80$ pC/N
(f) piezoelectric voltage constant $g_{33}=33\times10^{-3}$ V/m N
(g) planar electromechanical coupling factor $k_p=0.35$
(h) dielectric loss tg $\delta<2\times10^{-3}$

② The open circuit voltage gain, (V_2/V_1), of ac output of the piezoelectric ceramic transformer in the present study is 400, at the half-wavelength ($\lambda/2$) resonance mode.

③ The solar cell of the present study exhibits the following characteristics under light illumination; the open circuit voltage (V_{oc}) is 0.5 V, and the short circuit current density (I_{sc}) is 5–7 mA/cm^2, for the test conditions where the temperature is 25 °C and the irradiance is 80 mW/cm^2.

III. CONCLUSION

Our investigation into piezoelectric ceramic, which also exhibits pyroelectricity and an internal photoeffect, has shown good progress and results. Making use of this kind of piezoelectric semiconductor ceramic, an integrated photoelectric device, with a piezoelectric ceramic transformer, inverter, and solar cell, has been developed. However, in the development of integrating devices, there are many parts which need precision processing techniques to be applied. For the materials developed in our work some new requirements have become evident, such as the need to increase transparency of the piezoelectric ceramic.

To summarize:

(1) The first applicable aspects of the integrated photoelectric device will be a technique for utilization of solar energy to solve the needs of a power source, especially, where no other power source (including no movable power source) is available, and, an increased voltage will be made available for civil electronic apparatus and equipment.
(2) The material may be used as a pyroelectric detector (including a multielement detector array, etc.).
(3) The present work opens up and suggests further broad applied fields for multifunctional piezoelectric ceramics.

ACKNOWLEDGMENT

The author is grateful to senior engineer Hu Fan of Harbin-Chronar Solar Energy Electricity Corporation for his profitable discussion relating to the work.

[1] R. E. Newnham *et al.*, ''Smart electroceramics,'' J. Am. Ceram. Soc. **74**, 403 (1991).
[2] J. Narkiewicz, ''A survey applicability of electrically/mechanically stimulated materials to sensors and actuators on smart structures,'' Mech. Teor. Stosow. (Poland) **34**, 621 (1996).
[3] Li Quanlu, ''Proceedings of The First China International Conference on High-Performance Ceramics'' (Tsinghua University Press, Beijing, China, 1999), p. 305.
[4] F. Iannone, G. Noviello, and A. Sarno, ''Monte Carlo techniques to analyze the electrical mismatch losses in large photovoltaic generators,'' Sol. Energy **62**, 85 (1998).
[5] Kawada Masahiko, Electronic Materials (in Japanese) **9**, 219 (1970).

Investigation of active structural intensity control in finite beams: Theory and experiment

P. Audrain[a)]
L.A.U.M., I.A.M., Université du Maine, 72085 Le Mans Cedex 9, France and G.A.U.S, Université de Sherbrooke, Québec J1K 2R1, Canada

P. Masson[b)] and A. Berry[c)]
G.A.U.S, Mechanical Engineering Department, Université de Sherbrooke, Sherbrooke, Québec J1K 2R1, Canada

(Received 16 September 1999; revised 8 May 2000; accepted 12 May 2000)

An investigation of structural intensity control is presented in this paper. As opposed to previous work, the instantaneous intensity is completely taken into account in the control algorithm, i.e., all the terms are considered in the real-time control process and, in particular, the evanescent waves are considered in this approach. A finite difference approach using five accelerometers is used as the sensing scheme. A feedforward filtered-X least mean square algorithm is adapted to this energy-based control problem, involving a nonpositive definite quadratic form in general. In this respect, the approach is limited to cases where the geometry is such that the intensity component will have the same sign for the control source and the primary disturbance. Results from numerical simulations are first presented to illustrate the benefit of using a cost function based on structural intensity. Experimental validation of the approach is conducted on a free–free beam covered with viscoelastic material. A comparison is made between classical acceleration control and structural intensity control and the performance of both approaches is presented. These results confirm that using intensity control allows the error sensors to be placed closer to the control source and the primary disturbance, while preserving a good control performance.
[S0001-4966(00)03108-8]

PACS numbers: 43.40.At, 43.50.Ki, 43.40.Rj, 43.40.Dx [CBB]

INTRODUCTION

Active control of structural vibration is usually based on either a *modal* description, or a *wave* description of the structure. Modal-based active control aims at reducing the structural vibration globally, whereas the wave approach is more appropriate to control the energy transmission from one area of the structure to another area. A number of authors have investigated the active control of propagating waves in beams.[1–3] Evanescent waves originating from the control sources or from the beam boundary conditions generally dictate a low-frequency limit or a minimal separation between the control sources and the error sensors. Another approach to actively control wave propagation in beams is to minimize the structural intensity or power flow.[4–8] In 1987 Redman-White *et al.*[5] experimented with the active control of flexural wave power flow and concluded that this strategy is suitable for periodic disturbance. Several methods have been proposed to measure the structure-borne energy flow.[9,10] Pavic[11] formulated a finite difference scheme to evaluate the flexural structural intensity from local acceleration or velocity measurements. Schwenk *et al.*[7] developed a structural intensity control algorithm based on the assumption of propagating waves traveling in one direction only (semi-infinite beam). All previous investigations of intensity or power flow control assume far-field, propagating waves at the error sensor location. Such an assumption means that the error sensors must be located at a distance larger than approximately 0.75λ from the control sources or structural discontinuities, where λ is the structural wavelength.

As opposed to previous work, the instantaneous intensity is completely taken into account in the control algorithm, i.e., all the terms are considered in the real-time control process and, in particular, the evanescent waves are considered in this approach. A finite differences approach using five accelerometers is used as the sensing scheme. A feedforward filtered-X least mean square (LMS) algorithm is adapted to this energy-based control problem, involving a nonpositive definite function in general. In this respect, the approach is limited to cases where the geometry is such that the intensity component will have the same sign for the control source and the primary disturbance. Results from numerical simulations are first presented to illustrate the benefit of using a cost function based on structural intensity. Experimental validation of the approach is conducted on a free–free beam covered with viscoelastic material. The experimental setup includes a shaker to excite the primary disturbance while a piezoelectric patch is employed as the control source. A comparison is made between classical acceleration control and structural intensity control and the performance of both approaches is presented.

[a)]Electronic mail: pascal.audrain@laum.univ-lemans.fr; paudrain@linus.gme.usherb.ca
[b)]Electronic mail: Patrice.Masson@gme.usherb.ca
[c)]Electronic mail: Alain.Berry@gme.usherb.ca

I. FLEXURAL STRUCTURAL INTENSITY IN A BEAM

A. Definition of structural intensity

The structural intensity is the instantaneous rate of vibrational energy transfer, or energy flow, per unit area in a given direction.[12] The instantaneous energy flow in a beam, called the instantaneous structural intensity (subscript $_i$) in this article, originating from a flexural displacement $w(x,t)$, is the sum of a force term, the product of the shear force (T) and the transverse velocity (v), and a moment term, the product of the bending moment (M) and the rotational velocity (Ω):

$$\vec{I}_i=(T)(v)\vec{i}+(M)(\Omega)\vec{i}, \tag{1}$$

where $\vec{i}$ is the unit vector in the x direction.

The instantaneous structural intensity can be expressed as[11]

$$\vec{I}_i=\left(EI\frac{\partial^3 w}{\partial x^3}\right)\left(\frac{\partial w}{\partial t}\right)\vec{i}-\left(EI\frac{\partial^2 w}{\partial x^2}\right)\left(\frac{\partial^2 w}{\partial t\,\partial x}\right)\vec{i}, \tag{2}$$

which represents a conversion of Eq. (1) for the case of a Euler–Bernoulli isotropic beam and where E is the Young's modulus and I is the area moment of inertia of the beam.

The time average of the instantaneous intensity, called the active intensity (subscript $_a$), represents the net energy transport.[13] If the flexural displacement is assumed complex and time harmonic, the active intensity can be expressed as[11]

$$\vec{I}_a=\langle\vec{I}_i\rangle_T=\frac{1}{2}\Re\left\{\left(EI\frac{\partial^3 w}{\partial x^3}\right)\left(\frac{\partial w}{\partial t}\right)^*-\left(EI\frac{\partial^2 w}{\partial x^2}\right)\left(\frac{\partial^2 w}{\partial t\,\partial x}\right)^*\right\}\vec{i}, \tag{3}$$

where $\Re(\,)$ denotes the real part of a complex, * denotes the complex conjugate, and T is the time period.

B. Structural intensity in a finite damped beam

In this section the structural intensity in a finite beam is derived using a wave decomposition approach for the flexural harmonic displacement $w(x,t)$. The harmonic transverse displacement can be written as

$$w(x,t)=w(x)e^{j\omega t}, \tag{4}$$

where $j=\sqrt{-1}$ and w is the angular frequency.

In order to obtain an energy flow in the beam, a structural damping is introduced in the form of a complex Young's modulus of the beam:

$$\tilde{E}=E(1+j\eta), \tag{5}$$

where η is the structural loss factor. At any position, the flexural displacement can be expressed as the combination of four waves:

$$w(x,t)=(A_1e^{-jkx}+A_2e^{-kx}+A_3e^{jkx}+A_4e^{kx})e^{j\omega t}, \tag{6}$$

where k is the flexural wave number given by

$$k=\left(\frac{\rho S}{\tilde{E}I}\omega^2\right)^{1/4}, \tag{7}$$

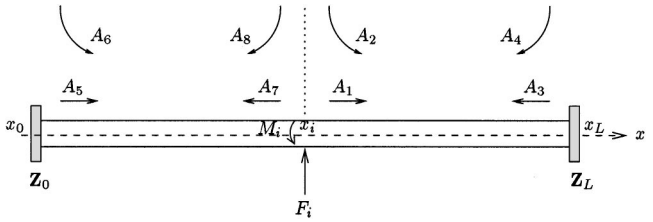


FIG. 1. Wave decomposition of the flexural displacement generated by a point force disturbance F_i and a line moment M_i located at x_i.

where ρ is the density and S is the area of the cross section. In Eq. (6), A_1 and A_3 represent waves traveling in the positive and negative directions, respectively, while A_2 and A_4 represent evanescent wave components.

Pan and Hansen[1] demonstrated that a 2×2 matrix $\mathbf{Z}$ can describe any dependence of boundary forces for flexural motion in a beam. At a boundary $x=x_b$ of the beam, this can be expressed as

$$\begin{bmatrix}M_b\\T_b\end{bmatrix}=\mathbf{Z}_b\begin{bmatrix}\dot{w}_b\\\dot{\theta}_b\end{bmatrix}, \tag{8}$$

where $\dot{w}_b=j\omega w(x_b,t)$ is the flexural velocity at the boundary, $\dot{\theta}_b=j\omega[\partial w(x_b,t)/\partial x]$ is the angular velocity, $M_b=\tilde{E}I[\partial^2 w(x_b,t)/\partial x^2]$ is the bending moment, $T_b=-\tilde{E}I[\partial^3 w(x_b,t)/\partial x^3]$ is the shear force, and the coefficients of the matrix Z_b depend upon the boundary conditions.

If a point force F_i and a line moment M_i are applied at $x=x_i$, the finite damped beam extending from x_0 to x_L is divided in two domains $x_i\leqslant x\leqslant x_L$ and $x_0\leqslant x\leqslant x_i$ as shown in Fig. 1. In each domain the four types of waves are present. Thus the out-of-plane displacement can be written as

$$x_i\leqslant x\leqslant x_L:\quad w(x)=A_1e^{-jkx}+A_2e^{-kx}+A_3e^{jkx}+A_4e^{kx}, \tag{9}$$

$$x_0\leqslant x\leqslant x_i:\quad w(x)=A_5e^{-jkx}+A_6e^{-kx}+A_7e^{jkx}+A_8e^{kx}. \tag{10}$$

At the excitation point $x=x_i$, the following continuity conditions apply:

$$w(x_i^+)-w(x_i^-)=0, \tag{11}$$

$$\frac{\partial w(x_i^+)}{\partial x}-\frac{\partial w(x_i^-)}{\partial x}=0, \tag{12}$$

$$\frac{\partial^2 w(x_i^+)}{\partial x^2}-\frac{\partial^2 w(x_i^-)}{\partial x^2}=-\frac{M_i}{\tilde{E}I}, \tag{13}$$

$$\frac{\partial^3 w(x_i^+)}{\partial x^3}-\frac{\partial^3 w(x_i^-)}{\partial x^3}=\frac{F_i}{\tilde{E}I}. \tag{14}$$

The system formed by Eq. (8) at each boundary ($x=x_0$ and $x=x_L$) and Eqs. (11)–(14) is written in matrix form as

$$\Gamma C=\Upsilon, \tag{15}$$

where

$$\mathbf{C}^T=[A_1\quad A_2\quad A_3\quad A_4\quad A_5\quad A_6\quad A_7\quad A_8], \tag{16}$$

$$\Upsilon^T=\left[0\quad 0\quad 0\quad 0\quad 0\quad 0\quad -\frac{M_i}{\tilde{E}Ik^2}\quad \frac{F_i}{\tilde{E}Ik^3}\right], \tag{17}$$

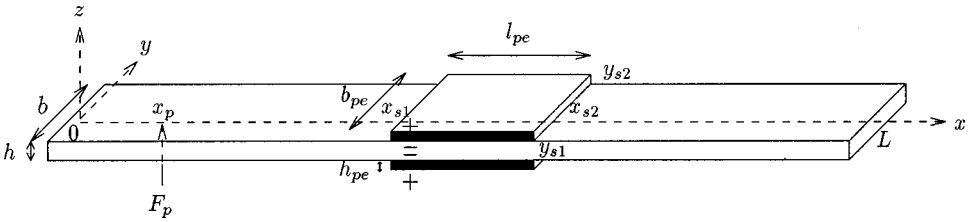


FIG. 2. Geometrical description and coordinate system for the beam.

and the boundary matrix Γ is detailed in the Appendix.

Point force disturbance: The response of the beam to a point force disturbance, $F_i = F_p e^{j\omega t}$, is a special case of the above equations when $M_i = 0$.

Piezoelectric patch actuator: A piezoelectric patch actuator is a pair of piezoelectric elements bonded symmetrically on opposite sides of the beam and subjected to opposite applied voltages (Fig. 2). Such a piezoelectric patch generates approximately two line moments at each free edge x_{s1} and x_{s2} of the patch:[14]

$$x = x_{s1}: \quad M_1 = b_{pe} M_x, \tag{18}$$

$$x = x_{s2}: \quad M_2 = -b_{pe} M_x, \tag{19}$$

where b_{pe} is the width of the piezoelectric material and M_x represents the bending moment per unit width generated in the beam by the actuator. If the bonding layer is very thin, its effect can be neglected[15] and the bending moment is

$$M_x = \frac{E_{pe} h^3 (h + h_{pe}) d_{31} V}{4[(h^3/4) + (E_{pe}/E) h_{pe} (\frac{3}{2} h^2 + 2h_{pe}^2 + 3h h_{pe})]}, \tag{20}$$

where $V = V_s e^{j\omega t}$ is the voltage applied to the piezoelectric elements, E_{pe}, h_{pe}, and d_{31} are the Young's modulus, thickness, and piezoelectric constant of the piezoelectric actuator, and h is the beam's thickness.

The transverse response to the piezoelectric patch is then divided in three domains:

$$x_{s2} \leqslant x \leqslant x_L: \quad w_s(x) = A_1^s e^{-jkx} + A_2^s e^{-kx} + A_3^s e^{jkx} + A_4^s e^{kx}, \tag{21}$$

$$x_{s1} \leqslant x \leqslant x_{s2}: \quad w_s(x) = A_9^s e^{-jkx} + A_{10}^s e^{-kx} + A_{11}^s e^{jkx} + A_{12}^s e^{kx}, \tag{22}$$

$$x_0 \leqslant x \leqslant x_{s1}: \quad w_s(x) = A_5^s e^{-jkx} + A_6^s e^{-kx} + A_7^s e^{jkx} + A_8^s e^{kx}. \tag{23}$$

The various wave coefficients A_i^s generated by the end moments are calculated from the equations

$$\Gamma C_{M1} = \Upsilon_{M1}, \tag{24}$$

$$\Gamma C_{M2} = \Upsilon_{M2}, \tag{25}$$

where

$$C_{Ml}^T = [A_1^{Ml} \quad A_2^{Ml} \quad A_3^{Ml} \quad A_4^{Ml} \quad A_5^{Ml} \quad A_6^{Ml} \quad A_7^{Ml} \quad A_8^{Ml}], \tag{26}$$

$$\Upsilon_{Ml}^T = \left[0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad -\frac{M_l}{\tilde{E} I k^2} \quad 0\right], \tag{27}$$

with $l = 1,2$, and

$$A_k^s = A_k^{M1} + A_k^{M2}, \quad k = 1,2,...,8, \tag{28}$$

$$A_9^s = A_1^{M1} + A_5^{M2}, \tag{29}$$

$$A_{10}^s = A_2^{M1} + A_6^{M2}, \tag{30}$$

$$A_{11}^s = A_3^{M1} + A_7^{M2}, \tag{31}$$

$$A_{12}^s = A_4^{M1} + A_8^{M2}. \tag{32}$$

Structural intensity distribution in a free-free damped beam: Figures 3 and 4 show active intensity distributions in the finite damped beam depicted in Fig. 2. The beam is made of aluminum, its characteristics are given in the first column of Table I; the boundary impedance matrix $\mathbf{Z}_b$ in the case of a free–free beam is the zero matrix. The active intensity [Eq.

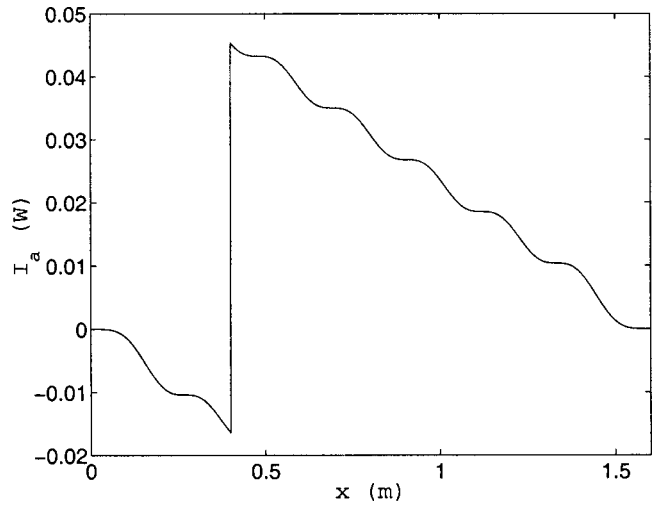


FIG. 3. Active intensity generated by a point force disturbance located at $x_p = 0.4$ m, at a frequency of $f = 315$ Hz.

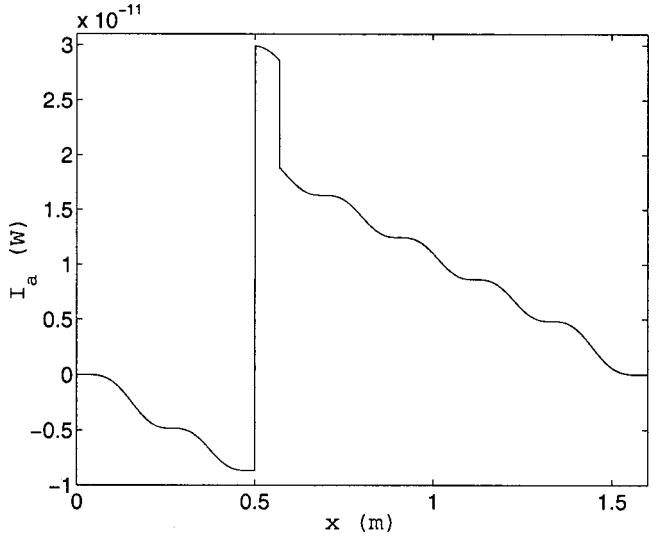


FIG. 4. Active intensity generated by a piezoelectric actuator extending from $x_{s1} = 0.5$ m to $x_{s2} = 0.57$ m, at a frequency of $f = 315$ Hz.

TABLE I. Component characteristics.

	Beam	Adhesive	Constraining layer
Material	aluminum	ISD830	aluminum
x dimension (m)	1.6	1.53	1.53
y dimension (m)	0.0254	0.0254	0.0254
Thickness h (m)	0.00636	0.000127	0.000305
Young's modulus E (GPa)	67	frequency dependent	67
Poisson's ratio ν	0.3	0.45	0.3
Density ρ (kg/m^3)	2710	1000	2710
Structural damping η	0.01	frequency dependent	0.01

(3)] was simulated numerically using the wave decomposition at the seventh resonant frequency of the structure, $f = 315\,\text{Hz}$. Figure 3 shows the intensity distribution generated by a point force of amplitude $F_p = 1\,\text{N}$ located at $x_p = 0.4\,\text{m}$. The active intensity is seen to have a smooth distribution with a sign indicating an energy flow in both directions originating from the excitation source. The energy falls off due to the structural damping and the active intensity is zero at the ends of the beam since there is no energy loss at the boundaries in this case.

Figure 4 shows the active intensity induced by a piezoelectric patch actuator with length $l_{pe} = 0.07\,\text{m}$, width $b_{pe} = 0.0254\,\text{m}$, placed between $x_{s1} = 0.5\,\text{m}$, and $x_{s2} = 0.57\,\text{m}$. The Poisson's ratio, Young's modulus, and thickness of the piezoelectric materials are respectively $\nu_{pe} = 0.29$, $E_{pe} = 61\,\text{GPa}$, and $h_{pe} = 0.2\,\text{mm}$, while the piezoelectric strain constant is $d_{31} = 1.71 \times 10^{-10}\,\text{m/V}$. The piezoelectric actuator generates energy flow emanating from the actuator in both directions and the active intensity is then discontinuous at the line moment application points.

Note that the discontinuity of active intensity at the point force or at the edge of the piezoelectric actuator arises from the application of the continuity equations for the bending moment [Eq. (13)] and the shear force [Eq. (14)].

The principle of active intensity control is to adjust the actuation voltage of the piezoelectric patch such that the combined intensity distribution (due to the primary disturbance and control input) is minimized past a given location on the beam.

C. Exact minimization of structural intensity

In the context of active control, the total structural displacement field is the sum of a primary displacement field (labeled p) created by the disturbance, and a secondary displacement field (labeled s) originating from n actuators exciting the structure. In this study, the primary disturbance is a point force while the control actuators are piezoelectric elements.[16] The total displacement field of the beam is given by

$$w(x) = w_p(x) + \mathbf{w}_s \mathbf{X}_s, \tag{33}$$

where w_p is the primary displacement, $\mathbf{w}_s$ is a n-dimensional line vector representing the displacements induced by the individual control actuators, and $\mathbf{X}_s$ is the n row-vector of the control variables representing the strengths of the individual control actuators.

The vector $\mathbf{X}_s$ is adjusted so as to minimize the active intensity at the error sensor located at x_e. Upon substitution of the total displacement field in Eq. (3), it is possible to express the flexural active intensity in a Hermitian quadratic form:

$$I_a(x_e) = \mathbf{X}_s^H \mathbf{A}(x_e) \mathbf{X}_s + \mathbf{X}_s^H \mathbf{b}(x_e) + \mathbf{b}^H(x_e) \mathbf{X}_s + c(x_e), \tag{34}$$

where

$$\mathbf{A}(x_e) = \frac{\omega I}{2} \Re\left\{ -j\tilde{E}\left[\mathbf{w}_s^{\mathrm{H}}(x_e) \frac{\partial^3 \mathbf{w}_s(x_e)}{\partial \mathbf{x}^3} - \left(\frac{\partial \mathbf{w}_s(x_e)}{\partial x} \right)^{\mathrm{H}} \frac{\partial^2 \mathbf{w}_s(x_e)}{\partial \mathbf{x}^2} \right] \right\}, \tag{35}$$

$$\mathbf{b}(x_e) = \frac{\omega I}{4} \left[-j\tilde{E}\left(\frac{\partial^3 w_p(x_e)}{\partial x^3} \mathbf{w}_s^{\mathrm{H}}(x_e) - \frac{\partial^2 w_p(x_e)}{\partial x^2} \times \left(\frac{\partial \mathbf{w}_s(x_e)}{\partial \mathbf{x}} \right)^{\mathrm{H}} \right) + j\tilde{E}^* \left(\left(\frac{\partial^3 \mathbf{w}_s(x_e)}{\partial \mathbf{x}^3} \right)^{\mathrm{H}} w_p(x_e) - \left(\frac{\partial^2 \mathbf{w}_s(x_e)}{\partial \mathbf{x}^2} \right)^{\mathrm{H}} \frac{\partial w_p(x_e)}{\partial x} \right) \right], \tag{36}$$

$$c(x_e) = \frac{\omega I}{2} \Re\left\{ -j\tilde{E}\left(\frac{\partial^3 w_p(x_e)}{\partial x^3} w_p^*(x_e) - \frac{\partial^2 w_p(x_e)}{\partial x^2} \left(\frac{\partial w_p(x_e)}{\partial x} \right)^* \right) \right\}, \tag{37}$$

and $^{\mathrm{H}}$ denotes the Hermitian transpose. The first term in Eq. (34) represents the active intensity generated by the control sources at the error sensor, and the scalar $c(x_e)$ represents the active intensity due to the disturbance alone. Provided the matrix $\mathbf{A}$ is positive definite, $I_a(x_e)$ has a unique global minimum, and the optimal control sources strengths are given by

$$\mathbf{X}_{s,\text{opt}} = -\mathbf{A}^{-1}\mathbf{b}. \tag{38}$$

Also, the minimized value of the intensity is $I_{a,\text{opt}}(x_e) = c(x_e) + \mathbf{b}^{\mathrm{H}} \mathbf{X}_{s,\text{opt}}$. It should be emphasized here that $\mathbf{A}$ is not *guaranteed to be positive* in general, since the active intensity is a signed quantity, which can take positive or negative values. This is especially true for a finite structure, where internal reflections can create sign fluctuations in the energy flow. Hence, the extremum condition given by Eq. (38) can imply a minimum or a maximum of the intensity. However, in the case considered here of a one-dimensional

damped system, a localized disturbance always creates a flow of energy emanating from the disturbance, such that **A** is a positive matrix. This property of the cost function imposes the intensity generated by the disturbance and the control actuator to have the same sign.[17] Thus, **A** is a positive matrix when the error sensor is located downstream of all control actuators. Another potential difficulty may arise with the control of the structural intensity: the value of the minimized intensity $I_{a,\mathrm{opt}}(x_e)$ is not necessarily positive. The minimization of the structural intensity can result in a large and negative value of the minimum, which translates into poor or negative control performance. This is typically the case when the error sensor is located between the primary disturbance and the control actuator. In other words, the active intensity originating from both sources, the disturbance and the control actuators, must have the same sign at the error sensors locations.

Note that the same quadratic minimization principles can be used to derive the exact minimization of the squared transverse acceleration at one point, or the sum of the squared transverse acceleration at several points downstream of the control sources. In this case the cost function has always a unique global minimum.

Total and partial mean square velocity: The total mean square velocity provides a global information on the flexural vibrations of the beam. It is defined as space and time average of the beam velocity squared:

$$\langle v^2(\omega)\rangle = \frac{1}{2L}\int_0^L \dot{w}(x,\omega)\dot{w}^*(x,\omega)\,dx, \tag{39}$$

where L is the length of the beam. Similarly, a partial mean square velocity can be calculated over the region of the beam extending past the error sensor:

$$\langle v_{x_e}^2(\omega)\rangle = \frac{1}{2(L-x_e)}\int_{x_e}^L \dot{w}(x,\omega)\dot{w}^*(x,\omega)\,dx. \tag{40}$$

The partial mean square velocity provides information on the vibratory state of the beam downstream of the error sensor, in order to evaluate the effectiveness of the control strategy in blocking energy transmission past the error sensor.

Optimal active control simulations: The setting used for the optimal active control simulations consists of the free–free aluminum beam described in Table I, excited harmonically by a point force disturbance of amplitude $F_p=1$ N located at $x_p=0.02$ m from one end. A PZT patch (PSI-5A-S2) located between $x_{s1}=0.41$ m and $x_{s1}=0.48$ m is used as the control actuator. Optimal control of flexural active intensity and acceleration at one and five points downstream of the PZT actuator are presented for a spanning of the excitation frequency between 0 and 1000 Hz. The control is performed at a single frequency by solving Eq. (38). Intensity and one-point acceleration control are performed at $x=x_e$, while control of acceleration at five locations is performed at $x=[x_e, x_e\pm\Delta, x_e\pm 2\Delta]$ with $\Delta=0.04$ m. In all cases, the partial mean square velocity $[\langle v_{x_e}^2\rangle$, Eq. (40)] is calculated past the last acceleration control point, $x_e+2\Delta$.

Figures 5 and 6 show the total mean square velocity of the beam and the mean square velocity past the error sensor

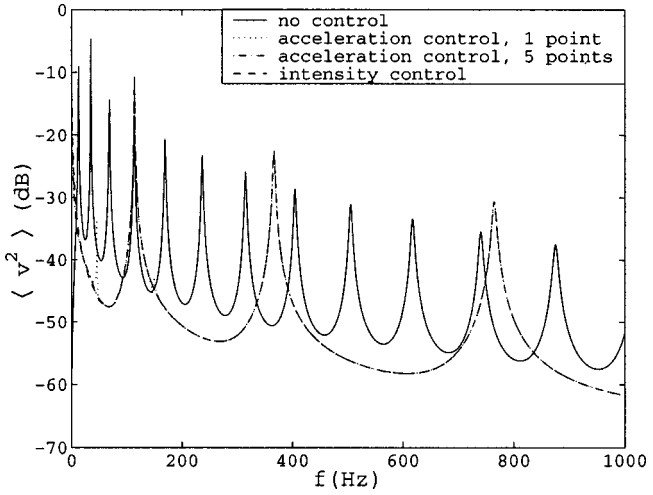


FIG. 5. Total mean square velocity of the beam for various optimal active control strategies, with the error sensor located at $x_e=0.9$ m.

for the three control strategies when the intensity error sensor is located at $x_e=0.9$ m. In this case, the control actuator–error sensor separation is 42 cm, which represents 0.56λ at 100 Hz, and 1.75λ at 1000 Hz (λ being the flexural wavelength). The error sensor is thus in the far field of the control actuator at most of the frequencies considered. In this case, controlling acceleration or active intensity has the same (and comparatively low) impact on the overall vibration of the beam (Fig. 5), and it has almost the same (but more significant) impact on the vibration of the beam past the error sensor (Fig. 6). The control of intensity provides a slightly better performance over the complete frequency range.

In the case of Figs. 7 and 8, the control actuator–error sensor separation is 9 cm, which represents 0.12λ at 100 Hz, and 0.38λ at 1000 Hz. In this case, the error sensor is in the near field of the control actuator, and therefore observes evanescent waves originating from the control actuator. The three control strategies still have approximately the same low impact on the overall beam vibration, with a slightly better performance using the acceleration control at five locations (Fig. 7). However, the results on the beam vibration past the

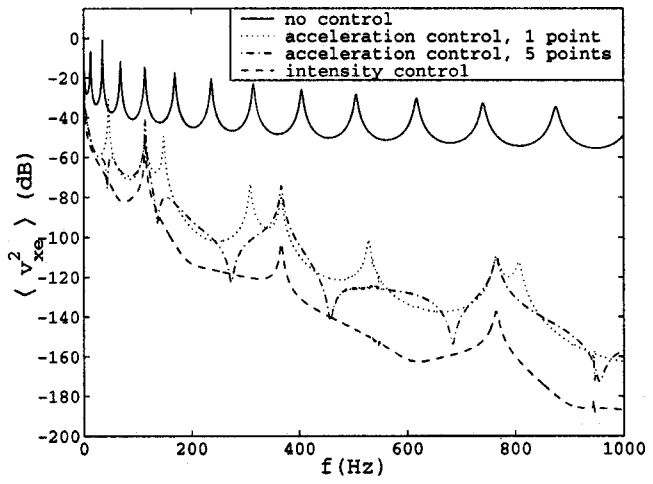


FIG. 6. Mean square velocity of the beam past the error sensor for various optimal active control strategies, with the error sensor located at $x_e=0.9$ m.

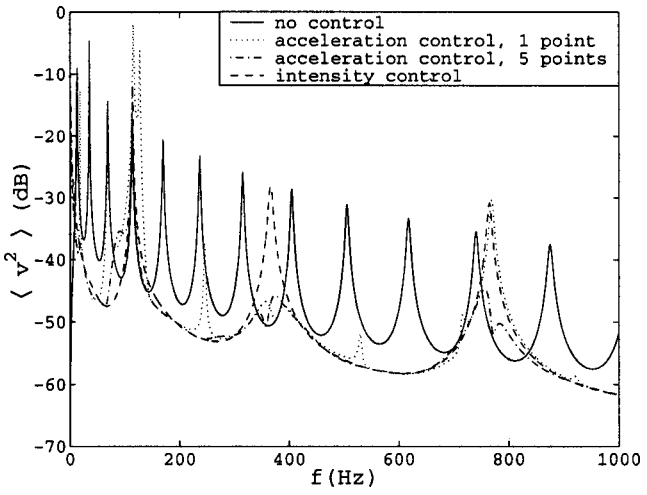


FIG. 7. Total mean square velocity of the beam for various optimal active control strategies, with the error sensor located at $x_e = 0.57$ m.

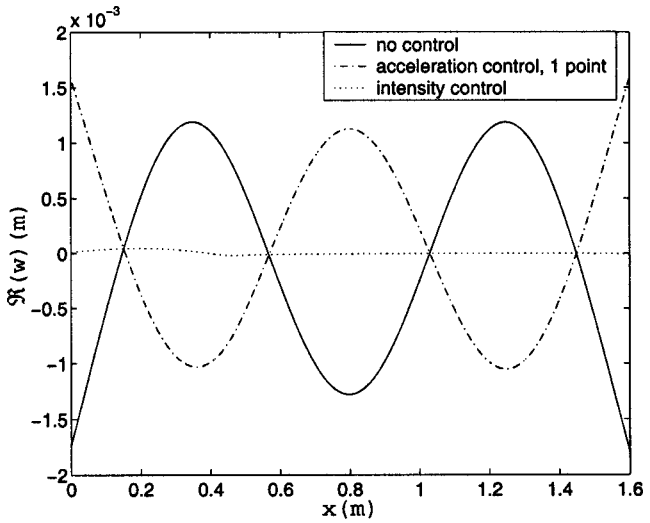


FIG. 9. Real part of the flexural displacement at $f = 68.6$ Hz and various optimal active control strategies, with the error sensor located at $x_e = 0.57$ m.

error sensor are very different (Fig. 8): minimizing the transverse acceleration is less efficient in this situation since the control spends unnecessary effort minimizing evanescent waves not related to the far-field flow of energy. Minimizing the acceleration at five locations provides generally better control performance but is also influenced by the presence of the near-field waves at the error sensor locations. In contrast, minimizing the structural intensity still provides significant vibration attenuation past the error sensor since the intensity sensor observes only propagating wave components. These results are consistent with earlier conclusions drawn in the case of infinite beams.[7,18]

Acceleration control at five points is more efficient in terms of the overall vibration level while intensity control provides better attenuation past the error sensor. Indeed, intensity control aims at controlling the energy flow downstream of the error sensor regardless of the vibration level upstream.

The comparison of Figs. 6 and 7 also shows that intensity control performance is less dependent on the error sensor location than acceleration control at one point. The frequencies for which intensity control provides less performance (approximately 120, 350, and 760 Hz in this case) correspond to situations where the control PZT cannot generate energy in the structure. Structural intensity has a smooth distribution over the length of the beam, having a maximum value at the excitation locations and vanishing at the ends of the beam. In contrast, the transverse acceleration distribution of a finite beam shows nodes which translate into observability problems for the control. Moreover, intensity control allows for the error sensor to be located close to the control actuator, which has the practical advantage of a physically compact system.

Finally, Figs. 9 and 10 show flexural displacement and active intensity distributions along the beam at $f = 68.6$ Hz for one-point acceleration and active intensity control strategies. This frequency corresponds to the third resonant frequency of the beam. The error sensor is located in the near field, relative to a wavelength, of the control actuator, at $x_e = 0.57$ m. This location corresponds to a node of the reso-

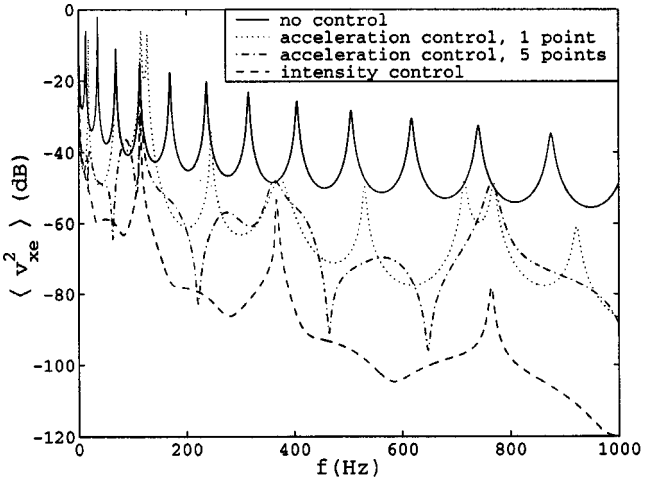


FIG. 8. Mean square velocity of the beam past the error sensor for various optimal active control strategies, with the error sensor located at $x_e = 0.57$ m.

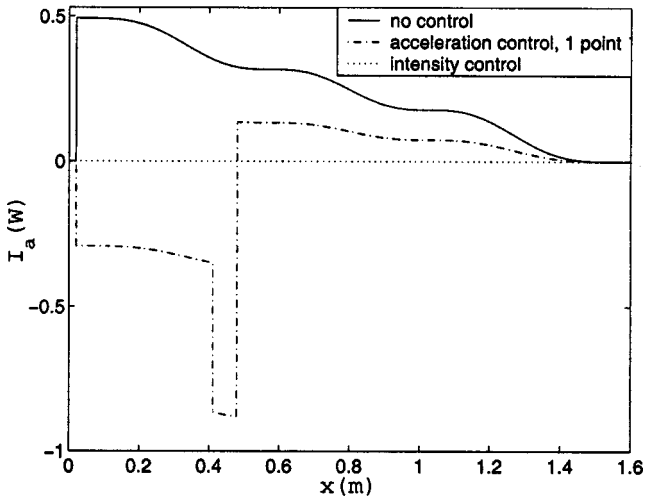


FIG. 10. Active intensity at $f = 68.6$ Hz and various optimal active control strategies, with the error sensor located at $x_e = 0.57$ m.

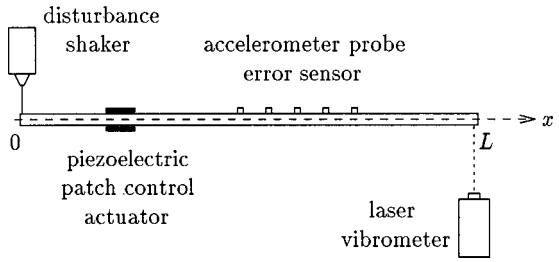


FIG. 11. Sketch of the experimental setup.

nant mode. Acceleration control in this case results in a very large control effort and a poor control performance, while intensity control still provides excellent attenuation past the error sensor. These results confirm that intensity minimization in finite, damped structures does not suffer from observability problems which occur in local acceleration control.

Optimal active control simulations have also been performed for a *simply supported* beam and the same tendencies were obtained. The advantage of using intensity minimization seems to apply regardless of the boundary conditions, as long as they are passive.

II. EXPERIMENTAL IMPLEMENTATION OF STRUCTURAL INTENSITY CONTROL

A. Experimental setup

In this section the setup used for the experimental validation of intensity and acceleration control is described. Figures 11 and 12 show a sketch of the experimental setup and a photograph of the experimental apparatus.

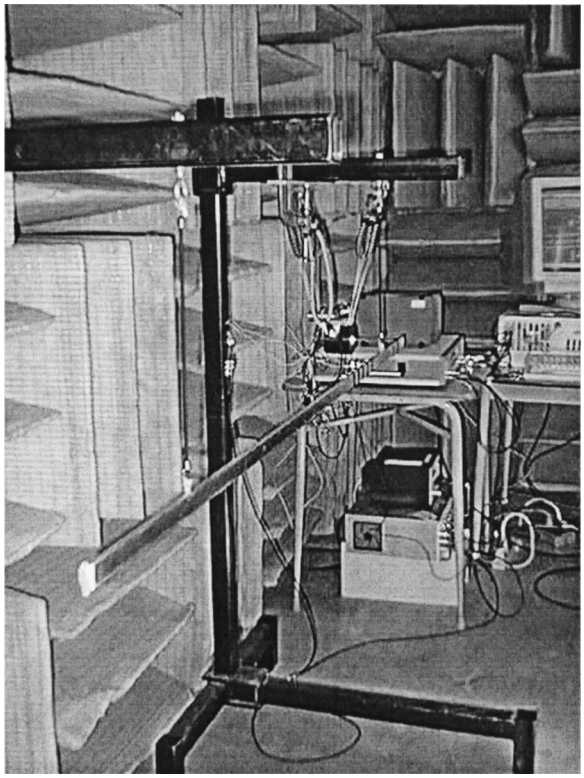


FIG. 12. Photograph of the experimental apparatus.

1. Mechanical setup

The structure is a free–free aluminum beam covered on one side by a 3M viscoelastic material combined with an aluminum constraining layer. The viscoelastic material was used to increase the structural damping and thus facilitate the measurement of structural intensity. The characteristics of the beam, the viscoelastic material (adhesive), and the constraining layer are given in Table I. The structure was suspended by two soft rubber springs attached at 0.25 m from each free end of the beam.

2. Sensors and actuators

The primary source was applied at one end of the beam, taken as the coordinate's origin, as a point force perpendicular to the surface of the beam. A Bruel & Kajer 4810 shaker was connected to the beam by a stinger and excited harmonically between 200 and 1000 Hz.

The five PCB accelerometers of the intensity probe were attached to the beam with a 4-cm spacing. A preliminary test was done to verify that they were correctly matched in phase. Two different sensor locations were experimented with for intensity and acceleration control. In the first series of experiments, the center of the intensity array, which corresponds to the intensity and acceleration measurement location, was at $x_e = 0.38$ m and for the second series at $x_e = 0.26$ m.

The piezoelectric control actuator has similar characteristics as in Sec. I C. The control actuator extends from $x_{s1} = 0.1$ m to $x_{s2} = 0.17$ m. The separation between the end of the piezoceramic and the center of the intensity probe is therefore 21 cm in the first series and 9 cm in the second series of experiments; in the experimental implementation, a 21-cm separation represents 0.44λ at 200 Hz and 0.95λ at 1 kHz, while a 9-cm separation represents 0.18λ at 200 Hz and 0.41λ at 1 kHz.

3. Measurement of structural intensity

Measurement of instantaneous structural intensity involves five accelerometers. This configuration allows intensity control to be compared to the control of acceleration at the same error sensor location, the center of the intensity probe. If the vibration is assumed to be time harmonic with frequency ω, the displacement $w_i(x,t)$ can be expressed in terms of acceleration $a_i(x,t)$ as $w_i(x,t) = -(1/\omega^2)a_i(x,t)$, where $i = 1,...,5$ is the accelerometer index (increasing with x).

The instantaneous intensity is evaluated at the third accelerometer position using finite difference techniques. The time differentials of the displacement are estimated using a backward finite difference scheme with error on the order of τ^2, where τ is the time increment between samples, and a central finite difference scheme is used to estimate the spatial differentials with error on order of Δ, where Δ is the spacing between two consecutive accelerometers. Thus the differentials used in the calculation of instaneous intensity are evaluated as follows:

$$\frac{\partial^3 w}{\partial x^3} = -\frac{1}{\omega^2}\frac{-a_1 + 2a_2 - 2a_4 + a_5}{2\Delta^3}, \tag{41}$$

$$\frac{\partial w}{\partial t} = -\frac{1}{\omega^2}\frac{3a_{3,t} - 4a_{3,t-1} + a_{3,t-2}}{2\tau}, \tag{42}$$

$$\frac{\partial^2 w}{\partial x^2} = -\frac{1}{\omega^2}\frac{a_2 - 2a_3 + a_4}{\Delta^2}, \tag{43}$$

$$\frac{\partial^2 w}{\partial t \partial x} = -\frac{1}{\omega^2} \times \frac{(3a_{4,t} - 4a_{4,t-1} + a_{4,t-2}) - (3a_{2,t} - 4a_{2,t-1} + a_{2,t-2})}{4\Delta\tau}, \tag{44}$$

and the instantaneous structural intensity is estimated as

$$I_i = EI\left[\left(\frac{\partial^3 w}{\partial x^3}\right)\left(\frac{\partial w}{\partial t}\right) - \left(\frac{\partial^2 w}{\partial x^2}\right)\left(\frac{\partial^2 w}{\partial t \partial x}\right)\right]. \tag{45}$$

Figure 13 shows a comparison between the experimental and theoretical active intensity for an intensity probe center located at $x_e = 0.38$ m. The experimental active intensity is the temporal mean value of the instantaneous intensity generated in the structure by the shaker excited harmonically between 200 Hz to 1 kHz. The sampling frequency was 3 kHz. The theoretical active intensity was obtained using an equivalent free–free beam model. The theoretical model takes into account the viscoelastic and the constraining layers by employing equivalent dimensions and frequency-varying Young's modulus and loss factor as the viscoelastic material characteristics are frequency dependent. Figure 13 shows a good agreement between the theoretical model and the experimental results for the frequency range studied, i.e., between 200 to 1000 Hz. At lower and higher frequency problems in intensity measurement arise due to the finite difference scheme and, in particular, to the spacing between the accelerometers and the sampling rate as discussed in detail by Schwenk *et al.* in Ref. 7.

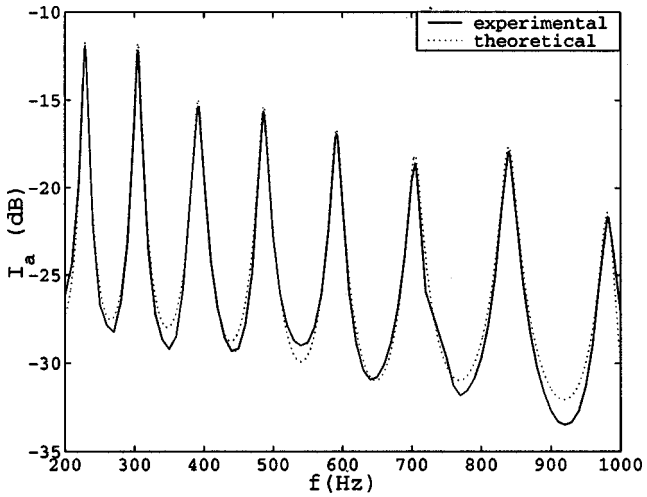


FIG. 13. Predicted and measured active intensity (obtained as the temporal mean value of the measured instantaneous intensity) with the error sensor probe center at $x_e = 0.38$ m.

B. Structural intensity control algorithm

This section describes the feedforward control approach employed to minimize structural intensity in a beam. The control algorithm is detailed here in the case of only one control actuator. At the error sensor location, the flexural displacement, $w(m)$, can be separated into a primary component, the displacement without control, $w_p(m)$, and a component due to the control actuator, $w_s(m)$, where m indicates a discrete-time index. This can be expressed as

$$w(m) = w_p(m) + \sum_{j=0}^{J-1} h_{0j}(m) y(m-j), \tag{46}$$

where h_{0j} is the jth coefficient of the finite impulse response (FIR) filter modeling the transfer function $\mathbf{H_0}$ between the control input $y(m)$ and the displacement originating from the control source.

For a FIR control filter, the control signal is obtained as

$$y(m-k) = \sum_{i=0}^{I-1} w_i(m) x(m-i-k), \quad k = 0,1,...,J-1, \tag{47}$$

where $w_i(m)$ is the ith coefficient of the FIR control filter, $\mathbf{W}$, and $x(m)$ is the reference signal.

A modified filtered-X least-mean-square (FXLMS) algorithm was used to update the control filter in order to minimize the instantaneous intensity at the error sensor. Considering the definition of the instantaneous intensity [Eq. (2)], the following time signals are defined:

$$e_1(m) = \frac{\partial^3 w}{\partial x^3}(m) = \frac{\partial^3 w_p}{\partial x^3}(m) + \sum_{j=0}^{J-1} h_{1j}(m) y(m-j), \tag{48}$$

$$e_2(m) = \frac{\partial w}{\partial t}(m) = \frac{\partial w_p}{\partial t}(m) + \sum_{j=0}^{J-1} h_{2j}(m) y(m-j), \tag{49}$$

$$e_3(m) = \frac{\partial^2 w}{\partial x^2}(m) = \frac{\partial^2 w_p}{\partial x^2}(m) + \sum_{j=0}^{J-1} h_{3j}(m) y(m-j), \tag{50}$$

$$e_4(m) = \frac{\partial^2 w}{\partial t \partial x}(m) = \frac{\partial^2 w_p}{\partial t \partial x}(m) + \sum_{j=0}^{J-1} h_{4j}(m) y(m-j), \tag{51}$$

where h_{ij} is the jth coefficient of the FIR filter $\mathbf{H_i}$ modeling the transfer function between the control input $y(m)$ and the component of the signal $e_i(m)$ ($i = 1$ to 4) originating from the control source. The various differentials involved in the signals e_i are measured in practice with Eqs. (41)–(44). Similarly, four filtered reference signals can be expressed as

$$r_i(m-k) = \sum_{j=0}^{J-1} h_{ij}(m) x(m-j-k), \quad i = 1,...,4; \quad k = 0,1,...,I-1. \tag{52}$$

The following vector notations are introduced:

$$\mathbf{W}^T(m) \equiv [w_0(m) \ldots w_{(I-1)}(m)], \tag{53}$$

$$\mathbf{R}_i^T(m) \equiv [r_i(m) \ldots r_i(m-I+1)], \quad i = 1,...,4, \tag{54}$$

$$\mathbf{R}^T(m) \equiv [\mathbf{R}_1 \ \mathbf{R}_2 \ \mathbf{R}_3 \ \mathbf{R}_4], \tag{55}$$

$$\mathbf{e}^T(m) \equiv [e_1(m) e_2(m) e_3(m) e_4(m)]. \tag{56}$$

Using Eqs. (47)–(54) and assuming that the coefficients of the control filter are time invariant, the signals e_i can be expressed as

$$e_1(m) = \frac{\partial^3 w_p}{\partial x^3}(m) + \mathbf{W}^T(m)\mathbf{R}_1(m), \tag{57}$$

$$e_2(m) = \frac{\partial w_p}{\partial t}(m) + \mathbf{W}^T(m)\mathbf{R}_2(m), \tag{58}$$

$$e_3(m) = \frac{\partial^2 w_p}{\partial x^2}(m) + \mathbf{W}^T(m)\mathbf{R}_3(m), \tag{59}$$

$$e_4(m) = \frac{\partial^2 w_p}{\partial t \partial x}(m) + \mathbf{W}^T(m)\mathbf{R}_4(m). \tag{60}$$

The instantaneous intensity is then obtained as

$$\begin{aligned} I_i(m) = EI\Bigg[& \left(\frac{\partial^3 w_p}{\partial x^3}(m) + \mathbf{W}^T(m)\mathbf{R_1}(m) \right) \\ & \times \left(\frac{\partial w_p}{\partial t}(m) + \mathbf{W}^T(m)\mathbf{R_2}(m) \right) \\ & - \left(\frac{\partial^2 w_p}{\partial x^2}(m) + \mathbf{W}^T(m)\mathbf{R_3}(m) \right) \\ & \times \left(\frac{\partial^2 w_p}{\partial t \partial x}(m) + \mathbf{W}^T(m)\mathbf{R_4}(m) \right) \Bigg]. \end{aligned} \tag{61}$$

The FXLMS algorithm is based on the concept of updating the control filter coefficients in the direction of the gradient of the mean-squared error with respect to the filter coefficients. The control filter converges to the optimal solution taken as the one which minimizes the mean-squared error, approximated by the instantaneous squared error signal $e^2(m)$. The update equation is therefore expressed as

$$\mathbf{W}(m+1) = \mathbf{W}(m) - \mu_0 \nabla_W e^2(m), \tag{62}$$

where μ_0 is a parameter taken to ensure convergence of the update equation. In the filtered-X least-mean-squares algorithm the squared error function $e^2(m)$ has to be quadratic in the control filter $\mathbf{W}$.[19] As the instantaneous intensity is quadratic in the filter control $\mathbf{W}$, the squared error $e^2(m)$ is set equal to the instantaneous intensity, and the update equation is given by

$$\mathbf{W}(m+1) = \mathbf{W}(m) - \mu_0 \nabla_W I_i(m). \tag{63}$$

The convergence of the algorithm to a minimum assumes that the second derivative of the instantaneous intensity with respect to the filter weights is positive (error surface with a positive curvature). The gradient of I_i with respect to $\mathbf{W}$ can be written as

$$\nabla_W I_i = EI(\mathbf{R}^T \mathbf{e_P}) = EI[\mathbf{R_1} e_2 + \mathbf{R_2} e_1 - \mathbf{R_3} e_4 - \mathbf{R_4} e_3], \tag{64}$$

where the reference to the time index m is omitted and the vector $\mathbf{e_P}$ is defined by

$$\mathbf{e_P}(m) = \mathbf{P}\mathbf{e}(m), \tag{65}$$

and the permutation matrix $\mathbf{P}$ is defined by

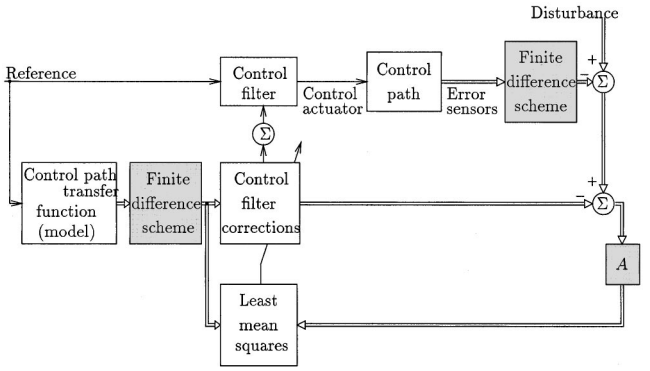


FIG. 14. Schematic representation of the control algorithm.

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{bmatrix}. \tag{66}$$

Note that the mean squared error signal $E(e^2)$ is estimated by the instantaneous squared error signal $e^2(m)$ at every sample. As convergence is found to occur in the mean,[20] minimizing the instantaneous squared error signal is equivalent to minimizing the mean squared error signal. Thus the algorithm developed is equivalent to controlling active intensity, the temporal mean value of the instantaneous intensity.

In contrast to the proposed algorithm, the update equation developed by Schwenk *et al.*[7] used an ''effective error signal'' (calculated as the sum of the transverse velocity times the wave number and the rotational velocity) and one filtered reference signal (the filtered transverse velocity), which makes their approach suitable to situations where the wave propagation is predominantly in one direction, and the near-field contribution of the control actuator can be neglected. The intensity cost function developed in this study is based on the general expression of structural intensity and uses five acceleration signals and a finite difference scheme. All the terms, i.e., the waves propagating in both directions and also the evanescent waves, are taken into account in the adaptation algorithm.

The schematic representation of this algorithm using four error functions e_i, and four filtered reference signals R_i, is shown in Fig. 14. The control path transfer functions are indentified before the control is applied using the same algorithm and injecting white noise to the control actuator. The algorithm was implemented on a Spectrum TMS 320C31 digital signal processing board, interfaced with three Spectrum AM/D16SA converters. The reference input signal was taken as the signal from the generator that drives the disturbance shaker. The sampling rate was set at 3 kHz in the experiments.

Note that when controlling acceleration, the error signal $e(m)$ is taken as the acceleration measured by the third accelerometer of the intensity probe.

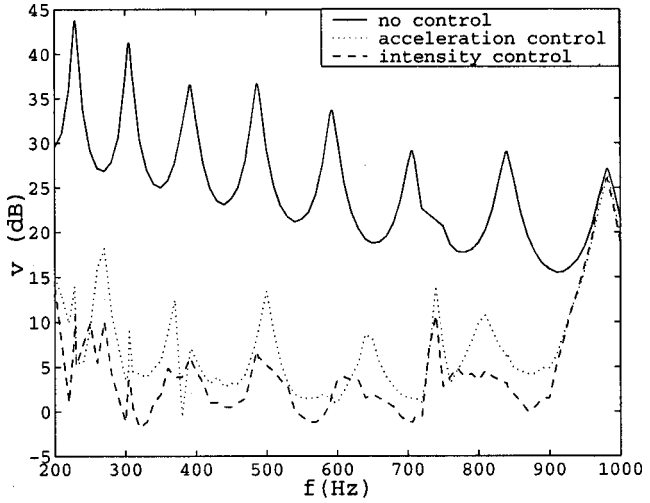


FIG. 15. Measured transverse velocity at the end of the beam for various control strategies, with the error sensor located at $x_e = 0.38$ m.

C. Experimental results

Figures 15 and 16 show the transverse velocity measured with a laser vibrometer at the end of the beam downstream of the error sensor (Fig. 11) for different control strategies and an harmonic excitation. Figure 15 shows active control results when the control PZT is located at $x_{s1} = 0.1$ m and the center of the intensity sensor is located at $x_0 = 0.38$ m. In this configuration the actuator–sensor separation is 0.44λ at 200 Hz and 0.95λ at 1 kHz. Controlling intensity gives better attenuations than controlling the acceleration. The peaks in the acceleration control correspond to vibration nodes at the accelerometer position; the absence of such nodes in the intensity produces a more uniform attenuation over the frequency range.

Figure 16 shows active control results when the control PZT is located at $x_{s1} = 0.1$ m and the center of the intensity sensor is located at $x_0 = 0.26$ m. In this configuration the actuator–sensor separation is 0.18λ at 200 Hz and 0.41λ at 1 kHz, and the probe center is in the near field, relative to a wavelength, of the control actuator in the frequency range investigated. In this case, intensity control provides a significantly larger vibration attenuation downstream of the error sensor than acceleration control. Acceleration control performance depends on the error sensor location with respect to vibration nodes; in contrast, intensity control performance is less affected by the error sensor location and produces a more uniform attenuation in frequency.

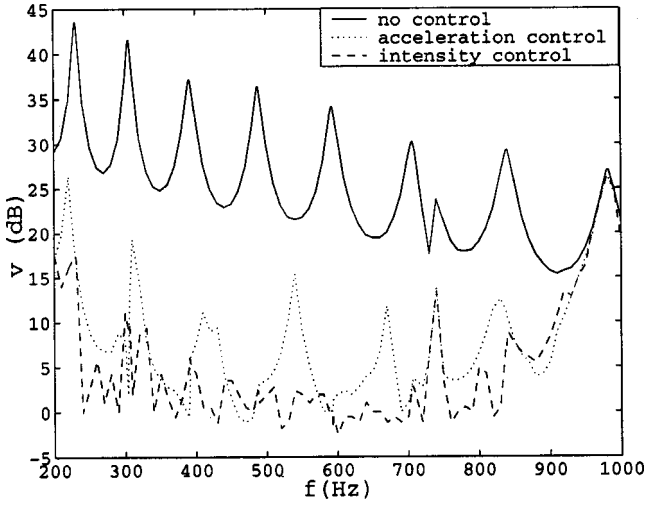


FIG. 16. Measured transverse velocity at the end of the beam for various control strategies, with the error sensor located at $x_e = 0.26$ m.

At low frequency, stability problems arise due to the finite difference scheme used to evaluate the error function, and in particular to the spacing between the accelerometers and the sampling rate.[7] Note that the mode of the structure at 980 Hz is not controlled. This is due to the control actuator location, the ends of which coincide with antinodes of this mode. A performance degradation is also noted at 740 Hz, which corresponds to the stinger resonance on the disturbance shaker. In this situation the shaker excites wave types other than flexural waves in the beam, which are not accounted for in the control strategy.

III. CONCLUSIONS

Active control of structural intensity in finite beams was investigated theoretically and experimentally in this article. The quadratic minimization problem associated to intensity control involves a nonpositive definite quadratic form in general. In this respect, the approach is limited to cases where the geometry is such that the intensity at the error sensor location will have the same sign for the control source and the primary disturbance. Optimal active control simulations demonstrate the advantage of choosing this energy-based strategy rather than the strategy of minimizing the transverse acceleration: the control performance is much less dependent on the error sensor location, and is not affected by the evanescent near-field waves emanating from the control actuators or structural discontinuities. This allows the sensors to be placed in the direct vicinity of the control actuators, with the advantage of a physically compact system. A finite difference approach using five accelerometers was used to experimentally measure and minimize structural intensity. A feedforward filtered-X LMS algorithm was adapted to this energy-based control problem. In contrast to previous work, the algorithm accounts for both evanescent and propagating waves in the minimization of structural intensity. The experimental results are consistent with the conclusions drawn from the theoretical analysis. The experimental control results show that intensity control is more effective than transverse acceleration control, especially when the control actuator–error sensor separation is smaller than a structural wavelength. However, low-frequency limitations related to the accurate measurement of structural intensity using the finite difference scheme were encountered. Future perspectives of this work include the use of PVDF sensors in order to measure structural intensity.

ACKNOWLEDGMENTS

We would like to thank the numerous people from the G.A.U.S and Jean-Claude Pascal and Bruno Gazengel from the L.A.U.M. who have helped in the preparation of this article.

APPENDIX: EXPRESSION OF THE BOUNDARY MATRIX

The matrix Γ in Eq. (15) is given by

$$\Gamma=\begin{bmatrix}
\frac{\Lambda_{1bL}}{\beta_{bL}^{j}} & \frac{\Lambda_{2bL}}{\beta_{bL}} & \Lambda_{3bL}\beta_{bL}^{j} & \Lambda_{4bL}\beta_{bL} & 0 & 0 & 0 & 0 \\
\frac{\Lambda_{5bL}}{\beta_{bL}^{j}} & \frac{\Lambda_{6bL}}{\beta_{bL}} & \Lambda_{7bL}\beta_{bL}^{j} & \Lambda_{8bL}\beta_{bL} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{\Lambda_{1b0}}{\beta_{b0}^{j}} & \frac{\Lambda_{2b0}}{\beta_{b0}} & \Lambda_{3b0}\beta_{b0}^{j} & \Lambda_{4b0}\beta_{b0} \\
0 & 0 & 0 & 0 & \frac{\Lambda_{5b0}}{\beta_{b0}^{j}} & \frac{\Lambda_{6b0}}{\beta_{b0}} & \Lambda_{7b0}\beta_{b0}^{j} & \Lambda_{8b0}\beta_{b0} \\
\frac{1}{\beta_i^{j}} & \frac{1}{\beta_i} & \beta_i^{j} & \beta_i & -\frac{1}{\beta_i^{j}} & -\frac{1}{\beta_i} & -\beta_i^{j} & -\beta_i \\
-\frac{j}{\beta_i^{j}} & -\frac{1}{\beta_i} & j\beta_i^{j} & \beta_i & \frac{j}{\beta_i^{j}} & \frac{1}{\beta_i} & -j\beta_i^{j} & -\beta_i \\
-\frac{1}{\beta_i^{j}} & \frac{1}{\beta_i} & -\beta_i^{j} & \beta_i & \frac{1}{\beta_i^{j}} & -\frac{1}{\beta_i} & \beta_i^{j} & -\beta_i \\
\frac{j}{\beta_i^{j}} & -\frac{1}{\beta_i} & -j\beta_i^{j} & \beta_i & -\frac{j}{\beta_i^{j}} & \frac{1}{\beta_i} & j\beta_i^{j} & -\beta_i
\end{bmatrix}, \tag{A1}$$

where

$$\Lambda_{1b}=j\omega\left(Z_{11b}-jkZ_{12b}+\frac{\tilde{E}Ik^2}{j\omega}\right), \tag{A2}$$

$$\Lambda_{2b}=j\omega\left(Z_{11b}-kZ_{12b}-\frac{\tilde{E}Ik^2}{j\omega}\right), \tag{A3}$$

$$\Lambda_{3b}=j\omega\left(Z_{11b}+jkZ_{12b}+\frac{\tilde{E}Ik^2}{j\omega}\right), \tag{A4}$$

$$\Lambda_{4b}=j\omega\left(Z_{11b}+kZ_{12b}-\frac{\tilde{E}Ik^2}{j\omega}\right), \tag{A5}$$

$$\Lambda_{5b}=j\omega\left(Z_{21b}-jkZ_{22b}+\frac{\tilde{E}Ik^3}{\omega}\right), \tag{A6}$$

$$\Lambda_{6b}=j\omega\left(Z_{21b}-kZ_{22b}-\frac{\tilde{E}Ik^3}{j\omega}\right), \tag{A7}$$

$$\Lambda_{7b}=j\omega\left(Z_{21b}+jkZ_{22b}-\frac{\tilde{E}Ik^3}{\omega}\right), \tag{A8}$$

$$\Lambda_{8b}=j\omega\left(Z_{21b}+kZ_{22b}+\frac{\tilde{E}Ik^3}{j\omega}\right), \tag{A9}$$

$$\beta_l=e^{kx_l}, \tag{A10}$$

$$\beta_l^{j}=e^{jkx_l}, \tag{A11}$$

with $l=i,0,L$ and $b=0,L$, meaning, respectively, the point force and moment location (x_i) and boundaries locations (x_0,x_L), and where the $Z_{ijb}\,(i,j=1,2)$ are the coefficients of the 2×2 matrix Z_b describing the beam's boundary conditions.

[1] X. Pan and C. H. Hansen, ''Effect of end condition on the active control of beam vibration,'' J. Sound Vib. **168**(3), 429–448 (1993).
[2] B. R. Mace, ''Active control of flexural vibrations,'' J. Sound Vib. **114**(2), 253–270 (1987).
[3] S. J. Elliott and L. Billet, ''Adaptive control of flexural waves propagating in a beam,'' J. Sound Vib. **163**(2), 295–310 (1993).
[4] M. J. Brennan, S. J. Elliott, and R. J. Pinnington, ''Strategies for the active control of flexural vibration a beam,'' J. Sound Vib. **186**(4), 657–688 (1995).
[5] W. Redman-White, P. A. Nelson, and A. R. D. Curtis, ''Experiments on the active control of flexural wave power flow,'' J. Sound Vib. **112**(1), 187–191 (1987).
[6] G. P. Gibbs and C. R. Fuller, ''Experiments on active control of vibrational power flow using piezoceramic actuators/sensors,'' AIAA J. **30**(2), 457–463 (1992).
[7] A. E. Schwenk, S. D. Sommerfeldt, and S. I. Hayek, ''Adaptive control of structural intensity associated with bending waves in a beam,'' J. Acoust. Soc. Am. **96**(5), 2826–2835 (1994).
[8] A. K. A. Pereira, F. J. O. Moreira, and J. R. F. Arruda, ''Active control of the structural intensity in beams using a frequency domain adaptive method,'' AIAA Pap. 98-1798, 841–849 (1998).
[9] D. U. Noiseux, ''Measurement of power flow in uniform beam and plates,'' J. Acoust. Soc. Am. **47**, 238–247 (1970).
[10] J.-C. Pascal, X. Carniel, V. Chalvidan, and P. Smigielski, ''Determination of structural intensity and mechanical excitation using holographic interferometry,'' in *Proceedings of the Fourth International Congress on Intensity Techniques* (CETIM, Senlis, 1993), pp. 137–143.
[11] G. Pavic, ''Measurement of structure borne wave intensity, part I: formulation of the methods,'' J. Sound Vib. **49**(2), 221–230 (1976).
[12] G. Pavic, ''Structure-borne energy flow,'' in *Encyclopedia of Acoustics*, edited by M. J. Crocker (Wiley-Interscience, New York, 1997), Vol. 2, pp. 881–891.
[13] K. S. Alfredsson, ''Influence of local damping on active and reactive

power flow,'' in *Proceedings of the Fourth International Congress on Intensity Techniques* (1993), pp. 71–78.
[14] C. H. Hansen and S. D. Snyder, ''Active control of noise and vibration,'' E&FN Spon (1997).
[15] S. J. Kim and J. D. Jones, ''Optimisation of piezo actuator/substructure coupling for active noise and vibration control,'' in *Proceedings of the Conference on Recent Advances in Active Control of Sound and Vibration* (1991), pp. 78–91.
[16] E. K. Dimitriadis, C. R. Fuller, and C. A. Rogers, ''Piezoelectric actuators for distributed vibration excitation of thin plates,'' J. Vibr. Acoust. **113**, 100–107 (1991).
[17] S. D. Sommerfeldt and P. J. Nashif, ''An adaptive filtered-x algorithm for energy-based active control,'' J. Acoust. Soc. Am. **96**, 300–306 (1994).
[18] X. Pan and C. H. Hansen, ''The effect of error sensor location and type on the active control of beam vibration,'' J. Sound Vib. **165**(3), 497–510 (1993).
[19] B. Widrow and S. D. Stearns, *Adaptive Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1985).
[20] C. R. Fuller and G. P. Gibbs, ''Simultaneous active control of flexural and extensional waves in beams,'' J. Intell. Mater. Syst. Struct. **1**, 235–247 (1990).

Experimental study of sound propagation in a flexible duct

Lixi Huang, Y. S. Choy, R. M. C. So, and T. L. Chong
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong

(Received 8 June 1999; accepted for publication 5 May 2000)

Propagation of sound in a flexible duct is investigated both theoretically and experimentally. Strong coupling of sound and flexural waves on the duct wall is found when the wall-to-air mass ratio is of the order of unity. The axial phase speed of sound approaches the *in vacuo* speed of flexural waves (subsonic in this case) at low frequencies. However, a speed higher than the isentropic sound speed in free space (340 m/s) is found beyond a critical frequency which is a function of the mass ratio. Experiments using a duct with a finite section of tensioned membrane are compared with the propagating modes pertaining to the infinite membrane model. Satisfactory quantitative agreement is obtained and the measured phase speed ranges from 8.3 to 1348 m/s. In the moderate frequency range, the theory predicts high spatial damping rate for the subsonic waves, which is consistent with the experimental observation that subsonic waves become increasingly undetectable as the frequency increases. Substantial sound reflection is observed at the interface between the rigid and the flexible segments of the duct without cross-section discontinuity, which, together with the high spatial damping, could form a basis for passive control of low-frequency duct noise. © *2000 Acoustical Society of America.* [S0001-4966(00)03308-7]

PACS numbers: 43.50.Gf, 43.20.Mv, 43.20.Hq [MRS]

INTRODUCTION

Strong sound–structure coupling normally occurs in underwater applications where the fluid density is of the same order of magnitude as the structural density (Fahy, 1997); the same in air is possible at the presence of substantial flow speed. This article reports strong air–structure coupling without flow in a chosen parameter range and the work is motivated by the prospect of devising smart structures for broadband passive control of low-frequency noise, which is hitherto more effectively dealt with by active noise control techniques. In traditional passive noise control (Beranek and Ver, 1992), sound reflection is achieved through changing the duct cross section, such as in a vehicle exhaust muffler, and sound attenuation is invariably achieved by forcing air through porous media or structures with the exception of some membrane sound absorbers (to be discussed later). Both measures may have detrimental aerodynamic effects, namely high pressure loss, and even causing extra noise called ''self-noise.'' Besides, they are either too bulky or very ineffective in the low-frequency range such as below 500 Hz (Ingard, 1994). To deal with low-frequency noise, there are two types of membrane absorbers whose working mechanisms depend on the strong sound–structure coupling. Ford and McCormick (1969) described an earlier type consisting of layers of thin membranes, such as 0.2-mm aluminum sheets, stacked at a certain distance from one another. They are used in broadcasting studios, concert halls, etc., and substantial sound absorption occurs only at narrow bands around a few resonance frequencies of the membrane-cavity structure. Fuchs and his colleagues developed another type of membrane absorber (Ackermann *et al.*, 1988) in which a perforated thin plate is glued to an all-metal honeycomb structure forming an array of Helmholtz resonators. The device shows two resonance peaks relating to the Helmholtz resonance and the resonance of the cover plate, respectively. Another smooth cover sheet, typically of 0.1-mm-thick aluminum, is then added in front of the perforated plate to protect the resonator from flow and dust. The separation distance is typically 1 mm. The result is an improved performance between the two resonance peaks. Membrane absorbers of this type have provided useful attenuation of low-frequency noise down to some 60 Hz when used in a papermill exhaust (Ackermann and Fuchs, 1989), but exact theoretical prediction proved rather difficult (Frommhold *et al.*, 1994).

Recently, Huang (1999) proposed another type of membrane absorber for low-frequency duct noise. A membrane of moderate length forms part of a duct wall which vibrates in response to grazing incident noise. The membrane experiences flexural waves with phase speed less than the *in vacuo* wave speed, which is in turn much less than the speed of sound in air. It is theoretically demonstrated that (a) substantial reflection of noise is possible without the change in duct cross section area, and (b) the slowly traveling coupled waves provide a powerful mechanism of energy dissipation at an attenuation rate (in dB per unit distance of travel) inversely proportional to the phase speed. Both (a) and, to a lesser extent, (b) are demonstrated experimentally in the present study. Theoretically, the flexural waves can be slowed down indefinitely by increasing the compliance of the membrane, but in practice there is a limit on the membrane thickness that may be used. The present study is motivated by the desire to further understand the coupled waves on a thin membrane under tension. The focus is on the mapping of the whole wave speed spectrum. At low frequencies, the phase speed is found to be less than the *in vacuo* wave speed of the membrane. At higher frequencies, however, waves of speed higher than the speed of sound in air, namely supersonic waves, are also found although not necessarily useful for noise control.

The fact that the speed of waves, including sound waves, is controlled by the compliance of passage walls is well known (Lighthill, 1978). Shallow water waves are the most familiar example where the free surface as a boundary controls the wave speed. Another example is wave propagation in the aorta of large mammals where wave speed ranges from 5 to 10 m/s (Lighthill 1978, p. 99). In the broader context of biological applications, Shapiro (1977) studied the incompressible flow through flexible tubes where the very low eigenwave speed replaces the usual sound speed in controlling the flow behavior, some of which is quite extraordinary. For example, when the flow speed exceeds the eigenwave speed, friction causes static pressure to rise instead of falling, similar to supersonic airflow in the contraction–expansion nozzle. Physical experiments using water have also been constructed to demonstrate such an effect (Kececioglu *et al.*, 1981). In the area of wave attenuation, Horne and Hansen (1982) demonstrated that as much as 60-dB reduction could be achieved by the coupling of water wave and the pipe wall of comparable acoustic impedance. In a separate work by Dunlop (1992), the speed of sound in such a coupled system was measured and the result was shown to agree with simple theory. A more general coverage of this topic was recently given by Fahy (1997).

Similar examples do not seem to exist in airflow through collapsible tubes in the sense that most such experiments were conducted in a starling-resistor-type configuration (e.g., Gavriely *et al.*, 1989). In such cases the tube collapses in one cross section, this can hardly be modeled as a long channel of uniform property. On the other hand, some theoretical attempts have been made in the context of submerged structure (Ko, 1994) to resolve the eigenvalues of structural vibrations coupled with either air or water flow inside a duct. The duct considered in the example is made of 2-in.-thick steel. The structure-to-fluid mass ratio is very high when the fluid is air, and is of the order of unity when the fluid is water. In both cases subsonic eigenwaves are found. Whether the eigenvalues found in the case of very high mass ratio represent anything significant in reality has not been validated experimentally. Supersonic eigenwaves are inadmissible in Ko's model because of the presence of external fluid which extends to infinity. The present study aims to explore the coupled waves in a laboratory experiment using a tensioned membrane whose mass ratio is of the order of unity and the external fluid is limited to a finite cavity. Acoustic excitation is used and the sound waves measured over the membrane shows dominating features of the eigenvalues found theoretically. As a first attempt to understand the physics the effect of flow has not been included. The possibility of flow-induced vibration is also postponed, but it is believed that a clear understanding of wave behavior without flow could benefit such instability studies in the future.

In what follows, the eigenwave speed in a duct of infinite length is first calculated, the main results being the prediction of both subsonic and supersonic phase speeds in air depending on frequency. The existence and significance of these wave speeds are then examined experimentally in a duct with finite length. Both subsonic and supersonic waves are found.

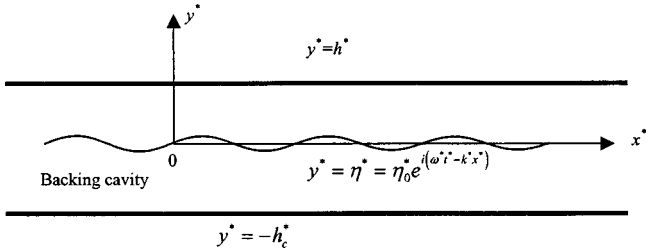


FIG. 1. Geometrical configuration of the theory.

I. THEORY

To illustrate the acoustic principle, a simple configuration of two-dimensional channel of height h^* with a single lower flexible wall at $y^*=0$ is considered, as shown in Fig. 1. The word ''duct'' will still be used in the text, but the three-dimensional effect is not considered theoretically. Some three-dimensional effects are inevitable in the experiment, but the agreement between the measured data and the two-dimensional prediction is satisfactory after taking special precautions to minimize these effects. It is assumed that the flexible wall is very thin and the only mechanical restoring force against displacement is the tensile stress applied. The flexible wall vibrates and radiates sound externally, and the so-called breakout noise will be important (Cummings, 1994) in the context of ventilation noise. In order to avoid such leaking, and more importantly to avoid the three-dimensional effect in the experiment, we consider an additional lower rigid wall at $y^*=-h_c h^*$, where h_c is the ratio of channel heights. The flexible wall has a mass per unit area of M^*, a damping coefficient D^*, and a tensile force T^* is applied. The flexible wall is subject to a traveling wave perturbation of displacement

$$\eta^* = \eta_0^* e^{i(\omega^* t^* - k^* x^*)}.$$

Under these conditions, what is the characteristic flexural wave speed $c^*=\omega^*/k^*$? Note that, in an infinite duct, there cannot be separate wave speeds for the structural and air motions although the same may not be true in the experiment where the structure is finite. A similar model excluding the effect of external fluid was briefly studied in a previous paper (Huang, 1999). The result of that model will now be used as a special case for comparison.

All variables are normalized by three basic quantities ρ_0^*, h^*, and c_0^*. Here ρ_0^* and c_0^* are, respectively, the air density and isentropic speed of sound in free space. After normalization these variables are denoted by

$$\begin{aligned}
&x=\frac{x^*}{h^*}, \quad y=\frac{y^*}{h^*}, \quad t=\frac{c_0^* t^*}{h^*}, \quad \eta_0=\frac{\eta_0^*}{h^*},\\
&c=\frac{c^*}{c_0^*}, \quad \omega=\frac{\omega^* h^*}{c_0^*}=k_0, \quad k=\frac{\omega}{c},\\
&M=\frac{M^*}{\rho_0^* h^*}, \quad c_T=\frac{\sqrt{T^*/M^*}}{c_0^*}, \quad D=\frac{D^*}{\rho_0^* c_0^*},\\
&\phi=\frac{\phi^*}{c_0^* h^*}, \quad p=\frac{p^*}{\rho_0^*(c_0^*)^2},
\end{aligned} \tag{1}$$

where M, D, and c_T are, respectively, the mass ratio, the dimensionless damping coefficient, and the *in vacuo* panel wave speed, and ϕ and p are, respectively, the dimensionless perturbation velocity potential and sound pressure inside the channel. Note that the dimensionless angular frequency ω is equal to the dimensionless wave number based on c_0^* and k_0, and is related to the ratio of the driving frequency, f^*, to the first cut-on frequency of sound in the rigid channel, f_1^*, by a factor of π. Thus

$$\omega = k_0 = \pi f, \quad \text{where } f = f^*/f_1^* \quad \text{and } f_1^* = c_0^*/2h^*.$$

In dimensionless form, the wave equation in fluid becomes

$$\left(-\frac{\partial^2}{\partial t^2} + \nabla^2 \right) \phi = 0,$$

which, for harmonic excitation, is reduced to the Helmholtz equation

$$\frac{\partial^2 \phi}{\partial y^2} = k_a^2 \phi,$$

where, for real wave number k,

$$k_a^2 = k^2 - k_0^2, \quad k_a = \begin{cases} \sqrt{k^2 - k_0^2}, & c<1,\ k>k_0, \\ i\sqrt{|k^2 - k_0^2|}, & c>1,\ k<k_0. \end{cases} \tag{2}$$

Here k_a is real for subsonic waves but imaginary for supersonic waves. The solution for the perturbation velocity potential inside the channel satisfying the following rigid wall condition and the kinematic condition on the flexible wall,

$$\left.\frac{\partial \phi}{\partial y}\right|_{y=1} = 0, \quad \left.\frac{\partial \phi}{\partial y}\right|_{y=0} = \frac{\partial \eta}{\partial t} = i\omega\eta,$$

is

$$\phi = -\frac{i\omega\eta}{k_a} \frac{e^{k_a(1-y)} + e^{-k_a(1-y)}}{e^{k_a} - e^{-k_a}}. \tag{3}$$

The pressure perturbation on the upper side of the flexible wall, $y = 0+$, is denoted by p_+,

$$p_+ = -\left.\frac{\partial \phi}{\partial t}\right|_{y=\eta} = -\frac{\omega^2 \eta}{k_a} \coth k_a. \tag{4}$$

The boundary conditions for the perturbation velocity potential in the lower channel are

$$\left.\frac{\partial \phi}{\partial y}\right|_{y=-h_c} = 0, \quad \left.\frac{\partial \phi}{\partial y}\right|_{y=0} = \frac{\partial \eta}{\partial t} = i\omega\eta,$$

so that

$$\phi = +\frac{i\omega\eta}{k_a} \frac{e^{k_a(h_c+y)} + e^{-k_a(h_c+y)}}{e^{k_a h_c} - e^{-k_a h_c}}.$$

Similarly, the pressure on the lower surface is found to be

$$p_- = -\left.\frac{\partial \phi}{\partial t}\right|_{y=0} = +\frac{\omega^2 \eta}{k_a} \coth k_a h_c. \tag{5}$$

The wave impedance of the fluid loading on the wall is defined and calculated as follows:

$$Z = \frac{p_+^* - p_-^*}{\rho_0^* c_0^* d\eta^*/dt^*} = \frac{p_+ - p_-}{i\omega\eta} = \begin{cases} +i\dfrac{k_0}{k_a}[\coth k_a + \coth k_a h_c], & \text{for } c<1; \\ -i\dfrac{k_0}{|k_a|}[\cot(|k_a|) + \cot(|k_a h_c|)], & \text{for } c>1. \end{cases}$$

For long waves without change of amplitude, $|k_a| \ll 1$, hence $\coth k_a$ and $\cot(|k_a|)$ are positive. The effect of fluid loading on the wall dynamics is the added mass for subsonic waves; however, for supersonic waves it is a spring stiffness. The presence of a trigonometric function in the supersonic case signifies oblique wave fronts similar to those in the high-order duct modes with rigid boundaries. The dynamics of the flexible wall is described by the equation

$$T\frac{\partial^2 \eta}{\partial x^2} - D\frac{\partial \eta}{\partial t} + p_- - p_+ = M\frac{\partial^2 \eta}{\partial t^2}, \tag{6}$$

where the damping coefficient D may be substituted by $M\omega\sigma$ after invoking the concept of complex Young's modulus. Here, σ is the loss factor of the material defined as the ratio of energy lost per radian to the maximum potential energy. Inserting the fluid pressures of Eqs. (4) and (5), Eq. (6) becomes

$$M\omega^2(1 - i\sigma) + \frac{\omega^2}{k_a}[\coth k_a + \coth k_a h_c] - Mc_T^2 k^2 = 0. \tag{7}$$

Mathematically, the complex eigenvalue $k = k_r + ik_i$ has to be sought on the entire complex plane. But in reality, only those in the fourth quadrant of the complex plane, $k_r > 0$, $k_i < 0$, are of physical interest. These eigenvalues represent propagating waves with decaying amplitude. For a lossless material, $\sigma = 0$, numerical search in this wave number quadrant indicates that eigenvalues exist only on the pure real axis, $k = k_r$, and on the pure imaginary axis, $k = ik_i$. The former represents free waves propagating in the axial direction, while the latter represents standing waves between the hard wall and the flexible wall, which we call the vertical mode. In both cases there is no amplitude decay in the direction of wave travel. This is understandable since there is no energy dissipation mechanism when $\sigma = 0$. When there is damping ($\sigma > 0$), however, the eigenvalues are expected to be complex. The numerical results show that there is no sudden jump in solutions when the damping approaches zero. The following numerical procedure is used to find the complex eigenwave number which then gives the eigenwave speed $c = \omega/k_r$ and the attenuation in decibels per unit distance of h^*,

$$\alpha = 20\log_{10} e^{-k_i} = -8.686 k_i.$$

Given the flexible wall parameters M, c_T, and σ, the lower channel height h_c, and angular frequency ω, c and k_i are the two independent variables in Eq. (7). The value of the left-hand side of Eq. (7), denoted as z, is calculated over a uni-

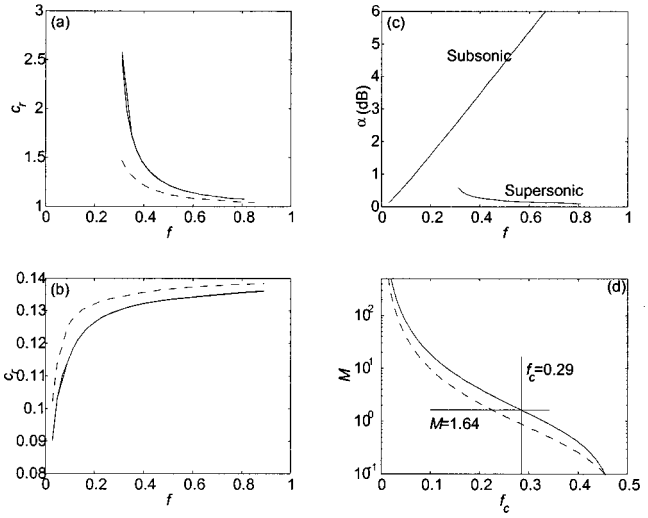


FIG. 2. Theory of sound propagation in a channel with a flexible wall with $M=1.64$ and $c_T=0.14$. Three cases are calculated: (I) lossless flexible wall without external fluid loading for which the results are shown in dashed lines; (II) lossless flexible wall with a lower channel of depth ratio $h_c=1.1$; and (III) flexible wall of loss factor $\sigma=0.1$ with lower channel of $h_c=1.1$. Note that cases II and III are shown in solid lines but the differences are too small to distinguish. (a) is the supersonic eigenwave speed. (b) is the subsonic eigenwave speed. (c) is the attenuation rate for case III. (d) is the functional relationship between mass ratio and the ''cut-on'' frequency for supersonic eigenwaves.

form mesh on the plane of c and k_i. The eigenvalues of c and k_i are found by the intersections of contour lines of $\mathrm{Re}(z)=0$ and $\mathrm{Im}(z)=0$ utilizing Matlab software.

Figure 2 is an example which covers the parameters used in the experiments: $M=1.64$ and $c_T=0.14$. Three cases are calculated; case I is a lossless wall without external fluid loading, namely with external vacuum, and is shown in the figures as dashed lines; case II is the lossless wall with a lower rigid wall forming a second channel of width $h_c=1.1$; case III is based on case II, but with a loss factor $\sigma=0.1$ which is of the order of magnitude expected of the membrane material used in the experiments. Figure 2(a) and (b) is the eigenwave speeds for the three cases mentioned. The two solid lines are for cases II and III, respectively, but they can hardly be distinguished from one another. This means that for the normal materials the eigenwave speed is hardly influenced by its friction. Figure 2(b) shows that subsonic eigenwaves approach $c_T=0.14$ as $f\to\infty$, which can be demonstrated as follows. As $k_a\to k_0\to\infty$, $\coth k_a/k_a$ and $\coth k_a h_c/k_a\to 0$. Equation (7) then gives $c\to c_T$ for the simple case of $\sigma=0$. Figure 2(b) shows the supersonic eigenwaves which are only found for frequencies above a certain value to be explained with Fig. 2(d).

Figure 2(c) gives the sound attenuation rate α (in dB per distance equal to the channel height h^*) for both subsonic and supersonic eigenwaves. Supersonic waves hardly decay, but subsonic waves experience attenuation compatible with that of sound propagating through a duct lined with porous sound absorption materials (Ingard, 1994). The reason for such a high attenuation in the present case is that the wave speed is much lower so that, for a unit travel distance, the resident time is longer and the frictional losses are greater.

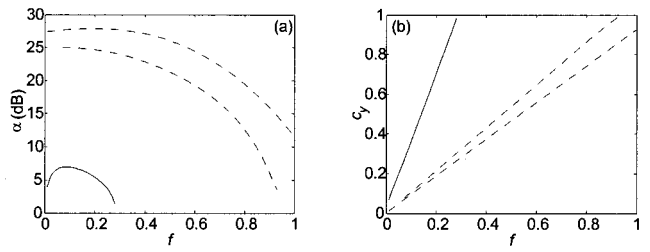


FIG. 3. Vertical mode for lossless membrane with $M=1.64$, $c_T=0.14$, and $h_c=1.1$. (a) is the axial attenuation rate and (b) is the vertical wave speed. The solid curve in (a) corresponds to the solid curve in (b), and similarly for the dashed curves, in the same clockwise order.

The attenuation is also proportional to frequency which is inherited from the simple damping model adopted, $D=M\omega\sigma$. Thus, at intermediate and high frequencies, say $f>0.3$, it would be impossible to observe these waves, although theoretically they are found to coexist with supersonic waves. This is consistent with our experimental observation that subsonic waves can hardly be measured beyond 500 Hz.

The frequency of the wave speed singularity ($c_r\to\infty$) shown in Fig. 2(a) may be called the critical frequency, f_c, above which supersonic waves may occur, namely supersonic cut-on frequency. f_c may be found by letting $k=\omega/c\to 0$, $k_a\to ik_{0c}$, and $k_{0c}=\pi f_c$, so that for $\sigma=0$ Eq. (7) immediately gives the following relationship:

$$M=[\cot k_{0c}+\cot k_{0c}h_c]/k_{0c}, \tag{8}$$

which is shown in Fig. 2(d). Here f_c reduces as M increases, and in this particular example $f_c=0.29$. The comparison between cases I and II in the figure means that the lower channel, or air cavity, adds virtual mass to the membrane. Note that $\cot k_{0c}$ and $\cot k_{0c}h_c$ are periodic functions of k_{0c}. That means for a single mass ratio M, there are multiple critical frequencies associated with the multiple solutions for k_{0c}. In the example given above, the range of frequency is restricted to the first cut-on of the duct and the backing cavity, $k_0<\pi$, $k_0h_c<\pi$. As a result, there is only one critical frequency k_{0c} found. In the experiment, this means a frequency ceiling of 5 kHz for a square duct of 3 cm in cross section.

The vertical mode mentioned earlier for a lossless membrane, $\sigma=0$, has a pure imaginary eigenwave number $k=ik_i$. In this case, $k_a=ik_{ai}$ where $k_{ai}=\sqrt{k_i^2+k_0^2}$, and $k_a^{-1}\coth k_a=-k_{ai}^{-1}\cot k_{ai}$. The vertical wave speed is $c_y=\omega/k_{ai}$. The result is shown in Fig. 3 for a configuration identical to that of Fig. 2. The range for the search is limited to $-k_i<3.5$. Two modes are found in this range. One is the slowly decaying mode (solid curves) which meets the no-decay limit of $\alpha=0$ at the critical frequency predicted earlier, $f_c=0.29$, and the other (dashed curves) has a rather high attenuation rate in the axial direction. Note that the vertical wave speed reaches unity, $c_y=1$, at the critical frequency for the less damped mode. The possible presence of this mode in a duct of finite length could confuse the axially propagating modes near the edges of the membrane, but the problem is expected to diminish after a distance equal to a few duct height h^*.

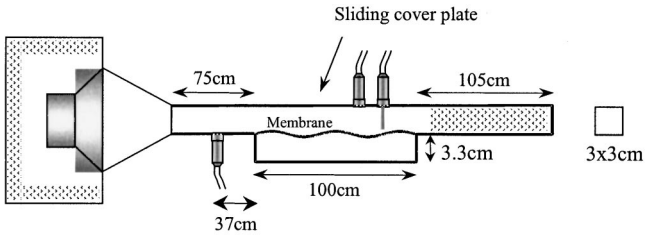


FIG. 4. The setup of the rig.

Mathematically speaking, the plane wave mode, $k=k_0$, is always a possible solution when the backing cavity is present. This solution is seen when the membrane experiences no displacement, $\eta=0$, and the two sides of the membrane experience equal pressure disturbance, $p_+=p_-$. This mode is mathematically trivial for the configuration of infinite duct, but it may become physically significant when the duct is finite. This is observed in the following experiment.

II. EXPERIMENTAL SETUP AND DATA PROCESSING

The rig is shown in Fig. 4. A loudspeaker driver unit of 6-in. diameter with 40-Hz resonance frequency is enclosed in a wooden box and is driven by a function generator and amplifier. The duct has a square cross section of 3×3 cm^2 and is coupled with the loudspeaker via a contraction cone. The duct has three regions: an upstream part of 75-cm length with rigid walls, a test section with a 1-m-long membrane on the lower side, and a downstream section with a 15-cm-long rigid wall and a 90-cm-long section filled with sound absorption material. Complete sound absorption is not expected since reflection occurs at the leading and trailing edges of the membrane anyway. The membrane is backed by a lower cavity 3.3 cm in depth to prevent breakout noise. A tensile force of 4.0 N, which produces 10% membrane elongation, is applied. The membrane used is 3M's transparent dispenser tape of thickness 0.065 mm and mass per unit area of $M^*=60\,\mathrm{g/m^2}$. In order to best model a two-dimensional configuration, a small clearance of less than 0.5 mm is allowed between the tensioned membrane and the side walls of the duct. Tests with the small gap sealed were also carried out (see later). Sound traveling over the membrane is expected to have a phase speed different from the normal speed of sound of about 340 m/s at a room temperature of 15 °C. The dispersion relation $c(f)$ is the main objective of the measurement.

Two B&K $\frac{1}{4}$-in. condenser-type microphones (type number 4135 with ±1-dB amplitude response from 6 Hz to 14 kHz) are installed flush with the duct walls: one at 37 cm upstream of the membrane and the other on the rigid wall opposite the membrane, which is designed to slide back and forth so that all axial locations can be covered with just a few microphone insertion holes. To get the sound field of the whole length over the membrane at repeatable loudspeaker excitation, the signal taken from the ''mobile'' microphone over the membrane section is compared with that from the upstream microphone which acts as a reference. The amplitude ratio and phase difference between the two pressure measurements are determined using the data acquisition and analysis system described below. A probe microphone (B&K type 4182) with a 5-cm stiff probe head of 1.24-mm diameter is inserted through the sliding plate next to the $\frac{1}{4}$-in. microphone. It moves along with the $\frac{1}{4}$-in. microphone and is used to measure the sound pressure right above the membrane surface. The frequency response of this probe is 3 dB down from dc to 2 kHz which is the maximum frequency required for the probe measurement. The probe microphone measurement is necessary because the sound amplitude decays exponentially in the y direction for subsonic waves and the ratio of pressure measured by the two microphones at the same axial coordinate can be estimated from Eq. (3), or

$$\frac{p|_{y=0}}{p|_{y=1}}=\frac{\phi|_{y=0}}{\phi|_{y=1}}=\frac{e^{k_a}+e^{-k_a}}{2}=\cosh k_a\,.$$

For example, according to the solid line shown in Fig. 2(b), at $f=0.2$, $c_r=0.1265$, $k_a=4.9278$, and $\cosh k_a=69.0$. Experimental data indeed shows that subsonic waves are more evident near the membrane surface than on the opposite wall.

The microphones are supported by the B&K's Nexus four-channel conditioning amplifier (type 2693) and the signals are fed into a Pentium 266MHz DELL PC installed with a National Instruments AD card (type PCI-MIO-16-E-1) controlled by the LabVIEW software. Three narrow-band frequency sweeps (or chirps) are generated from the function generator: 50–300 Hz, 300 Hz to 2.5 kHz, 2.5–5 kHz. The sweep is linear ''ramp-up'' and the durations are 4, 2, and 1 s, respectively. The AD card sampling frequencies are 4, 16, and 32 kHz for the three bands, respectively, and data sampling is carried out over one sweep cycle. The three-band separation is necessary because the conditioning amplifier for the microphone requires a different setting in order to maximize the true resolution of the A/D conversion. The spatial step for the ''sliding'' microphones is 2 cm for low frequencies and 1 cm for high frequencies. This step size is found to be sufficiently small to resolve the waveforms. For each position, the amplitude ratio and phase angle relative to the reference microphone are found at all frequencies. The data output is then sorted out in terms of amplitude and phase distribution along the membrane for all individual frequencies. A set of harmonic excitation tests were also carried out, which confirmed the validity of the chirp measurement.

Taking $\rho_0^*=1.225\,\mathrm{kg/m^3}$ and $c_0^*=340\,\mathrm{m/s}$, the dimensionless parameters M and c_T are calculated as follows,

$$M=\frac{60\times10^{-3}}{1.225\times0.03}=1.64,$$

$$c_T^*=\sqrt{\frac{4.0\ \mathrm{N}}{0.0603\ \mathrm{kg/m}\times0.03\ \mathrm{m}}}=47.0\ \mathrm{m/s},$$

$$c_T=\frac{c_T^*}{c_0^*}=0.14,$$

for which the supersonic cut-on frequency is predicted at

$$f_c=0.2854,\quad f_c^*=1617\ \mathrm{Hz}.$$

The real damping for the membrane is also a function of the applied tension and the measurement of such a parameter requires vacuum environment which is not available in the

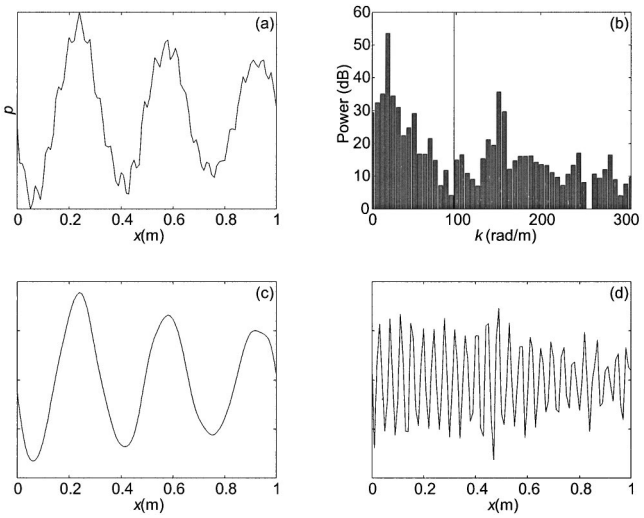


FIG. 5. Analysis of spatial sound pressure distribution for frequency 1030 Hz. (a) is the pressure waveform and (b) is the spectrum in which two peaks are identified within the two bands which are used to construct bandpass signals shown in (c) and (d). The speed of sound represented by the two peaks are 344.0 and 42.7 m/s, respectively.

present study. As a result the analysis on the damping of sound waves is only qualitative.

The amplitude and the phase angle of the pressure signals taken by the mobile microphone are compared with the simultaneous measurement taken by the upstream reference microphone. The amplitude ratio and the phase delay are found and used to construct the pressure distribution along the membrane. This distribution is equivalent to the simultaneous measurement for all points. Shown in Fig. 5 is an example taken at 1030 Hz. Figure 5(a) is the measured axial pressure distribution which shows long waves being modulated by short waves. Since the membrane length is limited, the number of wave cycles found is few and the wave number resolution is therefore limited to 1 m^{-1}. In most cases the true wave number may fall between two discrete wave numbers given by the discrete Fourier transform and the spectral leaking is corrected by taking the spectral center-of-gravity, $k_{\text{peak}}=\Sigma k_j P_j/\Sigma P_j$, where P and k are spectral power and wave number, respectively, over five wave number points, two on the left and two on the right of the apparent spectral peak. Wave speed c_r is then found by ω/k_{peak}. With this procedure two peaks can be identified in the power spectrum shown in Fig. 5(b), one at $k=18.8$ rad/m and the other at $k=151.6$ rad/m, corresponding to wave speeds of $c=344.0$ and 42.7 m/s, respectively. In order to see the spatial features of the two perceived wave components, the spectrum is divided into two parts shown by the thin solid line in Fig. 5(b). Inverse FFT is applied to each side and the decomposed waveforms are shown in Figs. 5(c) and 5(d), the latter being amplified for easier observation. The long wave propagates at sonic speed and the short wave has phase speed close to $c_T^*=47.0$ m/s. In this case the sonic wave dominates and both waves show apparent spatial decay. But the exact attenuation rate cannot be calculated directly from the amplitude envelope due to the presence of significant wave reflection. The spectral decomposition is similar to the harmonic wavelet transform found by Newland (1993), who chose to utilize the octave bands as the convenient segmentation and showed that a set of orthogonal harmonic wavelets may be defined as

$$\omega(x)=(e^{i4\pi x}-e^{i2\pi x})/i2\pi x$$

for dimensionless wave number from 2π to 4π. The magnitude of such a wavelet decays as $(2\pi x)^{-1}$, which is much less compact than other popular wavelet techniques. However, the latter suffer from the fact that a clear physical interpretation for the wavelets is not always possible. Therefore, the concept of harmonic wavelet is chosen with flexible choice of spectral bands adopted. The wider the spectral band, the more compact the wavelet becomes and the better the spatial resolution it gives. In the present case, only two bands are chosen in order to maximize the spatial resolution.

Note that an impulse-style measurement is not appropriate in this study because (a) reflections occur very close to the test sections, leaving no time window for low-frequency measurement, and (b) excitation of higher duct modes may seriously pollute the data at moderate to high frequencies.

III. RESULTS AND DISCUSSION

Before analyzing the experimental data, it has to be borne in mind that the theoretical predictions shown in Fig. 2 were made for a duct of infinite length. Thus, the sound radiated by the flexural waves definitely manifest themselves into traveling waves of a single phase speed. In a finite duct, however, such sound waves could interact with the scattering at the leading and trailing edges of the membrane to produce a complicated pattern. Nevertheless, the acoustic field can be tactically divided into two parts: one part is the normal standing wave pattern in a finite duct of partially reflecting ending if the membrane were replaced by a rigid wall, while the other part is the wave radiated by the membrane vibration which is itself induced by the incident sound. The present interest lies in the latter, namely, the second peak shown in Fig. 5(b), which may or may not dominate the spectrum.

Two factors need to be clarified before analyzing the results. One is the influence of the backing cavity, and the other is the edge scattering. As shown in Fig. 4, the lower rigid wall is added to prevent breakout noise. But more importantly, the backing cavity enables the rig to simulate wave propagation in two spatial dimensions for which a theory is easily constructed for comparison. It is likely that the very light and limp membrane partition is almost transparent to sound waves. A plane wave mode propagating at the normal speed of sound of 340 m/s may dominate in some frequency ranges. For this mode, the membrane does not experience any vibration and the result is a trivial solution for the membrane dynamics. This outcome does not necessarily imply weak or no coupling between the sound and the membrane for other configurations. The second factor addressed here is the effect of scattering by the two edges of the membranes. All physically admissible eigenmodes may be excited and our attention is focused on whether the propagating mode dominates. Vertical modes shown in Fig. 3 have moderate to high attenuation rates, and are therefore unlikely to dominate

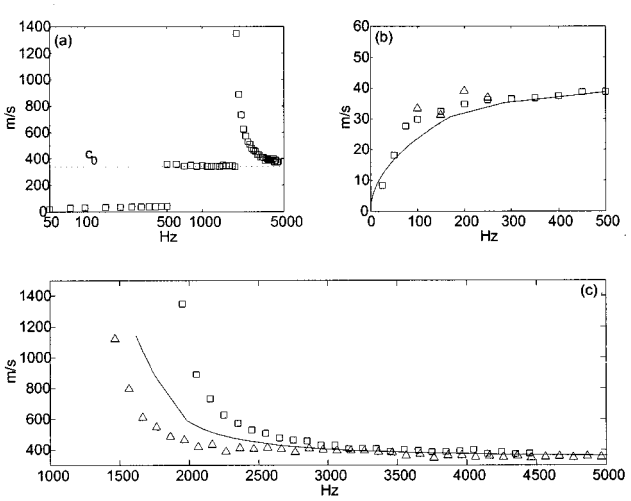


FIG. 6. Comparison of the dispersion relationship between theory and experiments. (a) is the overall distribution of experimental data, and (b) is the zoom-in view of the data in the subsonic regime. Theoretical prediction is given by the solid curve, squares are from the probe microphone measurement on the membrane surface, and triangles are from the $\frac{1}{4}$-in. microphone measurement of pressure on the hard wall. (c) compares the measured supersonic wave speeds with theoretical predictions (solid line). The squares are for the membrane with small sideway gaps simulating two-dimensional wave field and triangles are data taken when the gap is covered.

the waveforms over the bulk length of the membrane except for frequencies very close to the supersonic cut-on.

Including only the dominant propagating waves, Fig. 6(a) gives the dispersion relationship deduced from one set of experimental data using the sliding $\frac{1}{4}$ in. microphone flush with the rigid wall. The sound waves propagating at the normal speed of sound are always present, and are in fact dominant below 500 Hz. The reason why the normal speed of sound is not shown in the figure is that the test section is too short to have a complete waveform below 500 Hz. In spectral analysis, one finds that the corresponding spectral peak is located at $k \rightarrow 0$ as shown later in Fig. 7. Beyond 500 Hz, waves with $c_0 = 340$ m/s can be resolved and seen as an isolated peak in the wave number spectrum. The accompanying subsonic waves is rather weak, as shown in Fig. 5 for the probe microphone measurement near the membrane surface, and are not shown in Fig. 6(a) beyond 500 Hz. Note that 500 Hz happens to be where the spectral dominance changes from the subsonic wave number to the sonic wave number. It does not represent a cutoff frequency for the subsonic waves.

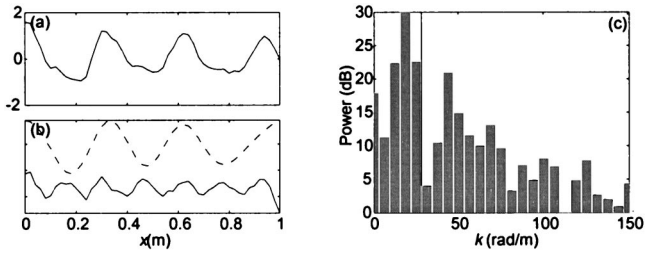


FIG. 7. Typical subsonic waveform found at frequency 250 Hz. (a) is the real part of the complex waveform measured by the probe microphone. (b) shows the wavelets representing the sonic wave (upper) and subsonic wave (lower) of speed 46 m/s. (c) is the spectrum showing the decomposition boundary.

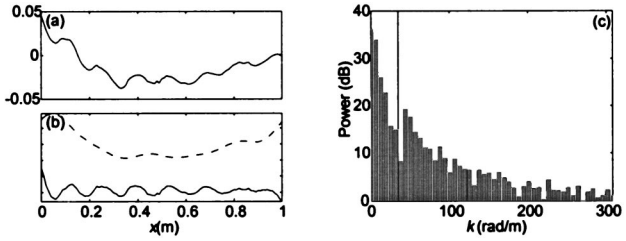


FIG. 8. Typical supersonic wave measured at frequency 2200 Hz. (a) is the measured pressure distribution, and (b) is the decomposed wavelets based on the spectral division shown in (c). The dashed line in (b) has speed 731 m/s.

Supersonic waves are found to be dominant beyond about 2 kHz, but, in this case, the frequency does represent the cut-on for the supersonic waves, although not a cutoff for the sonic waves either.

Figure 6(b) compares the prediction of subsonic waves with the experimental data. The quantitative agreement between theory and experiment is satisfactory. Note that subsonic waves can only be detected up to 250 Hz for the pressure on the rigid wall (data points indicated by △). This is consistent with earlier theoretical observation that pressure oscillation decays exponentially across the channel.

Figure 6(c) compares the theory with experimental data beyond the critical frequency. The predicted critical frequency is $f_c = 1617$ Hz. The data shows a maximum of 1348 m/s at 1950 Hz for the membrane with a side-gap allowance (□). In general, there is good agreement between theory and data, but the critical frequency is predicted under by about 333 Hz. It is suspected that this discrepancy could be caused by the gap between the membrane and the side walls. This conjuncture was tested by sealing up the gap with minimal extra membrane material without tensile force. The results are presented as the triangular (△) points. The maximum wave speed is 1120 m/s and it occurs at a frequency of 1465 Hz, which is 152 Hz lower than f_c predicted by the two-dimensional theory. Such results may be understood from the point of view of effective mass of the membrane vibration. When there is a gap, the pressure difference between the upper and lower surfaces of the membrane vanishes around the edges. As a result, the edges do not vibrate as much as elsewhere and the membrane as a whole appears lighter. This leads to a smaller equivalent M and a higher f_c (see Fig. 2). When the gap is sealed, the pressure difference across the width may be very uniform, but the restraint at the membrane edges makes the structure effectively heavier than it really is. Note that the maximum speed of sound that can be obtained with confidence in the signal processing is limited by the membrane length. For example, at the observed top wave speed in Fig. 6(c), the wavelength is nearly 0.7 m. However, wave speeds higher than 2000 m/s were deduced from the constant phase angle slope of a single traveling wave observed when the lower rigid wall was removed, but meaningful analysis of such configuration can only be carried out with a three-dimensional theory for the breakout waves.

Typical subsonic and supersonic waveforms are given in Figs. 7 and 8 using the method described earlier in Fig. 5.

The spectra are divided into two blocks and the decomposed ''wavelets'' correspond to sonic waves and those radiated by the membrane response. These examples differ from the earlier ones in that the sonic waves do not dominate the overall waveform. The spectral division for the two waves is not as clear-cut as in Fig. 5 since the two wave numbers are closer together and the wave number resolution is limited by the total length of the membrane section. Figure 7 shows that at 250 Hz, subsonic waves undulate the long wave, which is mainly made up of sonic waves plus the effects of scattering by the edges. The subsonic wave speed is found to be 46 m/s and the amplitude clearly decays downstream. However, the attenuation rate is not analyzed due to the difficulties of (a) the measurement of damping coefficient for the thin membrane under tension, and (b) the calculation of exponential decay rate at the presence of edge scattering and significant wave reflection. A typical supersonic wave is illustrated in Fig. 8 in which the long wave, the dashed line in Fig. 8(b), has a wave speed of 731 m/s.

The prevailing sound speed over the membrane section is also evident from the measurement of the ratio of sound amplitude to the reference point, which is located 37 cm upstream of the membrane. The wave amplitudes at the membrane leading edge are 0.08 and 1.64, respectively (both calibrated by the microphone responses), for the two examples. For the subsonic case, sound with a low amplitude is a clear indication that the interface (membrane leading edge) is nearly a pressure node like the open end of a duct, consistent with the expectation that the effect of the membrane may be similar to filling the duct with a gas medium with a low sound speed. For the supersonic case, the amplitude ratio is greater than unity and is indicative of solid-wall-style sound reflection. Such reflection could not have been caused by the boundary far downstream since the absorption ending works sufficiently well for frequencies above 2 kHz.

In the intermediate frequency range from 500 Hz to the supersonic cut-on frequency, the prevailing sound speed is the normal value of 340 m/s, which was earlier described as a ''trivial'' solution for the membrane dynamics. This is consistent with the theoretical observation of fast spatial decaying rate experienced by the subsonic waves on the membrane, cf. Fig. 2(c).

IV. CONCLUSIONS

Experimental studies have been carried out to demonstrate the strong coupling between sound and structure when the structure-to-fluid mass ratio is of the order of unity. A 1-m-long thin membrane under tension was used in the experiment and the data shows a close agreement with eigenvalue predictions for wave propagation in an infinite duct with flexible walls. Sound speeds as low as 8.3 m/s and as high as 1350 m/s have been measured. The departure from the normal value of 340 m/s over the membrane section results in a significant amount of wave reflection at the interface of the rigid and flexible sections of the duct. The speed of eigenwave over the membrane may be understood from the viewpoint of the mass-spring system of the fluid-loaded membrane. At low frequencies, the effect of fluid is purely masslike and the result is a membrane heavier than its own weight. The wave speed is slower than the *in vacuo* value, which, in this example is determined by the tensile force applied. At higher frequencies the effect of fluid loading is springlike and the eigenwave is found to be faster than the speed of sound in free space. At a particular frequency the effect of mass and spring balance each other, and singular sound speed is predicted for a lossless system. The frequency at which this singularity occurs depends only on the mass ratio and close agreement with experimental data is obtained. At supersonic speed, waves travel in a way similar to the high-order modes in a rigid duct where the sound waves advance in a zig-zag path with supersonic axial wave speed.

ACKNOWLEDGMENTS

The authors thank Professor F. P. Mechel for his many useful comments on the original paper. The work is part of a project (Grant No. G-S583) funded by The Hong Kong Polytechnic University which also supports the research studentships of Y.S.C. and T.L.C.

Ackermann, U., and Fuchs, H. V. (**1989**). ''Noise reduction in an exhaust stack of a papermill,'' Noise Control Eng. J. **33**(2), 57–60.

Ackermann, U., Fuchs, H. V., and Rambausek, N. (**1988**). ''Sound absorbers of a novel membrane construction,'' Appl. Acoust. **25**, 197–215.

Beranek, L. L., and Ver, I. L. (eds.) (**1992**). *Noise and Vibration Control Engineering* (Wiley, New York).

Cummings, A. (**1994**). ''Attenuation of sound in unlined ducts with flexible walls,'' J. Sound Vib. **174**, 433–450.

Dunlop, J. I. (**1992**). ''Measurement of acoustic attenuation in marine sediments by impedance tube,'' J. Acoust. Soc. Am. **91**, 460–469.

Fahy, F. J. (**1997**). ''Acoustic interaction between structures and fluids,'' in *Encyclopedia of Acoustics*, edited by Malcolm J. Crocker (Wiley, New York), Vol. 1, Chap. 10.

Ford, R. D., and McCormick, M. A. (**1969**). ''Panel sound absorbers,'' J. Sound Vib. **10**, 411–423.

Frommhold, W., Fuchs, H. V., and Sheng, S. (**1994**). ''Acoustic performance of membrane absorbers,'' J. Sound Vib. **170**(5), 621–636.

Gavriely, N., Shee, T. R., Cugell, D. W., and Grotberg, J. B. (**1989**). ''Flutter in flow-limited collapsible tubes: a mechanism for generation of wheezes,'' J. Appl. Phys. **66**, 2251–2261.

Horne, M. P., and Hansen, R. J. (**1982**). ''Sound propagation in a pipe containing a liquid of comparable acoustic impedance,'' J. Acoust. Soc. Am. **71**, 1400–1405.

Huang, L. (**1999**). ''A theoretical study of duct noise control by flexible panels,'' J. Acoust. Soc. Am. **106**, 1801–1809.

Ingard, K. U. (**1994**). *Notes on Sound Absorption Technology* (Noise Control Foundation, New York).

Kececioglu, I., McClurken, M. E., Kamm, R. D., and Shapiro, A. H. (**1981**). ''Steady, supercritical flow in collapsible tubes. Part 1. Experimental observations,'' J. Fluid Mech. **109**, 367–389.

Ko, S. H. (**1994**). ''Sound wave propagation in a two-dimensional flexible duct in the presence of an inviscid flow,'' J. Sound Vib. **175**, 279–287.

Lighthill, M. J. (**1978**). *Waves in Fluids* (Cambridge U.P., Cambridge).

Newland, D. E. (**1993**). ''Harmonic wavelet analysis,'' Proc. R. Soc. London, Ser. A **443**, 203–225.

Shapiro, A. H. (**1977**). ''Steady flow in collapsible tubes,'' Trans. ASME J. Biomech. Eng. **99**, 126–147.

Sound propagation and speech transmission in a branching underground tunnel

Hiroyuki Imaizumi, Sunao Kunimatsu, and Takehiro Isei
Safety Engineering Department, National Institute for Resources and Environment, 16-3 Onogawa, Tsukuba, Ibaraki 305-8569, Japan

(Received 14 November 1997; revised 18 April 2000; accepted 5 May 2000)

The characteristics of sound propagation and speech transmission along a tunnel with a ''T'' intersection were investigated. At receivers within sight of the sound source, low frequencies were mainly attenuated around the intersection than high frequencies. At receivers out of sight of the source, high frequencies were extensively attenuated. The overall pattern of sound attenuation along the different sections of tunnel, which was calculated by the conical beam method, agreed well with the measurements in this study. Numerical calculations of reflected and diffracted waves with minimum transmission paths in a two-dimensional plane showed that reflected waves were the primary contributors to sound fields out of sight of the source. The articulation scores measured at receivers within sight of the source were high, and most of the confusion concerned syllables that could easily be misheard, even if there were a high signal-to-noise ratio. The types of syllable confusions observed at the receivers out of sight of the source appeared to have been caused by the greater deterioration in speech signals along this part of the tunnel, especially at high frequencies. The evaluation by rapid speech transmission indices (RASTI) appeared to be overestimated at the receivers out of sight of the source. Taking into account the early decay times of impulsive sound and the calculation procedures used in RASTI, it is concluded that speech intelligibility may not have been evaluated correctly by RASTI.
[S0001-4966(00)02508-X]

PACS numbers: 43.50.Vt, 43.20.Mv, 43.20.Fn, 43.71.Gv [MRS]

INTRODUCTION

Sound propagation is influenced extensively by not only the environmental conditions in the sound field, but also the spread and shape of the field. Previous studies of sound propagation outdoors have examined the influence of atmospheric fluctuation, the acoustic characteristics of boundaries such as ground surfaces, and the topographic conditions.[1–7] Several prediction models for sound attenuation have been proposed,[2,3,8–10] and a long-range propagation model that takes into account the effects of turbulence in the atmosphere has also been presented.[11] In addition to these studies on sound propagation in a semi-free field, there has also been a large amount of research on sound propagation in enclosed spaces such as industrial workrooms.[12–19] The main aim of these studies was to protect employees from hearing loss at work, and several calculation models for predicting the noise exposure of workers in industrial environments have been developed. Ray-tracing and image-source methods are typical calculation models.

The relationships between speech intelligibility and many acoustic variables, such as speech levels, noise levels, and reverberation conditions, have been investigated in architectural and room acoustics.[20–31] Objective evaluation indices of speech transmission have also been developed, and attempts have been made to determine the most accurate method of evaluating acoustic conditions for speech communication.[32–34] The main purpose of this research has been to improve speech intelligibility in various types of rooms.

We have found only a few published studies[35–39] on sound propagation and speech transmission underground and/or in tunnels. In Japan and other countries, underground space is being increasingly used for public transport, industrial facilities, and others.[40] We have studied the propagation attenuation of sound as well as speech transmission characteristics under some environmental conditions in underground tunnels, with the aim of promoting safety underground.[41–43] In our earlier studies, we examined the characteristics of sound propagation and speech transmission in a straight tunnel with ventilation, and observed that the sound unexpectedly attenuated much more if the propagation were downwind rather than upwind. The hearing scores for speech were lower downwind than upwind. We also examined the characteristics of sound propagation and speech hearing in a curved reverberant tunnel, and showed the relationships between occurrences of syllable confusions and the attenuation of propagated sound. We also found a positive correlation between the hearing scores for several speech stimuli and rapid speech transmission indices (RASTI).

In this study we have considered a simple tunnel system with a ''T'' intersection as the simplest representation of a tunnel network. The characteristics of sound propagation and speech transmission in this underground tunnel have been examined. We have also determined the predominant syllable confusions, and investigated the occurrence of these confusions. We have investigated the relationships between speech hearing scores and methods of evaluating speech intelligibility, such as RASTI.

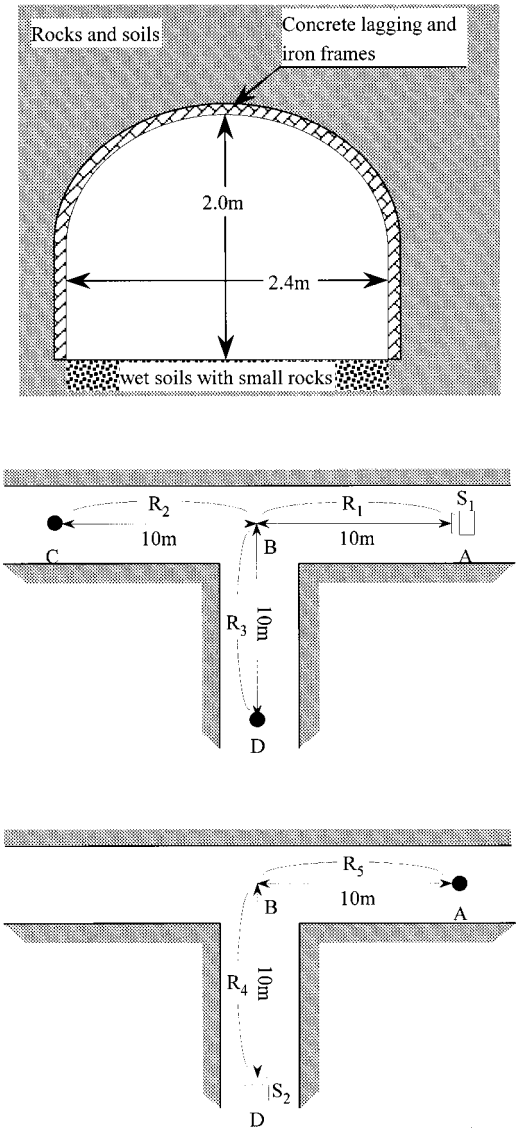


FIG. 1. Schematic plan of the experimental tunnel. Top: cross section of the tunnel. Center and bottom: tunnel plan showing the two source–receiver setups. The tunnels continued on in each direction.

I. EXPERIMENTS

A. Experimental tunnel

Our experiments on sound propagation and speech hearing in a ''T''-shaped underground tunnel were performed at the Coal Mining Safety Center, National Institute for Resources and Environment (NIRE), Kyushu. The cross sections of both branches of the ''T'' had the same arched shape, shown in Fig. 1, with a cross-sectional area of about 4.2 m^2. Both branches were 2.4 m wide and 2.0 m high. The floor consisted of wet soil with many small stones, and a track had been laid on the floor. The rocks and soil around the tunnels were supported by iron frames and lagging made of concrete (Fig. 2). The A-weighted sound pressure level was less than 36 dB during the experiments. There was a natural, ambient airflow of 0.3–0.5 m/s in the tunnels. The temperature remained almost constant at 18 °C, and the relative humidity at 90%, during the experiments.

The average reverberation times were measured at several different points and are shown in Fig. 3. The values were obtained from the slope between −5 and −25 dB of the decay curves of pink noise. Although the reverberation times became longer at low frequencies, the values at frequencies above 315 Hz were constant at about 0.5 s. The peak at 250 Hz was probably related to the standing waves formed in proportion to the cross-sectional area of the tunnels.

FIG. 2. Details of the tunnel lining and floor.

B. Apparatus

Pink noise was used as a sound source for measuring the sound propagation characteristics along the tunnels. The pink noise, from a generator, was emitted from a directional loudspeaker with a single cone, set up on a tripod 1.2 m above the floor. Transmitted pink noise was received by an omnidirectional microphone and recorded on magnetic tapes by a recorder. The microphone was set up on a tripod at the same height as the loudspeaker. Analyses of the recorded pink

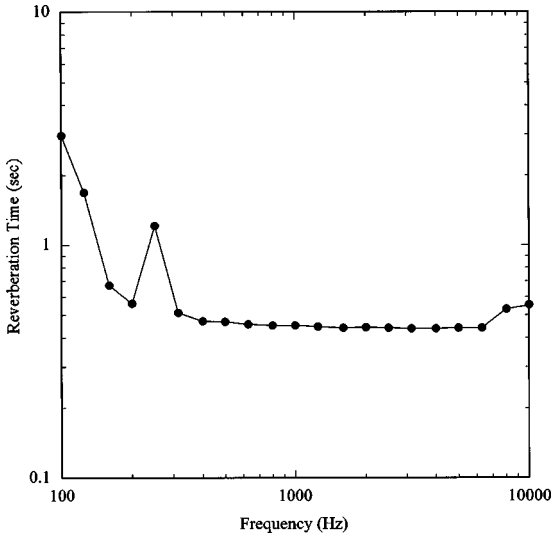


FIG. 3. Reverberation times in one-third octave bands in the tunnels.

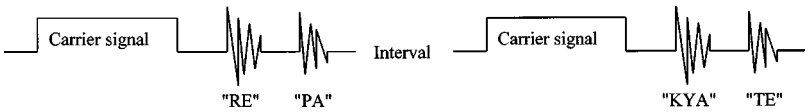


FIG. 4. Order in which the speech stimuli were presented. There was an interval of about 5 s between each pair of syllables for the subjects' responses.

noise were made in one-third octave bands by a real-time frequency analyzer. The sound pressure levels of each band were calculated by linear averaging over 2 min.

We also made RASTI measurements along the tunnels. The signal transmitter and receiver were set up in the same way as the sound source and the microphone, respectively. We used a measurement period of 32 s to minimize random errors during the experiments.

C. Locations of sound sources and receivers

Two sound-propagation and speech hearing setups were used. The sound source located on the cross arm of the ''T'' (at point A in the middle diagram, Fig. 1) was called S_1. The receivers for S_1 were located along lines *AB*, *BC*, and *BD* in the diagram, representing R_1 (up to 10 m from the source, within line of sight), R_2 (10–20 m from the source, within line of sight but after the tunnel intersection), and R_3 (10–20 m from the source and out of line of sight), respectively. The sound source located on the other arm of the ''T'' (at point *D* in the bottom diagram of Fig. 1) was called S_2. The receivers for S_2 were placed along lines *DB* and *BA* in the diagram, representing R_4 (up to 10 m from the source within line of sight) and R_5 (10–20 m from source and out of line of sight along the cross arm of the ''T''), respectively.

D. Subjects

Eight Japanese objects, three males and five females, participated in the speech hearing experiments. Their ages ranged from 22 to 29 years. All the subjects reported that they had no hearing problems. From audiometric measurements of their hearing threshold levels we determined that all the subjects had normal hearing.

E. Speech stimuli and hearing experiments in the laboratory

The articulation syllables defined by the Acoustical Society of Japan[44] were used as speech stimuli. The syllable list consisted of 100 Japanese nonsense syllables. Before the experiments, the syllables were pronounced in groups of two, with a short interval between each group, by a male professional announcer in an anechoic room. They were recorded on a magnetic tape with the same recorder used for the sound propagation experiments. The announcer was instructed to pronounce each syllable clearly at the same sound pressure level. The speech stimulus was played back through the loudspeaker, which was placed at S_1, and recorded through the microphone placed at the end of each of two tunnel paths (R_2 and R_3). Therefore, in each case the source and the receiver were 20 m apart in the tunnels. The sound pressure level of a reference signal at 1 kHz, recorded on the test tape with the articulation syllables, was kept at about 100 dB, at 1 m in front of the loudspeaker during the experiments.

The hearing experiments were carried out in a semi-anechoic room in a laboratory. The room was fully isolated from outside noises. The syllables, analog recorded under each source–receiver combination, were converted to digital form (A–D; 32.3-kHz sampling rate) one by one through the real-time frequency analyzer in the time domain; the digitized syllables were sent to a PC. The PC controlled the procedures used in the hearing experiments, and the syllables were randomly rearranged to prevent the subjects learning the presentation order of the syllables. Arbitrary pairs of two syllables, followed by a carrier signal of pure tone (signal duration 0.5 s), were continuously presented to both ears of the subjects through a stereo amplifier and a headphone as indicated in Fig. 4.

As a familiarization practice, groups of syllables were played to the subjects so that they could learn the procedures used in the hearing experiments before the measurements were made. The subjects were instructed to listen to and then write down each pair of syllables they had just heard. Each subject was examined twice for each group of speech stimuli recorded at the end of each of the two tunnel paths. The presentation levels of the speech stimuli were adjusted on the basis of the reference signal at 1 kHz.

II. RESULTS

A. Propagation attenuation of pink noise

Sound pressure levels were measured at receivers along R_1 and R_2 relative to the level measured at 1 m in front of the sound source S_1. The measurements are shown in Fig. 5. At receivers along R_1, sounds at frequencies above 500 Hz

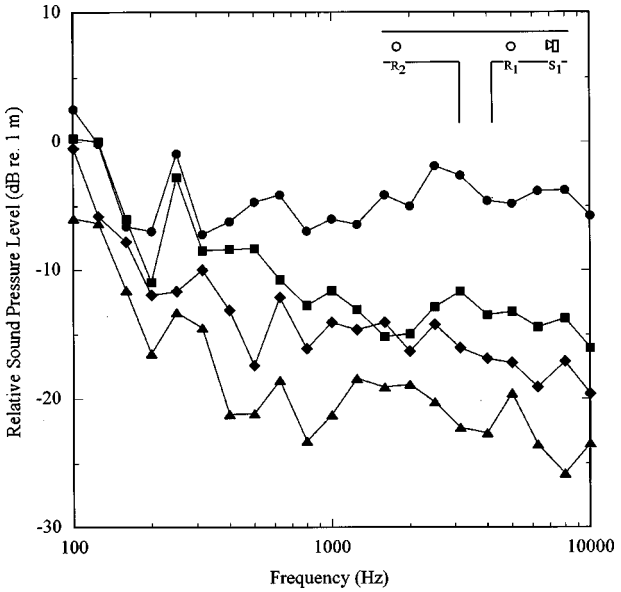


FIG. 5. Changes in sound pressure levels (relative to that measured at 1 m in front of the transmitter) with increasing propagation distance along R_1 and R_2. Filled circles, squares, diamonds and triangles show the relative levels at 2 (●), 8 (■), 12 (♦), and 20 m (▲) from S_1, respectively.

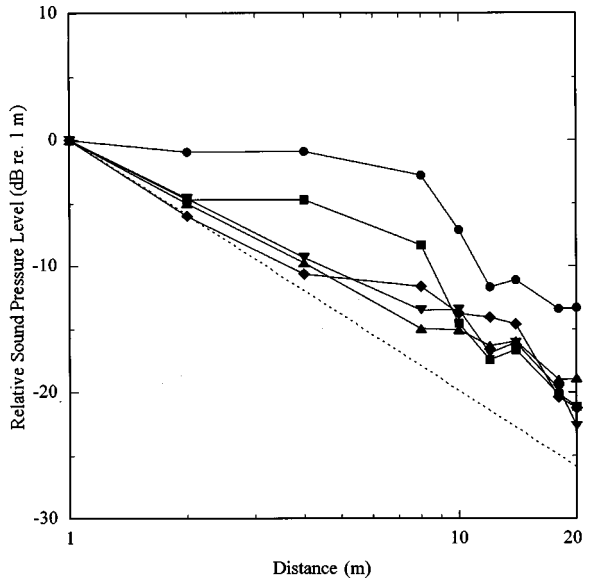


FIG. 6. Sound attenuation in distance of several one-third octave bands along R_1 and R_2. Filled circles, squares, diamonds, triangles, and reversed triangles show the relative values of 250 (●), 500 (■), 1000 (♦), 2000 (▲), and 4000 Hz (▼), respectively. The dotted line shows a decay of 6 dB per doubling of distance.

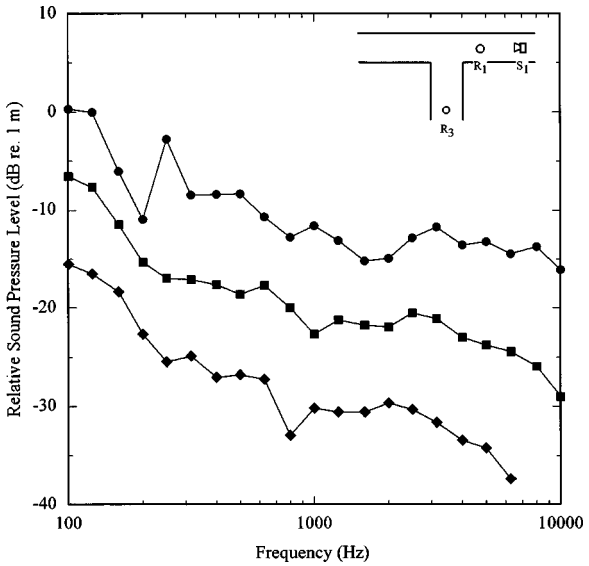


FIG. 7. Relative sound pressure levels with increasing propagation distance along R_1 and R_3. Filled circles, squares, and diamonds show the relative levels at 8 (●), 12 (■), and 20 m (♦) from S_1, respectively.

are mainly attenuated with increasing propagation distance. We observed small changes of attenuation at frequencies below 500 Hz at these receivers. The small attenuation at about 250 Hz is probably related to standing waves formed in proportion to the cross-sectional area of the tunnels. As the sound passed through the intersection and was measured at receivers along R_2, the frequency components from 250 to 500 Hz were attenuated more. Attenuation of other frequency components tended to increase with the propagation distance, and the relative sound pressure levels were approximately 20 dB or more at 20 m from the source. In Fig. 6, we plot curves of sound attenuation in distance for several one-third octave bands along R_1-R_2. Rapid attenuation is observed at 250 and 500 Hz at the receivers close to the intersection. High frequency bands are attenuated by slightly less than 6 dB per doubling of distance, with smaller influences from the intersection. Figure 7 shows the relative sound pressure levels at the receivers along R_3, which were out of sight of the sound source S_1. Here the changes in attenuation are relatively small at the lower frequencies, even when the propagation distance increased. The higher frequency components were uniformly decreased with increasing distance, and the values of relative sound pressure level were more than 30 dB at the most distant receiver. When we compared the relative sound pressure levels at two receivers along R_2 and R_3 equidistant from the sound source S_1, we found that the differences in attenuation at higher frequencies tended gradually to increase toward the distant measuring points, and the attenuation at the receivers along R_3 was larger.

The relative sound pressure levels measured for the sound source S_2 are plotted in Fig. 8. The attenuation at the receivers along R_4 is small, and the relative sound pressure level is less than 15 dB, even 8 m from S_2. The frequency components from 200 to 250 Hz were attenuated remarkably at the receivers along R_5. As the receivers became more distant from S_2, the attenuation of the high frequency components mainly increased, and the values of relative sound pressure level were 20 dB or more at the farthest receiver.

B. RASTI measurement

RASTI values were measured under both source–receiver combinations and are shown in Fig. 9. At receivers located along R_1, the RASTI values decreased up to the tunnel intersection with increasing distance from S_1. The RASTI values tended to increase slightly at the receivers along R_2. The RASTI values observed along R_3 were lower than those along R_2, and decreased with distance. For the S_2 sound source the distribution of RASTI values did not change remarkably from the values for the S_1. The RASTI values measured along both R_1 and R_4 were similar, al-

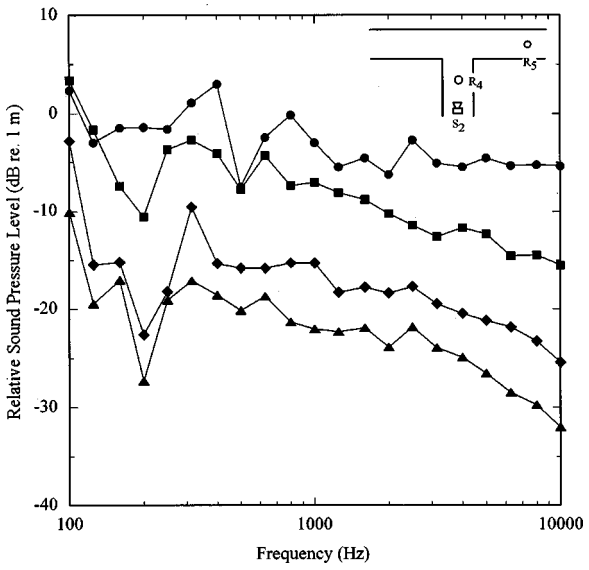


FIG. 8. Relative sound pressure levels with increasing propagation distance along R_4 and R_5. Filled circles, squares, diamonds, and triangles indicate relative levels at 2 (●), 8 (■), 12 (♦), and 15 m (▲) from S_2, respectively.

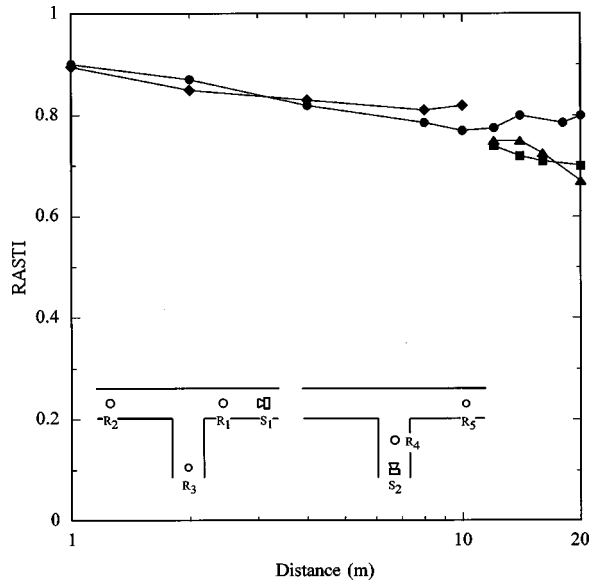


FIG. 9. Distribution of RASTI values along the tunnels. Filled circles and squares show RASTI values along R_1-R_2 (●) and R_3 (■), respectively. Filled diamonds and triangles show the values along R_4 (♦) and R_5 (▲).

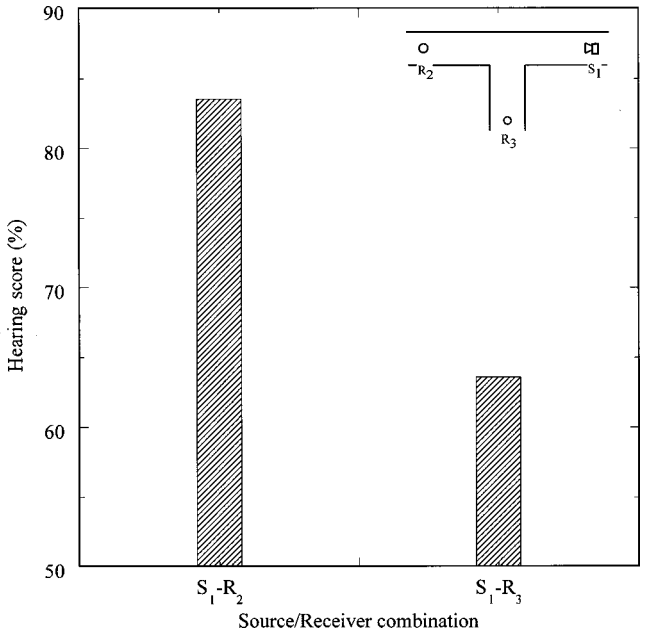


FIG. 10. Average correct hearing scores from recordings taken at the end of R_2 and the end of R_3.

though the values measured along R_4 were slightly higher than those along R_1 in the area of the intersection. The RASTI values measured at the receivers along R_3 and R_5 decreased in a similar way with distance.

All the RASTI values in the measurements were classified as ''excellent'' or ''good'' by the subjective evaluation scale for speech transmission.[34] The speech transmission measured at the farthest receivers for each sound source was classified as ''excellent,'' ''good,'' and ''good'' for R_2, R_3 and R_5, respectively.

C. Hearing experiments with speech stimuli

The average hearing scores for the articulation syllables recorded at the end of two tunnel paths are plotted in Fig. 10. The hearing score for recordings made at the end of R_2 was about 20% higher than for those made at the end of R_3.

Confusion matrices were constructed from the results of articulation tests for each tunnel path (Fig. 11). The radius of each circle represents the number of hearing mistakes on stimulus response. Overall, the patterns of the confusion matrices could be summarized as follows:

(1) Confusion occurred mostly in the consonant part of the syllables.
(2) Syllables with the phonemes /z/, /zy/, /py/, /w/, and /y/ were rarely confused with others.
(3) Obvious patterns of syllable confusion were observed.

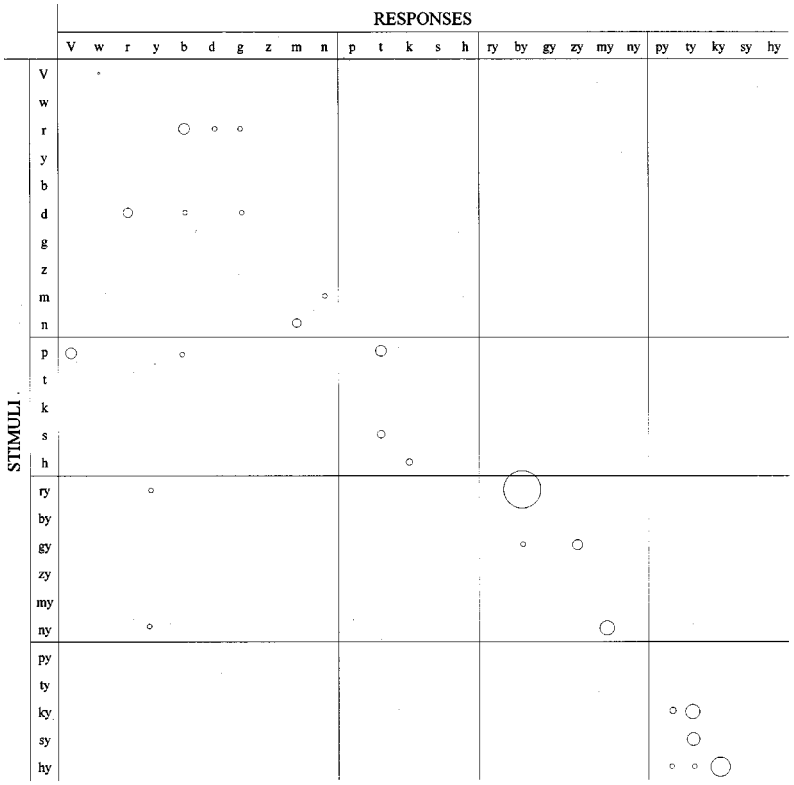
Most of the confusion for recordings made at the end of R_2 concerned voiced or voiceless consonants. A few confusions occurred from voiceless consonants to vowels, for example, from /pa/ to /a/. Most of the syllables with the phoneme /ry/ were mistaken for those with the phoneme /by/. Syllables with the phoneme /hy/ were frequently confused with those with the phoneme /ky/. The phoneme /ky/ was mainly confused with /ty/. Syllables with /ny/ were often confused with /my/. We found predominant mistakes from /pe/ to /te/ and /ri/ to /bi/. Confusion of /hi/ to /ki/, /ne/ to /me/, /si/ to /ti/, as well as /su/ to /tu/ was often observed along this tunnel path. Confusion between /mi/ and /ni/ occurred both ways.

Although the confusions for recordings made at the end of R_3 were generally similar to those for recordings made at the end of R_2, we observed additional patterns of confusion, such as from contracted sounds to syllables with the phoneme /y/, and from vowels to consonants. Confusions of the phonemes /h/, /my/, and /s/ remarkably increased. Contracted sounds with the phoneme /gy/ were misheard as contracted sounds with /by/, /py/, or sometimes as consonants with /y/ due to dropout of the phoneme /g/. Syllables with /ry/, /ny/, and /hy/ were identified partly in the same manner as /gy/. The vowels /i/, /u/, and /o/ were confused with consonants to some extent. Other notable mistakes were from syllables with /h/ to /p/, and /h/ to /k/ both ways. We observed several confusions of the voiced consonants with the contracted sounds recorded at the end of R_3. Voiceless consonants were sometimes confused with voiced consonants.

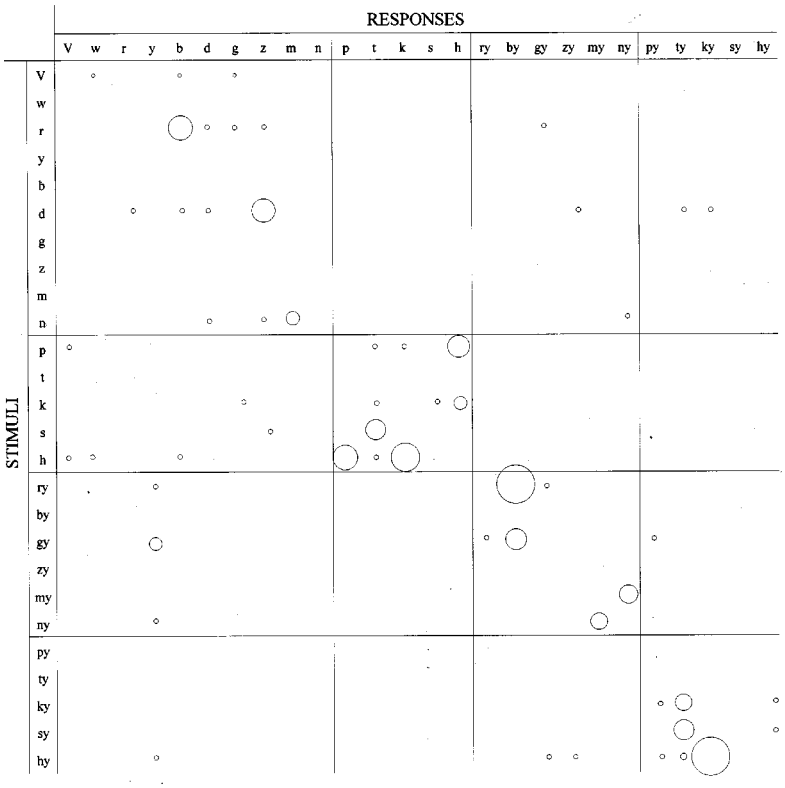
III. DISCUSSION

A. Sound propagation along the ''T''-shaped tunnel

In order to predict sound propagation in enclosed rooms such as industrial workrooms, ray tracing and an image-source method have been applied in many studies. Hodgson[18] reported that the ray-tracing method proposed by Ondet and Barbry[16] was very accurate as well as flexible for the prediction of sound propagation in enclosed spaces. Dance[14] developed a complete image-source method (CISM) based on the ray tracing by Ondet and Barbry, and showed that CISM was as accurate as the ray-tracing method. In our study, we used a conical beam method,[45] which combined the advantages of the ray-tracing and image-source methods,



(a) S_1-R_2



(b) S_1-R_3

FIG. 11. Confusion matrices for articulation syllables recorded at the end of R_2 and the end of R_3. The radius of each circle shows the number of the hearing mistakes.

to predict sound propagation along the ''T''-shaped tunnel. With this method, a large number of cones with vertices at the sound source were emitted, and the algorithms of the ray-tracing method in the axes of the cones were applied to the propagation of cones in the sound fields. The acoustic characteristics of the directional loudspeaker used in the field measurements were applied to the calculation. The average absorption coefficients of the tunnels were obtained by a

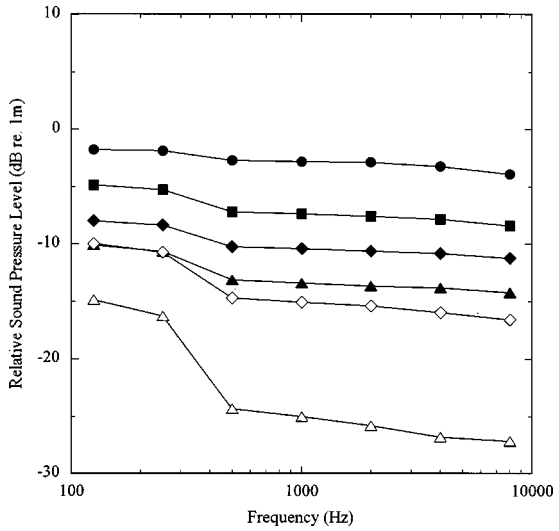


FIG. 12. Sound pressure levels along R_1-R_2 and R_1-R_3 relative to the level at 1 m in front of the sound source, which were calculated by the conical beam method. Filled circles, squares, diamonds, and triangles show the relative levels at 2 (●), 8 (■), 12 (♦), and 20 m (▲) from S_1, respectively. Open diamonds and triangles represent the relative levels at 12 and 20 m along R_3.

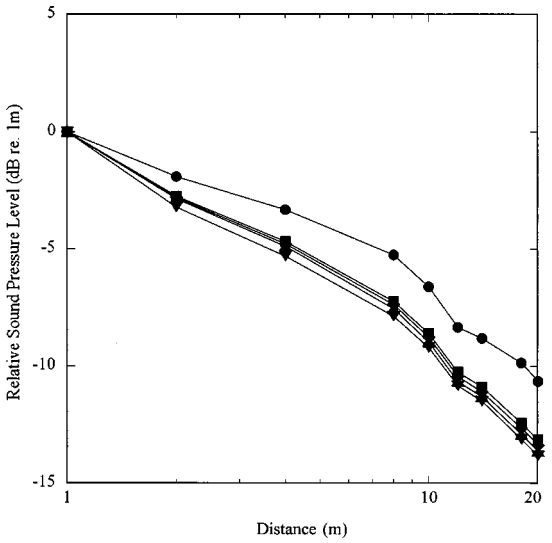


FIG. 13. Sound attenuation with distance of several one-octave bands obtained from numerical calculations along R_1 and R_2. Filled circles, squares, diamonds, triangles, and reversed triangles show the relative values of 250 (●), 500 (■), 1000 (♦), 2000 (▲), and 4000 Hz (▼), respectively. The dotted line shows a decay of 6 dB per doubling of distance.

method that used the measured reverberation times at one-third octave bands.[46] The number of rays and reflections were 5000 and 50, respectively, in our calculations.

We calculated the sound pressure levels relative to that at 1 m in front of the sound source for the R_1-R_2 and R_1-R_3 paths. These are shown in Fig. 12. At the receivers along R_1 and R_2, which were in sight of the sound source S_1, the sound pressure levels decreased with increasing propagation distance, and the attenuation at higher frequencies was larger than that at lower frequencies. At the receivers along R_3, which were out of sight of S_1, the attenuation was much larger than along R_1-R_2, and the relative sound pressure level at the most distant receiver was about 25 dB in the high frequency range. The sound attenuation with distance of 250, 500, 1000, 2000, and 4000 Hz frequencies in octave bands along R_1-R_2 is plotted in Fig. 13. The lower frequency components are attenuated more near the intersection of the tunnels, and the influence of sound attenuation at the higher frequencies is smaller than at the lower frequencies.

Our calculated values are generally lower than our direct measurements. Although the differences between both tend to increase at receivers distant from the sound source, the overall pattern of attenuation agrees well. Discrepancies between the calculations and the measurements may have led to errors in estimating the average absorption coefficients in the tunnels. The floor of the tunnels was relatively absorptive, although the inner walls were reflective (see Fig. 2). In addition, incident sound waves on the inner walls could have been reflected in a complex way, since the frames made the surfaces of the inner walls very rough. Therefore, the ostensible absorption characteristics became larger and may have led to errors between the calculations and the measurements.

The sound fields along R_3, which was out of sight of the source, consist of reflected and diffracted sounds, while the fields along R_1 and R_2 are formed by direct and reflected sounds. Numerical calculations were performed to examine the sound field at R_3 in a two-dimensional plane involving a sound source and a receiver (Fig. 14, top). A number of image sources, which were located around the real sound

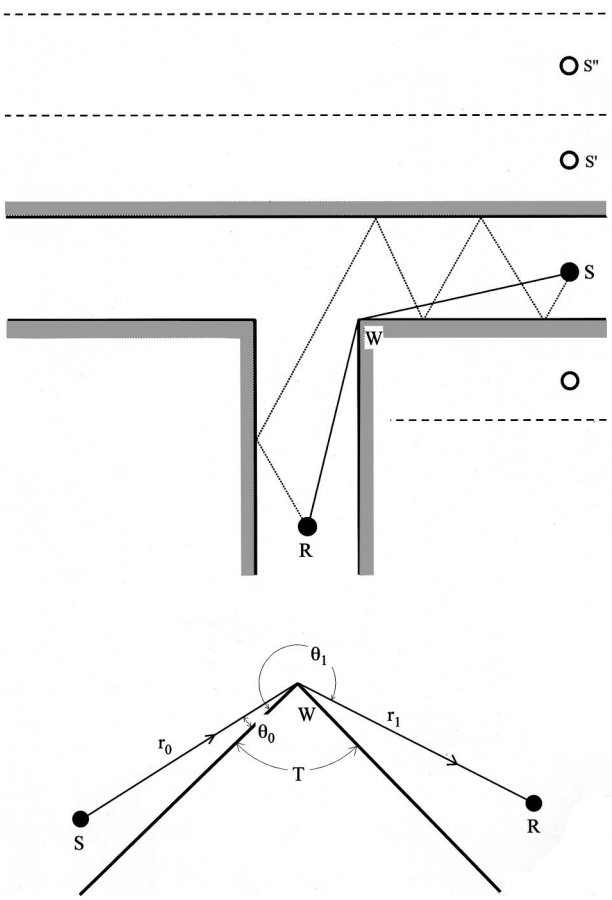


FIG. 14. Top: schematic two-dimensional plan used for numerical calculations, showing locations of a sound source *S*, image sources, and a receiver *R* in the tunnels. Bottom: incident and diffracted rays on a wedge *W*.

source, contributed to the sound field, and propagation paths with minimum reflection as well as diffraction on a wedge, *W*, were individually considered in this calculation. Therefore, we examined sound attenuation of the wave, which reflected only on the inner walls of the two-dimensional tunnel (for example, a dotted line, Fig. 14, top).

The pressure amplitude Φ at a receiver in the tunnel can be described by a direct wave Φ_d and a sum of the reflected waves. Letting N be the number of image sources, the pressure amplitude is given by the following equation:[38]

$$\Phi=\Phi_d+\sum_{j=1}^{N}\Phi_j, \quad \Phi_j=\Theta_{cj}\cdot\frac{e^{ikd_j}}{d_j}, \tag{1}$$

where Θ_{cj} is a jth image source strength,[47] d_j is the path length between the jth image source and the receiver, and k is the wave number. Here, the inner walls of the tunnels are assumed to be locally reacting.

Exact solutions for a diffraction problem on a wedge with an arbitrary opening angle have been given by Macdonald and others.[48,49] Kawai[50] numerically compared the approximation equation proposed by Kouyoumjian and Pathak[51] with the exact solutions for diffraction on a wedge, and found that the approximation model was sufficiently accurate and useful for predicting the diffraction fields. We applied the approximation solution to our study to calculate attenuation due to diffraction on the wedge. The geometry of a sound source and a receiver can be described diagrammatically (Fig. 14, bottom), and the pressure amplitude at R is described by the following equations:

$$\Phi=\frac{e^{ikR_1}}{kR_1}\left[V\left(\frac{r_0r_1}{R_1},1,\theta_1-\theta_0\right)+V\left(\frac{r_0r_1}{R_1},1,\theta_1+\theta_0\right)\right]. \tag{2}$$

Setting $A=r_0r_1/R_1$, we obtain

$$V^{\pm}(A,1,\theta)=-\frac{e^{i(\pi/4)}}{\sqrt{2\pi kA}}\cdot\frac{1}{2\nu}\cot\left(\frac{\pi\pm\theta}{2\nu}\right)\cdot F'(X^{\pm}(\theta)), \tag{3}$$

$$X^{\pm}(\theta)=2kA\cos^2\left(\frac{2N^{\pm}\nu\pi-\theta}{2}\right), \tag{4}$$

and

$$F'(x)=-2i\sqrt{x}e^{-ix}\int_{\sqrt{x}}^{\infty}e^{i\tau^2}d\tau=-2i\sqrt{x}e^{-ix}F(\sqrt{x}), \tag{5}$$

where

$$\theta=\theta_1\mp\theta_0, \tag{6}$$

$$\nu=2-\frac{T}{\pi}, \tag{7}$$

$$R_1=r_0+r_1, \tag{8}$$

and $F(x)$ is a Fresnel integral. N^+ and N^- are determined by following procedures:

$$N^+=\begin{cases}0 & \text{for } \theta\leqslant\pi-T,\\ 1, & \text{for } \theta>\pi-T,\end{cases} \tag{9}$$

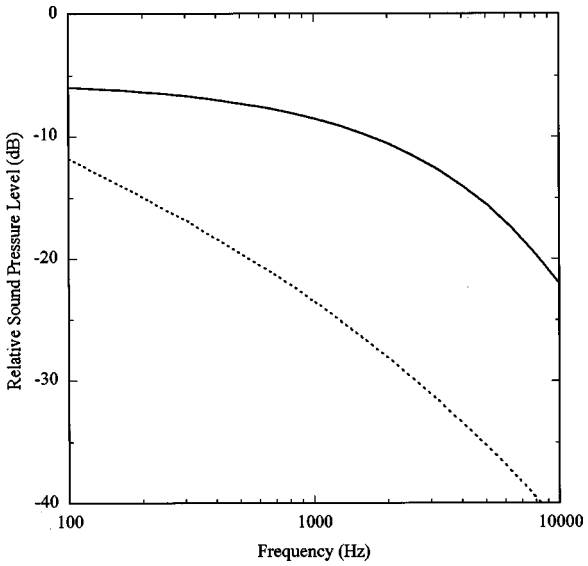


FIG. 15. Comparison of the sound pressure levels relative to the free field by reflected and diffracted waves with minimum transmission paths from *S* to *R* in Fig. 14. The solid and dotted lines represent the relative levels for reflected and diffracted waves, respectively.

$$N^-=\begin{cases}-1, & \text{for } \theta<T-\pi,\\ 0, & \text{for } T-\pi\leqslant\theta\leqslant 3\pi-T,\\ 1, & \text{for } \theta>3\pi-T.\end{cases} \tag{10}$$

We compare the excess attenuation of the reflected and diffracted waves relative to the free-field attenuation at the farthest receiver from the viewpoint of the individual wave path in Fig. 15. In the calculation, the acoustic impedance of the inner walls was calculated by the Wilson model using a flow resistivity of 3000 kPa s/m^2.[52] We estimated the flow resistivity from observation of the surfaces in the experimental tunnels and the classification of the ground surfaces by the flow resistivity.[5] Excess attenuation of the reflected and diffracted waves increases at higher frequencies, and the attenuation of the reflected waves is smaller than that of the diffracted waves. Although the actual propagation paths are a combination of both waves, it is possible to conclude that the reflected waves were the primary contributors to the sound field in those regions that were out of sight of the source.

B. Analyses of syllable confusions in the articulation tests

Sakata *et al.*[29] and Haruta[30,31] reported confusions of Japanese syllables under several noisy and reverberant conditions, and evaluated the transmission characteristics of speech by precisely analyzing the distribution of misheard syllables in articulation tests. Iida[53] also performed articulation tests with Japanese syllables in several rooms with various acoustic characteristics, and demonstrated some general patterns of syllable confusion, as well as the influence of different talkers. Saito[54] precisely analyzed confusion matrices obtained from the articulation tests, and investigated the phonetic characteristics of each Japanese syllable as well as the error patterns. The typical patterns of consonant confusion caused by distortion in time and frequency domain were

also described. Here, we compare the syllable confusions in our articulation tests with the results of these previous studies.

In the previous studies,[31,53,54] syllable confusions generally occurred in the consonant parts of the syllables, and the vowel parts of the syllables were rarely confused. Syllables with the phonemes /w/ and /y/ were classified as those that were not confused, and syllables with the phonemes /z/ and /zy/ were also identified very correctly under a high signal-to-noise ratio.[31,53] All these syllables retained intelligibility under our experimental conditions.

Most syllable confusions observed along the R_1-R_2 path occurred with syllables that had lower intelligibility to begin with, as described in the previous studies.[31] Syllables with the phonemes /ky/ were often confused with /py/ and /hy/, and confusion with /ty/ was also reported.[53] Confusion of the phonemes /ky/, /p/, /r/, and /gy/ was observed frequently with a decreasing signal-to-noise ratio in the high frequency range.[31] The confusion of /pa/ with /a/ suggested the influence of reverberation rather than a low signal-to-noise ratio as compared with the previous measurements.[31,53] The confusions we observed generally occurred among syllables with similar phonetic characteristics. Consequently, a slight deterioration of the speech signals tended to cause these syllable confusions observed along the R_1-R_2 path.

Confusions of syllables with /y/ by dropping out of the consonant part of contracted sounds, which was frequently observed along the R_1-R_3 path, mostly have occurred at a low signal-to-noise ratio in broadband frequencies.[31] Syllables with the phonemes /h/ have been confused with voiceless consonants because of deterioration of the consonant part.[54] The confusion we observed between the phonemes /h/ and /p/, as well as /h/ and /k/ both ways, agreed well with that found in the previous studies.[54] Changes in the acoustic characteristics of the consonant part may cause confusion of voiced consonants with contracted sounds when the phoneme /y/ is added.[31] Voiceless consonants have been misheard as voiced consonants when there was a large decline in the quality of the speech signals.[54] We observed this confusion pattern to some extent at the end of R_3. The increased confusion between the phonemes /my/ and /s/ also suggested that there was a decreased signal-to-noise ratio at high frequencies as reported in the published study.[31] Although vowels were rarely misidentified, the intelligibility of the vowels have been influenced by a frequency band at about 500 Hz.[31] Confusion of vowels with consonants could be generated by a change in the acoustic characteristics induced by decreased intelligibility. Most of the errors described above suggested that further deterioration of the speech signals during transmission along the tunnels, especially at high frequencies, caused these syllable confusions.

C. Relationships between speech hearing scores and several methods of evaluating speech transmission

A 75% articulation score generally corresponds to 100% intelligibility.[55,56] From these figures it can be seen that the 85% articulation score along the R_1-R_2 path was satisfactory, but the 65% along the R_1-R_3 path was relatively poor.

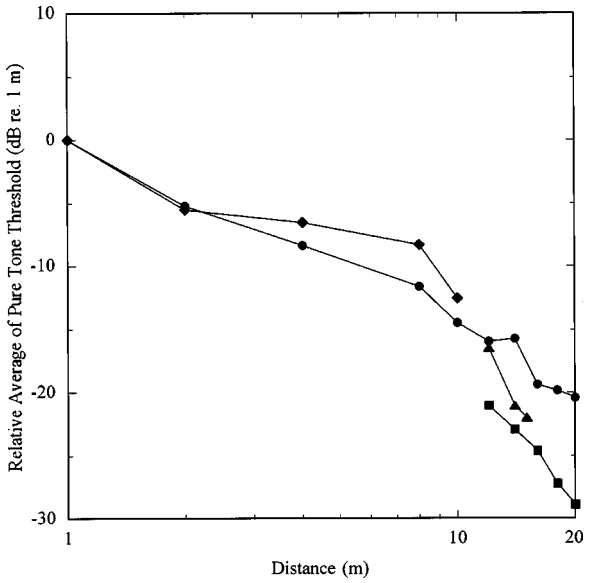


FIG. 16. Distribution of the average pure tone thresholds along the tunnels. Filled circles and squares show RASTI values along R_1-R_2 (●) and R_3 (■), respectively. Filled diamonds and triangles show the values along R_4 (◆) and R_5 (▲).

As an alternative method for evaluating speech transmission, we introduced the average pure tone threshold,[57] calculated from the results of propagation attenuation of pink noise along the tunnels. We calculated the average pure tone threshold from the arithmetic average of the sound pressure levels at frequencies of 500, 1000, and 2000 Hz. The average pure tone threshold along R_1 gradually drops with distance up to the intersection; see Fig. 16. Moreover, the average pure tone thresholds decrease with a slightly steeper gradient toward the most distant receiver along R_1-R_2. Along R_1-R_3 the average pure tone thresholds decrease more rapidly with increasing distance and drop to close to 30 dB. The patterns for S_2 were almost the same. We can surmise that speech intelligibility along R_1-R_3 was probably poor, since an average pure tone threshold of more than 30 dB is generally regarded as representing auditory disorder.

The RASTI measurements overestimated the speech intelligibility along R_1-R_3. The RASTI values may mainly reflect the reverberation characteristics of the experimental tunnel, since the background noise level was low. In calculating the RASTI values, echograms in the rooms have been assumed to be purely exponential decay, and the effect of reverberation on speech transmission has been interpreted as an equivalent signal-to-noise ratio.[32,34] The measured sound pressure–time waveforms of an impulsive sound at the most distant receivers along R_2 and R_3 are shown in Fig. 17. The sound pressure measured at the end of R_2 exhibited exponential decay, although the pressure at the end of R_3 did not attenuate exponentially. This was because the sound did not propagate directly, and reflected and diffracted waves contributed to the sound fields. Therefore, the measurement of RASTI along R_3 may not have been useful for correctly estimating speech intelligibility. The early decay times[58] calculated from the waveforms ranged from 100 to 130 ms, and there were not significant differences between the decay times. The modulation transfer functions,[32–34] which were

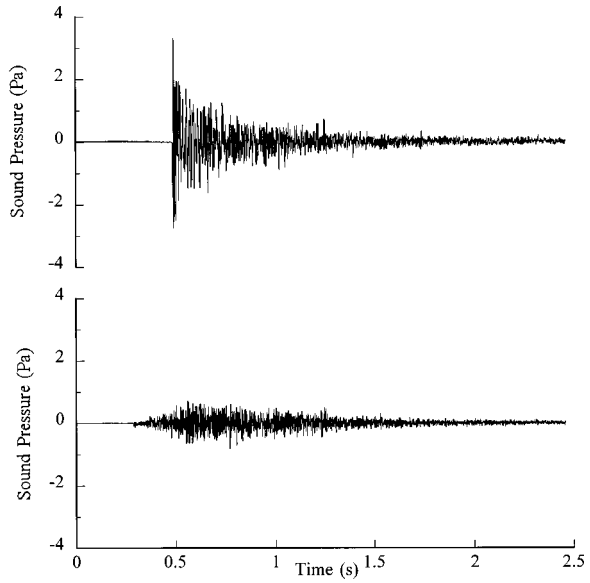


FIG. 17. Impulsive sounds measured at the receiver most distant from the source S_1 along R_2 (top) and R_3 (bottom).

closely related to the early decay times, were almost the same, so the calculated RASTI values may have been similar to each other. In general, speech intelligibility is significantly influenced by the initial decay of sound in rooms,[59] and intelligibility becomes lower as the decay becomes smaller. It may be difficult to accept that a 20% difference in hearing scores can be attributable to a slight difference in early decay times. Consequently, the difference in articulation scores may have resulted from a relatively low signal-to-noise ratio on the R_1-R_3 path, since the sound pressure levels of the speech signals transmitted to the receivers along R_3 were obviously smaller than those along R_2. STI (Speech Transmission Index) and RASTI are objective methods for evaluating speech transmission, and they tend to be widely used because of their practicality. However, some problems with the measurement principles for the Japanese language have been pointed out.[60,61] We need to consider the applicable ranges of RASTI measurement carefully.

IV. CONCLUSION

For sound propagated through a ''T'' intersection to receivers within sight of the sound source, frequency components from 250 to 500 Hz are considerably attenuated. At high frequencies the attenuation is unremarkable, and these frequency components gradually decrease by about 6 dB per doubling of the distance from the source. Greater attenuation occurs at high frequencies at distant receivers that are out of sight of the source. We found that the overall pattern of propagation attenuation of sound along the tunnels, as calculated by the conical beam method, agreed well with our measurements, although there was a quantitative discrepancy. From numerical calculations involving individual reflected and diffracted waves with minimum transmission paths in a two-dimensional plane, we found that reflected waves were the primary contributors to sound fields that were out of sight of the source.

Our hearing experiments, which used articulation syllables transmitted along the tunnels, showed that the hearing scores at receivers in sight of the sound source were higher than those at receivers that were out of sight. From the analyses of confusion matrices, we found that several syllables were frequently confused, both in and out of the line of sight. Most of the hearing mistakes observed at receivers within sight of the source occurred with syllables that had low intelligibility even under a high signal-to-noise ratio. We suggest that deterioration of the speech signals during transmission along the tunnels caused the syllable confusions at the receivers in sight of the source. There was a markedly increased confusion of syllables with the phonemes /h/, /my/, and /s/ at receivers out of sight of the source. Confusion about vowels, contracted sounds, and voiced consonants was also observed more frequently at receivers out of sight of the source. These syllable confusions may be caused by a greater deterioration in the speech signal at receivers along this path, especially at higher frequencies.

To some degree, the results of our articulation tests were consistent with the results of average pure tone thresholds, obtained from the results of propagation experiments of pink noise. The results of the empirical relationship between articulation score and intelligibility evaluation of transmitted syllables also agree with those of the articulation tests. However, the objective evaluation of RASTI overestimated speech intelligibility at receivers out of sight of the source. The difference between the early decay times calculated from impulsive sound measured at receivers within and out of sight of the source was unremarkable. RASTI measurement at receivers out of sight of the source was difficult, since the sound field consisted of only reflected and diffracted sounds, and the impulsive sound may not have been attenuated exponentially. It appears that the sound pressure levels of the transmitted speech signals may affect the articulation scores at receivers in both locations.

ACKNOWLEDGMENTS

We thank Dr. Yoshinori Toida for his discussion on the distribution of RASTI values in the tunnels, and the staff at the Kyushu branch of NIRE, who adjusted the experimental conditions for us.

[1] T. F. W. Embleton, ''Tutorial on sound propagation outdoors,'' J. Acoust. Soc. Am. **100**, 31–48 (1996).
[2] K. B. Rasmussen, ''Outdoor sound propagation under the influence of wind and temperature gradients,'' J. Sound Vib. **104**, 321–335 (1986).
[3] K. B. Rasmussen, ''Sound propagation over level terrain,'' Danish Acoustical Institute Report **33**, 1–147 (1982).
[4] T. Isei, M. Kinoshita, H. Imaizumi, and S. Kunimatsu, ''Effects of meteorological condition near ground surface on sound propagation,'' Pollution Control **26**, 33–42 (1991) (in Japanese).
[5] T. F. W. Embleton, J. E. Piercy, and G. A. Daigle, ''Effect of flow resistivity of ground surfaces determined by acoustical measurements'' J. Acoust. Soc. Am. **74**, 1239–1244 (1983).
[6] A. Berry and G. A. Daigle, ''Controlled experiments on the diffraction of sound by a curved surface,'' J. Acoust. Soc. Am. **83**, 2047–2058 (1988).
[7] G. A. Daigle, ''Diffraction of sound by a noise barrier in the presence of atmospheric turbulence'' J. Acoust. Soc. Am. **71**, 847–854 (1982).

[8] A. L'Espérance, Y. Gabillet, and G. A. Daigle, ''Outdoor sound propagation in the presence of atmospheric turbulence: Experiments and theoretical analysis with the fast field program algorithm,'' J. Acoust. Soc. Am. **98**, 570–579 (1995).
[9] A. L'Espérance, J. Nicolas, D. K. Wilson, D. W. Thomson, Y. Gabillet, and G. A. Daigle, ''Sound propagation in the atmospheric surface layer: Comparison of experiment with FFP predictions,'' Appl. Acoust. **40**, 325–346 (1993).
[10] Y. Gabillet, H. Schroeder, G. A. Daigle, and A. l'Espérance, ''Application of the Gaussian beam approach to sound propagation in the atmosphere: theory and experiments,'' J. Acoust. Soc. Am. **93**, 3105–3116 (1993).
[11] M. R. Stinson, D. I. Havelock, and G. A. Daigle, ''Simulation of scattering by turbulence into a shadow region using the GE-PE method,'' in *Proceedings of Sixth International Symposium on Long-Range Sound Propagation* (Ottowa, Canada), pp. 283–295 (1994).
[12] B. M. Gibbs and D. K. Jones, ''A simple image method for calculating the distribution of sound pressure levels within an enclosure,'' Acustica **26**, 24–32 (1972).
[13] S. M. Dance, ''The development of computer models for prediction of factory noise,'' J. Building Acoust. **2**, 437–454 (1995).
[14] S. M. Dance and B. M. Shield, ''The complete image-source method for the prediction of sound distribution in nondiffuse enclosed spaces,'' J. Sound Vib. **201**, 473–489 (1997).
[15] S. M. Dance and B. M. Shield, ''Modeling of sound fields in enclosed spaces with absorbent room surfaces. Part I: Performance spaces,'' Appl. Acoust. **58**, 1–18 (1999).
[16] A. M. Ondet and J. L. Barbry, ''Modeling of sound propagation in fitted workshops using ray tracing,'' J. Acoust. Soc. Am. **85**, 787–796 (1989).
[17] M. Hodgson, ''On the prediction of sound fields in large empty rooms,'' J. Acoust. Soc. Am. **84**, 253–261 (1988).
[18] M. Hodgson, ''On the accuracy of models for predicting sound propagation in fitted rooms,'' J. Acoust. Soc. Am. **88**, 871–878 (1990).
[19] K. H. Kuttruff, ''Sound decay in reverberation chambers with diffusing elements,'' J. Acoust. Soc. Am. **69**, 1716–1723 (1981).
[20] K. S. Pearsons, R. L. Bennett, and S. Fidell, ''Speech levels in various noise environments,'' Environmental Health Effects Research Series EPA-600/1-77-025 (1977).
[21] M. D. Wang and R. C. Bilger, ''Consonant confusions in noise: a study of perceptual features,'' J. Acoust. Soc. Am. **54**, 1248–1266 (1973).
[22] A. J. Duquesnoy, ''The intelligibility of sentences in quiet and in noise in aged listeners,'' J. Acoust. Soc. Am. **74**, 1136–1144 (1983).
[23] R. Plomp and A. M. Mimpen, ''Speech-reception threshold for sentences as a function of age and noise level,'' J. Acoust. Soc. Am. **66**, 1333–1342 (1979).
[24] K. L. Payton, R. M. Uchanski, and L. D. Braida, ''Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **95**, 1581–1592 (1994).
[25] A. K. Nabelek and P. K. Robinson, ''Monaural and binaural speech perception in reverberation for listeners of various ages,'' J. Acoust. Soc. Am. **71**, 1242–1248 (1982).
[26] K. S. Helfer and R. A. Huntley, ''Aging and consonant errors in reverberation and noise,'' J. Acoust. Soc. Am. **90**, 1786–1796 (1991).
[27] S. A. Gelfand and S. Silman, ''Effects of small room reverberation upon the recognition of some consonant features,'' J. Acoust. Soc. Am. **66**, 22–29 (1979).
[28] S. Watanabe and K. Inomoto, ''Effect of reverberation on speech intelligibility,'' J. Acoust. Soc. Jpn. **35**, 97–103 (1979) (in Japanese with English abstract).
[29] N. Sakata, T. Yoshida, H. Tokuyama, C. Haruta, and T. Ohtsuru, ''Fundamental analysis of transmission characteristics of articulation test,'' Trans. Architectural Inst. Jpn. **276**, 79–86 (1979) (in Japanese with English abstract).
[30] C. Haruta and N. Sakata, ''Effects of some factors on speech communication in sound field,'' Report of Kurume Institute of Technology **10**, 63–70 (1986) (in Japanese with English abstract).
[31] C. Haruta, ''Souonba niokeru onsei dentatsu tokusei ni kansuru kenkyu (Japanese title only),'' Doctorate thesis (1987) (in Japanese).
[32] T. Houtgast and H. J. Steeneken, ''The modulation transfer function in room acoustics,'' Brüel & Kjær Techn. Rev. **3**, 3–11 (1985).
[33] T. Houtgast, H. J. Steeneken, and R. Plomp, ''Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics,'' Acustica **46**, 60–72 (1980).
[34] T. Houtgast and H. J. M. Steeneken, ''A multi-language evaluation of the RASTI method for estimating speech intelligibility in auditoria,'' Acustica **54**, 185–199 (1984).
[35] A. G. Galaitsis and W. N. Patterson, ''Prediction of noise distribution in various enclosures from free-field measurements,'' J. Acoust. Soc. Am. **60**, 848–856 (1976).
[36] C. J. Rainsford and J. R. Brown, ''The propagation and attenuation of noise in coal mine roadways,'' The Mining Engineer **9**, 115–122 (1983).
[37] C. J. Rainsford, ''Propagation of noise in underground roadways,'' Doctorate thesis (1982).
[38] G. S. Green, ''Noise propagation in mine roadways and longwall faces,'' Doctorate thesis (1986).
[39] M. Fujimiya, N. Ichikawa, S. Kanno, H. Hamada, Y. Kogure, and T. Miura, ''On the relations between physical parameters and articulation in a tunnel,'' Paper of the Technical Group on Electroacoustics **82**, EA-82-54 (1982) (in Japanese).
[40] T. Hirai, N. Masaki, and K. Okada, ''The use of underground space and the underground shopping mall projects in Japan,'' ICUSESS'92 **5**, 32–41 (1992).
[41] H. Imaizumi, T. Isei, and M. Sagisaka, ''Sound propagation and speech information transmission along an underground gallery with ventilation,'' J. Jpn. Soc. Safety Engineer. **31**, 228–239 (1992) (in Japanese with English abstract).
[42] H. Imaizumi, T. Isei, S. Kunimatsu, A. Tanaka, M. Kinoshita, and M. Tanaka, ''Influence of ventilation on hearing of speech in an underground roadway,'' J. NIRE **6**, 369–380 (1994).
[43] H. Imaizumi, M. Kinoshita, S. Kunimatsu, and T. Isei, ''Sound propagation and speech hearing in a curved reverberant tunnel,'' J. Acoust. Soc. Jpn. (E) **18**, 129–137 (1997).
[44] Acoustical Society of Japan, *Meiryodo shikenho no kijyun* (Japanese title only) (Tokyo, Japan, 1957) (in Japanese), pp. 1–44.
[45] LMS Numerical Technologies, RAYNOISE USER MANUAL, Leuven, Belgium (1997).
[46] W. N. Patterson, G. G. Huggins, and A. G. Galaitsis, ''Noise of diesel-powered underground mining equipment: impact, prediction, and control,'' Bureau of Mines OFR **58–75**, D1-19 (1975).
[47] U. Ingard, ''On the reflection of a spherical sound wave from an infinite plane,'' J. Acoust. Soc. Am. **23**, 329–335 (1951).
[48] H. M. Macdonald, ''A class of diffraction problems,'' Proc. London Math. Soc. **14**, 410–427 (1915).
[49] H. S. Carslaw, ''Diffraction of waves by a wedge of any angle,'' Proc. London Math. Soc. **18**, 291–306 (1919).
[50] T. Kawai, ''On asymptotic solutions of the three-dimensional problems of diffraction by a wedge,'' J. Acoust. Soc. Jpn. **35**, 544–553 (1979) (in Japanese with English abstract).
[51] R. G. Kouyoumjian and P. H. Pathak, ''A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface,'' Proc. IEEE **62**, 1448–1461 (1974).
[52] D. K. Wilson, ''Simple, relaxational models for the acoustical properties of porous media,'' Appl. Acoust. **50**, 171–188 (1997).
[53] S. Iida, ''Tendency of mishearing in articulation test,'' J. Acoust. Soc. Jpn. **29**, 527–534 (1973) (in Japanese).
[54] S. Saito, ''Evaluation of articulation test and articulation score,'' Electrical Commun. Lab. Techn. J. **8**, 845–853 (1959) (in Japanese).
[55] H. Fletcher and J. C. Steinberg, ''Articulation Testing Methods,'' Bell Syst. Tech. J. **8**, 806–854 (1929).
[56] Acoustical Society of Japan, *Onsei* (Japanese title only) (Corona Publishing Co. Ltd., Tokyo, 1997), pp. 177–181 (in Japanese).
[57] Acoustical Society of Japan, *Onkyo Yougo Jiten* (Japanese title only), 3rd ed. (Corona Publishing Co. Ltd., Japan, 1994), p. 225 (in Japanese).
[58] Acoustical Society of Japan, *Onkyo Yougo Jiten* (Japanese title only), 3rd ed. (Corona Publishing Co. Ltd., Japan, 1994), p. 627 (in Japanese).
[59] B. Yegnanarayana and B. S. Ramakrishna, ''Intelligibility of speech under nonexponential decay conditions,'' J. Acoust. Soc. Am. **58**, 853–857 (1975).
[60] T. Nakajima, ''Measurements of the modulation transfer function (MTF) and speech transmission index (STI) in room acoustics,'' J. Acoust. Soc. Jpn. **49**, 103–110 (1993) (in Japanese).
[61] Y. Toida, ''Speech intelligibility in sound fields,'' J. Acoust. Soc. Jpn. **51**, 312–316 (1995) (in Japanese).

From a profiled diffuser to an optimized absorber

T. Wu, T. J. Cox, and Y. W. Lam
School of Acoustics and Electronic Engineering, University of Salford, Salford M5 4WT, United Kingdom

(Received 28 October 1999; revised 24 March 2000; accepted 23 April 2000)

The quadratic residue diffuser was originally designed for enhanced scattering. Subsequently, however, it has been found that these diffusers can also be designed to produce exceptional absorption. This paper looks into the absorption mechanism of the one-dimensional quadratic residue diffuser. A theory for enhanced absorption is presented. Corresponding experiments have also been done to verify the theory. The usefulness of a resistive layer at the well openings has been verified. A numerical optimization was performed to obtain a better depth sequence. The results clearly show that by arranging the depths of the wells properly in one period, the absorption is considerably better than that of a quadratic residue diffuser. © *2000 Acoustical Society of America.* [S0001-4966(00)01308-4]

PACS numbers: 43.55.Ev, 43.55.Dt [JDQ]

INTRODUCTION

Profiled diffusers were invented by Schroeder[1] in the 1970s. An example of the one-dimensional Schroeder diffusers is shown in Fig. 1. They are periodic surface structures with rigid construction; the elements of the structure are wells of the same width separated by thin fins. Within one period, the depths of the elements vary according to a pseudo-stochastic sequence.

The basis behind the Schroeder diffuser is as follows: When the sound is incident on the surface of the structure, plane waves travel down and up each wells. The returning waves at the entrances of the wells no longer have the same phases because of the different depths they have traveled. If the phase differences are appropriate, the diffuser will disperse the sound rather than specularly reflect it, and thus generate diffusion in the far field. For this purpose, Schroeder exploited the Fourier transform of the surface reflection factors to choose the depth sequence in one period, as it approximately gives the far field diffracted pressure distribution. The most popular Schroeder diffuser is the quadratic residue diffuser (QRD), which employs the quadratic residue sequence to determine the well depth. The Fourier transform of the QRD surface reflection factors gives a constant power spectrum leading to even energy diffraction lobes. The QRD has been widely used in concert halls, theatres, and studio monitor rooms.[2,3]

The QRD was designed to diffuse rather than absorb sound, although for some time there was anecdotal evidence of absorption. Dramatic levels of absorption from Schroeder diffusers were measured by Fujiwara and Miyajima[4] in 1992. They reported the unexpectedly high measured absorption of a poorly constructed QRD at low frequency, which they could not explain. It was later found[5] that the quality of construction was to blame for some of the excess absorption. Kuttruff[6] tried to explain the absorption by assuming that the total sound pressure on the diffuser plane was constant. This led to an absorption based on the same average surface admittance, which generated additional possible air flow between adjacent wells. This additional air flow was associated with the excess absorption of the diffusers, but this air flow is not explicitly included in the average admittance model. He could only find agreement with Fujiwara's data when he reduced the width of the wells to an unrealistic one-tenth of the width in the real diffuser. Mechel thoroughly investigated the Schroeder diffuser in his paper in Acustica[7] and in greater detail in his book.[8] He described the absorption effect for the near field as well as the directivity for the far field analysis, with and without viscous and thermal losses in and in front of the wells using a Fourier wave decomposition model. Mechel demonstrated that resistive layers at the well entrances turn these diffuser structures into potentially useful and practical absorbers. Furthermore, he discussed how using a primitive root sequence to determine the well depths of the structure could result in a better absorber than the more common quadratic residue sequence. Unfortunately, there were no corresponding experimental results to verify the findings, something which is added later in this paper.

In this paper, the sound absorption by the one-dimensional Schroeder diffuser is studied both theoretically and experimentally. The Schroeder structure can be used to make an effective absorber or a low loss diffuser depending on the geometry and construction of the device. Since absorption is more of a concern here than diffusion, the width of the wells has been reduced for this study compared to QRDs designed for diffusive purposes. The losses in the wells caused by the viscous and thermal conduction have been considered. To predict the surface sound absorption, the method used by Mechel is used. The use of a resistive layer on the surface to improve the absorption of the structure is investigated in some detail.

As shall be shown later, a optimum diffuser does not necessary produce the best absorption. It is possible to produce a configuration that is a better absorber than those based on the quadratic residue sequence. This is done by producing many well-tuned and well-distributed resonance frequencies by choosing the depth sequence in one period using an optimization approach.

I. PREDICTION METHOD

Two methods have been used to predict the absorption on the surface of profiled diffusers; one uses the average

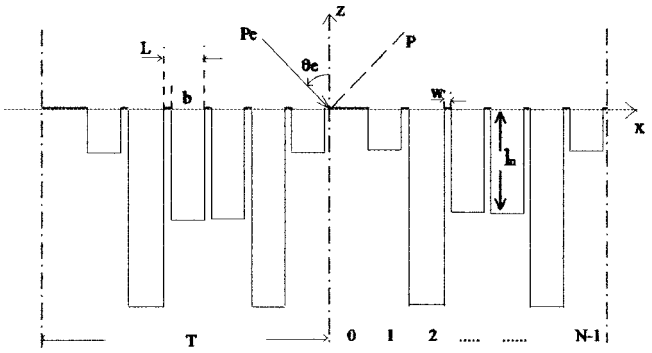


FIG. 1. One-dimensional quadratic residue diffuser for $N=7$, 2 periods shown.

surface admittance, another uses Fourier analysis. The former was introduced by Kuttruff by assuming uniform pressure on the entrance of the structure. In many cases, however, as he indicated, it is only an approximate method, as it does not properly model the mutual interference between the differently tuned wells caused by different depths in one period. The later method, i.e., Fourier analysis, more correctly considers the coupling between the wells, and thus is more rigorous. In this paper only Fourier analysis method will be discussed and used in the predictions. Figure 1 shows a sketch of a one-dimensional QRD structure, together with the coordinate system. The expression of the depth sequence, l_n, in one period of a one-dimensional QRD is well known[1] as

$$l_n=\frac{c \bmod(n^2,N)}{N(2\,f_r)}, \quad n=0,1,...,N-1. \tag{1}$$

where f_r is the design frequency, N is the prime number, and c is the speed of sound in air. As shown in Fig. 1, the diffuser is periodic with N wells in one period. The period is $T=N*L$ wide with the grid length of one well, $L=b+w$, where b is the well width and w is the wall thickness between the wells.

A. Energy losses in the wells

Normally the well width of the QRD is quite large to minimize absorption. However, in this study, the width of the wells needs to be reduced to get higher absorption. In this case, the losses caused by the viscous and thermal conduction in the narrow wells cannot be neglected. In general, in the middle frequency range of interest, the width of the wells b is usually quite narrow compared with the incident wavelength, i.e., $b \ll \lambda/2$, so that there is only fundamental mode propagating in each well.

When describing the sound propagation in a narrow slit, three waves will be of concern: propagating wave, thermal wave, and shear wave. In many cases of practical interest, the attenuation caused by the thermal wave and viscous wave only occurs in the boundary layers, which are of a fraction of a millimeter thick. Usually their effects can be incorporated into the boundary conditions on the propagating wave. Morse and Ingard[9] derived the complex wave number in a cylindrical tube. The method can also be used for a narrow slit.

Assuming that the slit has two infinite parallel rigid walls, then it is somewhat more convenient to run the z axis along the central line of the slit. Because the boundary surface is rigid, the particle velocity u and the temperature fluctuation τ at the surface are zero. The propagation wave for fundamental mode has acoustic pressure p_p, temperature τ_p, and normal u_{px} and tangential u_{pz} velocity:[9,10]

$$\begin{aligned}
p_p&=\cos\left(\frac{\pi q x}{b}\right)e^{-jk_t z}e^{j\omega t},\\
\tau_p&=\frac{\gamma-1}{\alpha\gamma}\cos\left(\frac{\pi q x}{b}\right)e^{-jk_t z}e^{j\omega t},\\
u_{px}&=\frac{1}{j\omega\rho}\,\frac{\pi q}{b}\sin\left(\frac{\pi q|x|}{b}\right)e^{-jk_t z}e^{j\omega t},\\
u_{pz}&=\frac{k_t}{\rho c k}\cos\left(\frac{\pi q x}{b}\right)e^{-jk_t z}e^{j\omega t},
\end{aligned} \tag{2}$$

where $k_t^2=k^2-(\pi q/b)^2$, $k=\omega/c$ is the wave number, ρ is the air density, and $\gamma=(C_P/C_V)$ is the ratio of the specific heat, $\approx 7/5$ for air. The constant q can be adjusted to fit the boundary conditions by incorporating the thermal wave and shear wave in the boundary layers. Following the method used by the Morse and Ingard,[9] for width $b \gg d_v, d_h$, where d_v, d_h are the thickness of the viscous and thermal boundary layers, q can be defined for a narrow slit:

$$(\pi q)^2=-(1-j)k^2 b[d_v+(\gamma-1)d_h]. \tag{3}$$

This gives the propagation number as

$$k_t=\sqrt{k^2-\left(\frac{\pi q}{b}\right)^2}\approx k+\frac{k}{2b}(1-j)[d_v+(\gamma-1)d_h] \tag{4}$$

for air at atmospheric pressure and room temperature, d_v, d_h can be determined[10] as

$$d_v=\sqrt{\frac{2\mu}{\rho\omega}}\approx 0.21\frac{1}{\sqrt{f}}, \quad d_h=\sqrt{\frac{2K}{\rho\omega C_p}}\approx 0.25\frac{1}{\sqrt{f}}\ \text{(cm)}, \tag{5}$$

where f is sound frequency, and μ, K, and C_p are the properties of gas. μ is the coefficient of viscosity, K is the thermal conductivity, and C_p is the heat capacity per unit mass at constant pressure.

Once the sound propagation in the slit is known, the inward impedance on the surface of well with length l_n can be easily derived as

$$Z(l_n)=-j\rho_e c\frac{k}{k_t}\cot(k_t l_n), \tag{6}$$

where ρ_e is the effective density of air in the slit:[11]

$$\rho_e=\rho[1+(1-j)d_v/b] \tag{7}$$

so that the inward normalized specific impedance of the well with length l_n is then

$$\begin{aligned}
\zeta(l_n)&=\frac{Z(l_n)}{\rho c}\\
&=-j\{1+(1-j)[d_v-(\gamma-1)d_h]/2b\}\cot(k_t l_n).
\end{aligned} \tag{8}$$

B. Prediction of absorption by QRD

The analysis below closely follows the method used by Mechel.[7] The sound field in front of the diffuser, shown in Fig. 1, is decomposed into the incident plane wave $p_e(x,z)$ and scattered field $p_s(x,z)$, which is made up of propagating and nonpropagating evanescent waves:

$$p(x,z)=p_e(x,z)+p_s(x,z), \tag{9}$$

$$p_e(x,z)=P_e\cdot e^{j(-xk_x+zk_z)},$$

$$\text{where } k_x=k\sin\theta_e, \quad k_z=k\cos\theta_e, \tag{10}$$

$$p_s(x,z)=\sum_{n=-\infty}^{+\infty}A_n e^{j(-x\beta_n-z\gamma_n)}. \tag{11}$$

Since the QRD is periodic, the scattered field is also periodic in x. Therefore the wave numbers in the x and z directions of the spatial harmonics are (the first from the periodicity, the latter from the wave equation)

$$\beta_n=k_x+n\frac{2\pi}{T};$$

$$\gamma_n^2=k^2-\beta_n^2\Rightarrow\gamma_n=-jk\sqrt{\left(\sin\vartheta_e+n\frac{\lambda}{T}\right)^2-1}, \tag{12}$$

where $\lambda=2\pi/k$ is the wavelength. The corresponding radiating harmonics indices n_s, which can propagate to the far field, are determined from

$$\left(\sin\theta_e+n_s\frac{\lambda}{T}\right)^2\leqslant 1. \tag{13}$$

Considering the outward particle velocity along the positive z direction, the relation of pressure and particle velocity on the surface is $\rho c v_z(x,0)=-G(x)p(x,0)$. This gives

$$\cos\theta_e P_e-\sum_{n=-\infty}^{+\infty}\frac{\gamma_n}{k}A_n e^{-jxn2\pi/T}$$

$$=G(x)\left[P_e+\sum_{n=-\infty}^{\infty}A_n e^{-jxn2\pi/T}\right]. \tag{14}$$

Since $G(x)$ is periodic with a period T, we apply a Fourier analysis:

$$G(x)=\sum_{n=-\infty}^{+\infty}g_n e^{-jn(2\pi/T)x},$$

$$g_n=\frac{1}{T}\int_0^T G(x)e^{+jn(2\pi/T)x}dx. \tag{15}$$

Equation (15) inserted into the boundary condition gives, after multiplication by $e^{jm(2\pi/T)x}$ and integration over T:

$$\sum_{n=-\infty}^{+\infty}A_n\left(g_{m-n}+\delta_{m,n}\left(\frac{\gamma_n}{k}\right)\right)=P_e(\delta_{m,0}\cos\theta_e-g_m);$$

$$m=-\infty,\ldots\ldots,+\infty, \tag{16}$$

where

$$\delta_{m,n}=\begin{cases}1, & m=n\\ 0, & m\neq n,\end{cases}$$

TABLE I. Depth sequences of different structures (in cm).

		Optimization			
		Without mirror-image		With mirror-image[a]	
QRD	Random sequence	1	2	1	2
0.0	0.0	3.6	2.5	10.0	4.8
2.5	1.0	10.0	10.0	9.2	3.4
10.0	3.0	5.6	3.8	8.5	3.7
5.0	5.0	9.1	9.4	7.2	8.6
5.0	7.0	6.7	5.9	5.8	5.7
10.0	9.0	8.2	7.3	3.9	8.3
2.5	10.0	4.5	4.8	4.3	10.0

[a]See Sec. II B.

and the infinite large system of equations will be terminated at the index limits $n,m=\pm 2*N$, where N is the number of wells in one period. By solving the above equations, the coefficients A_n can be obtained.

The absorption coefficient of diffuser is then

$$\alpha(\theta_e)=1-\left|\frac{A_0}{P_e}\right|^2-\frac{1}{\cos\theta_e}\sum_{n_s\neq 0}\left|\frac{A_{n_s}}{P_e}\right|^2\sqrt{1-(\sin\theta_e+n_s\lambda/T)^2}; \tag{17}$$

the summation runs over radiating spatial harmonics only. Evidently the second term is the specular reflection, and the third term is due to scattering.

If a resistive layer is applied on the surface of structure to enhance the absorption, the only replacement made in the above method is $G(x)\rightarrow 1/(R+\zeta_{\text{well}})$, where R is the normalized effective impedance of the resistive layer, and ζ_{well} is the input normalized impedance of a well.

II. EXPERIMENTAL AND THEORETICAL RESULTS

A. Methods of measurement

Two types of samples, one with constant length slits and the other a one-dimensional QRD, were built from a square tube that had a cross-section size 54 mm×54 mm. Because of tube's size, the QRD are limited to: prime number 7, design frequency $fr=980$ Hz, the maximum length of wells 10 cm, well width $b=6$ mm, and separation wall thickness $w=1$ mm. The length sequence has been listed in Table I, as ''QRD.'' The wells were terminated by MDF (Medium Density Fiberboard) which has been varnished three times. The wells are separated by aluminum sheets. In order to compare fairly, the corresponding constant length structure is composed of seven slits with the same width as the QRD and with all the wells having a length of 10 cm. The whole sample is sealed with petroleum jelly as good as possible.

A method similar to the standard two microphone method[12] is used to measure the surface impedance of the sample in the impedance tube. Consequently, the results presented here are restricted to normal incidence. Furthermore, the normal incidence absorption coefficient is computed for the impedance from the well-known formula:[13]

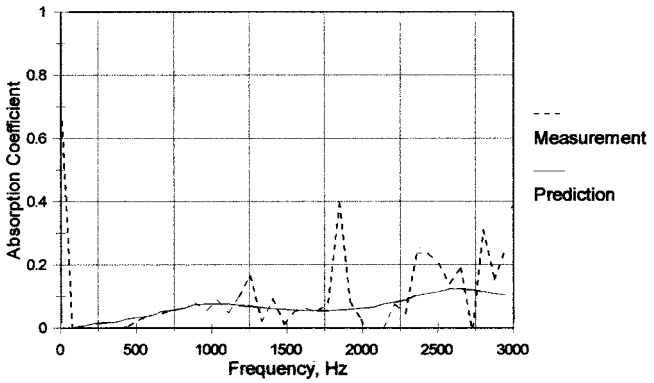


FIG. 2. Absorption coefficient of constant length slit sample.

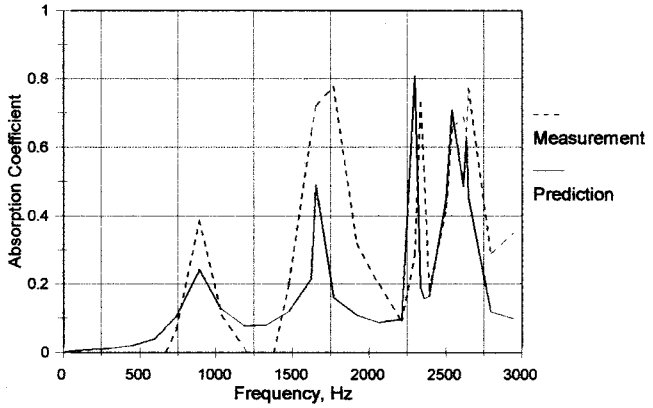


FIG. 3. Comparison between the prediction and experiment for the QRD.

$$\alpha = \frac{4\,\mathrm{Re}(\zeta)}{[1+\mathrm{Re}(\zeta)]^2+[\mathrm{Im}(\zeta)]^2}, \tag{18}$$

where $\mathrm{Re}(\zeta)$ and $\mathrm{Im}(\zeta)$ are the real part and imaginary part of the normalized impedance, respectively.

The effect of the resistive layers is also tested. A thin wire mesh with resistance 550 rayl (MKS), whose normalized specific acoustic impedance is 1.325, is applied in front of the sample.

B. Theoretical prediction

To predict the experimental results, there are a few things that should be noted:

First, because the sample is set up in the impedance tube with rigid walls, the prediction in Sec. I B cannot be used directly. The period must be doubled to take into account the mirror imaging effect of the side walls. For example, considering the tested QRD with $N=7$, the depth sequence in one period is 0 1 4 2 2 4 1. An image sequence 1 4 2 2 4 1 0 will be created by the rigid walls. So, in fact, this system has 14 elements in one period.

Second, when predicting the absorption coefficient of the sample with wire mesh, the mass effect of the wire mesh has to be considered. In most cases, wire mesh can be treated as rigid. But a thin wire mesh is flexible, and there will be an inertial mass contribution resulting from the induced motion of the sheet.[14] So the *normalized effective impedance* of the wire mesh is

$$R = \left[\frac{R_m(\omega m)^2}{R_m^2+(\omega m)^2} + j\,\frac{R_m^2(\omega m)}{R_m^2(\omega m)^2}\right] \Big/ \rho c, \tag{19}$$

where R_m and m are, respectively, the resistance and mass per unit area of wire mesh.

C. Results and discussions

For conciseness, the following abbreviations in the normalized impedance graphs will be adopted throughout this paper. The real part of normalized impedance will be *R*, and the imaginary part I.

Figures 2 and 3 show the predicted and measured sound absorption coefficient for the constant slits and QRD without wire mesh, and Figs. 4 and 5 show the results with wire mesh. The agreement is good, especially for the results with wire mesh. Without wire mesh, the absorptions are small and very difficult to measure, especially for the constant slits in Fig. 2. Figure 3 contains two important features. First, around the second peak, the absorptions are higher than prediction, it may due to both the gaps in the QRD structure and the resonant vibration of the aluminum sheet, which separated the wells; Second, there is a very sharp absorption peak around 2330 Hz. This only appears in the prediction if the mirror imaging of the rigid walls is modeled. It would not appear for a straight periodic QRD. Comparing Figs. 2 and 3, it is clear that the absorption of the QRD is much higher than the corresponding constant slits. In Figs. 4 and 5, it can be seen that the resonant frequencies are slightly shifted down when the wire mesh is added. This is due to the mass effect introduced by the thin flexible wire mesh. Comparing the absorption in Figs. 4 and 5, higher and more uniform absorption is found for the QRD compared with the constant slit surface.

III. ABSORPTION MECHANISM DISCUSSION

Comparing the absorption of the constant length structure in Fig. 2 and the absorption of the QRD in Fig. 3, although less cells of the QRD are resonating at a particular frequency, it generates higher absorption than the constant slits, even at the resonance frequencies of constant slits, where all slits contribute. This is because the variable depths of the wells create a nonuniform surface impedance that scatters the incident sound. The scattering enhances the propa-

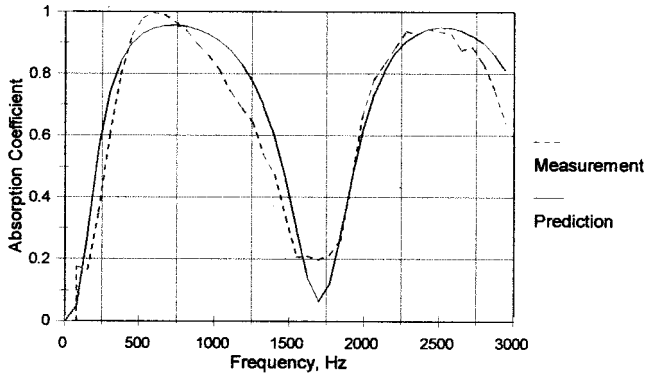


FIG. 4. Absorption coefficient from constant slit sample with 550 Rayls wire mesh on the entrance of the slit.

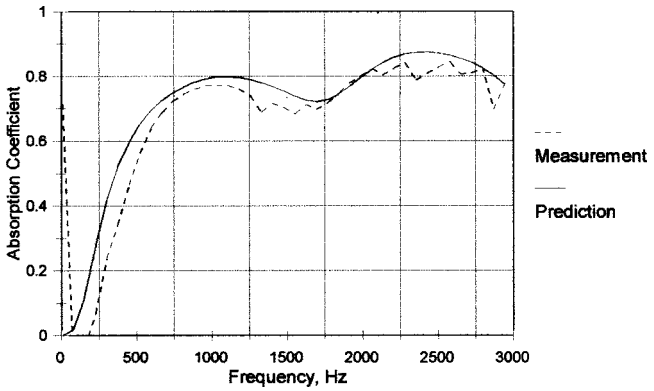


FIG. 5. Comparison between predicted and measured absorption for the QRD with facing wire mesh 550 Rayls.

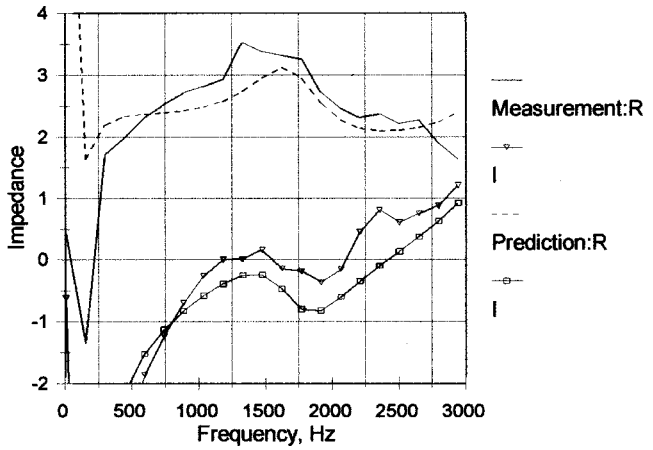


FIG. 7. Normalized specific impedance corresponding to Fig. 5.

gation of the sound wave between wells and hence increases the sound absorption. The impedance graphs of them are typical for pipe systems, except that the QRD has nonuniform resonant frequencies caused by depth variety and slightly higher resistance than constant slits because of the coupling of wells. Applying facing wire-mesh can improve the absorption coefficient significantly across the whole frequency range. Figure 6 shows the normalized specific acoustic impedance corresponding to the constant slits tested in Fig. 4, with a wire mesh glued on the opening of slits, and Fig. 7 shows that corresponding to the QRD tested in Fig. 5 with glued facing wire mesh. For constant length slits, applying wire mesh shifts down the resonance frequencies (the fundamental and first harmonic) because of the mass effect. The real part remains close to the resistance of the wire mesh at 1.325. There is no coupling between the wells that may effect the impedance of whole structure. The absorption is high at the resonance frequencies, but elsewhere is poor. For the QRD, Fig. 7 shows that applying the wire mesh not only shifts down the resonance frequencies, it also smooths the imaginary part and real part of impedance. The imaginary part stays close to zero. This is the reason why high absorption is achieved across the whole frequency range. However, the resistive part is also substantially increased away from unity, leading to a slightly smaller absorption coefficient at resonances. This indicates that the choice of wire mesh is important in optimizing the absorption.

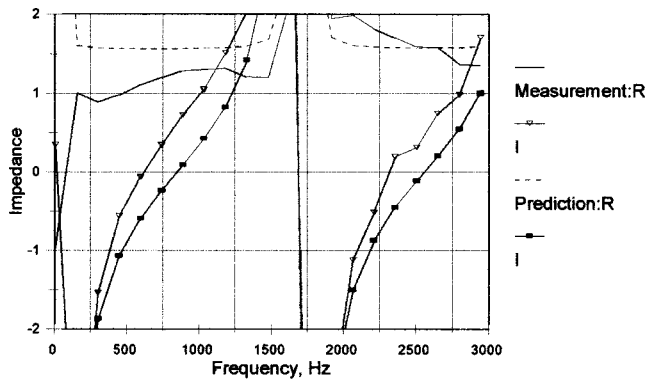


FIG. 6. Normalized specific impedance corresponding to Fig. 4.

IV. OPTIMIZATION

As mentioned before, the depth sequence of a Schroeder diffuser is very important to the absorption of whole structure. In this section, a numerical optimization method,[15] downhill simplex method, is used to tune this sequence. By optimizing the sequence, many well-tuned and well-distributed resonance frequencies can be generated, and higher absorption can be achieved.

A. Optimization process and "absorption parameter" discussion

The process to produce an optimum profiled absorber is based on an iterative process:

(1) An absorber with N wells in one period is constructed with a randomly determined depth sequence.

(2) Absorption coefficients of the absorber is calculated by the Fourier analysis method across the frequency range of interest.

(3) A single figure cost function is calculated which can measure the degree of broadband absorption.

(4) The well depths are altered according to the Downhill Simplex method.

(5) Steps (2)–(4) are repeated until a minimum in the cost function is found indicating an optimum absorber.

There are two cost functions which have been used in the optimization: one is $-\Sigma_{i=1}^{N}\alpha_i/N$, where α_i is absorption coefficient, N is the number of frequencies chosen in the frequency range interested, and the other is $\sqrt{(\Sigma_{i=1}^{N}X_i^2)/N}$, where X_i is the imaginary part of specific acoustic impedance. The first parameter measures the negative of the mean absorption, hence the minimum value gives the highest average absorption. The second one is used because absorption is strongly related to the impedance on the surface of the structure. When the imaginary part of the impedance stays close to zero, high absorption is usually obtained. But one thing that should be noticed here is that the coupling between the wells may cause real part of the impedance to increase and result in less absorption. It is necessary to run the optimization process many times with different starting conditions. The reason for this is that the minimization is being carried out within bounded space. The space hold many local minima within which the minimization routines could be-

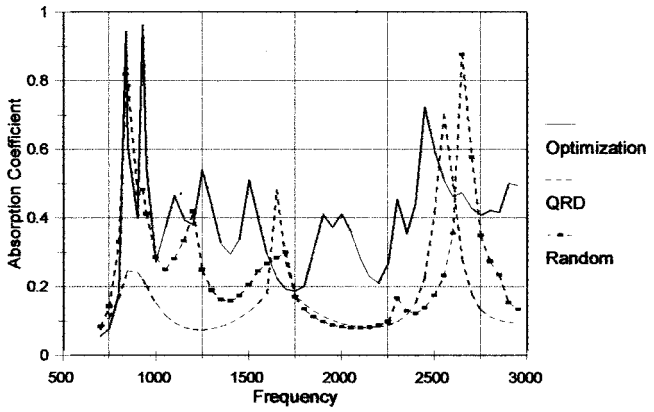


FIG. 8. Comparison of the absorption coefficient for the optimized surface, the QRD, and random sequence structure.

come trapped, particularly at the edges. The results shown below are the best results from many attempts of the iteration process. This does not exclude the possibility that from some particular starting point yet untried there might have a better minimum available.

B. Optimization examples: Theoretical results

As an example, the QRD with the same parameters as that tested in Fig. 3 has been optimized for normal plane wave incidence. In order to compare the results fairly, the maximum lengths of the wells are restricted to 10 cm in the optimization program. The first ''absorption parameter'' cost function, based on the mean absorption coefficient, is used. The frequency range for the optimization is 700 Hz–3000 Hz. The obtained depth sequence of optimized surface structures without and with wire mesh have been listed in Table I as ''Optimization-Without mirror-image'' 1 and 2, respectively.

The results of the comparison with optimized structure, QRD and structure employing random depth sequence, without facing wire mesh on the structure, is illustrated in Fig. 8. The random depth sequence is listed in Table I under ''random sequence.'' It clearly shows that the optimized structure improves the absorption coefficient significantly compared with other two. Its mean absorption coefficient is increased from 0.16 to 0.35 compared with the QRD. In the QRD, the zero depth well does not contribute directly to the absorption as there are no viscous losses due to progragation in this well, so its removal is useful. As there were more nonuniform depths in one period of the optimized structure than the original QRD, more resonance frequencies are generated within the whole frequency range, and the real part of the optimized surface impedance is also improved because of the better coupling of wells when compared to the QRD and random depth sequence. But as shown in Fig. 8, the absorption coefficient is not smooth with many peaks and troughs according to the resonance and anti-resonance frequencies, which is inevitable because of the small resistive part.

The result of the optimization with a facing wire mesh is shown in Fig. 9. This wire mesh has a small flow resistance 55 rayls (MKS), which is a normalized specific impedance of 0.1325. Figure 9(a) shows that the optimized structure is

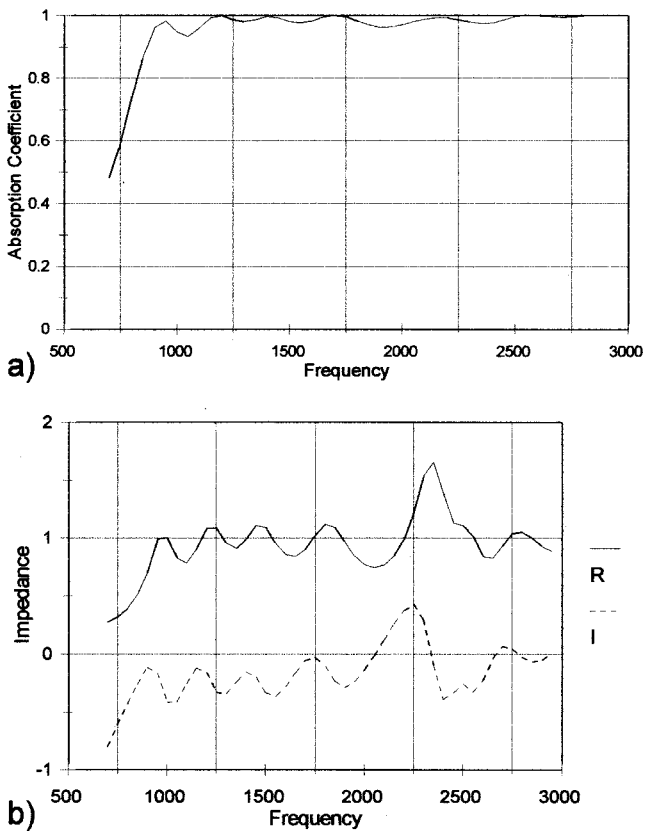


FIG. 9. Optimized structure with facing wire-mesh flow resistance 55 Rayls. (a) absorption coefficient; (b) normalized specific impedance.

obvious a good absorber, with a mean absorption coefficient of 0.96. And Fig. 9(b) shows that, not only does the imaginary part stay close to zero, but the real part of the specific impedance had been raised close to unity. This is more than the flow resistance of the wire mesh can achieve alone. This example demonstrates that optimized the depth sequence can improve the absorption of structure significantly, and downhill simplex method is fast and efficient for this purpose.

As mentioned before, there are two cost functions that can be used to measure the absorption, the following discussion is about the difference in results obtained by using them separately. The base structure is the same as above, but the facing flow resistance is now at 550 rayl (MKS), which is ten times as used before. The results are shown in Fig. 10.

Regarding Fig. 10(a), it is hard to say which cost function is better. The first cost function parameter, average absorption, generated a high absorption across the frequency range; but the second cost function parameter, minimum imaginary Z, produces higher absorption at low frequency. Figure 10(b) also shows other important points. The second, impedance based ''cost function,'' which is intended to produce the imaginary part close to zero, does work. But it also pushes the real part higher at some frequencies at the same time, this induces less absorption at these frequencies; The first, absorption based ''cost function,'' seems to optimize the real part and keep it close to 1, and therefore achieving the higher absorption. Although the imaginary part of impedance is not as close to zero as in the other case, there is no doubt it is optimized as well. Therefore, for the facing sheet

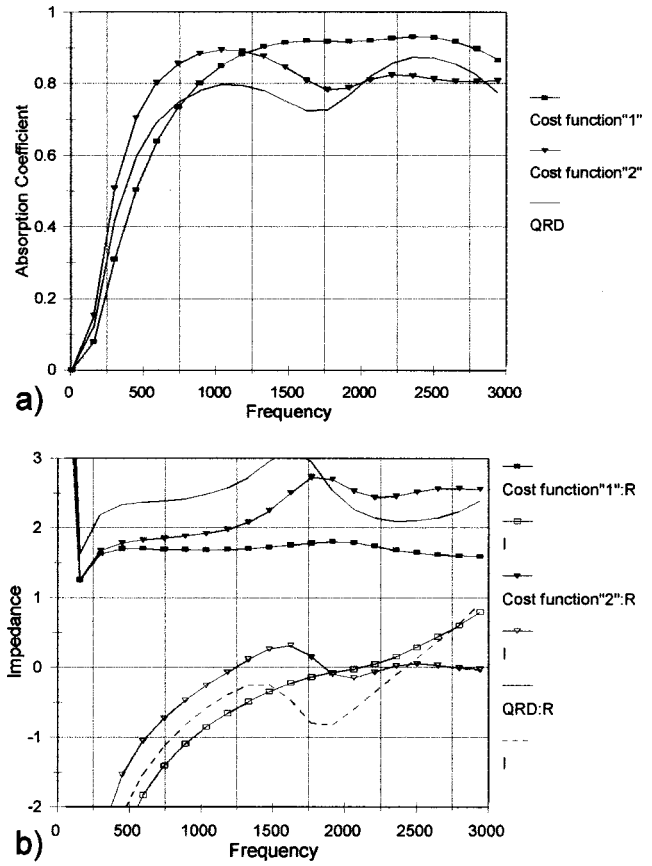


FIG. 10. Comparison of the results using different ''absorption cost functions.'' (a) absorption coefficient; (b) normalized specific impedance.

that has higher resistance, it is better to use the imaginary part cost function in the optimization process, where the real part cannot be reduced. On the other hand, the first cost function is better, for the lower facing resistance.

C. Experimental verification

Again the tests carried out in the impedance tube, and the optimizations had to be repeated with and without wire mesh to take into account of the mirror imaging effect of the impedance tube samples. The resistance of the wire mesh is Rayls (MKS), which is a normalized specific impedance of 0.434. The optimized structures used in the experiments have

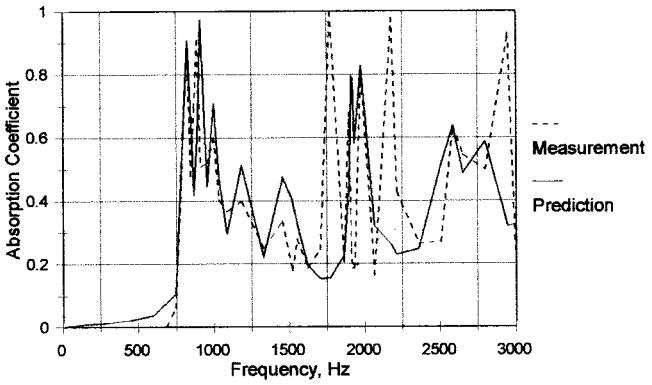


FIG. 11. Comparison of the predicted and measured absorption coefficient for optimized surface.

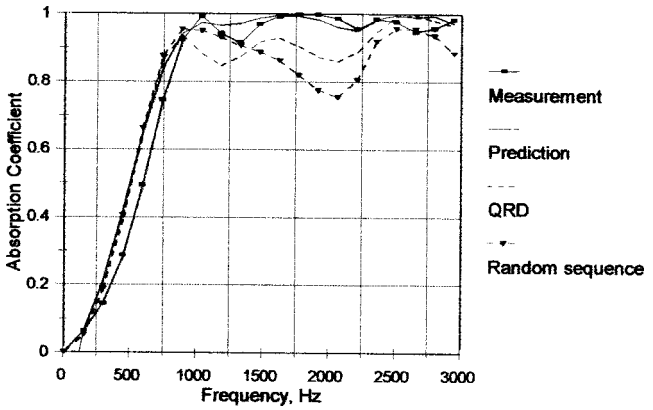


FIG. 12. Comparison between absorption coefficients for predicted and measured optimized surface with wire mesh and predicted the QRD and random sequence structure with wire mesh.

depth sequences shown in Table I, labeled as ''Optimization-with mirror-image'' 1 and 2, where ''1'' represents a depth sequence without wire mesh and ''2'' with wire mesh. The structures are built according to the above depth sequence, and other parameters are the same as tested before: seven wells, $b=6$ mm, $w=1$ mm.

Figure 11 is the comparison between the prediction and experiment absorption data without wire mesh, it shows that in generally they match each other. But there are two peaks which are not expected. It is suggested that they may be caused by the resonant vibration of the aluminum sheets, or higher modes generated in the wells.[16] Figure 12 shows the tested result of the optimized structure with wire mesh; the very good agreement can be clearly seen. The improvement compared with the QRD and random depth sequence structure is also illustrated.

V. CONCLUSIONS

The above study has shown that the variable depth sequence concept can be used to significantly improve the absorption and impedance characteristics of conventional constant depth design. A theory for the prediction of the enhanced absorption was presented and verified by normal incidence measurements on a variety of samples, with and without facing wire meshes. The prediction generally agrees well with experiments and the accuracy is particularly good in cases where a wire mesh is present. An optimization algorithm has been implemented. It was used to demonstrate the improvements in normal incidence absorption performance that can be achieved by optimizing the depth sequence. The optimized structure was tested in impedance tube measurements. The data generally agree with the predicted performance. Further work will be needed to extend the investigation to 2D and to test the optimised samples for the oblique incident sound.

ACKNOWLEDGMENTS

This work was funded by Engineering and Physical Sciences Research Council (EPSRC) of Britain, under Grant No. GR/L34396.

[1] M. R. Schroeder, ''Binaural dissimilarity and optimum ceilings for concert halls: More lateral sound diffusion,'' J. Acoust. Soc. Am. **65**, 958–963 (1979).
[2] P D'Antonio and T. J. Cox, ''Two decades of room diffusers. Part 1: Applications and design,'' J. Audio. Eng. Soc. **46**, 955–976 (1998).
[3] P D'Antonio and T. J. Cox, ''Two decades of room diffusers. Part 2: Measurement, prediction and characterisation,'' J. Audio. Eng. Soc. **46**, 1075–1091 (1998).
[4] K. Fujiwara and T. Miyajima, ''Absorption characteristics of a practically constructed Schroeder diffuser of quadratic-residue type,'' Appl. Acoust. **35**, 149–152 (1992).
[5] K. Fujiwara and T. Miyajima, ''A study of the sound absorption of a quadratic-residue type diffuser,'' Acustica **81**, 370–378 (1995).
[6] H. Kuttruff, ''Sound absorption by pseudostochastic diffusers (Schroeder diffusers),'' Appl. Acoust. **42**, 215–231 (1994).
[7] F. P. Mechel, ''The wide-angle diffuser—A wide-angle absorber?,'' Acustica **81**, 379–401 (1995).
[8] F. P. Mechel, *Schallabsorber* (S. Hirzel Verlag, Stuttgart, 1998), Vol. III, Chap. 5.
[9] P. M. Morse and K. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), Chap. 9, pp. 519–522.
[10] P. M. Morse and K. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), Chap. 6, pp. 285–291.
[11] J. F. Allard, *Propagation of Sound in Porous Media: Modeling Sound Absorbing Materials* (Elsevier Science, London, 1993), Chap. 4, pp. 48–53 and 59–62.
[12] International Standard, ISO 10534-2, ''Acoustics—Determination of sound absorption coefficient and impedance in impedance tubes.''
[13] H. Kuttruff, *Room Acoustics* (Elsevier Science, London, 1991), third edition, Chap. 2, pp. 30–33.
[14] U. Ingard, *Notes on Sound Absorption Technology* (Noise Control Foundation, Poughkeepsie, New York, 1994), Chap. 1, pp. 1–7.
[15] W. H. Press *et al.*, *Numerical Recipes* (Cambridge University Press, Cambridge, 1989), Chap. 10, pp. 289–293.
[16] F. P. Mechel, *Schallabsorber* (S. Hirzel Verlag, Stuttgart, 1998), Vol. III, Chap. 1.

On the combined effects of early- and late-arriving sound on spatial impression in concert halls

J. S. Bradley,[a] R. D. Reich, and S. G. Norcross
Institute for Research in Construction, National Research Council, Montreal Road, Ottawa IK1A 0R6, Canada

(Received 14 January 2000; revised 7 April 2000; accepted 15 May 2000)

This paper describes six new experiments involving subjective ratings of the listener envelopment, LEV, and the apparent source width, ASW, of simulated sound fields. Previous work has identified LEV and ASW as the principal components of spatial impression in concert halls and has shown that ASW is primarily influenced by the level of early lateral reflections and LEV by late-arriving lateral reflections. The new results in this paper show that LEV can result from nonlateral late-arriving sounds and demonstrate the conflicting effects of early- and late-arriving lateral sound on ASW and LEV when both are present, as would occur in real halls. While it is possible to create simulated sound fields with only either LEV or ASW, in typical concert halls, the balance between early- and late-arriving lateral sound will determine the relative importance of LEV and ASW. LEV and ASW are shown to be perceived when the critical components of the sound field are salient relative to other components. The results of the new subjective studies were used to estimate expected ASW and LEV in 16 halls. In these halls LEV is predicted to be the stronger component of spatial impression. [S0001-4966(00)04608-7]

PACS numbers: 43.55.Hy, 43.66.Pn [JDQ]

INTRODUCTION

Spatial impression is the multidirectional quality of musical sounds that one experiences when listening to music indoors compared to the more unidirectional experience of hearing music outdoors. It is well-established that spatial impression is one of the important acoustical aspects of a good concert hall. Early work tended to relate spatial impression to reverberation.[1] However, about 30 years ago, early-arriving lateral reflections were shown to be a major contributor to spatial impression.[2,3] After the importance of this work was fully appreciated, it had a significant impact on concert hall design and many new halls have been designed to included strong, early lateral reflections. More recent work has suggested that spatial impression is composed of more than one single aspect.[4,5] Work by Bradley and Soulodre[6] demonstrated that there are two subjective dimensions to spatial impression and that they are related to different physical aspects of the sound fields in halls.

As Barron and Marshall showed,[3,7] early lateral reflections are important contributors to spatial impression. It is now clear that increased early lateral reflection energy leads to the perception of source broadening or to an increased apparent source width (ASW). ASW has been defined as the width of a sound image fused temporally and spatially with the direct sound image.[5] Later-arriving lateral sound energy leads to a sense of listener envelopment (LEV). LEV can be defined as the sense of being surrounded by a diffuse array of sound images that are not associated with particular source locations. The division between early- and late-arriving sound is taken to be 80 ms.

Acoustical measures have been developed that can be related to the expected amount of ASW or LEV. The (early) lateral energy fraction (LF_E) is measured by comparing the responses from a figure-of-eight and an omnidirectional microphone over the first 80 ms of an impulse response.[7] The $LF_E(125\text{-}1K)$ measures the fraction of the early sound energy arriving from the side but ignores the important effect of sound level on ASW. Similarly, interaural cross correlations (IACC) have been used to indicate the expected ASW,[8] but again this measure ignores the important effect of sound level. The relative late lateral sound level G_{LL} has been shown to relate to judgments of LEV.[9] It measures the level of the late-arriving sound energy arriving from the side and so includes the effects of direction and time of arrival as well as level. Both ASW and LEV are thought to be influenced by low- to midfrequency sound and both $LF_E(125\text{-}1K)$ and $G_{LL}(125\text{-}1K)$ are usually measured over the octave bands from 125 to 1000 Hz.

While most previous subjective experiments have been performed with sounds arriving from sources only in the horizontal plane, Barron and Marshall showed that early-arriving reflections from above the horizontal plane had the same effects (varying with the cosine of the angle to the line through the two ears of the listener) as sounds from sources in the horizontal plane.[7] Recently, Evjen *et al.*[10] published the results of a similar experiment for LEV. They showed that late-arriving sound from above the horizontal plane had the same effect on LEV as does late-arriving sound from sources in the horizontal plane of the listener's ears, i.e., it varies with the cosine of the angle to the line through the two ears of the listener.

The various experiments, which have gradually developed our current understanding of spatial impression, have used increasingly more complex sound fields. The earlier studies in simulated sound fields by Barron mostly used only

[a]Electronic mail: John.Bradley@nrc.ca

a direct sound and two reflections.[3,7] By using more sound sources and including late-arriving sound, Bradley and Soulodre were able to demonstrate the importance of later-arriving sound. However, to date, most experiments have focused on the individual effects of either early-arriving or late-arriving sound on spatial impression. Thus, we know that increased early-arriving lateral sound energy leads to the perception of increased ASW and that increased late-arriving lateral sound energy gives us the sense of increased LEV. However, previous work had not completely explained how early- and late-arriving lateral sound energy combine to influence ASW and LEV. That is, with strong early reflections and a good sense of ASW, what is the effect of adding late lateral sound energy as would occur to various degrees in real concert halls? Similarly, with adequate late-arriving sound energy to create perceived LEV, what is the effect of varied early-arriving lateral reflections?

The present work was intended to take previous subjective experiments one step closer to reality by investigating the combined effects of varied early- and late-arriving lateral sound energy on perceived spatial impression. As in previous studies, it was assumed that the subjectively relevant characteristics of impulse responses can be represented by the separate sums of the early- and late-arriving sound energy, where early includes the direct sound and all reflections arriving up to 80 ms after the direct sound. It was expected that, to some extent, early- and late-arriving sound would mask or diminish the effects of each other. Previous experiments with early lateral reflections had shown that perceived ASW was substantially reduced when late-arriving sound was also included.[9] Thus, in real concert halls where both early- and late-arriving sound are present, perceived ASW is expected to be less than indicated by experiments in simulated sound fields with only early reflections. Similarly, it was hypothesized that increased early-arriving sound would mask or reduce the perceived LEV caused by the late-arriving sound.

The specific goals of this work included first verifying that some previous results concerning spatial impression are valid in more complex sound fields. It was further intended to quantify the influence of early-arriving sound on the relationship between late-arriving sound and LEV. Similarly, it was desired to quantitatively determine the effect of late-arriving sound on the relationship between early lateral energy and ASW. Room acoustics measurements from real halls were used to give an initial estimate of the degree of ASW and LEV that are typically experienced by listeners, as well as the relative importance of ASW and LEV. By extending subjective experiments to conditions closer to those in real halls, it was hoped that these new studies would give us a more complete understanding of spatial impression that would help us to better understand some concert hall design issues.

I. EXPERIMENTAL DETAILS

This paper describes six experiments in which subjects compared either the apparent source width (ASW) or the listener envelopment (LEV) of pairs of simulated sound fields with varied early- and late-arriving sound components. The sound fields were simulated using an eight-channel electroacoustics system in an anechoic room. The eight loudspeakers were located in a circle in the horizontal plane of the listener's ears. The listener sat at the center of this circle 1.8 m from the loudspeakers. The loudspeaker directly in front of the listener simulated the direct sound and all loudspeakers produced early-arriving reflections lower in amplitude than the direct sound and distributed over the early time interval. The music source was clearly localized on the loudspeaker directly in front of the listener. (Subjects were free to turn their heads but would normally focus on the loudspeaker directly in front of them.) If the loudspeaker directly in front of the listener is at 0°, the others were located at ±35°, ±90°, ±125°, and 180° for the LEV experiments. For the ASW experiments the other loudspeakers were located at ±45°, ±90°, ±135°, and 180°, and were again in the horizontal plane of the listener's ears. Although the differences between the two sets of angles may not be significant, they were each selected to exactly repeat those used in particular previous studies.

Each channel reproduced either the direct sound or a simulated reflection and a reverberant decay. This was created by summing the outputs of a digital equalizer with the output of a digital reverberator. The digital reverberators were the Lexicon Reflex model and the digital equalizers were Yamaha DEQ7s, having variable time delay to vary the arrival time of the discrete reflections. Both are MIDI controllable. The early reflections were distributed over the 80-ms early time interval after the arrival of the direct sound. The reverberant decays started approximately 80 ms after the direct sound. The exact start of each reverberant decay was varied by a few ms, and slightly different decay settings were used for each loudspeaker to avoid problems due to correlated decays. The output of each channel was amplified and fed to one of eight Paradigm Compact Monitor loudspeakers. These were equalized to have a flat response (±3 dB from 80 to 12 kHz).

The source signal was an 18-s section of an anechoic recording of Handel's Water Music.[11] The music repeated continuously during the tests. Subjects compared pairs of sound fields by using a small keypad to switch back and forth between the sound fields of each pair without interrupting the music. They could take as much time as they wished and eventually rated the difference in either ASW or LEV of the two sound fields on a 5-point scale. A score of 1 indicated no obvious difference and a score of 5 indicated the largest difference that the listener expected to hear based on their experience in practice sessions. The sound-field pairs were presented to the listener in random order.

Tests were completed by between 6 and 11 subjects depending on the test. They were all adults and varied in age from 20 to 63 years old. They had no known hearing problems and were intended to be representative of a broad range of typical listeners. Before each test, each subject had several practice sessions in which they listened to sound-field pairs that included the most extreme cases from the complete test.

Each sound field was measured using our RAMSOFT room acoustics measurement system. This measures impulse responses using a maximum length sequence signal and a fast Hadamard transform technique. The impulse responses

are filtered into octave bands and various room acoustics parameters including sound levels, decay times, and energy ratios are determined. In this paper, reference is made to the measured relative sound levels of the early-arriving, late-arriving, and total sound levels. The relative total sound level G, is the measured sound level relative to the level of the same source at a distance of 10 m in a free field. G_{80} is similarly the relative level of the direct and early-arriving sound (within 80 ms of the direct sound) and G_L is the relative level of the late-arriving sound arriving more than 80 ms after the arrival of the direct sound. Because spatial impression has been related to the level of sound arriving from the side, early- and late-arriving relative lateral sound levels were also determined. The lateral sound was measured using a figure-of-eight microphone with its null pointed towards the source of the direct sound. G_{EL} is the relative level of the early-arriving lateral sound (again within 80 ms of the arrival of the direct sound) G_{LL} is the relative level of the late-arriving lateral sound. All relative sound levels are relative to the same reference of the level of the same source at a distance of 10 m in a free field. In this paper, these relative sound levels are usually only presented for the sum of the 125- to 1-kHz octaves because this frequency range has been shown to be most important for spatial impression.

II. EFFECTS OF EARLY AND LATE SOUND ON PERCEIVED LEV

A. Interaction of the effects of early and late sound on LEV for realistic conditions

1. Experimental procedure

In the first experiment listeners judged the relative listener envelopment (LEV) of 18 sound fields consisting of combinations of varied early and late sound levels. These were obtained by using six different values of the balance between early- and late-arriving sound [C_{80}(125-1K)] and then repeating these for three different overall sound levels. [C_{80}(125-1K) is ten times the logarithm of the ratio of the sums of the early- and late-arriving sound energies and early is defined as the direct sound plus all reflections arriving within 80 ms after the direct sound.] Both C_{80}(125-1K) and the overall sound levels were varied in 3-dB steps, which is a readily detectable change in each quantity.[12,13] Varying C_{80}(125-1K) values for constant sound level requires that both early and late sound levels vary. Thus, as late-arriving sound levels are reduced to create higher C_{80}(125-1K) values, the early-arriving sound levels must be increased to maintain the same overall level. However, by varying the balance between early- and late-arriving sound for each sound level, the strong effects of overall level are easily separated.

To minimize the number of variables, the time and direction of arrival of early- and late-arriving sounds were kept constant. The relative amplitudes of the direct and early reflections were also not varied. Only the combined level of the direct sound and early reflections was varied. The late energy was distributed equally among all eight loudspeaker sources. Its distribution was kept constant and only the overall level of all late-arriving sound energy was varied.

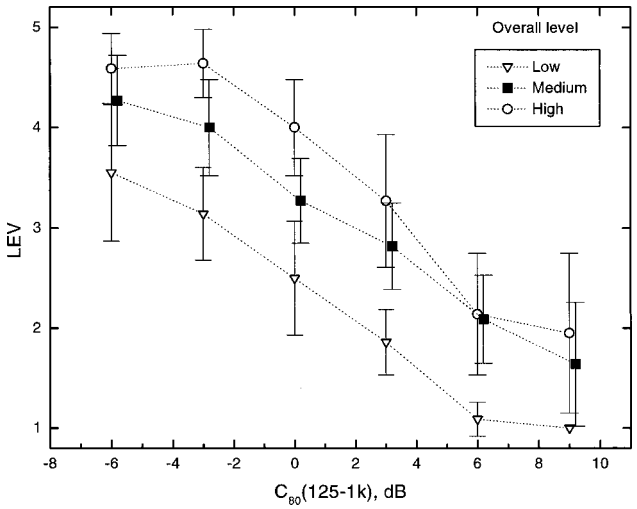


FIG. 1. Mean LEV scores with error bars showing 95%-confidence limits versus the early-to-late ratio, C_{80}(125-1K), for three overall sound levels, 65, 68, and 71 dBA.

The C_{80}(125−1K) values varied from −6 to +9 dB in 3-dB steps. [C_{80}(125-1K) values were averaged over the four octave bands from 125 to 1 kHz because these frequencies have been shown to be important for spatial impression.] The overall A-weighted levels of the music were 65, 68, and 71 dBA when integrated over the complete 18-s music sample. Subjects rated the LEV of each sound field relative to the sound field with the least LEV [i.e., the sound field with the lowest sound level and the highest C_{80}(125-1K)]. Eleven subjects each completed the test twice.

2. Results

The mean LEV scores and 95%-confidence limits are plotted versus measured C_{80}(125-1K) values in Fig. 1 for the three different overall sound levels. LEV scores clearly decrease with increasing C_{80}(125−1K) and with decreasing overall level. Increasing C_{80}(125-1K), for constant level, corresponds to decreasing late-arriving sound level. Thus, these results indicate that the perceived listener envelopment (LEV) decreases with decreasing late level and decreasing overall level, as has previously been shown for less complex sound fields.[6] An analysis of variance test of these results indicated that there were highly significant ($p<0.001$) main effects of both C_{80}(125-1K) and overall level but that the interaction effect was not significant.

The relative sound levels of the total early- and late-arriving sound were measured for each sound field as well as the early and late lateral sound levels. In this experiment the spatial distribution of early and late sound was not varied, and hence the results relate directly to changes in the total early- and late-arriving sound levels as measured with an omnidirectional microphone.

The basic underlying effects are better understood in terms of early- and late-arriving sound levels, and therefore Fig. 2 plots the same mean LEV scores versus measured late-arriving relative sound levels, G_L(125-1K), for the three different overall level cases. Thus, as one moves to the right for each of the three data sets, the level of late-arriving sound is increasing. As one moves vertically upwards from

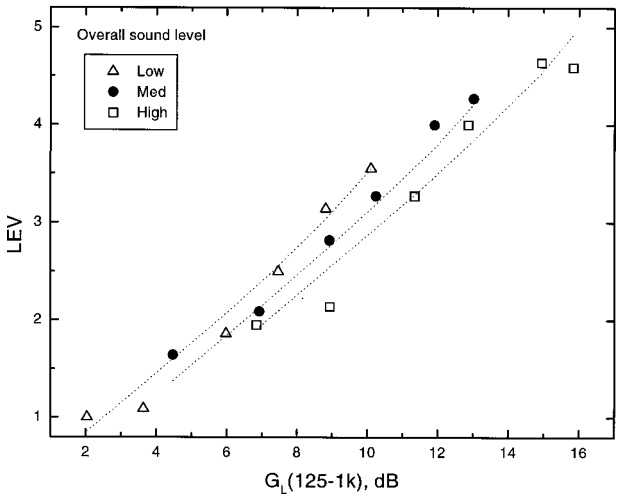


FIG. 2. Mean LEV scores as a function of the measured late-arriving sound levels, $G_L(125\text{-}1K)$, for three different overall levels, 65, 68, and 71 dBA. The dotted lines were obtained using Eq. (1).

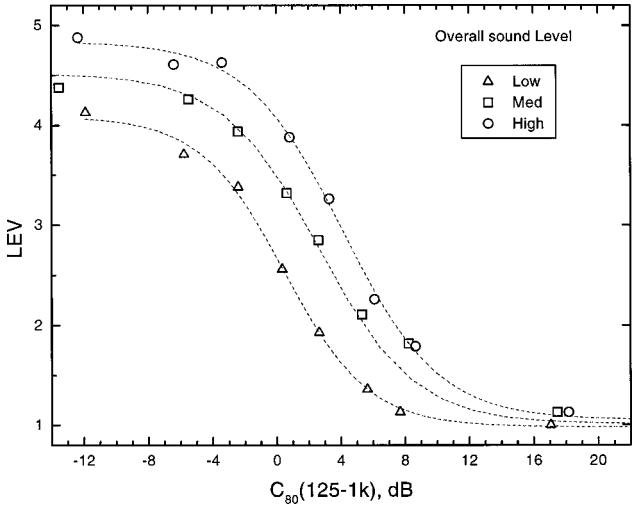


FIG. 3. Mean LEV scores for an extended range of the early-to-late ratio, $C_{80}(125\text{-}1K)$, and for three overall sound levels, 65, 68, and 71 dBA.

one data set to the next, the overall level is decreasing. Since moving vertically on the graph corresponds to constant late-arriving sound level, the decrease in overall level is due to a decrease in early level. Thus, the combined effects of early- and late-arriving sound on perceived LEV are illustrated in Fig. 2. Increasing late-arriving sound energy leads to increased LEV. Increasing early-arriving sound energy tends to decrease perceived LEV.

A multiple linear regression analysis of the combined effects of early and late sound levels on mean LEV scores produced the following result:

$$\mathrm{LEV} = 0.296\, G_L(125\text{-}1K) - 0.06181\, G_{80}(125\text{-}1K) + 0.818, \quad (1)$$

where $G_L(125\text{-}1K)$ is the late-arriving relative sound level, $G_{80}(125\text{-}1K)$ is the early-arriving relative sound level, and the related R^2 value was 0.973.

A similar multiple linear regression analysis in terms of early and late lateral relative sound levels produced the following equation:

$$\mathrm{LEV} = 0.275\, G_{LL}(125\text{-}1K) - 0.100\, G_{EL}(125\text{-}1K) + 2.406, \quad (2)$$

where $G_{LL}(125\text{-}1K)$ is the late-arriving relative lateral sound level and $G_{EL}(125\text{-}1K)$ is the early-arriving relative lateral sound level. For Eq. (2) the related R^2 value was 0.955. In both regression analyses both predictors were highly significantly related to LEV scores ($p<0.001$).

These results, and in particular the regression equations derived from them, clearly show how early- and late-arriving sound combine to influence our perceptions of LEV. While LEV does increase with increasing late-arriving (lateral) sound, in agreement with previous studies, increased early-arriving sound can diminish the perceived LEV. It may therefore be possible to have too much early-arriving sound and hence have conditions in which a high degree of envelopment is not possible.

B. Extended range effects of early and late sound on LEV

The results in Fig. 1 indicate an approximately linear relationship between perceived listener envelopment (LEV) and $C_{80}(125\text{-}1K)$ values over most of the range included in this experiment. This range was chosen to include the complete range of $C_{80}(125\text{-}1K)$ values likely to be encountered in actual concert halls. The results suggest that this linear trend may not extend to the extremes of this range. To more fully understand the effects of early and late sound levels on LEV, a second experiment was carried out that included an extended range of $C_{80}(125\text{-}1K)$ values. While the added more-extreme cases are not likely to be found in concert halls, they allow us to better understand our perceptions of envelopment when either early- or late-arriving sound is very dominant over the other.

1. Experimental procedure

In this experiment, subjects rated the relative LEV of 21 sound fields. These consisted of seven levels of $C_{80}(125\text{-}1K)$ varying from -12 to $+18$ dB and the same three overall sound levels as in the previous experiment. Again, the sound field with the highest $C_{80}(125\text{-}1K)$ and the lowest sound level (i.e., with the lowest LEV) was the reference sound field that subjects compared with each of the other sound fields. As in the first experiment, the time and direction of arrival of early- and late-arriving sound were kept constant.

2. Results

The mean LEV scores are plotted versus measured $C_{80}(125\text{-}1K)$ values for the three different overall levels in Fig. 3. Over the range of $C_{80}(125\text{-}1K)$ values from -6 to $+9$ dB, the results are very similar to the previous experiment shown in Fig. 1. For $C_{80}(125\text{-}1K)$ values between -4 and $+6$ dB, there is again an approximately linear relationship with LEV responses. However, there are now different trends at either extreme of the $C_{80}(125\text{-}1K)$ scale. For very high $C_{80}(125\text{-}1K)$ values, where the early-arriving sound

very much predominates, LEV scores approach a value of 1, indicating little sense of envelopment regardless of the overall level. For very low $C_{80}(125\text{-}1K)$ values the LEV scores seem to reach a different plateau for each overall sound level case. The very low $C_{80}(125\text{-}1K)$ cases correspond to situations where the late energy is dominant, and hence the overall level is largely due to the level of the late-arriving sound. Thus, there are three different plateaus for the LEV scores at very low $C_{80}(125\text{-}1K)$ values because there are three different late-arriving sound levels.

These results show that the tradeoff between early- and late-arriving sound on LEV scores, identified in the first experiment, is limited to the range where both have perceptible contributions. Where one component is 10 or 12 dB greater than the other, only the dominant component influences judgments. These results also suggest that in larger halls where sound levels would naturally tend to be lower, maintaining adequate LEV would require greater relative amounts of late-arriving sound energy. This may be one of the reasons that preferred reverberation times are usually required to be longer in larger halls.

C. Effect of varied spatial distribution of late sound on LEV

In the first two experiments, LEV was varied by varying the level of the total late-arriving sound. The late-arriving sound was always equally distributed among the eight loudspeakers and so arrived equally from all directions. It is known that perceived listener envelopment (LEV) varies with the angle of arrival of the late-arriving sound.[6] The third experiment was intended to examine the combined effects of levels and the spatial distribution of late-arriving sound on perceived LEV.

1. Experimental procedure

In this experiment, subjects listened to seven different distributions of the late-arriving sound which were repeated for three different overall sound levels. The three levels corresponded to the levels of the music being 65, 68, and 71 dBA when integrated over the 18-s sample of music. The portion of the late energy that was arriving from the side was lowest when all of the late energy arrived from only the front and rear loudspeakers which were directly in front and directly behind the listener. (The contributions to the late energy of the front and rear loudspeakers were equal.) As there was then no late-arriving energy from the side, these were expected to lead to judgments of minimal LEV. Late lateral sound levels were systematically increased by having a greater portion of the late lateral sound arrive from the other six loudspeakers on either side of the listener. However, the total late-arriving sound (and the total early-arriving sound) was kept constant for the seven sound fields of each overall sound level group. When the overall sound level was increased, the early and late sound levels were increased by similar amounts, and the same variation in the spatial distribution of late-arriving sound was maintained. The $C_{80}(125\text{-}1K)$ values of all 21 sound fields were all -0.5 dB. A total of seven subjects each performed the test twice.

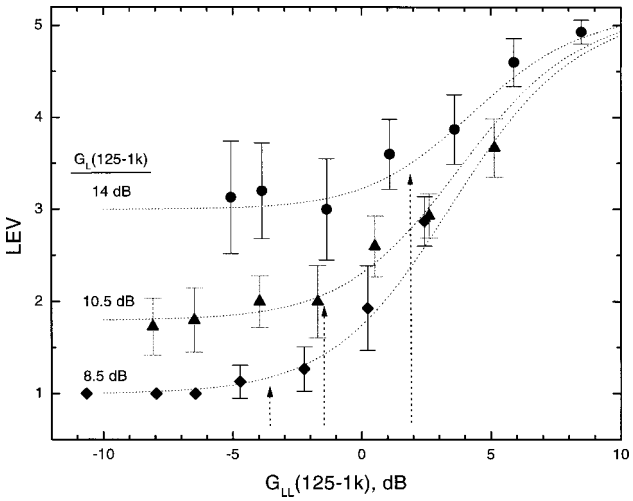


FIG. 4. Mean LEV scores versus late lateral sound level, $G_{LL}(125\text{-}1K)$ for three overall sound-level cases. Each curve is labeled according to its late sound level $G_L(125\text{-}1K)$ and the vertical arrows indicate where the $G_{LL}(125\text{-}1K)$ value is 12 dB less than the $G_L(125\text{-}1K)$ value.

2. Results

Figure 4 shows the mean LEV scores and 95%-confidence limits as a function of the measured relative late lateral sound levels $G_{LL}(125\text{-}1K)$. The perceived LEV increases with increasing $G_{LL}(125\text{-}1K)$ on the right-hand side of the plot, but on the left side of the plot LEV values seem to reach a different plateau for each overall sound level case. There is some lower $G_{LL}(125\text{-}1K)$ value below which LEV does not vary. This residual level of LEV depends on the overall sound level.

Although this was not expected, it is in accord with previously unexplained results by Evjen *et al.*[10] Since all sound fields have the same $C_{80}(125\text{-}1K)$ value, changing the overall sound level does not influence the balance between early- and late-arriving sound or the interacting effects of early-and late-arriving sounds observed in the first two experiments. Increasing the overall sound level does increase the total late-arriving sound level. The left-most data points of each of three sound-level cases in Fig. 4 correspond to late-arriving sound from only directly in front of and behind the listener. It appears therefore, that with only late-arriving sound from the front and rear, LEV can still vary with the level of the late-arriving sound. A similar effect can be seen in the results of Evjen *et al.*[10] Unfortunately, it is not possible to determine from these results whether late-arriving sound from the front and rear is equally important for LEV as they were not separately varied.

Two different approaches give essentially equivalent explanations of the results of this experiment. In both cases the results can be assumed to indicate that perceived LEV is determined by the combination of the total late-arriving sound and that portion of the late-arriving sound that reaches the listener from the side. One approach is to assume that when the late-arriving energy from the side is very much below the level of other late-arriving sound energy, it will not be perceptible and hence cannot further influence LEV. Thus, in Fig. 4, LEV scores do not increase with increasing $G_{LL}(125\text{-}1K)$ when late lateral energy $[G_{LL}(125\text{-}1K)]$ is

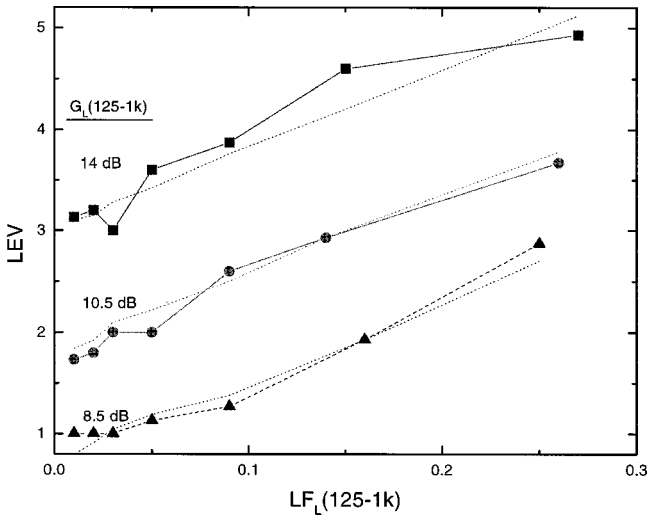


FIG. 5. Mean LEV scores versus the late lateral energy fraction, $\mathrm{LF}_L(125\text{-}1K)$, for three different overall sound levels. The dotted lines were calculated using Eq. (4).

much more than 10 to 12 dB below the total late-arriving sound energy $[G_L(125\text{-}1K)]$. The vertical arrows in Fig. 4 indicate where the late lateral energy is about 12 dB below the total late energy and confirm that this explanation is a reasonable approximation to the measured results. To the left of these arrows, a negligible portion of the late-arriving sound comes from lateral directions, but some LEV is still experienced.

An alternative approach is to consider the results of Fig. 4 to be due to the combined effects of the total late-arriving sound and the fraction of the total late energy that arrives from the side. This fraction can be described by the late lateral energy fraction and can be calculated as

$$\mathrm{LF}_L(125\text{-}1K) = 10^{(G_{LL}(125\text{-}1K) - G_L(125\text{-}1K))/10}. \tag{3}$$

One might expect that in addition to the total late-arriving sound energy, above some minimum $\mathrm{LF}_L(125\text{-}1K)$ value LEV scores would increase with increasing $\mathrm{LF}_L(125\text{-}1K)$ value. The measured LEV scores are plotted versus $\mathrm{LF}_L(125\text{-}1K)$ values for each of the three different sound-level cases in Fig. 5. A multiple linear regression analysis was performed relating LEV scores to a combination of $G_L(125\text{-}1K)$ and $\mathrm{LF}_L(125\text{-}1K)$ values. The resulting highly significant relationship ($p<0.001$, $R^2=0.98$) produced the following regression equation:

$$\mathrm{LEV} = 0.397 G_L(125\text{-}1K) + 7.4 \mathrm{LF}_L(125\text{-}1K) - 2.530. \tag{4}$$

The dotted lines shown in Fig. 5 are based on this equation. This figure indicates that Eq. (4) is a reasonable fit to the measured LEV scores and can be used to estimate expected LEV. The slight irregularities in the calculated lines are because the actual measured $G_L(125\text{-}1K)$ values were used in calculating LEV values with this equation and, although intended to be constant, they varied by small fractions of a decibel.

When considered in this way, LEV scores seem to increase with increasing late lateral energy when $\mathrm{LF}_L(125\text{-}1K)$ values exceed about 0.05. However, this is approximately equivalent to the points marked in Fig. 4 where the late lateral energy was 12 dB below the total late-arriving sound energy. Thus, both explanations are completely equivalent.

An alternative suggestion is that for very low LEV, subjects may confuse ASW with LEV. The three different overall sound-level cases would lead to three corresponding levels of early arriving sound that would be perceived as three different levels of ASW, and hence might explain the three plateaus of LEV in Fig. 4. However, previous experiments[6] have shown that subjects can differentiate between LEV and ASW and the results of Fig. 3 can also be used to refute this suggestion. In the results of the previous experiment illustrated in Fig. 3, perceived LEV does decrease to zero as the level of the late-arriving sound energy is decreased, even though there is substantial early-arriving sound energy and presumably a sense of ASW[$\mathrm{LF}_E(125\text{-}1K)=0.20$]. There was apparently no confusion of ASW and LEV in that experiment. The difference in this third experiment (Fig. 4) is that total late-arriving sound energy is not decreased but its spatial distribution is modified. Thus, the suggestion that the plateaus in Fig. 4 are due to subjects confusing ASW and LEV is not supported.

The results of this experiment show that even when the ratio of total early- to total late-arriving sound is kept constant, the perceived LEV varies with both the level of the total late energy and the portion of that late-arriving energy that arrives at the listener from the side. It is also seen that in the absence of late lateral energy, nonlateral late-arriving sound can create a sense of envelopment.

III. EFFECTS OF EARLY AND LATE SOUND ON PERCEIVED ASW

A previous experiment by Bradley and Soulodre[6] showed that the magnitude of perceived ASW (apparent source width) was considerably reduced when late-arriving sound was added to the direct sound and early reflections. Although this was not extensively pursued, it was evident that both early-and late-arriving sound influenced perceived ASW, and that while early-arriving lateral reflections tended to increase perceived ASW, later-arriving lateral reflections tended to diminish perceived ASW. This previous work had been successful in extending the complexity of earlier experiments but was itself limited because it used simulated sound fields with only three loudspeaker sources.

The new experiments reported here were intended to extend this previous work to more complex sound fields and were planned to quantify the combined effects of early- and late-arriving sound on judgments of ASW. This was done in three steps. First, the previous experiment was replicated using the same experimental setup as in the original experiment to confirm the previous results and to be sure of the validity of the new work. Then, this was extended to more complex sound fields with eight different sound sources. As in the experiments reported in the previous sections of this paper, complete room acoustics measurements were made of each condition to enable the derivation of quantitative relationships between room acoustics measures and mean ASW scores.

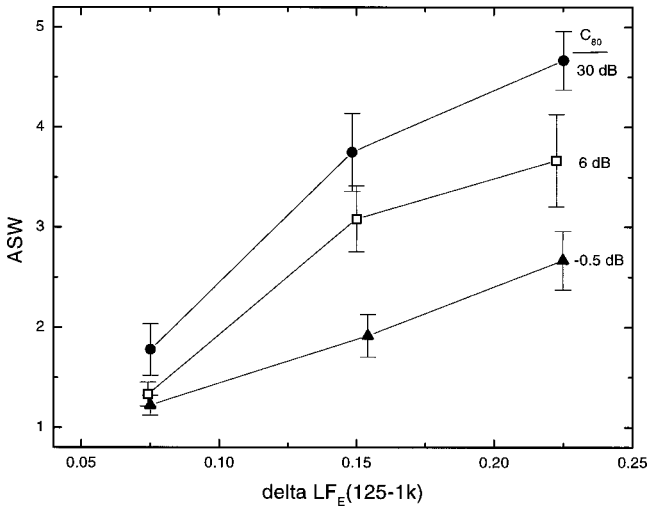


FIG. 6. Mean ASW scores versus differences in the early lateral energy fraction, $LF_E(125\text{-}1K)$ for sound fields created with only three loudspeakers.

A. Replication of previous results with three sources

To replicate the previous experiment,[6] simulated sound fields were created using only three loudspeaker sources. One was directly in front of the listener and the other two were at angles of ±45 deg relative to straight ahead of the listener. Twelve different sound fields were created consisting of the combination of four different values of the early lateral energy fraction, $LF_E(125\text{-}1K)$, and three levels of $C_{80}(125\text{-}1K)$. $C_{80}(125\text{-}1K)$ values, averaged over the octave bands from 125 to 1 kHz, were approximately −0.5, +6, and +30 dB. The −0.5-dB case would correspond to a quite reverberant concert hall, while the +6-dB $C_{80}(125\text{-}1K)$ value would relate to a hall with quite high clarity. The very high +30-dB $C_{80}(125\text{-}1K)$ case corresponded to simulated sound fields with no added late-arriving sound energy. To replicate the previous experiment, subjects rated the differences in apparent source width (ASW) for all combinations of the four different sound fields for each $C_{80}(125\text{-}1K)$ value. That is, they compared 6 pairs of sound fields for each of 3 $C_{80}(125\text{-}1K)$ values for a total of 18 comparisons. Six subjects completed the test on two different occasions.

The mean ASW scores and 95%-confidence limits are shown in Fig. 6. The ASW scores are plotted against the difference in $LF_E(125\text{-}1K)$ between each sound-field pair. There are disadvantages to this approach, which are discussed below, but it was desired that this experiment replicate the previously published work as closely as possible. The results in Fig. 6 do show the same trends as the earlier work. Perceived ASW increases with increasing changes in the early lateral energy fraction, $LF_E(125\text{-}1K)$. When there was no late-arriving sound $[C_{80}(125\text{-}1K) = +30\,\text{dB}]$, the variation of ASW with differences in $LF_E(125\text{-}1K)$ was greatest. As late-arriving energy was added, the perceived ASW for a particular difference in $LF_E(125\text{-}1K)$ decreased. An analysis of variance test of these results showed highly significant ($p<0.001$) main effects of both $C_{80}(125\text{-}1K)$ and delta $LF_E(125\text{-}1K)$. There was also a highly significant ($p<0.001$) interaction effect of $C_{80}(125\text{-}1K)$ and delta $LF_E(125\text{-}1K)$ on the ASW scores.

TABLE I. Equalization of direct, early- and late-arriving sound components relative to the 500- and 1000-Hz midfrequency values.

	Attenuation, dB			
	2 kHz	4 kHz	8 kHz	6 kHz
Direct	0	0	−1.6	−2
Early	−1	−3	−9	−13
Late	−2	−4	−10	−14

Preliminary tests showed the importance of an appropriate equalization of the spectra of the simulated early- and late-arriving sound to produce realistic simulations. Because music recordings are usually made with relatively close microphone locations, they do not represent the frequency response of early- and late-arriving sounds at typical seats in concert halls. The attenuations relative to midfrequency (500 and 1000 Hz) values of the higher frequencies that were used are given in Table I. The direct sound, early reflections, and later-arriving sound were attenuated by different amounts. The attenuations in the 2- and 4-kHz octave bands were selected to be representative of measured results in concert halls. However, such measurements are not available for the 8- and 16-kHz octave bands, and these values were estimated to be representative of conditions in concert halls. Achieving realistic values of these equalizations is critical to simulating realistic spatial impression. Spatial impression, including ASW, is related to lower frequency sounds in the octave bands from 125 to 1 kHz. As has been demonstrated by Hartmann *et al.*,[14] higher-frequency sounds best enable us to localize sound sources. In some respects spatial impression is the inverse of localization and thus too much high-frequency sound creates the perception of new sound sources rather than the broadening of the central source. Such additional sound sources could confuse listeners and be interpreted as something like envelopment and would not be a realistic simulation of a typical concert hall, where air and porous materials tend to absorb higher-frequency sounds.

B. Extension to eight sources

The second ASW experiment extended the previous experiment to include simulated sound fields with eight loudspeaker sound sources. These were located in a circle around the listener in the horizontal plane of the listener's ears as in the LEV experiments. As before there were loudspeakers directly in front (0°) and directly behind the listener (180°) as well as on either side (±90°). However, the other four loudspeakers were located at ±45° from in front of and behind the listener to be more similar to the previous ASW experiment. The experimental procedure was the same as in the first ASW experiment described in the previous section. Subjects rated the differences in ASW of all 6 combinations of the sound fields for each of the 3 different $C_{80}(125\text{-}1K)$ values to give a total of 18 comparisons. Six subjects completed the test on two different occasions. The average level of the music test samples was 71 dBA.

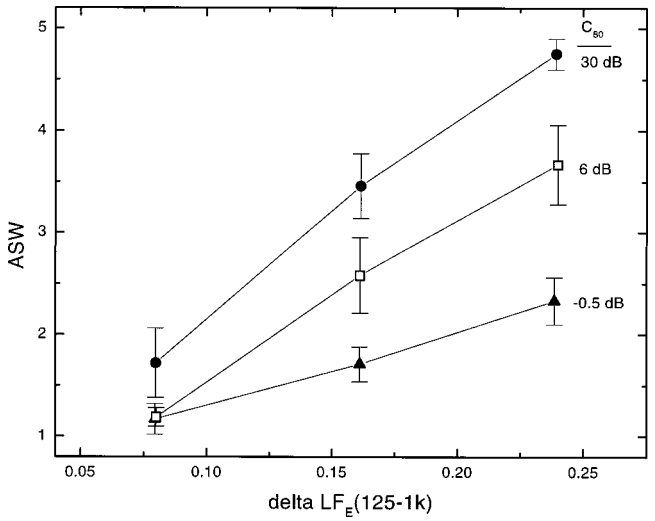


FIG. 7. Mean ASW scores versus differences in the early lateral energy fraction, $LF_E(125\text{-}1K)$ for sound fields created with eight loudspeakers.

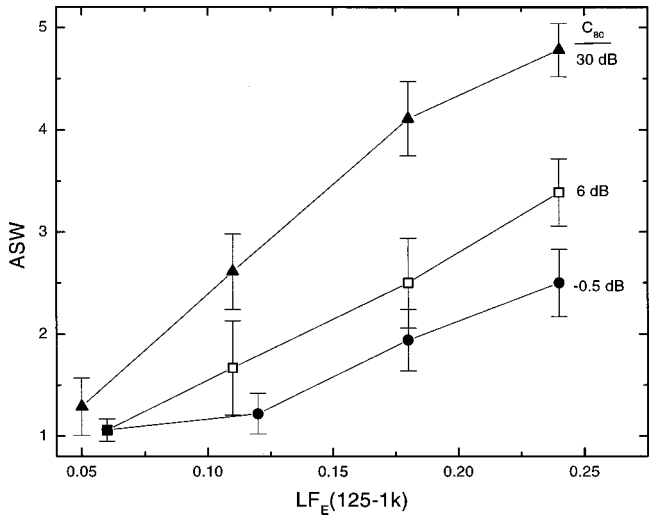


FIG. 8. Mean ASW scores versus measured early lateral energy fractions, $LF_E(125\text{-}1K)$ for sound fields created with eight loudspeakers.

The mean ASW scores and 95%-confidence limits are plotted versus the differences in early lateral fractions, delta $LF_E(125\text{-}1K)$, in Fig. 7. The results show a very similar trend to the previous experiment where there were only three loudspeaker sound sources. Perceived ASW increased with increasing delta $LF_E(125\text{-}1K)$ values, but the effect was markedly diminished as more late-arriving sound was added. An analysis of variance of these results showed highly significant ($p<0.001$) main effects of both $C_{80}(125\text{-}1K)$ and delta $LF_E(125\text{-}1K)$ as well as a highly significant ($p<0.001$) interaction effect of $C_{80}(125\text{-}1K)$ and delta $LF_E(125\text{-}1K)$ values on the ASW scores. Thus, the observed combined effect of early-and late-arriving sound on perceived ASW remains when the complexity of the simulated sound fields is increased from three to eight loudspeaker sources.

By plotting the results versus the difference in early lateral fraction for each sound-field pair, delta $LF_E(125\text{-}1K)$, one is assuming that a particular delta $LE_E(125\text{-}1K)$ will have the same effect no matter what the actual $LF_E(125\text{-}1K)$ values. That is, in Figs. 6 and 7, a 0.075-delta $LF_E(125\text{-}1K)$ is assumed to have the same subjective effect whether it is between $LF_E(125\text{-}1K)$ values of 0.05 and 0.125 or between 0.2 and 0.275. This may not be exactly correct and one may be averaging out the effects of the magnitude of the $LF_E(125\text{-}1K)$ values. It also does not directly illustrate the effect of early and late sound levels on perceived ASW. The final ASW experiment described below was designed to relate perceived ASW directly to the levels of the early- and late-arriving lateral sound energy.

C. Interaction of the effects of early and late sound on ASW

In the final ASW experiment, listeners judged the ASW of 12 sound fields consisting of the combinations of 4 values of $LF_E(125\text{-}1K)$ and 3 values of $C_{80}(125\text{-}1K)$. However, the sound fields for each $C_{80}(125\text{-}1K)$ value were compared to a reference sound field with the same $C_{80}(125\text{-}1K)$ and a very low $LF_E(125\text{-}1K)$ value of only 0.02. Thus, subjects rated the effect of the actual magnitude of the $LF_E(125\text{-}1K)$ of the sound field. Complete room acoustics measurements of each sound field were made, including the relative early- and late-arriving lateral sound levels $G_{EL}(125\text{-}1K)$ and $G_{LL}(125\text{-}1K)$. Nine subjects completed this test on two different occasions. The average sound level of the music samples was 72 dBA.

The mean ASW scores and 95%-confidence limits are plotted versus the measured $LF_E(125\text{-}1K)$ values for each sound field in Fig. 8. The trends are similar to the previous ASW experiments, and an analysis of variance test indicated highly significant ($p<0.001$) main effects of both $C_{80}(125\text{-}1K)$ and $LF_E(125\text{-}1K)$ as well as a highly significant ($p<0.001$) interaction effect of these two predictor variables.

When the same ASW scores were plotted versus the measured early lateral relative sound levels, $G_{EL}(125\text{-}1K)$, they indicated that perceived ASW increased approximately linearly with increasing $G_{EL}(125\text{-}1K)$ above a $G_{EL}(125\text{-}1K)$ value of about -2 dB. Below this value there was only a very minimal effect of early lateral sound on perceived ASW. This point corresponded to the level of the early lateral sound level $[G_{EL}(125\text{-}1K)]$, being about 10 dB below the level of the total sound $[G(125\text{-}1K)]$. This indicates that when the level of the early lateral sound is very small compared to the total sound level, it does not influence a listener's perception of ASW.

In the region of $G_{EL}(125\text{-}1K)$ values greater than -2 dB, it was possible to perform a multiple linear regression analysis to relate ASW scores to both $G_{EL}(125\text{-}1K)$ and $G_{LL}(125\text{-}1K)$ values. The resulting equation for the eight data points where $G_{EL}(125\text{-}1K)>-2$ dB was

$$\mathrm{ASW}=0.484G_{EL}(125\text{-}1K) - 0.0715G_{LL}(125\text{-}1K)+2.191. \quad (5)$$

Both predictors contributed at a highly significant level ($p<0.001$) and the associated R^2 value was 0.981.

The ratio of the coefficients of $G_{EL}(125\text{-}1K)$ and $G_{LL}(125\text{-}1K)$ in the above equation is 0.15. A new variable,

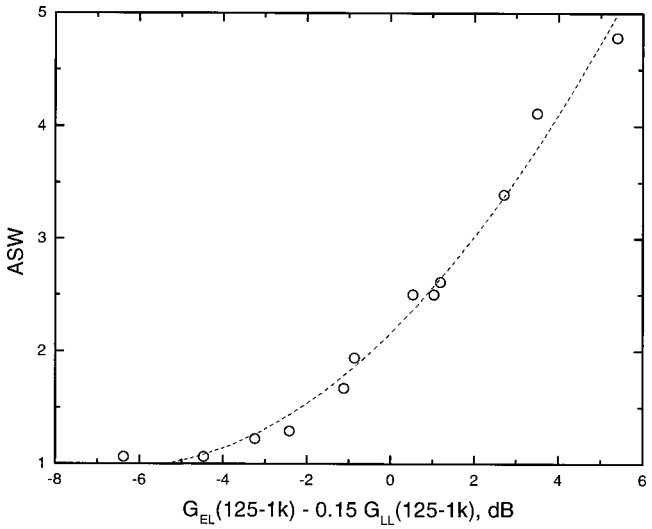


FIG. 9. Mean ASW scores versus the difference $G_{EL}(125\text{-}1K) - 0.15G_{LL}(125\text{-}1K)$. The dotted line indicates the best fit line of Eq. (6).

$X = (G_{EL}(125\text{-}1K) - 0.15G_{LL}(125\text{-}1K))$, was created and a second-order polynomial was used to relate this variable to ASW scores for all 12 data points from this experiment. The resulting equation was

$$\text{ASW} = 0.0285X^2 + 0.370X + 2.16. \quad (6)$$

The measured data and this regression equation are given in Fig. 9. The equation is a good fit ($R^2 = 0.987$, $p < 0.001$) to the measured data over the complete range of $G_{EL}(125\text{-}1K)$ and $G_{EL}(125\text{-}1K)$ values and can be used to predict the expected degree of ASW. The intermediate variable X is proposed merely as a convenient method of obtaining a good fit to the data of this particular experiment.

IV. APPLICATION OF THE RESULTS TO CONCERT HALLS

Both the experiments with perceived listener envelopment (LEV) and perceived apparent source width (ASW) indicate that early and late lateral sound components have conflicting influences. It is clearly of interest to determine how early- and late-arriving lateral sound combine to produce LEV and ASW in real concert halls. Is there a tendency for one to dominate over the other, and hence for either LEV or ASW to be more frequently experienced in concert halls? For example, an excess of early lateral sound might lead to strong perceptions of ASW but little sense of LEV. Conversely, too much late-arriving sound might create a strong sense of LEV but might mask the effects of the early lateral energy on ASW. What is the expected magnitude of LEV and ASW in actual concert halls?

Data from 16 North American halls were available to consider these questions. The names of the halls and their volumes are listed in Table II. While some data from other halls are available, measured values of late lateral sound levels are quite scarce. However, the data from these 16 halls are representative of a wide range of conditions and they include a range of hall sizes and shapes. Hall average $G_{LL}(125\text{-}1K)$ values varied from -6.0 to $+6.0$ dB as has previously been reported.[15] The 16 halls had average

TABLE II. The 16 halls, their volumes, and the calculated LEV and ASW values.

Hall	Volume, m^3	LEV	ASW
Mechanics Hall, Worcester	10 760	3.5	1.8
Troy Music Hall	11 320	3.9	2.3
Western Manitoba Centennial Auditorium, Brandon	12 390	2.8	1.5
Massey Hall, Toronto	14 190	3.1	1.6
Academy of Music, Philadelphia	15 700	1.7	1.1
Severance Hall, Cleveland	15 700	2.6	1.4
Symphony Hall, Detroit	15 700	3.0	1.5
Queen Elizabeth Theatre, Vancouver	16 760	1.5	1.0
Kleinhans Music Hall, Buffalo	18 220	2.5	1.1
Boston Symphony Hall	18 740	3.1	1.4
Kennedy Center Concert Hall, Washington	19 300	2.7	1.3
Thomas Hall, Akron	19 800	2.3	1.0
Myerhoff, Baltimore	21 500	3.3	1.1
Orpheum Theatre, Vancouver	22 000	2.6	1.4
Boettcher Hall, Denver	37 237	2.9	1.1
Tanglewood Music Shed, Lennox	42 450	2.7	1.2

$\text{LF}_E(125\text{-}1K)$ values varying from 0.07 to 0.23 compared to the value of 0.20 used in the simulations of the LEV experiments in the current research, and similar to the range of used in the ASW experiments as illustrated in Fig. 8. Average $\text{LF}_L(125\text{-}1K)$ values in the 16 halls varied from 0.17 to 0.34, which does not include the very low values that were included in the present simulation studies as indicated in Fig. 5.

For the average measured $G_{EL}(125\text{-}1K)$ and $G_{LL}(125\text{-}1K)$ values from each hall, the expected average ASW and LEV were calculated and are also listed in Table II. Expected LEV values were calculated using Eq. (2). Expected ASW values were calculated using Eq. (6). The average of the calculated ASW values for the 16 halls was 1.30 with a standard deviation of ± 0.33. The average LEV value was 2.66 with a standard deviation of ± 0.60. There was an approximate trend for halls with larger ASW values to also have larger LEV values, as illustrated in Fig. 10. These values would correspond to points on the 5-point rating scales used in the subjective experiments for ASW and LEV. A scale value of 1 indicated no obvious effect and a value of 5 indicated the maximum difference in either ASW or LEV.

These results suggest that the sense of LEV will be stronger than the sense of ASW in these halls. The average estimated LEV is close to the midpoint of the 5-point LEV scale. The average estimated ASW score is only a little above the low end of the scale and this would suggest that there will not be much sense of ASW in these halls. Similarly, the range of LEV values among the halls is larger than the range of ASW values. However, the fact that halls with larger LEV values also have larger ASW values does indicate that a hall can give a strong sense of both LEV and ASW. The two effects are not mutually exclusive. Further

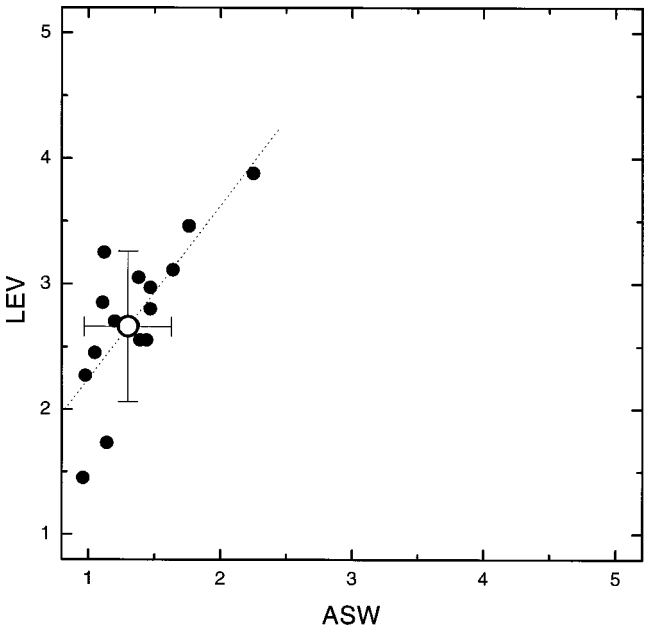


FIG. 10. Estimated LEV versus estimated ASW for 16 halls. LEV values were estimated using Eq. (2) and ASW values with Eq. (6). The large open circle symbol and error bars indicate the overall mean LEV and ASW of the 16 halls and the standard deviations of these values.

studies are required to determine whether the relative importance of ASW and LEV suggested by the results of Fig. 10 is more generally applicable. This first estimate is, of course, limited because it is derived from subjective studies in a very limited number of simulated sound fields and tested in terms of measurements in only 16 halls.

V. CONCLUSIONS

The magnitudes of both components of spatial impression, ASW and LEV, are determined by the particular combination of early-and late-arriving sound energy. Early-and late-arriving sound energy have opposing effects on perceived spatial impression. ASW increases with increasing early lateral sound levels, $G_{EL}(125\text{-}1K)$, but is decreased when late-arriving lateral sound is added. The ratio of the coefficients of $G_{EL}(125\text{-}1K)$ and $G_{LL}(125\text{-}1K)$ in Eq. (5) is $0.484/0.0715=6.77$. Thus, a 6.77-dB increase in $G_{LL}(125\text{-}1K)$ would be required to completely cancel the effect of a 1-dB increase in $G_{EL}(125\text{-}1K)$ on perceived ASW. LEV increases with increasing late-arriving sound energy, but is decreased when early-arriving sound is added. The ratio of the coefficients for $G_{LL}(125\text{-}1K)$ and $G_{EL}(125\text{-}1K)$ in Eq. (2) is $0.275/0.100=2.75$. Thus, a 2.75-dB increase in $G_{EL}(125\text{-}1K)$ would be required to completely cancel the effect of a 1-dB increase in $G_{LL}(125\text{-}1K)$ on perceived LEV. By comparing these two results, it is seen that early-arriving sound is a more effective masker of late sound for LEV than is late-arriving sound for ASW.

Although it is generally accepted that lateral energy is required to create a sense of source broadening, these new results show that this is not completely true for envelopment. In these experiments, late-arriving sound arriving from directly in front of and directly behind the listener were perceived as enveloping, and LEV increased with the level of these late-arriving nonlateral sounds. This may explain the early studies of spatial impression that focused on the importance of reverberant sound.[1] It has been shown that all reverberant or late-arriving sound can influence LEV, but that late-arriving sound from the side of the listener is more important for creating a strong sense of LEV.

In these experiments various effects only occurred when the levels of the relevant sound component were high enough to be detectable in the presence of other sound components. Typically, it was observed that when the critical component was 10 to 12 dB or more below the level of other components, then the effect of the critical component was not perceived. For example, LEV increases with the level of the late sound as long as it is no more than about 10 dB less than the level of the early-arriving sound. Similarly, LEV increases with increasing late lateral level, $G_{LL}(125\text{-}1K)$, when $G_{LL}(125\text{-}1K)$ is above a level of about 12 dB below the total late sound level. It was also observed that perceived ASW increased with increasing early lateral level, $G_{EL}(125\text{-}1K)$, when the $G_{EL}(125\text{-}1K)$ values were greater than about 10 dB below the total sound level.

However, over a wide range of conditions likely to be found in concert halls both ASW and LEV can coexist. It is only under quite extreme conditions that only one or the other of ASW and LEV is possible. These extreme conditions are not likely to be found in typical concert halls. The balance between early-and late-arriving lateral sound will determine the balance between LEV and ASW.

Calculations of expected ASW and LEV in 16 halls suggest that while substantial values of LEV occur, the estimated ASW is often relatively negligible. Both LEV and ASW would vary with the level of the music, but the relative importance would remain the same. If these halls are approximately representative of other halls, then one is more likely to experience envelopment than source broadening. The data from the 16 halls also indicated a trend for halls with higher LEV values to also have higher ASW values. The highest values of both quantities are found in more reverberant smaller halls that would tend to have both strong early and late lateral sound energy. The lowest LEV will occur in larger halls with relatively strong early energy and high clarity $[C_{80}(125\text{-}1K)]$. These larger halls will tend to have lower overall levels but relatively strong early sound leading to low LEV. This provides an argument for requiring longer reverberation times in larger halls to maintain an adequate level of reverberant (late-arriving) sound to achieve good LEV.

The acoustical design of concert halls is a complex balance among many important quantities. It is hoped that these new results have helped to improve our understanding concerning the balance between early- and late-arriving lateral sound energy and its effects on spatial impression.

ACKNOWLEDGMENTS

The authors would like to acknowledge the invaluable assistance of the listeners who participated in these tests. We hope we have not diminished their appreciation of Handel's Water Music.

[1] W. Kuhl, ''Spaciousness (spatial impression) as a component of total room impression,'' Acustica **40**, 167–181 (1978).
[2] A. H. Marshall, ''A note on the importance of room cross-section of concert halls,'' J. Sound Vib. **5**(1), 100–112 (1967).
[3] M. Barron, ''The subjective effects of first reflections in concert halls: The need for lateral reflections,'' J. Sound Vib. **15**(4), 475–494 (1971).
[4] M. Morimoto and Z. Maekawa, ''Auditory spaciousness and envelopment,'' in Proceedings of the 13th International Congress on Acoustics, Belgrade, 1989 (SAVA Centar), pp. 215–218.
[5] M. Morimoto and K. Iida, ''A new physical measure for psychological evaluation of a sound field; front/back energy ratio as a measure of envelopment,'' J. Acoust. Soc. Am. **93**, 2282(A) (1993).
[6] J. S. Bradley and G. A. Soulodre, ''The influence of late-arriving energy on spatial impression,'' J. Acoust. Soc. Am. **97**, 2263–2271 (1995).
[7] M. Barron and A. H. Marshall, ''Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure,'' J. Sound Vib. **77**(2), 211–232 (1981).
[8] M. Moromoto and K. Iida, ''Relation between auditory source width in various sound fields and degree of interaural crosscorrelation,'' International Symposium on Computer Modelling and Prediction of Objective and Subjective Properties of Sound Fields in Rooms, Copenhagen, August, 1991 (unpublished).
[9] J. S. Bradley and G. A. Soulodre, ''Objective measures of listener envelopment,'' J. Acoust. Soc. Am. **98**, 2590–2597 (1995).
[10] P. Evjen, J. S. Bradley, and S. G. Norcross, ''The effect of late reflections from above and behind on listener envelopment,'' Appl. Acoust. (to be published).
[11] ''Anechoic orchestral music recording,'' Denon CD PG-6006, 1988.
[12] J. S. Bradley, R. Reich, and S. G. Norcross, ''A just noticeable difference in C_{50} for speech,'' Appl. Acoust. (submitted).
[13] T. J. Cox, W. J. Davies, and Y. W. Lam, ''The sensitivity of listeners to early sound field changes in auditoria,'' Acustica **79**, 27–41 (1993).
[14] W. M. Hartmann, B. Rakerd, J. B. Gaalaas, T. vander Velde, W. R. Thorpe, and M. M. Oh, ''Localization of sound in rooms: Broadband noise,'' Michigan State University, Psychoacoustics Report 97, February 1999.
[15] J. S. Bradley, ''The sound field for listeners in concert halls and auditoria,'' in *Computational Architectural Acoustics* (WIT, Southampton, United Kingdom, 1999).

Variation of measured sound speeds in gaseous and liquid air with temperature and pressure

George S. K. Wong, Lixue Wu, and Kam Leung
Institute for National Measurement Standards, National Research Council, Ottawa, Ontario K1A 0R6, Canada

(Received 13 March 2000; revised 24 April 2000; accepted 27 April 2000)

Based on sound speeds in gaseous and liquid air measured by Younglove and Frederick [Int. J. Thermophys. **13**(6), 1033–1041 (1992)], empirical equations for the computation of sound speeds in the above media at relatively smaller temperature and pressure ranges were derived. For gaseous air, over a temperature range from 200 to 300 K and pressure from 0.614 to 10.292 MPa, the maximum deviation between the measured sound speeds and those computed with the empirical equation is 56 ppm. For liquid air, over the ranges from 90 to 110 K for temperature and from 0.763 to 13.823 MPa for pressure, the corresponding deviation is 173 ppm. [S0001-4966(00)01508-3]

PACS numbers: 43.58.Dj [SLE]

LIST OF SYMBOLS

$c(T,P)$	sound speed as a function of temperature and pressure, m/s
P	pressure, MPa
T	thermodynamic temperature, K
$T(68)$	temperature (based on ITS-68), K
$T(90)$	temperature (based on ITS-90), K

Coefficients for gaseous air (see Table I):

$a_0,\ a_1,\ b_0,\ b_1,\ c_0,\ c_1,\ d_0,\ d_1,\ e_0,\ e_1,\ f_0$ and f_1.

Coefficients for liquid air (see Table II):

$a_0,\ldots,a_2;\ b_0,\ldots,b_2;\ c_0,\ldots,c_2;\ d_0,\ldots,d_2;$ $e_0,\ldots,e_2$; and $f_0,\ldots,f_2$.

INTRODUCTION

An empirical equation for gaseous nitrogen sound speeds measured[1] at temperatures from 80 to 350 K over a pressure range from 0.028 75 to 1.506 85 MPa was derived.[2] The aim here is to derive similar empirical equations for gaseous and liquid air. The variations of sound speeds with temperature and pressure measured by Younglove and Frederick[3] are shown in Figs. 1 and 2, respectively. The air mixture[3] used for the measurement consisted of 0.7811 N_2+0.2097 O_2+0.0092 Ar, and the temperatures shown in Figs. 1 and 2 are based on $T(68)$ that has been redefined[4] in 1990. For this investigation, the $T(68)$ temperatures were converted to $T(90)$ with the eight-order polynomial formulas[4] published by the Bureau International des Poid et Mesures [the relevant formula is Eq. (1.3) on p. 17 of the BIPM document]. It is relatively difficult to deduce measured sound speeds for the above media at other pressures and temperatures that do not coincide with those measured[3] above. The procedure[2] used for multi-column coefficient two-dimensional curve fitting is applied to the above measured data,[3] to derive empirical equations that enable one to compute gaseous and liquid air sound speeds as functions of pressure and temperature over selected temperature and pressure ranges.

I. IMPLEMENTATION

The theory of the multi-column coefficient curve-fitting method had been described.[2] For gaseous air, it is relatively difficult to obtain an empirical equation for all the temperature groups shown in Fig. 1. However, there is substantial interest in the measured sound speed over the given measured temperature range from 200 to 300 K. By applying the above multi-column coefficient fit method[2] (order 2) to the gaseous air sound speeds $c(T=200)$ and $c(T=300)$, measured by Younglove and Frederick[3] at pressures from 0.614 to 10.292 Pa, an empirical equation was obtained:

$$c(T,P)_{\text{gaseous air}}=(a_0+a_1T)+(b_0+b_1T)P+(c_0+c_1T)P^2+(d_0+d_1T)P^3+(e_0+e_1T)P^4+(f_0+f_1T)P^5. \quad (1)$$

Gaseous air numerical coefficients for Eq. (1) are listed in Table I. By substituting pressures in megapascals and temperatures in kelvins into Eq. (1), the maximum deviation between the calculated gaseous air sound speeds and the measured values[3] is 56 ppm. Over 72% of the calculated data points deviate by less than 30 ppm from the corresponding

TABLE I. Coefficients of the polynomial for calculating the measured sound speed in gaseous air as determined by the multicolumn coefficient curve fitting method.

Suffix	Coefficients		
x	a_x	b_x	c_x
0	1.557 620E+2	−7.695 485E+0	−1.363 525E−1
1	6.379 939E−1	2.971 727E−2	6.963 378E−4
	d_x	e_x	f_x
0	1.348 640E−1	−7.750 967E−3	2.723 743E−4
1	−4.216 049E−4	2.299 385E−5	−7.984 355E−7

TABLE II. Coefficients of the polynomial for calculating the measured sound speed in liquid air as determined by the multicolumn coefficient curve fitting method.

Suffix	Coefficients		
x	a_x	b_x	c_x
0	6.213 136E+2	3.373 019E+2	−6.523 822E+1
1	1.287 298E+1	−7.287 247E+0	1.388 323E+0
2	−1.265 255E−1	4.067 853E−2	−7.405 012E−3
	d_x	e_x	f_x
0	7.279 536E+0	−4.242 993E−1	1.007 089E−2
1	−1.534 943E−1	8.911 230E−3	−2.109 911E−4
2	8.079 223E−4	−4.663 011E−5	1.099 737E−6

measured data that deviates from theoretical computation[3] by +0.07% and −0.05%.

It is interesting to point out that at a temperature of 273.15 K and pressure at 101.325 kPa, which is much lower than the low-pressure limit of 0.614 MPa for Eq. (1), the sound speed for gaseous air calculated with Eq. (1) is 330.07 m/s, which deviates by 0.37% from the calculated standard air sound speed[5] of 331.29 m/s.

Similarly, an empirical equation for liquid air sound speeds based on data from $c(T=90)$, $c(T=100)$, and $c(T=110)$, measured by Younglove and Frederick[3] at pressures from 0.763 to 13.823 MPa (Fig. 2), is given by

$$\begin{aligned} c(T,P)_{\text{Liquid air}} = {} & (a_0+a_1T+a_2T^2)+(b_0+b_1T+b_2T^2)P \\ & +(c_0+c_1T+c_2T^2)P^2 \\ & +(d_0+d_1T+d_2T^2)P^3 \\ & +(e_0+e_1T+e_2T^2)P^4 \\ & +(f_0+f_1T+f_2T^2)P^5. \end{aligned} \tag{2}$$

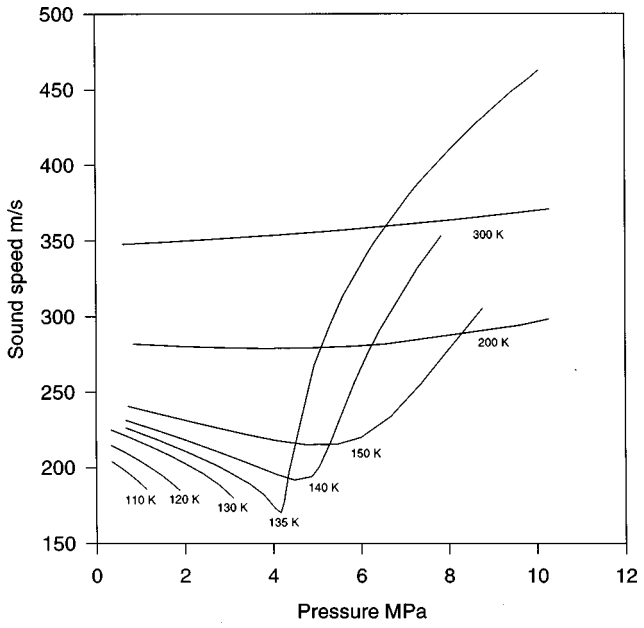


FIG. 1. Gaseous air sound speeds measured by Younglove and Frederick (Ref. 3) over a temperature range from 110 to 300 K and at pressures from 0.338 to 10.292 MPa.

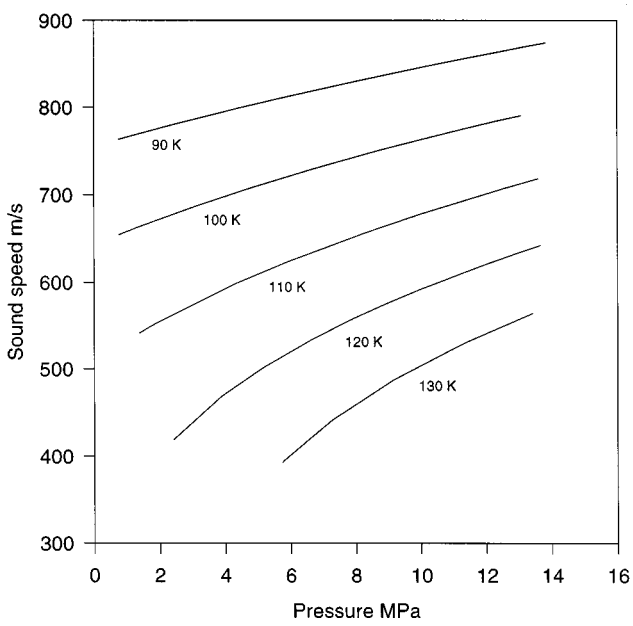


FIG. 2. Liquid air sound speeds measured by Younglove and Frederick (Ref. 3) over a temperature range from 90 to 130 K and at pressures from 0.763 to 13.823 MPa.

Numerical coefficients for Eq. (2) are listed in Table II. By substituting pressures in megapascals and temperatures in kelvins into Eq. (2), the maximum deviation between the calculated liquid air sound speeds and the measured values[3] is less than 173 ppm. Over 68% of the calculated data points deviate by less than 100 ppm from the corresponding measured data that deviates from theoretical computation[3] by +0.39 and −0.36%.

II. CONCLUSIONS

Based on published experimental data[3] on the sound speeds of gaseous air and liquid air, empirical equations are obtained to enable calculation of gaseous and liquid air sound speeds in terms of relatively limited temperature and pressure ranges. The uncertainty of the measured data is quoted[3] as 0.05%. When the empirical equations are applied over the temperature and pressure ranges, from 200 to 300 K and from 0.614 to 10.292 MPa for gaseous air and from 90 to 110 K and 0.763 to 13.823 MPa for liquid air, the maximum deviations of the computed sound speeds from the measured data are 56 and 173 ppm, respectively.

[1] B. A. Younglove and R. D. McCarty, ''Speed-of-sound measurements for nitrogen gas at temperatures from 80 to 350 K and pressures to 1.5 MPa,'' J. Chem. Thermodyn. **12**, 1121–1128 (1980).

[2] G. S. K. Wong and L. Wu, ''Variation of measured nitrogen sound speed with temperature and pressure,'' J. Acoust. Soc. Am. **102**, 650–651 (1997).

[3] B. A. Younglove and N. V. Frederick, ''Speed-of-sound measurements in liquid and gaseous air,'' Int. J. Thermophys. **13**(6), 1033–1041 (1992).

[4] *Supplementary Information for the International Temperature Scale of 1990*, edited by H. Preston-Thomas and T. J. Quinn (Bureau International des Poids et Mesures Sèvres, France, 1990); see also H. Preston-Thomas, ''The International Temperature Scale of 1990 (ITS-90),'' Metrologia **27**, 3–10 (1990).

[5] G. S. K. Wong, ''Speed of sound in standard air,'' J. Acoust. Soc. Am. **79**, 1359–1366 (1986).

Medial efferent effects on auditory-nerve responses to tail-frequency tones II: Alteration of phase

Konstantina M. Stankovic
Eaton-Peabody Laboratory of Auditory Physiology, Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Harvard-MIT Division of Health Sciences and Technology, and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

John J. Guinan, Jr.[a)]
Eaton-Peabody Laboratory of Auditory Physiology, Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114, Harvard-MIT Division of Health Sciences and Technology, and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, and Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts 02115

(Received 19 July 1999; accepted for publication 24 April 2000)

It is often assumed that at frequencies in the tuning-curve tail there is a passive, constant coupling of basilar-membrane motion to inner hair cell (IHC) stereocilia. This paper shows changes in the phase of auditory-nerve-fiber (ANF) responses to tail-frequency tones and calls into question whether basilar-membrane-to-IHC coupling is constant. In cat ANFs with characteristic frequencies $\geqslant$10 kHz, efferent effects on the phase of ANF responses to tail-frequency tones were measured. Efferent stimulation caused substantial changes in ANF phase ($\Delta\Phi$) (range $-80°$ to $+60°$, average $-15°$, a phase lag) with the largest changes at sound levels near threshold and 3–4 octaves below characteristic frequency (CF). At these tail frequencies, efferent stimulation had much less effect on the phase of the cochlear microphonic (CM) than on ANF phase. Thus, since CM is synchronous with basilar-membrane motion for low-frequency stimuli in the cochlear base, the efferent-induced change in ANF phase is unlikely to be due entirely to a change in basilar-membrane phase. At tail frequencies, ANF phase changed with sound level (often by $90°-180°$) and the $\Delta\Phi$ from a fiber was *positively* correlated with the slope of its phase-versus-sound-level function at the same frequency, as if $\Delta\Phi$ were caused by a 2–4 dB *increase* in sound level. This correlation suggests that the processes that produce the change in ANF phase with sound level at tail frequencies are also involved in producing $\Delta\Phi$. It is hypothesized that both efferent stimulation and increases in sound level produce similar phase changes because they both produce a similar mix of cochlear vibrational modes. [S0001-4966(00)00808-0]

PACS numbers: 43.64.Pg, 43.64.Kc, 43.64.Ld [LHC]

INTRODUCTION

In our previous paper (Stankovic and Guinan, 1999) we demonstrated that medial olivocochlear (MOC) efferents can inhibit the firing *rate* of auditory-nerve fibers responding to tones in the low-frequency, broadly tuned ''tail'' region of tuning curves. In this paper we explore efferent effects on auditory-nerve-fiber response *phase* and *synchrony* in the tail region. As before, our focus is on fibers with high characteristic frequencies (CFs), fibers for which the distinction between tip and tail is clear.

There are several reasons for interest in efferent effects on auditory-nerve-fiber (ANF) response phase and synchrony to tail-frequency tones. Phase measures are dominated by the ac receptor potentials of inner hair cells (IHCs), whereas rate measures are dominated by the dc receptor potentials (Palmer and Russell, 1986; Cheatham and Dallos, 1992; reviewed by Ruggero, 1992). This suggests that efferent effects on ANF phase may be due, at least in part, to different causes than efferent effects on rate. The efferent inhibition of firing rate at tail frequencies is equivalent to an attenuation of sound (Stankovic and Guinan, 1999). Determining whether the efferent effects on synchrony and phase are also equivalent to an attenuation of sound will indicate the extent to which common mechanisms shape both measures and should also provide insight into where these effects take place. Finally, a description of the efferent effects on response phase is an important component of any attempt to understand the system behavior. Phase changes must be accounted for in models of efferent effects, or in any cochlear model.

A major unresolved question from the previous paper is whether efferent inhibition of auditory-nerve responses to tail-frequency tones is produced by an efferent-induced reduction of basilar-membrane motion. For CF tones, efferent inhibition appears to be predominantly due to an efferent reduction of basilar-membrane motion (Murugasu and Russell, 1996), presumably produced by an action of medial-efferent synapses on outer hair cells (OHCs) that reduces the gain of the cochlear amplifier (Guinan, 1996). If this were

[a)]Electronic mail: jjg@epl.meei.harvard.edu

the only way in which medial efferents inhibit, then medial efferents would be expected to have little or no effect at frequencies in the tails of tuning curves because active cochlear mechanisms are thought to have little or no effect on basilar-membrane motion an octave or more below CF (Rhode, 1973; Ruggero and Rich, 1991; Ruggero *et al.*, 1996b, 1997; Nuttall and Dolan, 1996). Although our demonstration of an efferent-induced reduction of firing rate at tail frequencies clearly shows that efferents do inhibit at these frequencies (Stankovic and Guinan, 1999), the few published data on efferent effects on basilar-membrane motion at tail frequencies (Murugasu and Russell, 1996; Russell and Murugasu, 1998) leave open the question of whether this inhibition comes about by a change of basilar-membrane motion, or by some other mechanism. Here, we pursue the related question of whether efferent changes of auditory-nerve response phase are due to underlying changes in basilar-membrane phase. The question is more readily answered for phase than for amplitude (i.e., firing rate), because for low-frequency stimulation, the phase of the cochlear microphonic (CM) provides a monitor of the phase of basilar-membrane motion in the base of the cochlea (Dallos *et al.*, 1974; Dallos and Cheatham, 1976; Patuzzi *et al.*, 1989; Ruggero *et al.*, 1986b), whereas the amplitude of CM does not serve this function for firing rate (e.g., Fex, 1959).

The relationship between ANF response phase and basilar-membrane response phase is complex, but understanding this relationship is crucial to achieving a basic understanding of how the cochlea works. For tones far below CF (i.e., tail-frequency tones), ANF phase varies along the length of the cochlea relative to basilar-membrane phase in the chinchilla (Ruggero and Rich, 1983, 1987; Ruggero *et al.*, 1996a) and to some extent in the gerbil, guinea pig and cat (Sokolich *et al.*, 1976; Konishi and Nielsen, 1978; Oshima and Strelioff, 1983; Cai and Geisler, 1996). Cheatham and Dallos (1999) have identified three factors that may account for the variation in ANF phase relative to basilar membrane phase: (1) IHC-membrane low-pass filtering, (2) the dynamics of IHC ion channels, and (3) extracellular potentials changing the IHC transmembrane potential and thus IHC transmitter release. The relationship between ANF response phase and basilar-membrane phase may also depend on medial-efferent activation, but this has not been previously explored. Since medial efferents form synapses directly on outer hair cells, knowing the extent to which medial efferents change ANF response phase should provide important insights both into medial-efferent effects and the extent to which OHC properties can influence the relationship of ANF phase to basilar-membrane phase.

I. METHODS

The surgical preparation, methods for single-fiber recording, and the stimulation paradigms were as in Stankovic and Guinan (1999). Highlights and important differences in methods are presented below.

We used adult cats anesthetized by diallyl barbiturate in urethane. Middle-ear-muscle tendons were cut and the auditory nerve was approached dorsally as in Kiang *et al.* (1965). Efferents were stimulated with an electrode along the midline of the floor of the fourth ventricle using trains of shocks at 200/s. Treatment of experimental animals was in accordance with protocols approved by the Committees on Animal Care at the Massachusetts Institute of Technology and the Massachusetts Eye and Ear Infirmary.

A. Single-fiber recording and data analysis

On each auditory-nerve fiber we obtained: (1) a tuning curve from which fiber CF was determined, (2) 20 s of spontaneous activity from which spontaneous rate (SR) was calculated, and (3) level series with and without efferent stimulation in which both sound level and efferent stimulation were randomized. Auditory-nerve fibers were grouped into low ($SR \leqslant 0.5$), medium ($0.5 < SR < 18$), and high ($SR \geqslant 18$ spikes/s) categories (Liberman, 1978). In randomized level series, points were acquired every 3 s, and at each point the firing rate was obtained by averaging responses from ten sequential tone bursts. The tone bursts were presented every 100 ms, had 2.5-ms rise/fall times, and 50-ms duration (rise onset to fall onset) and 47.5 ms between bursts (end of fall to beginning of next rise). Efferent stimulation, when present, began 100 ms before the first tone burst, lasted 1.1 s, and ended after the last tone burst. ANF response phase was calculated from period histograms obtained from the steady-state part of each response, i.e., from 6 ms after the onset of each tone burst to approximately 1 ms after the beginning of the offset of the tone burst (it takes about 2 ms for the effect of the offset to become apparent in ANF firing). The sampling window was adjusted across sound frequencies so that it included an integer number of cycles.

The stored spike times were used in off-line analysis to construct histograms of action potentials relative to the zero crossing of the tone waveform, i.e., ''post-zero-crossing'' or ''period'' histograms, each with 400 bins per tone cycle. For a given tone frequency and sound level, two period histograms were constructed—one with, and one without efferent stimulation. Period histograms were used to determine (1) synchrony as measured by the synchronization index (also known as vector strength), which is the amplitude of the fundamental Fourier component of the histogram normalized by the mean firing rate (Goldberg and Brown, 1969; Johnson, 1980), and (2) response phase, i.e., the phase of the fundamental Fourier component of the period histogram. For Figs. 1–4, data were rebinned to 25 bins per cycle to make the histograms smoother.

Data were used only if they met criteria for quality single-fiber recordings and criteria for accurate determination of phase. As in Stankovic and Guinan (1999), spikes had to have excellent triggering, no effects of shock artifacts, no evidence of lost short-interval spikes in inter-spike-interval histograms and, for tail-frequency tones, no harmonic distortion that fell within the tuning-curve tip. The data reported here are from 13 cats out of the 17 reported in Stankovic and Guinan (1999) (we did not record phase timing markers during the initial experiments). A vast majority of the data is from high-SR fibers with $CFs > 10$ kHz stimulated with tail-frequency tones of 1–2 kHz. Although tones at 3 or 4 kHz are still within the tail of fibers with $CFs \geqslant 10$ kHz, they are less useful for studying efferent effects on response phase

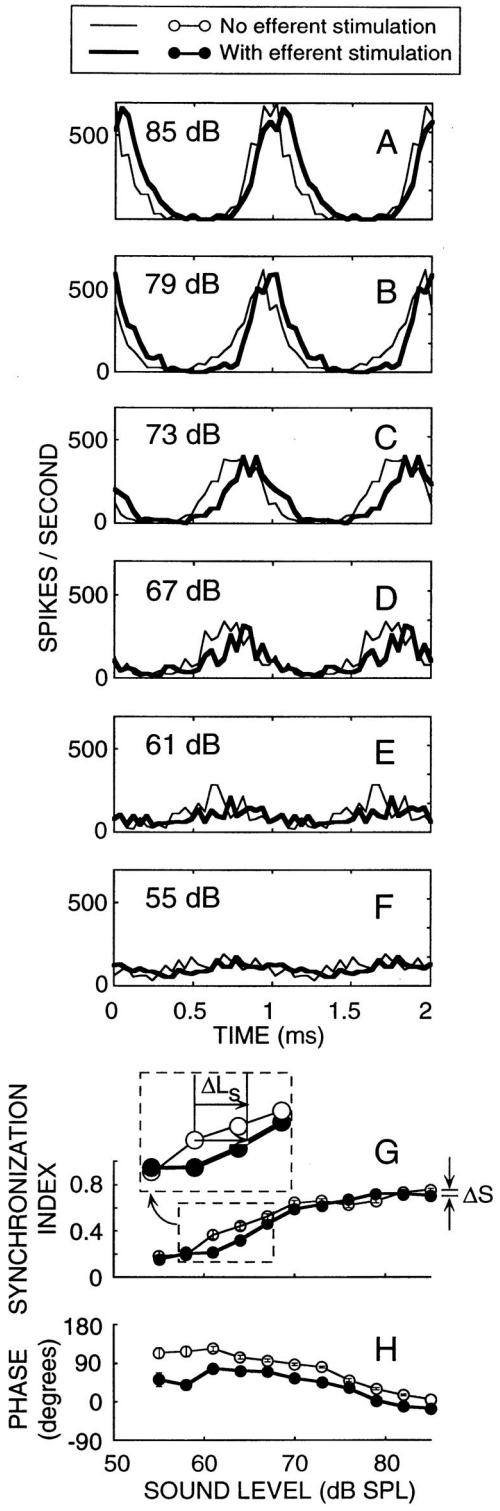


FIG. 1. Data from a unit showing the typical pattern of efferent effects for 1-kHz tones. Panels (A)–(F) show period histograms that, for clarity, are shown over two cycles by duplicating the one-period response. At each sound level (indicated in the upper left of the panel), responses in the absence (thin line) and presence (thick line) of efferent stimulation are shown. Each histogram combines data from six separately randomized runs. Panel (G) synchronization index derived from the above histograms (and others) without (open circles and thin line) and with (closed circles and thick line) efferent stimulation. The bars are standard deviations. Panel (H) response phase derived from the above histograms [symbols as in panel (G)]. The bars are standard error of the mean. Most error bars in (G) and (H) are smaller than the symbol size. Unit 37-15, CF=17 782 Hz, SR=118.9 spikes/s.

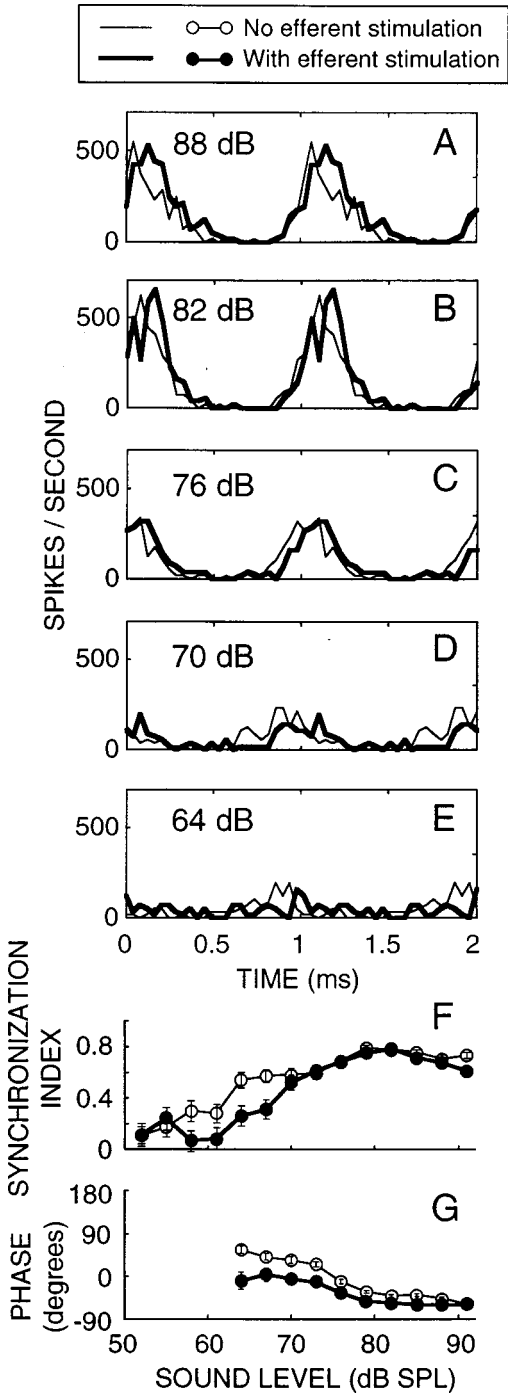


FIG. 2. Data from another unit showing the typical pattern of efferent effects for 1 kHz tones. Layout as in Fig. 1. Data from three separately randomized runs. In panel (G), phase is not shown below 65 dB SPL because it did not meet the accuracy criterion. Unit 37-13, CF=16 692 Hz, SR=57.0 spikes/s.

because responses at these frequencies have lower synchrony (Johnson, 1980).

B. Determining the accuracy of phase and synchrony data

To determine which measured response phases were sufficiently accurate to be used, we applied a criterion based on the estimated standard error of the phase. Specifically, data were accepted only if the standard error of the phase, Φ_{SE},

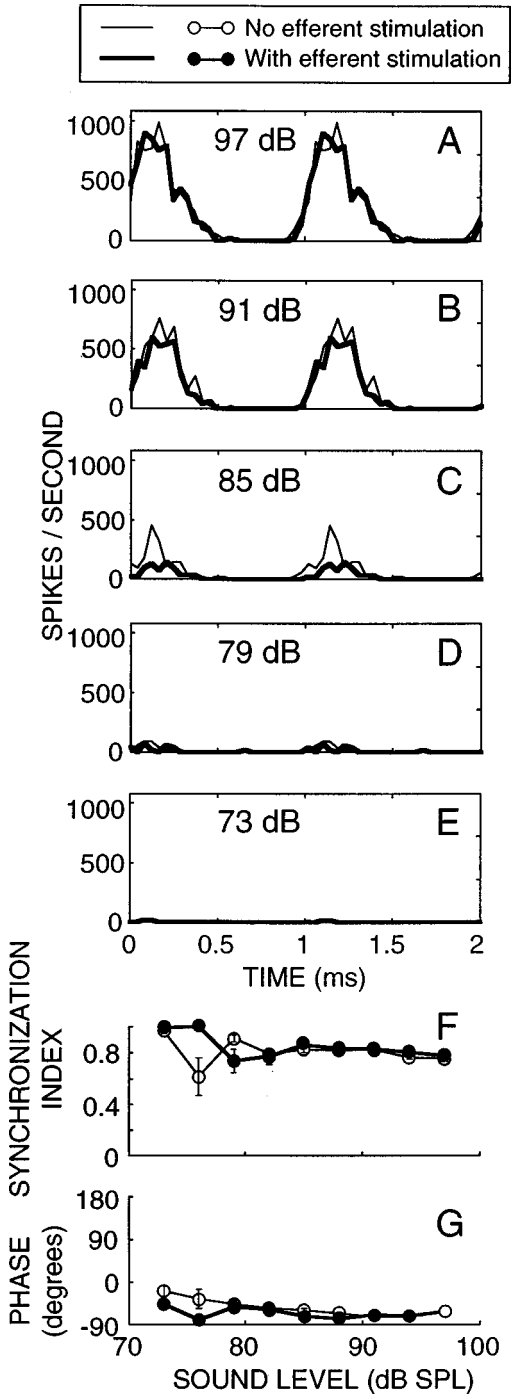


FIG. 3. Data from a fiber whose response phase at 1 kHz was affected very little by efferents. Layout as in Fig. 1. Data from three separately randomized runs. Unit 37-22, CF=13 412 Hz, SR=0.2 spikes/s.

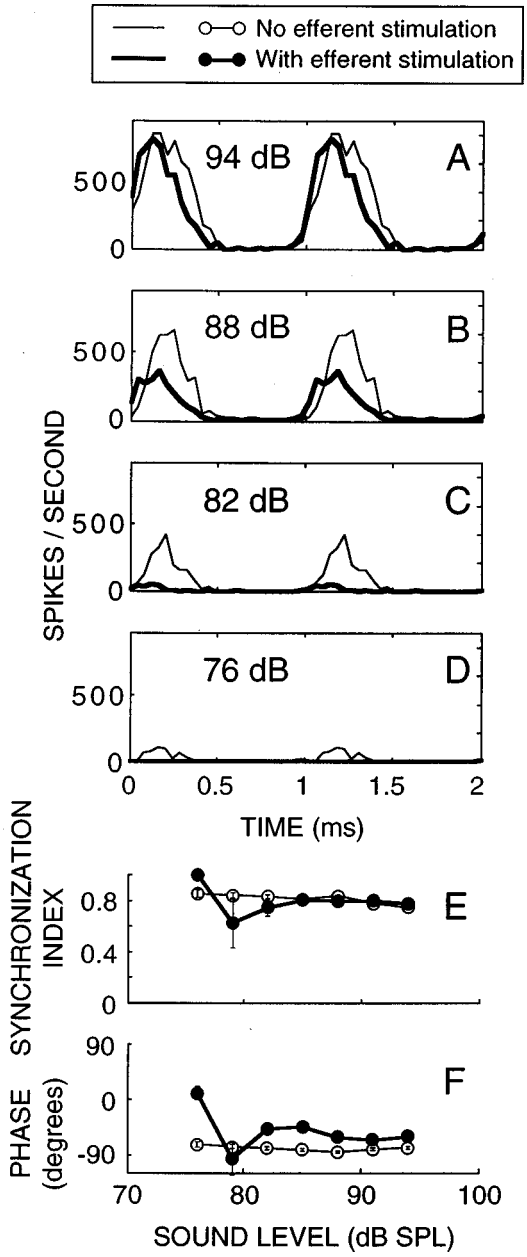


FIG. 4. Data from a fiber in which efferents produced a phase lead in the response at 1 kHz. Layout as in Fig. 1. Data from five separately randomized runs. Unit 37-17, CF=10 592 Hz, SR=1.2 spikes/s.

was $|\Phi_{SE}| \leqslant 30°$ for at least two adjacent points in the sound-level series both with and without efferent stimulation. To obtain the standard error of the phase, the 95%-confidence interval of the phase was calculated according to the procedure described by Mardia (1972). The data are assumed to be described by a von Mises distribution, which is similar to a normal distribution except that it describes directional data. A key step is to estimate the von Mises ''concentration parameter,'' κ (κ is similar to the inverse of ''standard deviation'' in a normal distribution). We approximated the maximum likelihood estimate of κ, $\hat{\kappa}$, as $\hat{\kappa} = 10^{(-1.402+7.310\hat{S}-12.9972\hat{S}^2+8.4733\hat{S}^3)}$ from the table in Appendix 2.3 of Mardia (1972), which applies for histograms with significant synchrony, $\overline{S}$. The 95%-confidence interval of phase in degrees, I_{95}, was then calculated for a histogram of n spikes as: $I_{95}=(360°/2\pi)(1.96/\sqrt{\kappa'})$ where $\kappa'=n\overline{S}\hat{\kappa}$. This formula is appropriate for $n>30$, but for $n<30$ and $\kappa'<4$, the formula is an underestimate and I_{95} is always large, so I_{95} was arbitrarily set to the maximum possible, i.e., 180°. Finally, we applied the criteria $|\Phi_{SE}| \leqslant 30°$ from Φ_{SE} calculated as $\Phi_{SE}=I_{95}/1.96$ by analogy to the normal distribution for which the standard error is the {95% confidence interval}/1.96.

An estimate of the standard deviation of synchrony, $\sigma_{\overline{S}}$, was calculated using the empirical approximation of Johnson (1974): $\sigma_{\overline{S}}=[(1-e^{-10\overline{S}})/4+(\sqrt{4-\pi})/2]/\sqrt{n}$. Data were ac-

cepted if they met the arbitrary criterion of synchrony, $\overline{S} \geq 0.3$ and $\sigma_{\overline{S}} < 0.1$. For Figs. 2 and 3, this criterion was relaxed to show more data.

To determine if several apparent trends in the data were statistically significant, without assuming that the data were normally distributed, we used permutation tests (Efron and Tibshirani, 1993; see Stankovic and Guinan, 1999, for a detailed explanation). In a few cases we tested statistical significance with analysis of variance (ANOVA) using DATADESK™. The loess fit (Cleveland, 1993) was used in some scatter plots to aid the eye in detecting trends in data.

C. Measurements of the cochlear microphonic

To measure efferent effects on CM, tone bursts (50-ms long, 2.5-ms rise time, 100-ms repetition period) were presented and recordings were made from an electrode on or near the round window referenced to a distant electrode. After amplification, in-phase and quadrature outputs of an EG&G 5210 lock-in (time constant 3 ms) were sampled over the last 45 ms of each burst, and 32 responses were averaged. Sound-level series were run from 40–52 dB SPL up to 85 dB SPL either sequentially (one cat) or by randomizing the presentation of sound level and efferent stimulation (two cats). As expected, randomization did not significantly affect the response because CM shows little, if any, adaptation.

To avoid contamination of CM by evoked response from the auditory nerve, we used forward masking with broadband noise bursts (47-ms long, 2.5-ms rise time) interleaved with the tone bursts. To avoid electronic artifacts at the recording electrode produced by the efferent shocks, (1) the shock frequency (227.27 Hz) was chosen to be incommensurate with the tone frequencies, and/or (2) efferents were stimulated mostly between the tone bursts (i.e., for 66.6 ms, starting 42.5 ms before the onset of a tone burst). Since the efferent effect has a long time constant, the strength of the efferent effect was changed little by interleaving efferent stimulation. This was confirmed in two experiments where CM was recorded mostly between, and then completely overlapping with, efferent stimulation.

II. RESULTS

A. Efferent effects on the shapes of period histograms

Efferent stimulation did not produce much change in the shapes of period histograms but it did change the time at which the peak occurred. This is illustrated in Figs. 1 and 2 with data from two fibers which show typical responses but were contacted for a long time so that we have more data than usual on them. The period histograms in Figs. 1 and 2 are typical in that there is only one peak per cycle, the peak resembles a rectified sine wave, and efferent stimulation delayed the peak to later in the cycle. Sometimes, however, efferent stimulation did not delay the peak (Fig. 3) or shifted the peak to earlier in the cycle (Fig. 4). In all of these histograms, the shapes are similar to those in previous reports of ANF responses without efferent stimulation (e.g., Rose *et al.*, 1967; Anderson, 1971; Johnson, 1980).

Although a vast majority of our period histograms had a single peak per tone cycle, we occasionally found period histograms with two peaks per cycle—the behavior known as ''peak splitting'' (Kiang *et al.*, 1965). We saw peak splitting only in fibers stimulated at high sound levels with 500-Hz tones (in 7 of 14 fibers stimulated at 500 Hz), but not for tones of 1 kHz or higher. This is consistent with published data on high-CF fibers (Kiang, 1990; Ruggero *et al.*, 1996a; Cai and Geisler, 1996). In most period histograms with peak splitting, efferent stimulation did not have a dramatic effect on histograms' shape. However, there are too few data to draw definitive conclusions.

Since, with a few exceptions, our period histograms were single-peaked, and their shape was not obviously changed by efferent activation, characterizing these histograms by their first Fourier-transform component should be adequate for determining the effect of efferents. The first Fourier component has two dimensions, phase and synchrony, with synchrony typically measured by the synchronization index (Goldberg and Brown, 1969; Johnson, 1980). Although we are most interested in phase, we deal first with synchrony because phase cannot be measured unless there is adequate synchrony.

B. Efferent effects on synchrony

An important question is whether efferent effects on synchrony are equivalent to an attenuation of sound. To answer this question we must measure efferent effects in the range where synchrony changes with sound level, but we have only a few accurate data in this range. An accurate determination of synchrony requires many spikes if there is little synchronization, so usually the synchrony-accuracy criterion (see Sec. I) was reached only when synchrony was high. Nonetheless, some fibers had adequate synchrony in the rising portion of synchrony-versus-sound-level functions.

As a measure of the efferent effect on synchrony in the range where synchrony increases as sound level increases, we used the synchrony level shift, ΔL_S, the amount by which the sound level must be increased with efferent stimulation to produce the same synchrony as obtained without efferent stimulation [inset in Fig. 1(G)]. In 12 of 13 cases ΔL_S was positive, consistent with an efferent attenuation of sound. When ΔL_S was plotted versus the efferent-induced level shift in rate, ΔL_R (ΔL_R from Stankovic and Guinan, 1999), there was considerable scatter (Fig. 5). If both ΔL_S and ΔL_R were due to the same efferent-induced attenuation, the points would be along the dashed line in Fig. 5. The low correlation ($r = 0.33$) suggests that the attenuations of ΔL_S and ΔL_R were often not the same. However, there are few data, there are errors in both measurements, and ΔL_S and ΔL_R were not measured over the same sound-level ranges. All considered, a strong conclusion cannot be reached from these data.

At high sound levels where synchrony is close to its maximum value, efferent activation had only a minor effect on synchrony [Figs. 1(G), 2(F), 3(F), and 4(E)]. To quantify this effect we used the efferent-induced synchrony change, ΔS, which is: {synchrony without efferent stimulation} minus {synchrony with efferent stimulation} [Fig. 1(G) right].

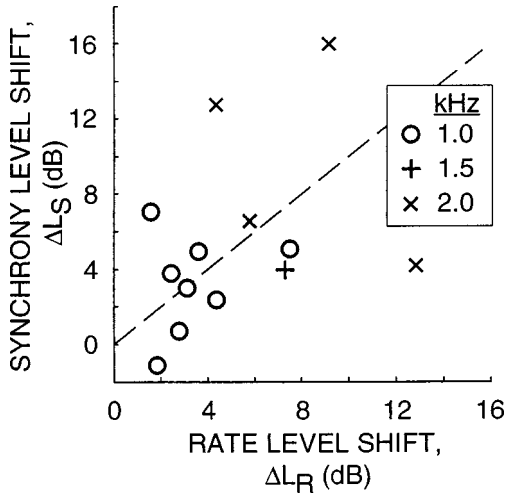


FIG. 5. The level shift for synchrony, ΔL_S, versus the level shift for rate, ΔL_R. Both ΔL_S and ΔL_R were normalized to a compound action potential (CAP) inhibition of 20 dB as in Stankovic and Guinan (1999). ΔL_S is the average of all ($\leqslant 3$) ΔL_S measurements in the region where ΔL_S increases with sound level. ΔL_R here is the average normalized ΔL ($\Delta L_{n,,\text{Ave}}$) of Stankovic and Guinan (1999). Dashed line indicates slope of 1.

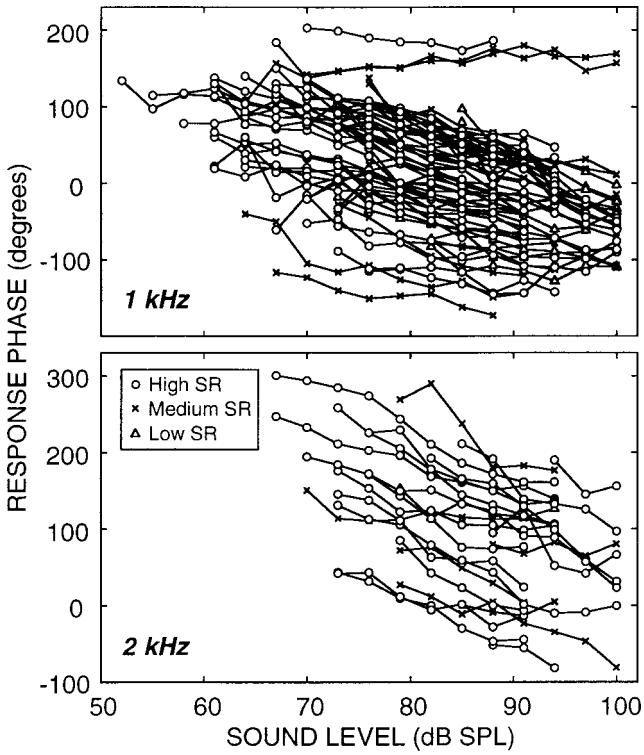


FIG. 6. Auditory-nerve-fiber response phase as a function of sound level for 1-kHz tones (top) and 2-kHz tones (bottom). All fibers had CFs>10 kHz. (The phase reference was not necessarily the same across cats, so the absolute values of the phase may not be reliable.)

A single ΔS was obtained for each fiber by averaging the points from three sound levels near the maximum synchrony (usually 79–85 dB SPL for high-SR fibers and 85–91 dB SPL for medium-SR and low-SR fibers, or sometimes the three highest levels available). For tones between 500 Hz and 3 kHz (mostly 1 kHz) and fibers with CFs 10–30 kHz, the mean ΔS was not statistically different from zero (*t*-test $p=0.12$) and there were no statistically significant differences across SR groups (ANOVA $p=0.83$). The mean ΔS, standard error of the mean, and number of fibers were: high SR: -0.0098 ± 0.0056, $N=69$; medium SR: -0.0044 ± 0.0100, $N=25$; low SR: -0.0029 ± 0.0161, $N=17$. Since synchrony is high throughout most of the range where phase can be accurately determined, one advantage of the very small efferent effect on synchrony at high levels is that changes in synchrony do not need to be considered when analyzing the efferent-induced changes in phase.

C. Efferent effects on response phase at tail frequencies

First, we consider phase properties without efferent stimulation. Without efferent stimulation, auditory-nerve-fiber responses to tail-frequency tones of 1 or 2 kHz typically had slowly increasing phase lags as sound level increased, with changes of as much as 90°–180° (Fig. 6). There are few other comparable data from the cat. Liberman and Kiang (1984, Fig. 3) characterized 1-kHz phase-versus-level functions for four fibers with CFs>10 kHz as being less than 90°, which appears to differ from our finding that many fibers have changes greater than 90°. The reason for this difference is not known. Cai and Geisler (1996) found relatively little change in phase with sound level for much lower frequencies (e.g., 100 Hz) and a wide range of CFs. In contrast, results from chinchilla using tail-frequency tones (but frequencies <1 kHz) show small phase changes at moderate sound levels and then abrupt phase reversals at high sound levels (Ruggero and Rich, 1983; Ruggero *et al.* 1996a).

Efferent stimulation caused a phase lag in most (86/108 =80%) fibers [e.g., Figs. 1(H), 2(G)]. However, efferent stimulation had little effect on some (13/108=12%) fibers [e.g., Fig. 3(G)] and produced a phase lead in some (9/108 = 8%) fibers [e.g., Fig. 4(F)].

Since efferent effects on firing rate and synchrony can be characterized as being equivalent to an attenuation of the sound stimulus, one can ask if the efferent effects on phase are also equivalent to an attenuation of sound. For most high-CF fibers stimulated with low-frequency tones, efferent stimulation produced a phase lag [e.g., Fig. 7(A)], but the phase-versus-sound-level curves show that an attenuation of sound would produce a phase advance [Figs. 6, 7(A)]. Even for the few fibers in which efferent stimulation produced a phase lead, the efferent effect was not necessarily equivalent to an attenuation of sound. For example, the data in Fig. 4(F) show that there was an efferent-induced phase lead but almost no change in phase with sound level at high levels. Thus, for the great majority of fibers, the efferent effect on phase was not equivalent to an attenuation of sound.

One consequence of this result is that the usual metric for describing efferent effects, the equivalent attenuation, does not work for phase. As a measure of the efferent effect on auditory-nerve response phase, we used the phase change, $\Delta\Phi$, defined as: {the phase with efferent stimulation} minus {the phase without efferent stimulation} [Fig. 7(A)]. Figure 7(B) shows a typical measurement of $\Delta\Phi$ along with its

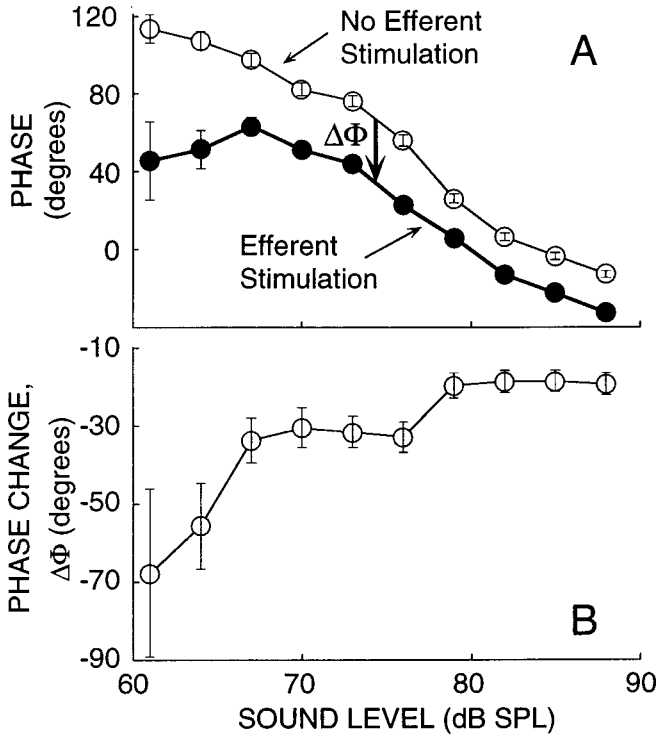


FIG. 7. A typical measurement of efferent-induced phase change, $\Delta\Phi$, and its error. (A) Response phase as a function of sound level without (open circles) and with (closed circles) efferent stimulation. Arrow labeled ''$\Delta\Phi$'' shows the measurement of $\Delta\Phi$ at one point. (B) $\Delta\Phi$ derived from the data in (A). Bars on individual points indicate the standard error of $\Delta\Phi$ ($\delta\Delta\Phi$), determined from the standard errors of the individual phase measurements without, $\delta\Delta\Phi_{\mathrm{NoShs}}$, and with, $\delta\Delta\Phi_{\mathrm{Shs}}$, efferent stimulation as $\delta\Delta\Phi = \sqrt{(\delta\Delta\Phi_{\mathrm{NoShs}})^2+(\delta\Delta\Phi_{\mathrm{Shs}})^2}$. Tone frequency: 1 kHz. Unit 37-14, CF =17 179 Hz, SR=111.9 spikes/s.

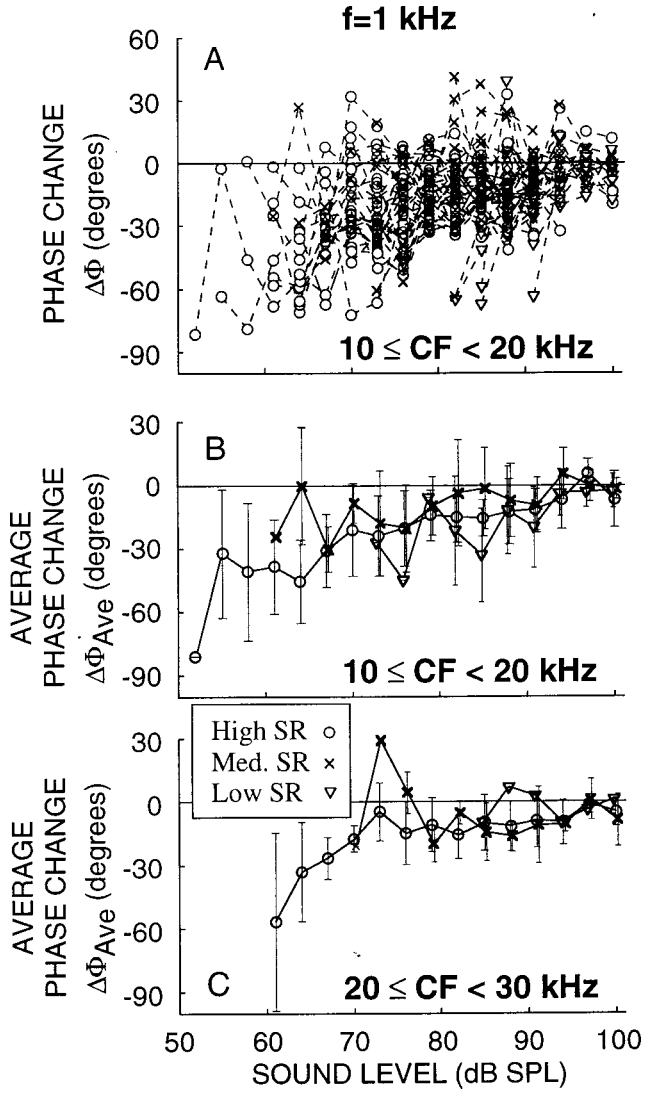


FIG. 8. The efferent-induced change of phase, $\Delta\Phi$, as a function of sound level for 1-kHz tones and two CF regions: 10–20 kHz (A), (B), and 20–30 kHz (C). (A) Superposition of $\Delta\Phi$-versus-sound-level curves from many individual fibers. Responses from a single fiber are connected by dotted lines. Differences in the lowest initial sound level reflect differences in synchrony thresholds, i.e., the levels at which phase error is $<30°$. (B) Average $\Delta\Phi$, $\Delta\Phi_{\mathrm{Ave}}$, for each SR group, based on fibers from (A). Bars indicate standard deviations; points without bars are from one fiber only. (C) Average $\Delta\Phi$ for fibers with $20 \leqslant CF \leqslant 30$ kHz, for each SR group. The number of fibers in each SR group are: (A), (B) 29 high, 13 medium, 9 low; (C) 11 high, 4 medium, 1 low.

calculated error. Such error bars were used in statistical tests described below but, for clarity, have been omitted from later plots of $\Delta\Phi$ [Figs. 8(A), 9(A)].

For tones of 0.5, 1, and 2 kHz, the efferent-induced changes in phase, $\Delta\Phi$, were significantly different from zero $p<0.0001$ at each frequency), as judged by *t*-tests applied to all of the data at that frequency (i.e., all sound levels and all fibers with CFs 10–30 kHz). The average $\Delta\Phi$ was $-10.5°$ (a phase lag) for 500-Hz tones (8 fibers), $-15.3°$ for 1-kHz tones (67 fibers), and $-16.2°$ for 2 kHz tones (24 fibers). For the three fibers stimulated at 3 kHz, the average $\Delta\Phi$ was not significantly different from zero ($p=0.98$).

These average $\Delta\Phi$ values, while significant, ignore variations in $\Delta\Phi$ that depended on both tone variables (intensity and frequency), and fiber characteristics (CF and spontaneous rate). These dependencies are explored in the following two sections.

D. Dependence of $\Delta\Phi$ on sound level and spontaneous rate

To investigate the dependence of $\Delta\Phi$ on sound level and spontaneous rate, one tone frequency was considered at a time. Since there were many fibers stimulated by 1-kHz tones, they were divided into two CF bands (10–20 kHz and 20–30 kHz, Fig. 8). The smaller number of fibers stimulated by 2-kHz tones were considered in a single band (Fig. 9) which is listed as 10–20 instead of 10–30 kHz because none of the fibers had a CF above 20 kHz. Fibers stimulated by 500-Hz tones were not analyzed in detail because there were only 14 fibers with data, and half of them showed peak splitting.

To give a full picture, the $\Delta\Phi$ data are shown in two ways. Figures 8(A) and 9(A) are superimposed sound-level functions from each fiber. These show the range of the data and trends with sound level. Figures 8(B), 8(C), and 9(B) are plots of the average $\Delta\Phi$ for each SR group versus sound level. These provide comparisons across SR groups, sound levels, and tone frequencies.

A visual examination of Figs. 8 and 9 suggests that $\Delta\Phi$ depended on sound level, but it is not clear whether $\Delta\Phi$

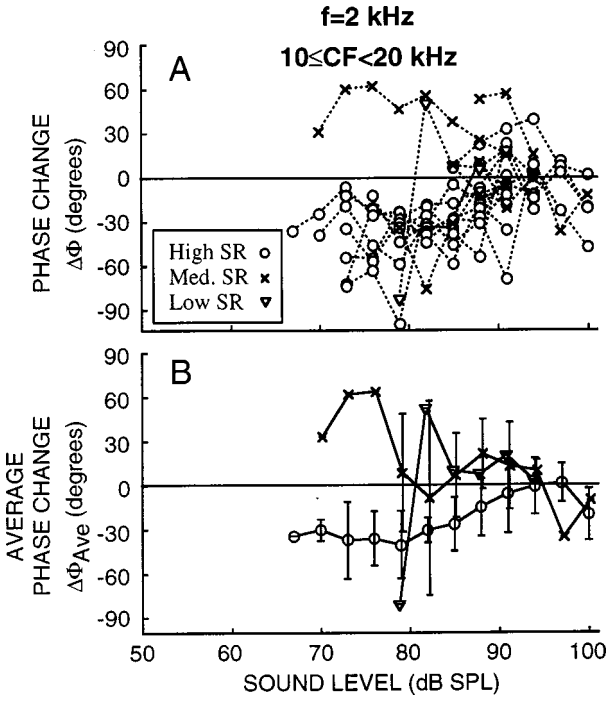


FIG. 9. The efferent-induced change of phase, $\Delta\Phi$, as a function of sound level for 2-kHz tones and fibers with CFs 10–20 kHz. Symbols as in Fig. 8. The number of fibers in each SR group are: 12 high, 4 medium, 1 low.

depended on spontaneous rate. Possible dependencies of $\Delta\Phi$ on sound level and spontaneous rate were tested statistically using permutation tests (Efron and Tibshirani, 1993; see also Stankovic and Guinan, 1999) with separate tests for each of the three data sets used in Figs. 8(B), 8(C), and 9(B). These tests were complicated by the fact that $\Delta\Phi$ from individual fibers, and for SR groups, do not all span the same sound-level range. To overcome this, $\Delta\Phi$-versus-sound-level data were represented by a straight line (fit using a least-squares criterion with points weighted by the errors of individual measurements). To test for differences across SR groups, for each SR group a single straight line was fit to *all* of the $\Delta\Phi$-versus-sound-level points from fibers of that SR group. The differences in slopes (or intercepts) between two SR groups were then used as comparison statistics. A permutation test trial consisted of shuffling fibers into new ''pseudo-SR groups'' and calculating ''pseudo'' comparison statistics; 1000 trials were used to form probability histograms against which the original statistic was compared. In most cases, statistically significant differences were not detected between SR groups at the $p=0.05$ level. The only exception was a difference between high-SR and medium-SR fibers for stimulation with 2-kHz tones ($p<0.01$ for both slope and intercept). However, this result needs to be interpreted with caution because there were only four medium-SR fibers in this data set.

Next, we wanted to determine whether the apparent dependence of $\Delta\Phi$ on sound level was statistically significant. Although the previous analysis revealed a possible difference between two SR groups at 2 kHz, for simplicity this difference was ignored and data from all SR groups were pooled. Permutation tests were done by multiple trials with, on each trial, shuffling the data from each sound level across ''pseudo sound-level groups'' and calculating the slope of a line fit to the result (fit by minimizing the least-square error with points weighted by their errors). These tests revealed that $\Delta\Phi$ significantly depended on sound level ($p<0.001$) for all three data sets considered [Figs. 8(B), 8(C), and 9(B)].

E. Dependence of $\Delta\Phi$ on sound frequency

Ideally, the frequency dependence of $\Delta\Phi$ would be studied with many points obtained from one fiber. However, there were practical limitations that prevented such a study, so we pooled data across fibers to explore how $\Delta\Phi$ depended on tone frequency. To focus on sound levels where $\Delta\Phi$ tended to be large, we averaged $\Delta\Phi$ across sound levels $\leqslant 85$ dB SPL and obtained a single average $\Delta\Phi$ ($\Delta\Phi_{\mathrm{Ave}}$) for each fiber. The exact range of sound levels included in $\Delta\Phi_{\mathrm{Ave}}$ varied across fibers depending on fiber synchrony and the availability of data. The averaging produced a bias across SR groups because high-SR fibers typically had data extending to lower sound levels than medium-SR and low-SR fibers, and $|\Delta\Phi|$ tended to be larger at lower sound levels. Because of this bias, comparisons across SR groups are not warranted and fibers of all SR groups were considered together.

Figure 10(A) shows $\Delta\Phi_{\mathrm{Ave}}$ as a function of tone frequency. Figure 10(B) shows averages and standard errors at each tone frequency. Figures 10(A) and 10(B) suggest that $\Delta\Phi_{\mathrm{Ave}}$ tends to become more negative (i.e., efferents produce an increasing phase lag) as frequency increases from 500 Hz to 2.5 kHz.

We also analyzed the frequency dependence of $\Delta\Phi_{\mathrm{Ave}}$ by expressing tone frequency relative to fiber CF [Fig. 10(C)]. This aligns tuning-curve tips so that data at a few tone frequencies in fibers of many CFs show $\Delta\Phi_{\mathrm{Ave}}$ at many different locations within the tuning-curve tails (note that we did not find any dependence of $\Delta\Phi_{\mathrm{Ave}}$ on fiber CF). As expected, there is a trend similar to that in Figs. 10(A) and 10(B): $\Delta\Phi_{\mathrm{Ave}}$ shows a slowly increasing phase lag as tone frequency increases from six to three octaves below CF. In addition, $\Delta\Phi_{\mathrm{Ave}}$ appears to become a phase lead as tone frequency increases further. However, the apparent reversal at high tone frequencies in Fig. 10 should be interpreted with caution, given the scatter in the data and the small number of fibers with data at the highest frequencies.

F. The efferent-induced change in phase, $\Delta\Phi$, versus the slope of the tail-frequency phase-versus-sound-level function

We noticed that auditory-nerve fibers with large efferent-induced phase lags tended to have large negative slopes in their phase-versus-level functions [Figs. 1(H), 2(G)], whereas those with small efferent-induced phase delays or phase advances did not [Figs. 3(G), 4(F)]. To explore this, the efferent-induced change in phase, $\Delta\Phi$, was plotted versus the slope of a fiber's phase-versus-level function without efferent stimulation, both measured at the same tail frequency (Fig. 11). Although the data show considerable scatter, correlations were present ($r=0.39$ at 1 kHz; $r=0.47$ at 2 kHz). Regression lines through these data had slopes of 2.4

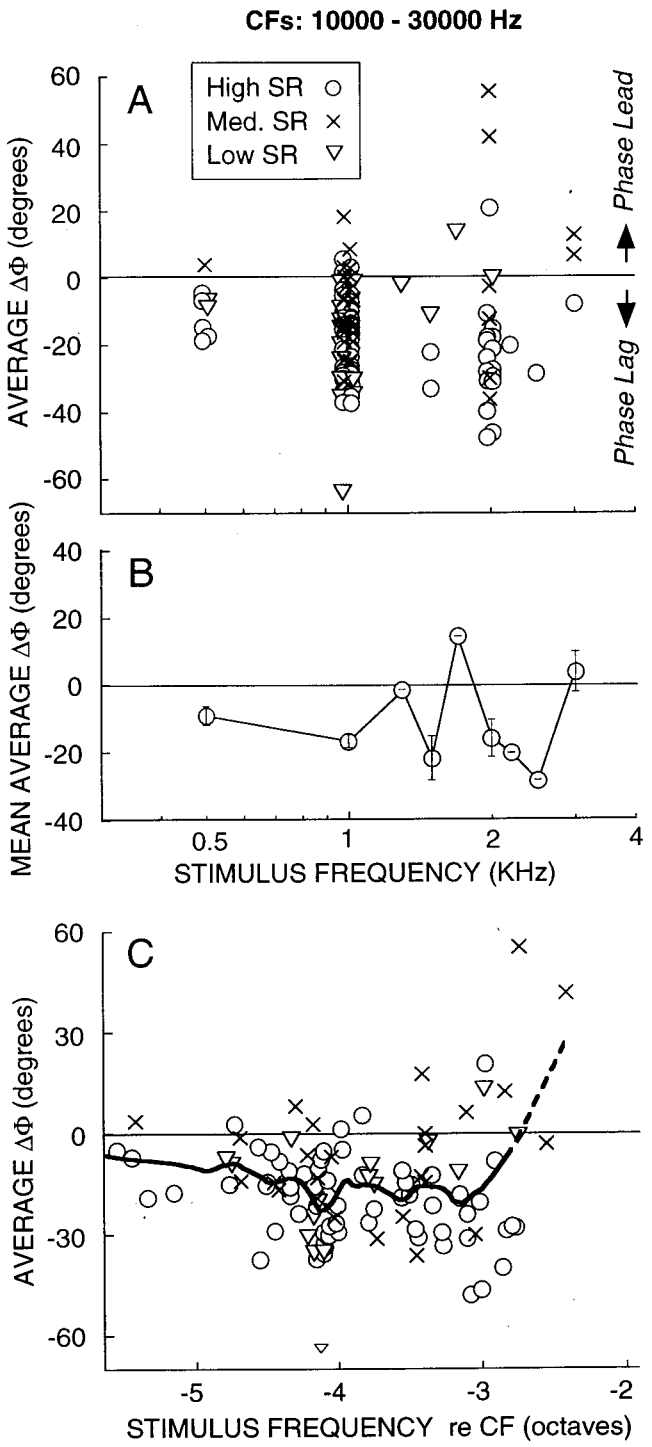


FIG. 10. Efferent-induced phase changes as a function of tone frequency. (A) $\Delta\Phi$ averaged across sound levels ($\Delta\Phi_{Ave}$) up to 85 dB SPL with each point from one fiber. Data from 68 fibers (CFs 10–30 kHz) produced 108 points (66 high-SR, 26 medium-SR, and 16 low-SR) because some fibers were stimulated at more than one tone frequency. (B) Each point is the mean of the $\Delta\Phi_{Ave}$ at a given frequency of the data in panel (A); bars indicate standard error of the mean. (C) $\Delta\Phi_{Ave}$ versus tone frequency relative to CF. Same data as in panels (A) and (B). The thick line is a loess fit (smoothing factor = 0.2) to the data. The dashed part of the line indicates uncertainty in the trend due to paucity of data.

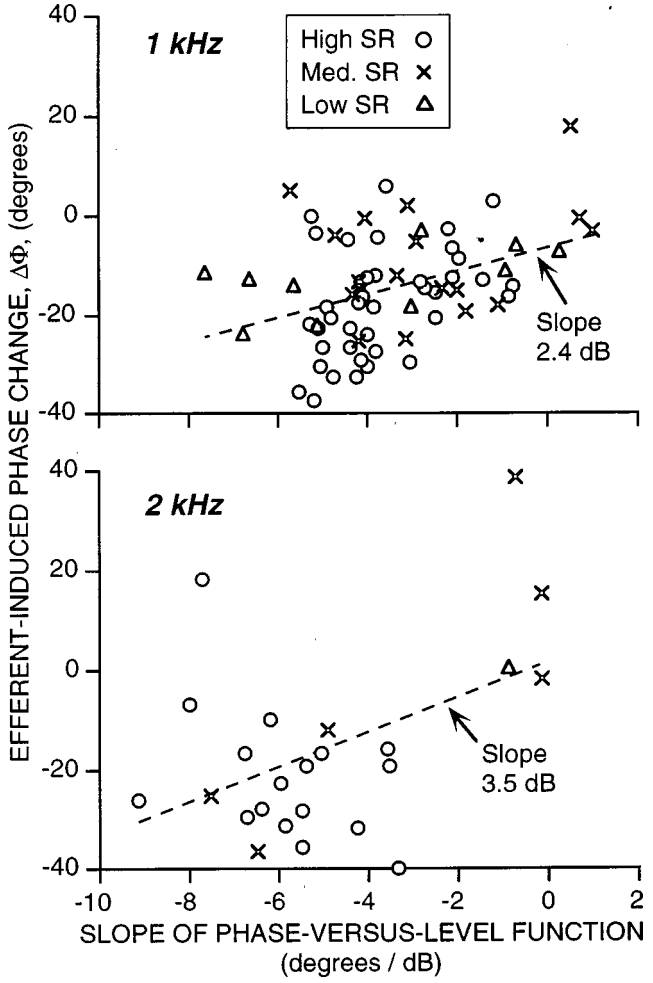


FIG. 11. The efferent-induced change of phase, $\Delta\Phi$, versus the slope of the phase-versus-level function from the same fiber at the same tail frequency in the absence of efferent stimulation. Top: 1-kHz tones. Bottom: 2-kHz tones. Symbols code for spontaneous rate (SR). The dashed lines are regression lines. $\Delta\Phi$ was calculated as explained for $\Delta\Phi_{Ave}$ in the caption for Fig. 10(A), except that here $\Delta\Phi$ included points up to 100 dB SPL in the average to be comparable to the level range over which the slope of the phase-versus-level function was calculated.

dB at 1 kHz and 3.5 dB at 2 kHz. Note that the slopes of the regression lines have units of {degrees}/{degrees per dB} =dB. Thus, the 2.4-dB slope of the 1 kHz regression line means that if efferents produced a 2.4 dB *increase* in the sound signal, then, for the slopes of the phase-versus-sound-level functions, the resulting efferent-induced changes in phase (plus a −6.4° offset at the origin) would fall along the regression line. Thus, it would take an efferent effect in the opposite direction from an attenuation to produce the data shown in Fig. 11.

G. Cochlear microphonic

To help resolve the question of whether efferent effects at tail frequencies are due to efferent-induced changes in basilar-membrane motion, in three cats on which we measured $\Delta\Phi$, we also measured efferent effects on cochlear microphonic recorded with an electrode on or near the round window. The efferent-induced changes in CM in response to tones at four frequencies are shown in Fig. 12. Efferent stimulation increased CM magnitude and had a small effect (a few degrees, on average) on the CM phase. Similar small efferent-induced changes in CM phase were found in guinea pigs by Mountain *et al.* (1980) using intracochlear elec-

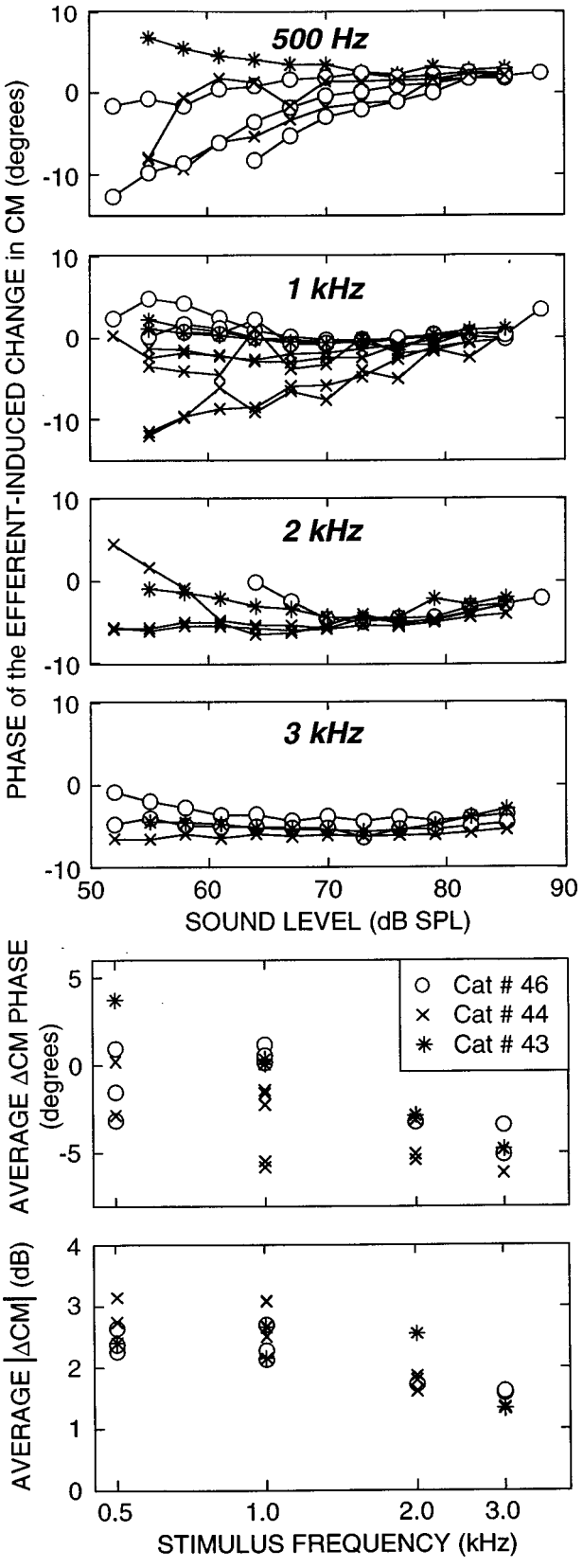


FIG. 12. Efferent effects on the cochlear microphonic (CM) response to tones at four frequencies. The data are from three cats, some with runs done at different times; symbols in inset. In the top four panels, phase data are shown as a function of sound level. In the bottom two panels, phase and amplitude data are shown as a function of sound frequency, with each point being the average across all sound levels of one set of data.

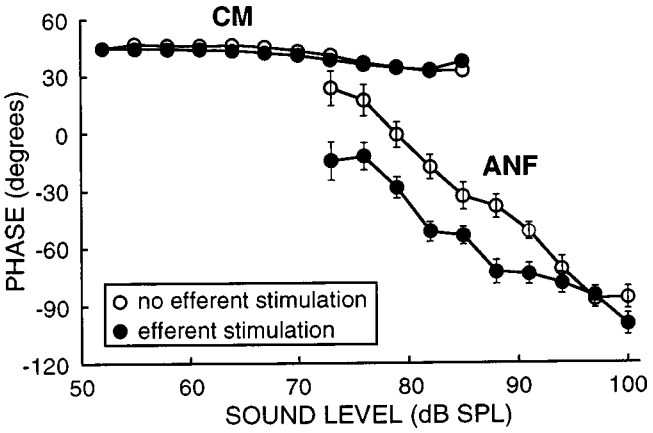


FIG. 13. A comparison between efferent effects on cochlear-microphonic (CM) phase and auditory-nerve-fiber (ANF) phase in the same cat (TS44). The CM phase is from the recording closest in time to (3 h before) the ANF data. The phase references for the two sets of data are not comparable. Tone frequency: 1 kHz.

trodes. The small efferent effect on CM phase contrasts with the much larger efferent-induced changes in auditory-nerve-fiber phase. This is effectively demonstrated from measurements made in the same cat about 3 h apart (Fig. 13). Although there is only a partial overlap of the sound levels for which data were obtained, both the sound-level dependence and the efferent effects on response phase are much larger on the auditory-nerve fiber than on CM.

III. DISCUSSION

Our data show that there are efferent-induced changes in the phase of auditory-nerve-fiber responses to tail-frequency tones. Although errors in phase data are more difficult to quantify than errors in rate, our statistical tests insure that the measured changes in phase are sufficiently accurate. Thus, the large scatter of the phase data, showing both positive and negative phase changes, is real and indicates that there are large fiber-to-fiber differences in efferent effects on phase.

There are no previously published data on efferent-induced changes in auditory-nerve-fiber phase or synchrony at tail frequencies. Although several papers provide data on efferent-induced changes in firing rate in response to tones well below CF, none of these gives phase data (Wiederhold, 1970; Kiang *et al.*, 1970; Guinan and Gifford, 1988b). Similarly, papers on efferent-induced effects on IHC responses and basilar-membrane motion do not provide data on efferent-induced changes in phase (Brown and Nuttall, 1984; Murugasu and Russell, 1996; Russell and Murugasu, 1998).

A. Dependence of $\Delta\Phi$ on tone level and frequency

At tail frequencies, the efferent-induced change in phase, $\Delta\Phi$, varied with tone level and frequency with parametric dependencies similar to those of the efferent-induced change in rate, ΔL_R (see Stankovic and Guinan, 1999). As sound level was lowered toward threshold, both $|\Delta\Phi|$ and ΔL_R increased. Furthermore, as frequency increased, both $|\Delta\Phi|$ and ΔL_R increased slowly up to a few octaves below CF, and decreased sharply at higher frequencies, with ΔL_R

reaching a peak at a somewhat higher frequency than $|\Delta\Phi|$. However, the high-frequency reversals in $|\Delta\Phi|$ and ΔL_R should be regarded with caution because: (1) these reversals were based on data from only a few fibers, (2) the reversal in $|\Delta\Phi|$ was not seen in the data from the few fibers with CFs of 5–10 kHz, and (3) the plots of $\Delta\Phi$ and ΔL_R do not include exactly the same fibers. In particular, the fibers that show the downturn in the ΔL_R plot (Fig. 9 of Stankovic and Guinan, 1999) are not included in the $\Delta\Phi$ plot [Fig. 10(C)] because phase could not be accurately determined at these frequencies (4–5 kHz). If the ΔL_R plot is restricted to include only the fibers that were used in the $\Delta\Phi$ plot of Fig. 10(C), ΔL_R would rise with frequency up to 2.5 octaves below CF; then, the plot would end.

The similar dependencies of $|\Delta\Phi|$ and ΔL_R on tone level and frequency suggest that both efferent-induced changes are due to similar underlying causes. However, there was very little correlation between ΔL_R and $\Delta\Phi$ in scatter plots of the data from single fibers ($r=0.15$). These two observations, taken together, suggest that there is similarity but also differences in the efferent-induced cochlear changes that produce $\Delta\Phi$ and ΔL_R.

B. Is the efferent effect equivalent to an attenuation of sound?

Our data indicate that, at tail frequencies, efferent effects on synchrony and firing rate are equivalent to attenuating the sound, but efferent effects on phase are *not* (Figs. 1–5, 13; Stankovic and Guinan, 1999). A similar pattern was reported for low-CF fibers stimulated with tones at frequencies within their tuning-curve tips (Gifford and Guinan, 1983). Are the different effects on phase versus synchrony and firing rate, and the low correlation between $\Delta\Phi$ and ΔL_R, evidence that different processes produce these effects? Different processes cannot be ruled out but the most parsimonious explanation is that one process produces the efferent-induced changes in phase, rate, and synchrony, and that this process is *not* entirely equivalent to an attenuation of sound. In addition, we do not know if efferent effects at tail frequencies are due to the processes that produce efferent effects at CF, or if different processes are involved.

The finding that efferents have no significant effect on synchrony at high sound levels is consistent with the view that synchrony is ''saturated'' at high sound levels, and that the processes that produce this saturation are not affected by efferents. This, and the attenuation-like efferent changes in rate and synchrony, are consistent with the interpretation that efferents act at a stage before the spike-generation mechanism of the afferent synapse, which presumably sets rate and synchrony.

C. Is the efferent-induced phase change, $\Delta\Phi$, due to a change in basilar-membrane motion?

There are two lines of evidence that strongly suggest that the efferent-induced phase change at tail frequencies, $\Delta\Phi$, is *not* entirely due to a change in the phase of basilar-membrane motion. First, at frequencies an octave or more below CF, the phase of basilar-membrane motion appears to change very little with changes in sound level (Nuttall and Dolan, 1996; Ruggero *et al.*, 1997). Furthermore, at these frequencies basilar-membrane motion (phase and amplitude) is insensitive to a variety of insults (Rhode, 1973; Ruggero and Rich, 1991; Ruggero *et al.*, 1996b, 1997; Nuttall and Dolan, 1996). Both of these observations suggest that the phase of basilar-membrane motion at low frequencies is not readily changed by alterations of cochlear properties. Thus, it seems unlikely that the phase of basilar-membrane motion in response to low-frequency tones would be changed by efferent stimulation. A caveat is that basilar membrane phase is usually plotted on a scale that makes it difficult to see changes of the size produced by efferent stimulation. In contrast, the phase changes in auditory-nerve-fiber responses produced by variation of sound level (which are often 90°–180°, Fig. 6) would be readily visible in plots of basilar-membrane (BM) motion. Such large phase changes have not been found for basilar-membrane motion at tail frequencies, although they have been found at tip frequencies (Nuttall and Dolan, 1996; Ruggero *et al.*, 1997 see also Cooper and Rhode, 1992). Thus, there is good evidence that, at tail frequencies, auditory-nerve-fiber phase does not mimic basilar-membrane phase (see also Ruggero and Rich, 1983, 1987; Ruggero *et al.*, 1996a).

A second line of evidence that $\Delta\Phi$ is *not* entirely due to a change in the phase of basilar-membrane motion is that at sound levels at which efferent stimulation produced large changes in the tail-frequency phase of high-CF auditory-nerve fibers, efferent stimulation produced much smaller changes in the phase of CM (Figs. 12 and 13). The phase of CM recorded from the round window indicates the phase of OHC receptor currents and considerable evidence supports the interpretation that *for low-frequency sounds*, CM phase also indicates the phase of basilar-membrane displacement in the basal turn (Dallos *et al.*, 1974; Dallos and Cheatham, 1976; Patuzzi *et al.*, 1989; Ruggero *et al.*, 1986b). Although it is possible that the presence of the masking tone might influence the efferent effect on CM, or that efferents may change the phase of BM motion and somehow make a compensating change in the CM produced by OHCs in response to the changed BM motion, such a cancellation of effects seems unlikely. Together, the two lines of evidence make a strong case that $\Delta\Phi$ is *not* entirely due to a change in basilar-membrane motion. However, direct measurements of efferent effects on basilar-membrane phase are necessary to determine what part, if any, of the efferent effect on phase is present in basilar-membrane motion.

From the above, the efferent-induced change in auditory-nerve-fiber phase, $\Delta\Phi$, appears to be due mostly to changes that take place after basilar-membrane motion but before synchrony and firing rate are determined at the point of spike initiation in auditory-nerve fibers. This conclusion lends support to the hypothesis that the efferent change in rate, ΔL_R, is also due *mostly* to changes which take place after basilar-membrane motion (Stankovic and Guinan, 1999).

D. What are the mechanisms by which medial efferents change the response phase of auditory-nerve fibers?

If the efferent-induced change in phase, $\Delta\Phi$, of responses to tail-frequency tones is not due to changes in basilar-membrane motion, what is it due to? First, we point out that our shocks selectively stimulate medial olivocochlear efferents and that $\Delta\Phi$ is almost certainly due to fast inhibition from medial efferents (detailed reasons in Stankovic and Guinan, 1999). In the basal turn of the cat cochlea, medial efferents terminate almost exclusively on OHCs (Guinan *et al.*, 1984). Fast medial-efferent effects appear to be due to activation of alpha-9 acetylcholine receptors on OHCs producing an increase in local intracellular OHC calcium concentration that activates Ca^{2+}-dependent K^+ channels in the OHC basolateral membrane (Housley and Ashmore, 1991; Doi and Ohmori, 1993; Kakehata *et al.*, 1994; Fuchs, 1996; Vetter *et al.*, 1999). This activation of potassium channels is the only known efferent effect fast enough to produce the efferent-induced phase and rate changes reported here and by Stankovic and Guinan (1999) [the increase in OHC calcium concentration may produce other effects (Sridhar *et al.*, 1995, 1997; Sziklai *et al.*, 1996; Dallos *et al.*, 1997) but these appear to be more than an order of magnitude too slow to account for our data].

The efferent-induced change in the OHC synaptic conductance does three things that might act as agents causing a phase change. It produces: (1) an increase in CM (Fex, 1959), (2) the MOC potential (Fex, 1967), and (3) an OHC hyperpolarization (this hyperpolarization has never been directly measured, but the indirect evidence for it is strong—reviewed by Guinan, 1996). The increase in CM and the MOC potential are extracellular potentials, so it is clear that they might have effects outside of OHCs (e.g., at IHCs), but the OHC hyperpolarization is an intracellular potential and might be thought to act only within OHCs. However, since the cochlea is a bidirectionally coupled system (Weiss, 1982; Kiang *et al.*, 1986; Mountain and Cody, 1989; Patuzzi, 1996), changes in OHCs (e.g., mechanical changes) may produce effects outside of the OHCs.

A useful way to approach the problem is to ask what are the factors that affect auditory-nerve-fiber phase (relative to basilar-membrane phase), and then to consider which of these might be influenced by efferent stimulation. Three factors have been identified by Cheatham and Dallos (1999) as influencing the phase of basal-turn fibers responding to low-frequency sound in guinea pigs: (1) IHC-membrane low-pass filtering, (2) the dynamics of IHC ion channels, and (3) extracellular potentials affecting the IHC transmembrane potential and transmitter release. To this list we add (4) changes in cochlear micromechanics (i.e., the mechanical coupling of basilar-membrane motion to the bending of IHC stereocilia).

It seems unlikely that the first two of these, i.e., IHC-membrane low-pass filtering, and the dynamics of IHC ion channels, can be primarily responsible for the efferent-induced phase change at tail frequencies. It is possible that the phase of IHC receptor potentials might be affected by changes in the corner frequency of the IHC membrane filter, and that corner-frequency changes are produced by IHC conductance changes (from opening transduction or other ion channels). In addition, the phase of IHC receptor potentials at low frequencies might be changed by the dynamic properties of IHC ion channels (e.g., by the opening or closing of IHC channels lagging the IHC receptor potential; Kros and Crawford, 1990; van Emst *et al.*, 1998). If a significant number of IHC ion channels are opened or closed by receptor potentials, then channel dynamics could have a strong influence on the phase of low-frequency receptor potentials. Since receptor potential amplitude depends on sound level, such an effect would be expected to produce a sound-level dependence of phase. If the sound-level dependence of auditory-nerve-fiber phase at tail frequencies were produced by such a mechanism, then medial-efferent inhibition, which must reduce IHC receptor potentials, would produce a phase change by this mechanism. However, at tail frequencies the change in phase produced by medial efferents is in the opposite direction of the change expected from the efferent reduction of IHC receptor potentials. This argues strongly against $\Delta\Phi$ being produced by the dynamics of IHC ion channels, and to some extent against the sound-level dependence of phase at tail frequencies being produced by this mechanism. While this argument does not rule out such level-dependent effects taking place, it does mean that they cannot be the only way that efferents change the phase of auditory-nerve responses at tail frequencies.

The third element in the list, the influence of extracellular potentials on IHCs, can be expanded into the influences of dc and ac extracellular potentials. Initially we hypothesized that efferent inhibition at tail frequencies was due to the MOC potential, the dc (i.e., slow) extracellular voltage change produced by efferent activation (Fex, 1967). How could a dc potential change the *phase* of auditory-nerve firing? Perhaps the MOC potential produces a small change in IHC transmembrane potential and this changes the operating point of voltage-sensitive IHC channels, which changes receptor potential phases by affecting IHC membrane low-pass filtering, or the dynamics of IHC ion channels (as in the previous paragraph). Alternately, the MOC potential acting on the dendrites of auditory-nerve fibers might affect the production of action potentials and produce a phase change, but it is difficult to see how this might account for the various parametric dependencies of $\Delta\Phi$. All considered, it seems unlikely that the MOC potential is a major factor in producing $\Delta\Phi$.

The extracellular ac potential, the cochlear microphonic, is a viable candidate for producing $\Delta\Phi$. The efferent effect on CM is unusual in that during efferent stimulation CM *increases*, while basilar-membrane motion and other cochlear potentials *decrease* (Galambos, 1956; Fex, 1959; Brown and Nuttall, 1984; Murugasu and Russell, 1996). The regression lines in Fig. 11 indicate that 2.4-dB (at 1 kHz) and 3.5-dB (at 2 kHz) efferent-induced *increases* of drive to the mechanism that creates the phase change with level would be enough to produce the measured $\Delta\Phi$s. These values are similar to, or greater than, the measured increases in CM of 2.6 dB (at 1 kHz) and 1.9 dB (at 2 kHz) (Fig. 12). Alternately, the relevant factor may be the ratio of CM to some other signal derived from the sound stimulus, e.g., the IHC

receptor potential. Since we do not know the value of the efferent-induced decrease in the IHC receptor potential, we will use the efferent-induced inhibition of firing rate, ΔL_R, (a 2.3-dB decrease at 1 kHz, or 4.8 dB decrease at 2 kHz from the data of Stankovic and Guinan, 1999) as an estimate of the efferent effect on the signal that CM interacts with. The resulting CM/signal ratios are 4.9 and 6.7 dB (at 1 and 2 kHz, respectively) which are larger than the needed increases. However, if CM and the other signal have different phases (as presumably they would in order to produce the phase change), the effective change in CM/signal ratio would not be a simple sum. All considered, these data show that the efferent-induced increase in CM is in the correct range to produce the data in Fig. 11. Furthermore, if CM produces $\Delta\Phi$, it seems likely that CM is also involved in producing the sound-level dependence of auditory-nerve-fiber phase at tail frequencies because the data of Fig. 11 show that $\Delta\Phi$ and the sound-level dependence of phase are related.

The above reasoning suggests that CM may be involved in producing $\Delta\Phi$ but does not tell us where the interaction takes place. One possibility is that CM may influence auditory-nerve firing by changing the transmembrane potential at the IHC synapse (Sellick *et al.*, 1982; Russell and Sellick, 1983; Ruggero and Rich, 1983; Ruggero *et al.*, 1986a; Guinan and Gifford, 1988a; Cody and Mountain, 1989; Cheatham and Dallos, 1999). Presumably this putative electric coupling between OHCs and IHCs would have a significant effect near threshold at tail frequencies because (1) the mechanical response of the cochlea is small near threshold, and (2) the electrical response (i.e., CM) is large because basilar-membrane motion is in phase over a considerable length of the cochlea. In the guinea-pig basal turn, CM is bigger than the IHC receptor potential for very low frequency tones, although not for midfrequency tones (Russell and Sellick, 1983). However, even if CM is less than the IHC receptor potential, it can still have an influence on ANF response phase (Cheatham and Dallos, 1999).

A second way that CM might produce $\Delta\Phi$ is by producing a mechanical change in OHCs. The OHC mechanical change would have to change auditory-nerve-fiber phase while producing only a small change in CM phase, and would also have to depend on tone level and frequency. Frequency-dependent phase changes might be produced if efferents changed a mechanical resonance; level-dependent changes would be produced if a mechanical nonlinearity was involved. One possibility is that efferents cause a fast mechanical change in OHCs that is sensitive to the ratio of extracellular CM to OHC intracellular receptor potential (see Dallos and Evans, 1995, for one such mechanism). Another possibility is that efferent activity changes OHC stereocilia stiffness and that this ultimately changes the coupling between basilar membrane motion and the bending of IHC stereocilia (this might be accomplished by CM—or the MOC potential—changing the Ca^{2+} entry into stereocilia; see Lumpkin and Hudspeth, 1998).

A final possibility, the one we think is most likely, is that efferents produce $\Delta\Phi$ by producing a mechanical change in OHCs, but without CM being involved. This hypothesis has most of the advantages and disadvantages discussed above for the CM-mechanical hypothesis. Recent work indicates that the excitation of auditory-nerve fibers is through at least three excitation drives that might correspond to cochlear vibrational modes, and that the mix of these excitation drives changes with sound level (Lin and Guinan, 2000). In addition, efferent stimulation appears to affect some excitation drives more than others (Lin and Guinan, 1999). Since efferent stimulation appears to reduce cochlear active processes (Guinan, 1996) it seems reasonable to think that efferent stimulation might change the mixture of modes by reducing the importance of modes that depend on active processes, and that increasing the sound level might also do this. Thus, both efferent activation and high-level sound might cause the cochlea to respond with a similar mix of modes.

In summary, there are two main candidates for producing $\Delta\Phi$: (1) the efferent-induced increase in CM, and (2) an efferent-induced mechanical change. However, whatever the actual mechanisms involved in producing $\Delta\Phi$, both the $\Delta\Phi$ data, and the large change in auditory-nerve-fiber phase with sound level at tail frequencies, indicate that the common conception that auditory-nerve-fiber tuning-curve tails represent a passive, constant coupling of basilar-membrane motion to the drive to auditory-nerve fibers is no longer tenable (see also Ruggero and Rich, 1983, 1987; Ruggero *et al.*, 1996a).

IV. CONCLUSIONS

(i) Medial-efferent activation produces a change in phase, $\Delta\Phi$, in auditory-nerve-fiber responses to tail-frequency tones. The largest changes are at sound levels near threshold and for tones 3–4 octaves below CF.

(ii) Efferent stimulation had only small effects on the phase of CM at tail frequencies, which indicates that the efferent-induced changes in auditory-nerve-fiber phase are unlikely to be due entirely to efferent-induced changes in basilar-membrane phase.

(iii) Auditory-nerve fiber responses to tail-frequency tones showed large level-dependent phase changes (90°–180°). The $\Delta\Phi$ from a fiber was correlated with the slope of its phase versus sound-level function as if $\Delta\Phi$ were produced by an *increase* in sound level. This correlation suggests that the processes involved in producing $\Delta\Phi$ are also involved in producing the change of phase with sound level at tail frequencies.

(iv) Both the large change in auditory-nerve-fiber phase with sound level and the efferent-induced change in phase indicate that the common conception that auditory-nerve-fiber tuning-curve tails represent a passive, constant coupling of basilar-membrane motion to the drive of auditory-nerve fibers needs revision.

ACKNOWLEDGMENTS

We thank Dr. M. C. Brown, Dr. M. C. Liberman, and Dr. C. A. Shera for comments on the manuscript and Dr. D. K. Eddington, Dr. D. C. Mountain, Dr. J. B. Nadol, Jr., Dr.

W. F. Sewell, and Dr. C. A. Shera for comments on the thesis that led to this paper. Supported (in part) by Research Grant No. 5 RO1 DC 00235 and Training Grant No. T32 DC00038, from the National Institute of Deafness and Other Communication Disorders, and by Medical Scientist Training Program Grant No. T3GM 07753, National Institutes of Health.

Anderson, D. J., Rose, J. E., Hind, J. E., and Brugge, J. F. (**1971**). ''Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: Frequency and intensity effects,'' J. Acoust. Soc. Am. **49**, 1131–1139.

Brown, M. C., and Nuttall, A. L. (**1984**). ''Efferent control of cochlear inner hair cell responses in the guinea pig,'' J. Physiol. (London) **354**, 625–646.

Cai, Y., and Geisler, C. D. (**1996**). ''Temporal patterns of the responses of auditory-nerve fibers to low-frequency tones,'' Hear. Res. **96**, 83–93.

Cheatham, M. A., and Dallos, P. (**1992**). ''Two-tone suppression in inner hair cell responses: Correlates of rate suppression in the auditory nerve,'' Hear. Res. **60**, 1–12.

Cheatham, M. A., and Dallos, P. (**1999**). ''Response phase: A view from the inner hair cell,'' J. Acoust. Soc. Am. **105**, 799–810.

Cleveland, W. S. (**1993**). *Visualizing Data* (AT&T Bell Laboratories, Murray Hill, NJ).

Cody, A. R., and Mountain, D. C. (**1989**). ''Low-frequency responses of inner hair cells: Evidence for a mechanical origin of peak splitting,'' Hear. Res. **41**, 89–99.

Cooper, N. P., and Rhode, W. S. (**1992**). ''Basilar membrane mechanics in the hook region of cat and guinea-pig cochleae: Sharp tuning and nonlinearity in the absence of baseline position shifts,'' Hear. Res. **63**, 163–190.

Dallos, P., and Cheatham, M. A. (**1976**). ''Production of cochlear potentials by inner and outer hair cells,'' J. Acoust. Soc. Am. **60**, 510–512.

Dallos, P., Cheatham, M. A., and Ferraro, J. (**1974**). ''Cochlear mechanics, nonlinearities, and cochlear potentials,'' J. Acoust. Soc. Am. **55**, 597–605.

Dallos, P., and Evans, B. N. (**1995**). ''High-frequency motility of outer hair cells and the cochlear amplifier,'' Science **267**, 2006–2009.

Dallos, P., He, D. Z., Lin, X., Sziklai, I., Mehta, S., and Evans, B. N. (**1997**). ''Acetylcholine, outer hair cell electromotility, and the cochlear amplifier,'' J. Neurosci. **17**, 2212–2226.

Doi, T., and Ohmori, H. (**1993**). ''Acetylcholine increases intracellular Ca^{2+} concentration and hyperpolarizes the guinea-pig outer hair cell,'' Hear. Res. **67**, 179–188.

Efron, B., and Tibshirani, R. J. (**1993**). *An Introduction to the Bootstrap* (Chapman & Hall, New York).

Fex, J. (**1959**). ''Augmentation of cochlear microphonic by stimulation of efferent fibers to the cochlea,'' Acta Oto-Laryngol. **50**, 540–541.

Fex, J. (**1967**). ''Efferent inhibition in the cochlea related to hair-cell dc activity: Study of postsynaptic activity of the crossed olivo-cochlear fibers in the cat,'' J. Acoust. Soc. Am. **41**, 666–675.

Fuchs, P. A. (**1996**). ''Synaptic transmission at vertebrate hair cells,'' Curr. Opin. Neurobiol. **6**, 514–519.

Galambos, R. (**1956**). ''Suppression of auditory activity by stimulation of efferent fibers to the cochlea,'' J. Neurophysiol. **19**, 424–437.

Gifford, M. L., and Guinan, Jr., J. J. (**1983**). ''Effects of crossed-olivocochlear-bundle stimulation on cat auditory nerve fiber responses to tones,'' J. Acoust. Soc. Am. **74**, 115–123.

Goldberg, J. M., and Brown, P. (**1969**). ''Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization,'' J. Neurophysiol. **32**, 613–636.

Guinan, Jr., J. J. (**1996**). ''The Physiology of Olivocochlear Efferents,'' in *The Cochlea*, edited by P. J. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 435–502.

Guinan, Jr., J. J., and Gifford, M. L. (**1988a**). ''Effects of electrical stimulation of efferent olivocochlear neurons on cat auditory-nerve fibers. II. Spontaneous rate,'' Hear. Res. **33**, 115–128.

Guinan, Jr., J. J., and Gifford, M. L. (**1988b**). ''Effects of electrical stimulation of efferent olivocochlear neurons on cat auditory-nerve fibers. III. Tuning curves and thresholds at CF,'' Hear. Res. **37**, 29–46.

Guinan, Jr., J. J., Warr, W. B., and Norris, B. E. (**1984**). ''Topographic organization of the olivocochlear projections from the lateral and medial zones of the superior olivary complex,'' J. Comp. Neurol. **226**, 21–27.

Housley, G. D., and Ashmore, J. F. (**1991**). ''Direct measurement of the action of acetylcholine on isolated outer hair cells of the guinea pig cochlea,'' Proc. R. Soc. London Ser. **244**, 161–167.

Johnson, D. H. (**1974**). ''The response of single auditory-nerve fibers in the cat to single tones: Synchrony and average discharge rate,'' Ph.D. thesis, Massachusetts Institute of Technology.

Johnson, D. H. (**1980**). ''The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones,'' J. Acoust. Soc. Am. **68**, 1115–1122.

Kakehata, S., Nakagawa, T., Takasaka, T., and Akaike, N. (**1994**). ''Cellular mechanism of acetylcholine-induced response in dissociated outer hair cells of guinea-pig cochlea,'' J. Physiol. (London) **463**, 227–244.

Kiang, N. Y. S. (**1990**). ''Curious oddments of auditory-nerve studies,'' Hear. Res. **49**, 1–16.

Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT Press, Cambridge, MA).

Kiang, N. Y. S., Moxon, E. C., and Levine, R. A. (**1970**). ''Auditory-nerve activity in cats with normal and abnormal cochleas,'' in *Ciba Foundation Symposium on Sensorineural Hearing Loss*, edited by G. E. W. Wolstenholme and J. Knight (Churchill, London), pp. 241–273.

Kiang, N. Y. S., Lieberman, M. C., Sewell, W. F., and Guinan, Jr., J. J. (**1986**). ''Single unit clues to cochlear mechanisms,'' Hear. Res. **22**, 171–182.

Konishi, T., and Nielsen, D. W. (**1978**). ''The temporal relationship between basilar membrane motion and nerve impulse initiation in auditory nerve fibers of guinea pigs,'' Jpn. J. Physiol. **28**, 291–307.

Kros, C. J., and Crawford, A. C. (**1990**). ''Potassium currents in inner hair cells isolated from the guinea-pig cochlea,'' J. Physiol. (London) **421**, 263–291.

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 Kiang, N. Y. S. (**1984**). ''Single-neuron labeling and chronic cochlear pathology. IV. Stereocilia damage and alterations in rate- and phase-level functions,'' Hear. Res. **16**, 75–90.

Lin, T., and Guinan, Jr., J. J. (**1999**). ''Organ of Corti vibration in modes: Supporting evidence from auditory-nerve-fiber responses to clicks and clicks with efferent stimulation,'' *Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Takasaka, K. Ikeda, and K. Ohyama (World Scientific, Singapore).

Lin, T., and Guinan, Jr., J. J., (**2000**). ''Auditory-nerve-fiber responses to high-level clicks: Interference patterns indicate that excitation is due to the combination of multiple drives,'' J. Acoust. Soc. Am. **107**, 2615–2630.

Lumpkin, E. A., and Hudspeth, A. J. (**1998**). ''Regulation of free Ca^{2+} concentration in hair-cell stereocilia,'' J. Neurosci. **18**, 6300–6318.

Mardia, K. V. (**1972**). *Statistics of Directional Data* (Academic Press, New York).

Mountain, D. C., and Cody, A. R. (**1989**). ''Mechanical coupling between inner and outer hair cells,'' in *Cochlear Mechanisms. Structure, Function and Models*, edited by J. P. Wilson and D. T. Kemp (Plenum, New York), pp. 153–160.

Mountain, D. C., Geisler, C. D., and Hubbard, A. E. (**1980**). ''Stimulation of efferents alters the cochlear microphonic and the sound induced resistance changes measured in scala media of the guinea pig,'' Hear. Res. **3**, 231–240.

Murugasu, E., and Russell, I. J. (**1996**). ''The effect of efferent stimulation on basilar membrane displacement in the basal turn of the guinea pig cochlea,'' J. Neurosci. **16**, 325–332.

Nuttall, A. L., and Dolan, D. F. (**1996**). ''Steady-state sinusoidal velocity responses of the basilar membrane in guinea pig,'' J. Acoust. Soc. Am. **99**, 1556–1555.

Oshima, W., and Strelioff, D. (**1983**). ''Responses of gerbil and guinea pig auditory nerve fibers to low- frequency sinusoids,'' Hear. Res. **12**, 167–184.

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.

Patuzzi, R. (**1996**). ''Cochlear Micomechanics and Macromechanics,'' in *The Cochlea*, edited by P. J. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 186–257.

Patuzzi, R. B., Yates, G. K., and Johnstone, B. M. (**1989**). ''The origin of the low-frequency microphonic in the first cochlear turn of guinea-pig,'' Hear. Res. **39**, 177–188.

Rhode, W. S. (**1973**). ''An investigation of postmortem cochlear mechanics

using the Mössbauer effect,'' in *Basic Mechanisms of Hearing*, edited by A. R. Møller (Academic, New York), pp. 49–67.
Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (**1967**). ''Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey,'' J. Neurophysiol. **30**, 769–793.
Ruggero, M. A. (**1992**). ''Physiology and Coding of Sound in the Auditory Nerve,'' in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay (Springer, New York), Vol. 2, pp. 34–93.
Ruggero, M. A., and Rich, N. C. (**1983**). ''Chinchilla auditory-nerve responses to low-frequency tones,'' J. Acoust. Soc. Am. **73**, 2096–2018.
Ruggero, M. A., and Rich, N. C. (**1987**). ''Timing of spike initiation in cochlear afferents: Dependence on site of innervation,'' J. Neurophysiol. **58**, 379–403.
Ruggero, M. A., Robles, L., and Rich, N. C. (**1986a**). ''Cochlear microphonics and the initiation of spikes in the auditory nerve: Correlation of single-unit data with neural receptor potentials recorded from the round window,'' J. Acoust. Soc. Am. **79**, 1491–1498.
Ruggero, M. A., Robles, L., and Rich, N. C. (**1986b**). ''Basilar membrane mechanics at the base of the chinchilla cochlea. II. Responses to low-frequency tones and relationship to microphonic and spike initiation in the VIII nerve,'' J. Acoust. Soc. Am. **80**, 1375–1383.
Ruggero, M. A., and Rich, N. C. (**1991**). ''Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane,'' J. Neurosci. **11**, 1057–1067.
Ruggero, M. A., Rich, N. C., Shivapuja, B. G., and Temchin, A. N. (**1996a**). ''Auditory-nerve responses to low-frequency tones: Intensity dependence,'' Aud. Neurosci. **2**, 159–185.
Ruggero, M. A., Rich, N. C., and Recio, A. (**1996b**). ''The effect of intense acoustic stimulation on basilar-membrane vibrations,'' Aud. Neurosci. **2**, 329–345.
Ruggero, M. A., Rich, N. C., Recio, A. Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Russell, I. J., and Murugasu, E. (**1998**). ''Efferent suppression of basilar membrane vibration depends on tone frequency and level: Implications for the active control of basilar membrane mechanics,'' in *Psychophysical and Physiological Advances in Hearing. Proceedings of the 11th International Symposium on Hearing*, edited by A. R. Palmer, A. Rees, and A. Q. Summerfield (Whurr, London), pp. 19–24.
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.
Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). ''Modulation of responses of spiral ganglion cells in guinea pig cochlea by low frequency sound,'' Hear. Res. **7**, 199–221.
Sokolich, W. G., Hamernik, R. P., Zwislocki, J. J., and Schmiedt, R. A. (**1976**). ''Inferred response polarities of cochlear hair cells,'' J. Acoust. Soc. Am. **59**, 963–974.
Sridhar, T. S., Liberman, M. C., Brown, M. C., and Sewell, W. F. (**1995**). ''A novel cholinergic 'slow effect' of olivocochlear stimulation on cochlear potentials in the guinea pig,'' J. Neurosci. **15**, 3667–3678.
Sridhar, T. S., Brown, M. C., and Sewell, W. F. (**1997**). ''Unique postsynaptic signaling at the hair cell efferent synapse permits calcium to evoke changes on two different time scales,'' J. Neurosci. **17**, 428–437.
Stankovic, K. M., and Guinan, Jr., J. J. (**1999**). ''Medial efferent effects on auditory-nerve responses to tail-frequency tones I. Rate reduction,'' J. Acoust. Soc. Am. **106**, 857–869.
Sziklai, I., He, D. Z. Z., and Dallos, P. (**1996**). ''Effect of acetylcholine and GABA on the transfer function of electromotility in isolated outer hair cells,'' Hear. Res. **95**, 87–99.
van Emst, M. G., Giguere, C., and Smoorenburg, G. F. (**1998**). ''The generation of DC potentials in a computational model of the organ of Corti: Effects of voltage-dependent K+ channels in the basolateral membrane of the inner hair cell,'' Hear. Res. **115**, 184–196.
Vetter, D. E., Liberman, M. C., Mann, J., Bahanin, J., Boulter, J., Brown, M. C., Saffiote-Kolman, J., Heinemann, S., and Elgoyhen, A. B. (**1999**). ''Role of alpha 9 nicotinic ACh receptor subunits in the development and function of cochlear efferent innervation,'' Neuron **23**, 93–103.
Weiss, T. F. (**1982**). ''Bidirectional transduction in vertebrate hair cells: A mechanism for coupling mechanical and electrical processes,'' Hear. Res. **7**, 353–360.
Wiederhold, M. L. (**1970**). ''Variations in the effects of electric stimulation of the crossed olivocochlear bundle on cat single auditory-nerve-fiber responses to tone bursts,'' J. Acoust. Soc. Am. **48**, 966–977.

A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones

Bernhard Roß,[a] Christian Borgmann, and Rossitza Draganova
Center of Biomagnetism, Institute of Experimental Audiology, University of Münster, Kardinal-von-Galen-Ring 10, D-48129 Münster, Germany

Larry E. Roberts
Department of Psychology, McMaster University, Hamilton, Ontario L8S 4K1, Canada

Christo Pantev
Center of Biomagnetism, Institute of Experimental Audiology, University of Münster, Kardinal-von-Galen-Ring 10, D-48129 Münster, Germany

(Received 3 November 1999; revised 31 January 2000; accepted 18 May 2000)

The cerebral magnetic field of the auditory steady-state response (SSR) to sinusoidal amplitude-modulated (SAM) tones was recorded in healthy humans. The waveforms of underlying cortical source activity were calculated at multiples of the modulation frequency using the method of source space projection, which improved the signal-to-noise ratio (SNR) by a factor of 2 to 4. Since the complex amplitudes of the cortical source activity were independent of the sensor position in relation to the subject's head, a comparison of the results across experimental sessions was possible. The effect of modulation frequency on the amplitude and phase of the SSR was investigated at 30 different values between 10 and 98 Hz. At modulation frequencies between 10 and 20 Hz the SNR of harmonics near 40 Hz were predominant over the fundamental SSR. Above 30 Hz the SSR showed an almost sinusoidal waveform with an amplitude maximum at 40 Hz. The amplitude decreased with increasing modulation frequency but was significantly different from the magnetoencephalographic (MEG) background activity up to 98 Hz. Phase response at the fundamental and first harmonic decreased monotonically with increasing modulation frequency. The group delay (apparent latency) showed peaks of 72 ms at 20 Hz, 48 ms at 40 Hz, and 26 ms at 80 Hz. The effects of stimulus intensity, modulation depth, and carrier frequency on amplitude and phase of the SSR were also investigated. The SSR amplitude decreased linearly when stimulus intensity or the modulation depth were decreased in logarithmic steps. SSR amplitude decreased by a factor of 3 when carrier frequency increased from 250 to 4000 Hz. From the phase characteristics, time delays were found in the range of 0 to 6 ms for stimulus intensity, modulation depth, and carrier frequency, which were maximal at low frequencies, low intensities, or maximal modulation depth. [S0001-4966(00)04708-1]

PACS numbers: 43.64.Ri, 43.64.Qh, 43.64.Yp [LHC]

INTRODUCTION

Depending upon the time course of the auditory stimulation, cerebral-evoked responses can be generally classified as transient or steady state. Responses evoked by auditory stimuli that are presented at interstimulus intervals long enough to allow intervening brain activity to subside are called transient-evoked responses.[1] If the interstimulus interval is shortened such that successive responses begin to overlap, a more complex compound response is generated. This response is referred to as a steady-state response (SSR).[2,3] Regan[4] has described the SSR as responses ''whose constituent discrete frequency components remain constant in amplitude and phase over an infinitely long time period'' (p. 35). Despite an increasing number of animal[5–9] and human[10–18] studies of the SSR during the last 15 years, there is still no general agreement with respect to the mechanisms or neural sources underlying SSR genesis. A cortical origin of the SSR seems most probable, although subcortical contributions from polysensory thalamic areas can not be excluded. Epidural recordings taken from the temporal cortex of rats and cats[5,8,19–21] are consistent with the conclusion that steady-state potentials evoked by auditory stimuli arise within the auditory cortex. Epidural recordings in humans[22,23] and electroencephalographic (EEG) and magnetoencephalographic (MEG) field patterns evoked by auditory stimuli[11,12,15,24–26] also suggest an active source for the auditory SSR in the supratemporal plane.

The stimuli most often used for eliciting the SSR are click trains, trains of short tone bursts, or amplitude modulated tones. Because the SSR is a periodic waveform, it has harmonic structure and can be characterized by the response parameters amplitude and phase. These response parameters depend strongly on stimulus parameters such as the stimulus rate or modulation frequency, stimulus intensity, carrier frequency, and modulation depth. This dependence is usually described by input–output characteristics (IOCs) which depict response amplitude or phase as stimulus parameters are

[a]Author to whom correspondence should be addressed. Electronic mail: bernhard.ross@ieee.org

manipulated over their dynamic range. Unfortunately, in all previous SSR studies conducted in humans, IOCs for different stimulus and response parameters have been determined for the most part on different subjects, owing to the impracticality of measuring multiple IOCs in the same subject within a single experimental session. A qualitative description of IOCs has been gained by this method, but a precise description of the neural dynamics underlying SSR generation requires measurement of IOCs within the same subjects and has heretofore been lacking.

The goal of the present study was to measure a range of IOCs within the same subjects, for auditory SSRs elicited by amplitude-modulated pure tones and recorded by MEG. IOCs relating SSR amplitude and phase to the stimulus parameters modulation rate, modulation depth, stimulus intensity, and carrier frequency were investigated. To overcome limitations of previous research imposed by session length, multiple sessions (16 per subject) were used. In addition, IOCs were based, not only on direct measurements of SSR amplitude and phase, but also on measurements of the behavior of cortical sources underlying SSR generation as estimated by the method of source space projection. An important advantage of this approach is that the activity of cortical sources estimated by source space projection are not affected by millimeter variations in the placement of MEG sensors with respect to the brain that inevitably occur between recording sessions. Hence, it was possible to collapse measurements across sessions, so that a high signal-to-noise ratio was achieved for each subject tested.

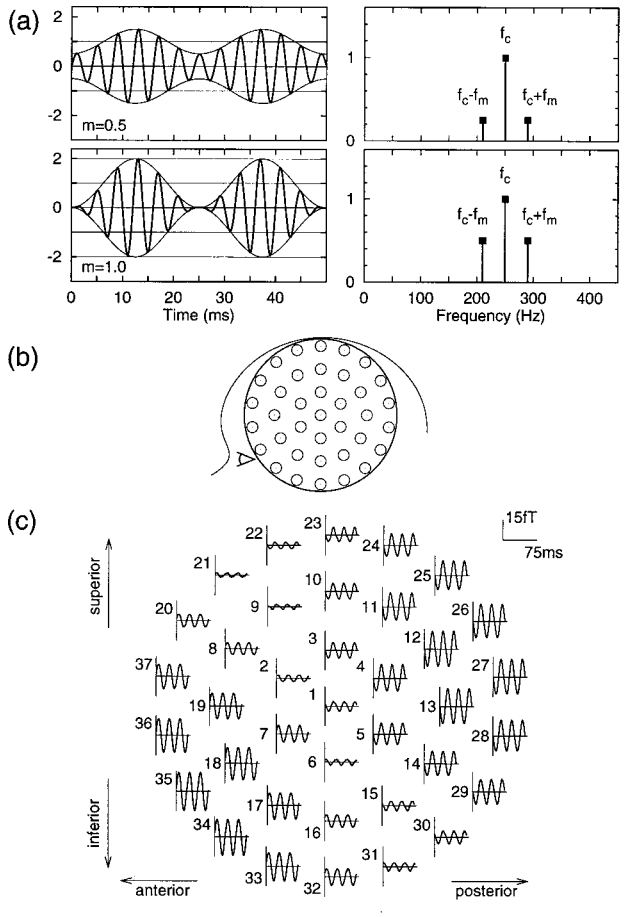


FIG. 1. (a) Time series and amplitude spectrum of sinusoidal modulated tones for modulation depth $m=0.5$ and $m=1$; (b) schematic of the sensor placement over the left hemisphere; and (c) waveforms of magnetic field obtained from 37 channels.

I. METHODS

A. Subjects

Four female and four male right-handed adults aged 22 to 32 years and free of otological or neurological disorders participated in these experiments. Normal audiological status (air conduction hearing threshold of no more than 10 dB hearing level between 250 Hz and 4 kHz) was verified by pure-tone audiometry. All subjects were right-handed according to the Edinburgh handedness questionnaire.[27] Informed consent was obtained from each subject after explaining to her/him the nature of the study. The experimental procedures were conducted in accordance with the Ethics Commission of the University of Münster and the Declaration of Helsinki. Subjects were paid for their participation.

For each subject, 16 experimental sessions of 1-h measurement time were scheduled. No more than two sessions per day were carried out with a half-hour break in between.

B. Stimulation

Sinusoidal amplitude-modulated (SAM) tones were continuously presented during an interval of 200 s. The SAM waveform $y(t)$ was defined by

$$y(t)=a\cdot\sin(2\pi f_c t)\cdot(1-m\cos(2\pi f_m t)), \qquad (2.1)$$

in which a denoted the amplitude of the signal, f_c denoted the carrier frequency, and f_m the modulation frequency. The last term describes the signal's envelope. The modulation depth m is equal to the ratio $(y_{\max}-y_{\min})/(y_{\max}+y_{\min})$, where $y_{\max}$ and $y_{\min}$ are the maximum and minimum of the waveform's envelope, respectively. The modulation depth ranged from zero to 1, and was expressed in percent. Resolving the sine–cosine product into a sum of terms yields the spectral representation of the SAM signal

$$y(t)=a\sin(\omega_c t)-m\frac{a}{2}\sin(\omega_1 t)-m\frac{a}{2}\sin(\omega_2 t), \qquad (2.2)$$

which consists of three distinct spectral lines at $\omega_c=2\pi f_c$, $\omega_1=2\pi(f_c-f_m)$, and $\omega_2=2\pi(f_c+f_m)$. Both the time series and the amplitude spectrum of the stimulus signal are shown in Fig. 1(a). The dependencies of the SSR on the four stimulus parameters, f_m, f_c, a, and m, were investigated in this study. Only integer values in Hz were chosen for the carrier frequency as well for the modulation frequency. Therefore, 1 s was a common multiple of the periodicity of $y(t)$ for all combinations of f_c and f_m, and permitted calculation of the stimulus waveform for periods of 1-s duration. The software for stimulus calculation and presentation, and the graphical user interface for controlling the experiments, were run on a Sun SPARC computer which also served as the front end for data acquisition. A 16-bit digital-to-analog converter (ADDA1418, Analyx, Inc.) was used to convert the stimulus waveform into an audio signal at a rate of 20 000 samples/s. A trigger at the onset of each 1-s stimulus

period was provided by the stimulus device for synchronization with the data acquisition system.

Magnetically silent delivery of the stimuli was provided by a special sound delivery system consisting of speakers (1 in. compression driver, Renkus-Heinz Inc.) mounted outside the magnetically shielded room which were connected to a silicon ear piece through 6.3 m of echo-less plastic tubing (16 mm inner diameter). The frequency characteristic of this system deviated less than ± 10 dB between 200 and 6000 Hz. Since the stimulus was defined in a frequency band no wider than 200 Hz width, stimuli were not distorted by the frequency characteristic of the sound delivery system. The transmission delay of about 19 ms was compensated by an appropriate shift of the trigger signal. Before carrying out the experiments, both the signal spectrum of the stimulus and its correct timing were checked by means of a 2-cm^3 ear simulator (Brüel & Kjær model 4157) that was equipped with an 1/2'' condenser microphone (Brüel & Kjær model 4134) and connected to the silicon ear piece at the end of the sound delivery system.

Four experiments were carried out in which auditory evoked steady-state responses were investigated in detail while stimulus intensity, carrier frequency, modulation frequency, and the modulation depth were manipulated. Within experimental sessions these stimulus parameters were varied one at a time while the others were kept constant. In all experiments the stimuli were presented continuously for 200 s monaurally to the subject's right ear. Stimulus intensity was referred to the individual sensation threshold. For this purpose, the subject's hearing threshold was measured prior to each experimental session by applying tonebursts of 500 ms duration of the relevant stimulus type through the sound delivery system.

1. Stimulus intensity

In order to obtain an optimal signal-to-noise ratio, the carrier frequency f_c of the SAM stimulus signal was set to 250 Hz, and the modulation depth m to 100%. The modulation frequency f_m was set to 39 Hz instead of 40 Hz because 39 Hz had no common factor with the power-line frequency of 50 Hz. The SAM stimulus signal was presented at seven intensity steps of 10 dB each, ranging from 30- to 90-dB sensation level (SL). Each intensity was presented at least twice. For the two lowest intensity steps (30- and 40-dB sensation level), a smaller response amplitude was expected. Therefore, these intensities were repeated four times, in order to increase the signal-to-noise ratio. For each subject, three experimental sessions of 1 h each were carried out.

2. Modulation depth

For the investigation of the effect of changing modulation depth on the SSR amplitude, SAM stimuli with $f_c = 250$ Hz, $f_m = 39$ Hz at an intensity of 70-dB sensation level were presented. The modulation depth m was changed from 100% to 10% in 10%-steps, and additionally $m = 5\%$ was investigated. The stimuli were presented twice for 200 s at $m \geqslant 30\%$ and four times at $m \leqslant 20\%$. Four sessions of 1-h duration each were carried out for each subject.

3. Carrier frequency

The carrier frequency of the SAM-tone signal was investigated at 250, 500, 1000, 2000, and 4000 Hz with a constant modulation frequency of $f_m = 39$ Hz, a modulation depth of 100%, and an intensity of 70-dB sensation level. Experimental blocks of 200-s duration were repeated five times in random order within three experimental sessions.

4. Modulation frequency

The effect of modulation frequency on the SSR was investigated between 10 and 48 Hz in steps of 2 Hz, and between 55 and 100 Hz in steps of 5 Hz. The power-line frequency (50 Hz) and its harmonics and subharmonics were excluded from the study to avoid line frequency noise. Ninety-eight Hz was included as an alternative for 100 Hz. Thus, a total of 30 different modulation frequencies was investigated. The carrier frequency was set to 250 Hz, the modulation depth to 100%, and the intensity to 70-dB sensation level. At the begining of each experimental session, the hearing threshold was determined at $f_m = 39$ Hz. Stimulus amplitudes were kept constant when changing the modulation frequency. Each stimulus was presented at least twice for 200 s, and the different stimuli were presented in random order. For each subject, six sessions of 1 h each were carried out.

C. Data acquisition

Recordings were performed in a magnetically and acoustically shielded room. The subjects rested in right lateral position on a vacuum cast, with their heads lying on molds to permit stable fixation throughout the whole experimental session.

The MEG was recorded with a 37-channel neuromagnetometer (MAGNES, Biomagnetic Technologies). The detection coils of the neuromagnetometer were arranged in a circular concave array with a diameter of 144 mm and a spherical radius of 122 mm. The distance between the centers of the coils was 22 mm; the coil diameter was 20 mm. The sensors were configured as a first-order axial gradiometer with a baseline of 50 mm. The spectral density of the intrinsic noise of each channel was between 5 and 7 $fT/\sqrt{Hz}$ in the frequency range above 1 Hz.

The sensor array was placed over the left temporal plane, centered over a point about 1.5 cm superior to the position T3 of the 10–20 system for electrode placement, as close to the subject's head as possible [Fig. 1(b)]. A sensor-position indicator system determined the spatial locations of the sensors relative to the head and indicated if head movements occurred during the recordings. No head movements sufficient to require discarding of data were observed in the study. During the MEG session, subjects watched cartoon videos that were projected via fiber cable optic into a nonmagnetic display in order to fixate their attention. The subjects were instructed to stay in a relaxed state to reduce the influence of myogenic activity on the MEG signals and their compliance was verified by video monitoring. The reason for

maintaining wakefulness was that substantially decreased SSR amplitudes during sleep have previously been reported.[28]

During the experimental blocks of 200 s, 37 channels of magnetic field data (bandwidth from 1 to 200 Hz, sampling rate 520 Hz, resolution 16 bits) were continuously recorded with stimulus triggers and stored for off-line analysis.

D. Data analysis

1. Amplitude and phase of the SSR

The data analysis was a threefold procedure involving data reduction and enhancement of the signal-to-noise ratio. First, the data stream of 200-s duration was split into segments of 1 s each, starting at the onset of the trigger signal. These segments were averaged in the time domain. Segments with signal variations of more than 3 pT were considered to contain artifacts and were rejected on a single-channel basis. This first step of data analysis reduced the amount of data by a factor of 200 and improved the signal-to-noise ratio by a factor of up to $\sqrt{200}$. The second step of the data analysis collapses the time series of the 37 MEG sensors into a single-channel signal in order to achieve an improved signal-to-noise ratio as compared to each of the 37 input signals. The method that was used (called ''source space projection'') estimates the activity in a certain brain area by linear combination of activities measured outside the brain. The resulting signal is a time series of the magnetic dipole moment. It reaches maximum values only for the typical dipolar magnetic field pattern of a single current source in a previously specified brain region. Therefore, the method is spatial sensitive. Spontaneous brain activity from other regions is diminished by source space projection. In addition, uncorrelated system noise is canceled out which otherwise would be a limiting factor for measuring frequencies above 40 Hz. The source space projection method is based on the linear relationship that exists between each current source $\mathbf{q}(\mathbf{R})$ at the source positions $\mathbf{R}$ in the brain and the magnetic field $b_i(\mathbf{r})$ which is measured with the ith sensor at position $\mathbf{r}$ outside the head. This relation is given by the equation $\mathbf{b}(\mathbf{r}) = \mathbf{L}(\mathbf{r},\mathbf{R}) \cdot \mathbf{q}(\mathbf{R})$.[29] The lead field matrix $\mathbf{L}(\mathbf{r},\mathbf{R})$ depends on the source position and the sensor position, as well as upon the properties of the volume conductor and the sensors. In general, the described relation maps each current source into a multidimensional *signal space*.[30] Vice versa, a linear spatially and orientation selective filter can be derived as a projection operator in signal space, which does a reverse mapping into the source space. The filter output signal is a waveform that would be seen by a *virtual sensor*[31] which responds maximally to the region of interest in the brain, attenuating signal from other regions or signals with different source current direction. Such a filter is the dot product of the measured magnetic field $\mathbf{b}(\mathbf{r})(t)$ with a weighting vector $\mathbf{W}(\mathbf{q}(\mathbf{R}))$.[32]

In this study, the pseudo inverse $\mathbf{L}^{-1}(\mathbf{r},\mathbf{R})$ of the lead field matrix and the assumed orientation $\mathbf{q}(\mathbf{R})/|\mathbf{q}(\mathbf{R})|$ of the underlying source were used to define the filter $\mathbf{W}$. Using the model of a single dipole in a homogeneous sphere, the waveform $\hat{q}(t)$ of magnetic dipole moment was calculated by

$$\hat{q}(t) = \mathbf{b}(\mathbf{r})(\mathbf{t}) \cdot \mathbf{L}^{-1}(\mathbf{r},\mathbf{R}) \cdot \mathbf{q}(\mathbf{R})/|\mathbf{q}(\mathbf{R})|. \tag{2.3}$$

Therefore, units of measurement which depend on the position of the sensor (magnetic field in fT) were transformed into a source strength estimate (dipole moment in nA·m) that did not depend on the slightly different positions of the measurement system with respect to the subject's head during the multiple sessions.

The third step in the data analysis was the calculation of the amplitude and phase of the spectral components of $\hat{q}(t)$ at the modulation frequency and some of its harmonics. The results were used to calculate the input–output characteristics (IOCs) for the different stimulus conditions. From the $N = 520$ data points of the magnetic dipole moment $\hat{q}(i)$, the discrete Fourier coefficients at the frequency k were calculated by

$$a(k) = \frac{2}{N} \sum_{i=1}^{N} \hat{q}(i) \sin \frac{ik}{N} \omega_k, \tag{2.4}$$

$$b(k) = \frac{2}{N} \sum_{i=1}^{N} \hat{q}(i) \cos \frac{ik}{N} \omega_k. \tag{2.5}$$

The discrete calculation of Fourier coefficients, which is also known as quadrature demodulation, had the advantage that amplitudes of the spectral components of $\hat{q}(i)$ were calculated at the exact frequencies of interest even when the number of data points (520) did not equal an integer power of 2. The phase of the SSR is usually measured with respect to the zero-phase angle of a sine approximation to the response waveform.[33] A delayed sine wave (positive time shift) has negative zero phase. Therefore, an increasing delay of the response waveform results in a negative increasing phase value if the carrier and modulation frequencies do not change. The phase $\varphi(f)$ was calculated from the complex Fourier coefficient at frequency f considering the signs of the real and imaginary parts, and was hence defined on the interval $[-\pi, \pi)$. In order to keep $\varphi(f)$ monotonically decreasing, -2π was added to it when needed.

For the analysis of the effect of modulation frequency on SSR phase, the group delay τ was calculated as

$$\tau(f_0) = -\frac{1}{2\pi} \left. \frac{\partial \varphi}{\partial f} \right|_{f_0}. \tag{2.6}$$

In a linear system the group delay $\tau(f_0)$ has the meaning that a group of frequencies centered around f_0 is delayed by $\tau(f_0)$ in the sense that its contribution to the impulse response $h(t)$ is maximum at $\tau(f_0)$.[34] In the context of SSR analysis, $\tau(f_0)$ is often called the *apparent latency*.[4]

2. The SSR waveform

The stationary part of power-line frequency noise (multiples of 50 Hz) was calculated by the discrete Fourier transform and, after time-series expansion, was subtracted from the waveform of the magnetic dipole moment. After the line frequency noise was thus eliminated, subsequent sections of length equal to one period of the modulation frequency f_m were averaged and formed one period of the time series of

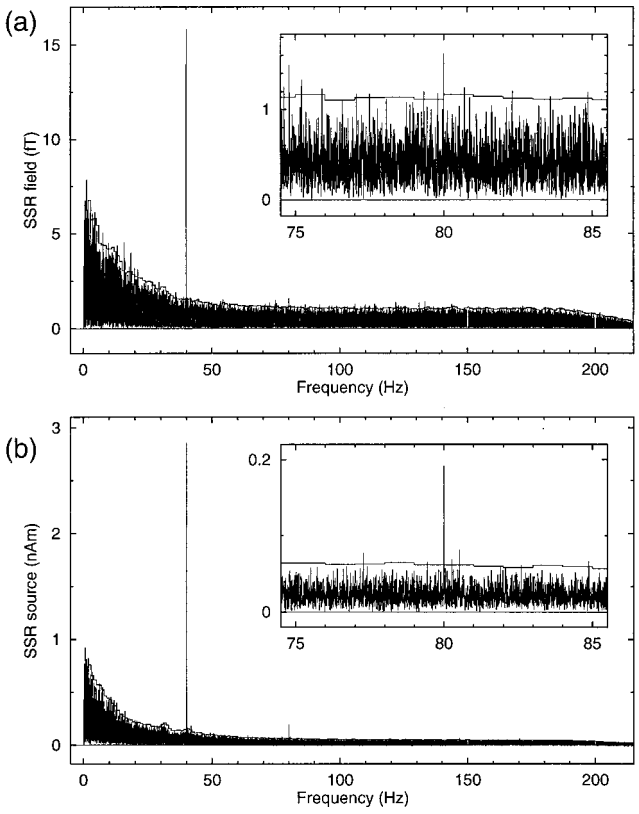


FIG. 2. (a) Amplitude spectrum of the magnetic field obtained from the maximally responding sensor channel; (b) amplitude spectrum of the source waveform after source space projection. Enlarged displays of both spectra are shown in the insets for a 10-Hz frequency range centered at 80 Hz. The thin horizontal lines denote the 99%-confidence limits (2.33 times the standard deviation) for response detection.

the SSR. Linear interpolation between sampled data points was necessary because $1/f_m$ did not equal an integer number of sampling intervals. Thus, the waveform of the SSR, which was measured for a total epoch length of T s, was the result of averaging of $N=T\cdot f_m$ stimulus related epochs. The number of averaged epochs, and, correspondingly, the improvement of the signal-to-noise ratio as well, depended on the modulation frequency and ranged from $N=2000$ at 10 Hz to $N=18\,600$ at 98 Hz. In contrast, these properties of data analysis in the frequency domain were independent of f_m, and the signal-to-noise ratio was improved evenly over the whole frequency range. For this reason, the waveforms of the SSR were calculated for visualization purpose only.

3. Spectral analysis

Source space projection was also applied to the raw (nonaveraged) data, and time series of magnetic moment (200 s×520 samples/s=104 000 samples) were calculated. High-resolution amplitude spectra were obtained from these time series by Fourier transform, giving 200 lines per Hz. Examples of the high-resolution amplitude spectrum are shown in Fig. 2 for the magnetic field data [panel (a)] and the dipole moment data [panel (b)], respectively. The purpose of this analysis was to estimate the non-stimulus-locked MEG background activity. Both the evoked-response amplitudes and the line frequency noise were expected to be concentrated around integer frequencies. Therefore, for each frequency bin starting at the integer frequency f_i between 1 and 200 Hz, the mean power was calculated from 180 spectral lines in the interval from $f_i+0.05$ Hz to $f_i+0.95$ Hz. The square root of the mean power gave an estimate of the standard deviation $\sigma_n(f)$ of the MEG background noise. For a measured response amplitude $a(f)$ at the frequency f, the signal-to-noise ratio SNR(f) was defined as the ratio $\mathrm{SNR}(f)=a(f)/\sigma_n(f)$. The value of $2.33\sigma_n$ was calculated for all frequency bins as an estimate of the 99%-confidence level for the detection of response signal amplitudes (Fig. 2). SSRs with amplitudes below the confidence level were rejected from further data analysis.

4. Source localization

A prerequisite for the source space projection was the estimation of the position $\mathbf{R}$ and the orientation $\mathbf{q}(\mathbf{R})/|\mathbf{q}(\mathbf{R})|$ of the underlying sources in a head-based coordinate system. In principle, either anatomical information or results from MEG source estimation carried out on the actual data could be used. The latter was the case in this study.

The origin of the head-based coordinate system was set at the midpoint of the mediolateral axis (y axis) which joined the center points of the entrances to the acoustic meatii of the left and right ears (positive towards the left ear). The postero–anterior axis (x axis) was oriented from the origin to the nasion (positive towards the nasion) and the inferior–superior axis (z axis) was perpendicular to the x–y plane (positive towards the vertex).

Magnetic field data epochs corresponding to single periods of the modulation signal were averaged, and the point of time at which the global field power (the mean square sum over all MEG channels) reached its maximum was determined. At this time point the field of a single moving dipole was fitted to the averaged data in order to estimate the magnetic source localizations. A homogeneous spherical model was used for the dipole fits. The shape of each subject's head was digitized during preparation of the experiments, and the best-fitting sphere to the curvature of the head above the estimated source was calculated. For each subject and experimental condition, the median values of the estimated source coordinates and dipole orientations were used as a reference for the source space projection. In contrast to the magnetic field data $\mathbf{b}(\mathbf{r})$, which depend on the position of the sensor array relative to the subject's head, the magnetic moment data $\hat{q}(t)$ refer to a fixed area in the head only. Therefore, the time series of the magnetic moment was a measure which enabled the comparison of data from subsequent experimental sessions without requiring any precise repositioning of the MEG sensor array.

5. Grand averages

IOCs were obtained for all single subjects and all experimental conditions and were grand averaged over all subjects for each experimental condition. In order to reduce the influence of interindividual variations, the individual IOCs were scaled to unity at an appropriate value of the stimulus parameter. In addition, the mean phase was subtracted from individual phase characteristics. The 99%-confidence interval of

each grand average was calculated by the bootstrap resampling method.[35] The modulation frequency IOC was smoothed by weighted averaging over adjacent values by the function $\overline{x_i} = 0.2x_{i-1} + 0.6x_i + 0.2x_{i+1}$.

II. RESULTS

Steady-state responses with amplitudes statistically significantly different from MEG background activity were detected at the modulation frequency of the auditory stimulus and some of its harmonics as well for all subjects and all experimental conditions. No peaks of oscillatory MEG activities at other frequencies were observed in the frequency range from 1 to 200 Hz.

A. Example of individual data

The 37-channel SSR data obtained from Subject A0010 when steady-state stimulation ($f_c = 250$ Hz, $fm = 40$ Hz, and $m = 100\%$) was applied at 70 dB_{SL} are displayed in Fig. 1(c). The measured magnetic field reveals a pronounced dipolar pattern with a polarity reversal between posterior and anterior channels. The field distribution reaches its posterior maximum in channel 27 with an amplitude of 16 fT. The amplitude spectrum of this channel, which is shown in Fig. 2(a), consists of a distinct spectral line at the modulation frequency and smaller lines at its harmonics, and the MEG background activity, which is widely spread over the whole frequency range. The MEG background spectrum reveals a $1/f$-shape for low frequencies and approaches a minimum constant spectral density near 50 Hz. The spectral power above 50 Hz represents mainly measurement system noise, which is reduced above 200 Hz by the antialiasing low-pass filter. The standard deviation of the background noise is 0.5 fT, which equals the sensor noise (7 $fT/\sqrt{Hz}$) divided by the square root of the length of the time window ($\sqrt{200\ s}$).

Figure 2(b) displays the SSR spectrum after source space projection, where a reduction of the MEG background activity is evident. Although Figs. 2(a) and 2(b) are on different scales (fT or $nA \cdot m$, respectively) the dimensionless SNR allowed a quantitative comparison. The SNR was enhanced by a factor of 1.5 at 10 Hz, by a factor of 2 at 50 Hz, and by a factor of 4 at 100 Hz and above. The intersection point between the low-frequency range in which the brain activity is mostly prevalent and the frequency range which contains mostly system noise shifted from about 50 Hz to about 100 Hz with source space projection. The SSR corresponding to the modulation frequency was again represented by a single spectral line at 40 Hz after source space projection. However, the spectral peak at the second harmonic (80 Hz) was enhanced compared to the untransformed field [Fig. 2(a)]. The frequency range around 80 Hz in both data sets is shown with enlarged scale in the insets of Figs. 2(a) and 2(b), with the thin horizontal line denoting the 99%-confidence level. In the untransformed data, the spectral line at 80 Hz exceeds this level with a SNR of 3.32 [Fig. 2(a)] compared to a SNR of 7.45 after source space projection [Fig. 2(b)]. The SNR at 40 Hz was increased by a factor of 1.88 from 24.5 to 46.0.

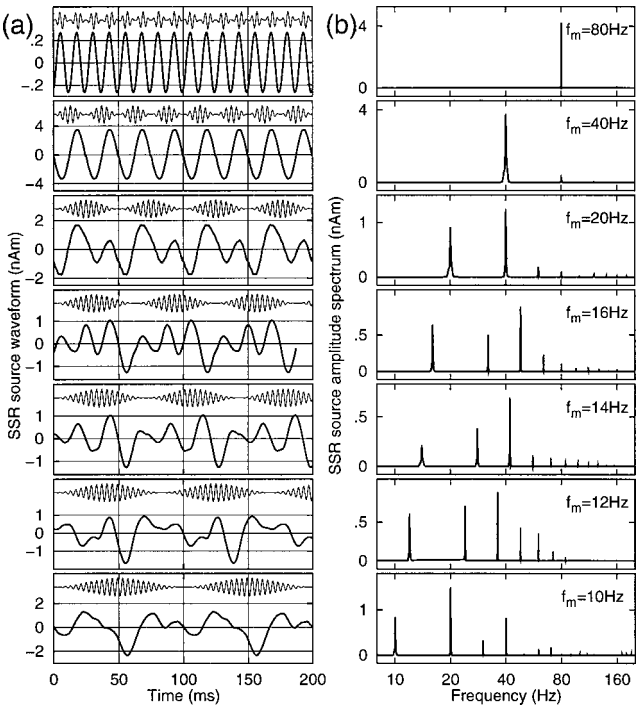


FIG. 3. Examples of individual SSR data; (a) Source waveforms obtained with sinusoidal amplitude modulated tones ($f_c = 250$ Hz, $m = 100\%$, 70 dB) at various modulation frequencies. Thin lines represent the stimulus waveforms. (b) Corresponding amplitude spectra of the source waveforms.

B. Source localization

Although source analysis of the human auditory SSR was not the objective of this study, source localization was a prerequisite for the source space projection procedure. The grand average of the median values of the estimated source coordinates of the eight subjects was 1.29 cm in the posterior–anterior direction (x), 4.30 cm in the mediolateral direction (y), and 6.12 cm in the inferior–superior direction (z). When the carrier frequency was varied between 250 and 4000 Hz, a tonotopic cortical representation was found which extended over a range of 5 mm in the mediolateral direction, with higher frequencies situated closer to the sagittal midline.[15]

C. Modulation frequency

As reported above, a clearly detectable SSR was observed at each tested modulation frequency in every subject. The source waveforms of the SSR for Subject A0010 at various modulation frequencies are shown in the left panel of Fig. 3. SAM stimulus signals are also depicted at each modulation frequency. At 10 Hz, the lowest modulation frequency, the response waveform consisted of a number of positive and negative deflections which repeated periodically corresponding to the stimulation rate. Increasing the modulation frequency reconfigured the SSR source waveform, and at about 40 Hz and above, the waveform revealed an almost sinusoidal shape. At modulation frequencies below 40 Hz, deflections occurring at the same latency as obtained from the 40-Hz response could be identified for most waveforms with respect to the rising edge of the stimulus signal. At modulation frequencies of 10 to 40 Hz, all waveforms exhibited a

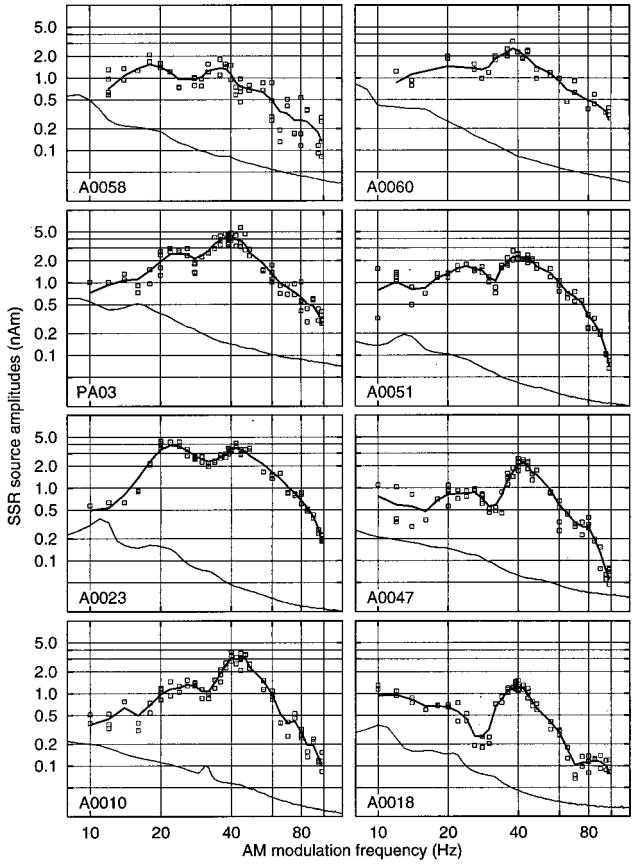


FIG. 4. SSR amplitudes versus modulation frequency for all subjects. The square symbols denote the results from single measurements of 200-s duration. The thick lines represent a smoothed average over single measurements. Thin lines denote the standard deviation of the MEG background activity.

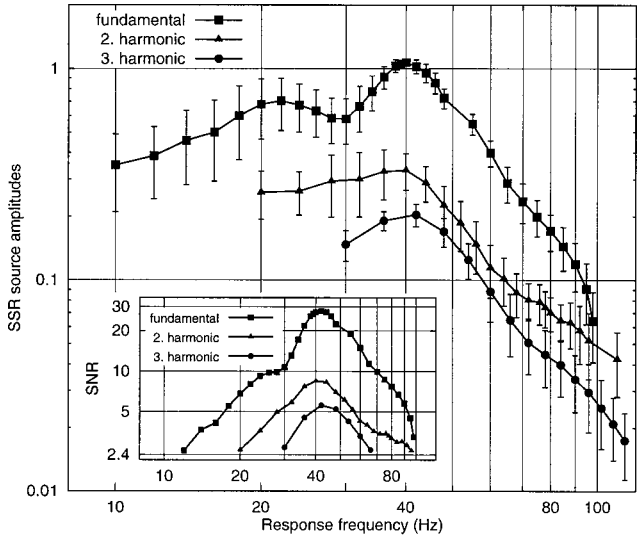


FIG. 5. Normalized spectral amplitudes of SSR at the modulation frequency and second and third harmonic as grand averages over eight subjects. The error bars denote the 99%-confidence interval obtained from bootstrap resampling. The inset displays the corresponding signal-to-noise ratio characteristics.

negative (downward) deflection at 56 ms. For f_m of 12 to 40 Hz, a positive (upward) deflection occurred at 44 ms, whereas a negative deflection at 31 ms was observed for f_m of 14 to 40 Hz, and so on.

A more quantitative description of the SSR at various modulation frequencies is given by the Fourier expansions of the source waveforms displayed in the right panel of Fig. 3. The spectrum, corresponding to $f_m = 10$ Hz, demonstrated a clear component at f_m, but the amplitude of the second harmonic (20 Hz) was almost double in size. The spectral component of the SSR at 40 Hz had about the same amplitude as the fundamental 10-Hz component. The SSR evoked by $f_m = 12$ Hz consisted of a component at 12 Hz, a somewhat bigger component at 24 Hz, a most prominent component at 36 Hz, and components with decreasing amplitudes above 40 Hz. The spectrum of the SSR at $f_m = 14$ Hz had a dominant peak at the third harmonic (42 Hz), which was twice the size of the second harmonic (28 Hz) and four times the size of the fundamental (14 Hz). At $f_m = 16$ Hz, the third harmonic (48 Hz) of the SSR was more pronounced than the fundamental. When the stimulus was modulated with $f_m = 20$ Hz, the SSR spectrum showed its largest peak at the second harmonic (40 Hz). In contrast, the SSR spectrum at $f_m = 40$ Hz was dominated by the fundamental peak, which was 17.3 times bigger than the second harmonic (80 Hz). Nevertheless, the amplitude of the peak at 80 Hz was measured with high accuracy and was 7.5 times the standard deviation of the background noise. The 80-Hz response spectrum showed a single peak at the fundamental frequency only (the amplitude of the second harmonic was not significantly different compared to the background noise).

IOCs describing the relation between the magnitude of the response and the modulation frequency are widely known as modulation transfer functions (MTFs). MTFs are presented in Fig. 4 (thick lines) for each of the eight subjects. The thin lines denote the standard deviation of the background MEG amplitude spectrum. In some cases, at frequencies below 20 Hz the fundamental amplitudes did not reach the significance level of 2.33 times the standard deviation ($p \leqslant 0.01$). However, because the second or higher harmonics exceeded the significance level, these frequencies were not excluded from further analysis. Inspection of Fig. 4 shows that MTFs peaked around 40 Hz in each subject, with amplitude varying between 1 and 4 nA·m among the subjects. A second peak around 20 Hz was also seen in all subjects, although the ratio of the 20-Hz peak to the 40-Hz peak differed between subjects. For two subjects (A0023 and A0058), the 20-Hz and 40-Hz peaks reached the same size; for all other subjects the 20-Hz peak was smaller than the 40-Hz peak.

In order to reduce the between-subject variation before calculating the grand average MTF, individual MTFs were scaled, setting the mean amplitude to unity in the frequency range of 36 to 44 Hz. The resulting grand average MTF is shown in Fig. 5. The 99%-confidence interval obtained from bootstrap resampling also is added. The amplitude characteristic showed a significant peak around 40 Hz. In contrast, the 20-Hz peak, which was clearly pronounced in some subjects, did not reach significance in the grand average. Above 40 Hz, the MTF decreased with increasing modulation frequency and followed a linear slope (18 dB/octave) in a log–log scaled diagram. Additionally, the frequency characteris-

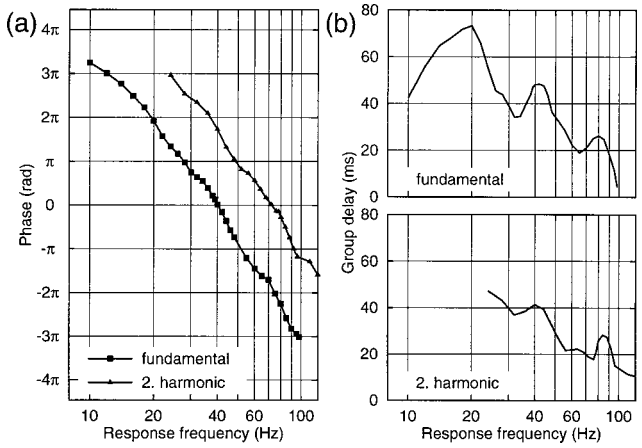


FIG. 6. (a) Phase characteristics of SSR at the modulation frequency and the second harmonic (shifted 2π upwards for clarity); (b) group delay (apparent latency) obtained from the phase characteristics of the fundamental and second harmonic response.

tics of the second and third harmonic of the SSR are shown in Fig. 5 (e.g., the third harmonic at 42 Hz is a response to 14-Hz stimulation). Both characteristics reached maximum values around 40 Hz and decreased for higher frequencies. Between 40 and 80 Hz, the shape of both characteristics was similar to the characteristic of the fundamental SSR amplitude. Above 80 Hz the variability increased due to the lower signal-to-noise ratio. The inset of Fig. 5 shows the MTF in relation to the MEG background activity (signal-to-noise ratio, SNR). This graph demonstrates that in terms of SNR the second harmonic of the SSR is predominant over the fundamental in the modulation frequency range between 10 and 20 Hz. Also, the SNR of the third harmonic is higher than the fundamental SSR in the modulation frequency range between 10 and 15 Hz.

The modulation frequency-phase IOC is shown in the left panel of Fig. 6 for the fundamental SSR, and for its second harmonic as well. Since the absolute phase angle is uncertain for additional multiples of 2π, the phase characteristic of the second harmonic was shifted by 2π upwards for the reason of clarity. Both characteristics showed similar shapes, decreasing monotonically with increasing modulation frequency. In order to obtain the corresponding group delay (apparent latency), the first derivative of the phase characteristic was calculated and displayed in the right panel of Fig. 6. Both the group delay of the fundamental SSR and of the second harmonic tended to shift toward shorter delays with increasing modulation frequency. The group delay IOC of the fundamental SSR showed three distinct peaks at 20, 40, and 80 Hz with apparent latencies of 72, 48, and 26 ms, respectively. The IOC for the second harmonic also suggested peaks around 40 and 80 Hz with group delay values in the same range.

D. Stimulus intensity

The effect of stimulus intensity on the SSR was investigated with a carrier frequency of 250 Hz and the modulation rate set to 39 Hz. For all subjects, an increase in SSR amplitude with increasing intensity was observed. In order to reduce between-subject variation, the amplitude values of each subject were scaled by the observed mean response amplitude in the 80- to 90-dB stimulus intensity range. The scaled SSR amplitudes were averaged over the eight subjects, and 99%-confidence intervals were calculated by the bootstrap resampling method. The resulting SSR intensity-amplitude IOC [Fig. 7(a)] was almost linear in the intensity range from 40 to 70 dB. The IOC reached its maximum value at 80-dB stimulus intensity, and no further increase was obtained with an intensity of 90 dB. Therefore, all following experiments were carried out at the intensity of 70-dB sensation level.

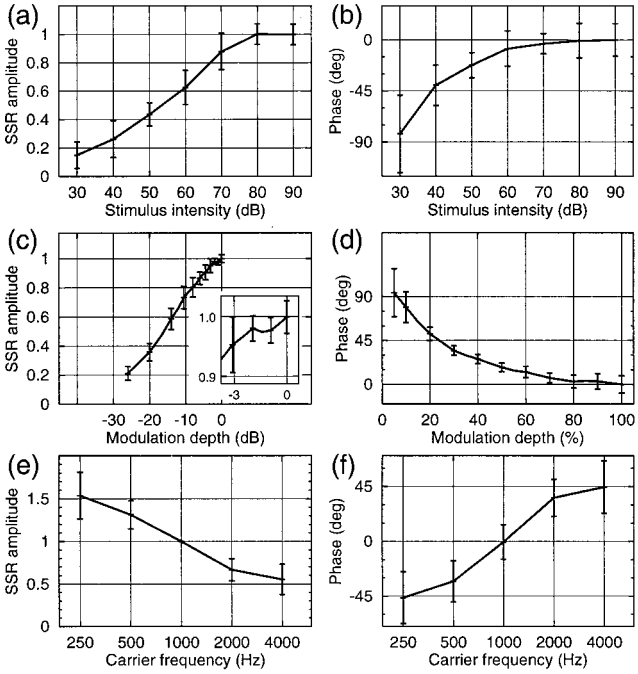


FIG. 7. Amplitude- and phase characteristics of SSR at the modulation frequency obtained by averaging over eight subjects (a), (b) as function of the stimulus intensity (c), (d) as function of the modulation depth, and (e), (f) as function of the carrier frequency. The error bars denote the 99%-confidence interval obtained from bootstrap resampling.

In order to calculate the stimulus intensity-phase IOC, the individual mean phase for each subject was subtracted first. The phase data were then averaged over the eight subjects and the 99%-confidence intervals were calculated. Finally, the phase value at 90 dB, which represented the highest intensity, was fixed to zero. The resulting stimulus intensity-phase IOC is shown in Fig. 7(b). Phase increased with increasing intensity, corresponding to a shorter delay at higher intensity. For the 30-dB sensation level, an average phase of $\pi/2$ was obtained and corresponded to a delay of 6 ms compared to the SSR at 90-dB sensation level. For all eight subjects, delay time was shorter at the highest intensity of 90 dB than the lowest intensity of 30 dB.

E. Modulation depth

All subjects showed a decrease in SSR amplitude when the modulation depth was reduced. In order to diminish the effect of between-subject variation, each individual IOC was transformed to a relative scale by setting the amplitude obtained at $m = 100\%$ to 1. The modulation depth-amplitude

IOC averaged over the eight subjects is displayed in Fig. 7(c). The modulation depth is given on a dB scale referenced to $m=100\%$, whereas the amplitude is given on a linear relative scale. In order to facilitate comparison between Figs. 7(c) and 7(a), the axes of both figures are scaled at the same size. The two characteristics showed little variation when m was changed between 100% and 80% (-2 dB) but decreased linearly when modulation depth was varied between 80% and 5% (-2 to -26 dB).

Alternatively, modulation depth can be described as a stimulus with constant carrier amplitude but with sideband amplitudes that change proportionally to m [cf. Fig. 1(a)]. Following this interpretation, the modulation depth-amplitude IOC decreased linearly when the sideband intensity decreased on a dB scale. The SSR amplitude decreased by a factor of 3, from $m=80\%$ to $m=8\%$ (corresponding to a 20-dB intensity variation). For comparison, the SSR amplitude decreased only by a factor of 2 when stimulus intensity changed from 70 to 50 dB at constant modulation depth of 100%. Thus, the slope of the modulation depth-amplitude IOC appeared to be steeper than the slope of the amplitude intensity IOC.

In contrast to the effect of stimulus intensity on the phase angle, all subjects showed an increasing delay with increasing modulation depth. The modulation depth-phase IOC, which is shown in Fig. 7(d), was calculated in the same way as the amplitude intensity characteristic.

F. Carrier frequency

In order to reduce the effect of between-subject variability in the absolute values of SSR amplitude, relative amplitude variations were referred to the SSR amplitude for the 1000-Hz carrier frequency. Five of the seven subjects participating in this experiment showed a monotonic decrease of the SSR amplitude when the carrier frequency was changed from 250 to 4000 Hz, whereas two subjects showed monotonic decreases between 500 and 2000 Hz and reversed behavior at the extreme frequencies. The carrier frequency-amplitude IOC obtained by averaging over all subjects is shown in Fig. 7(e). Inspection shows that SSR amplitude decreased linearly with increasing carrier frequency on a logarithmic scale. From 250 to 4000 Hz, SSR amplitude decreased by a factor of 2.8. The corresponding phase characteristic is displayed in Fig. 7(f) and exhibited an increasing phase angle with increasing carrier frequency. This is equivalent to a time delay 6.4 ms shorter at 4000 Hz than at 250 Hz.

III. DISCUSSION

Steady-state evoked magnetic fields were detected from all subjects in all experimental conditions and were analyzed in the time and frequency domains. Representation of the fundamental of the SSR by a single spectral line without any noticeable sidelobes indicates that the SSR had a fairly stable amplitude throughout the measurement time. The mean source coordinates ($x=1.29$ cm, $y=4.30$ cm, $z=6.12$ cm) obtained for single equivalent dipoles fitted to the evoked magnetic field for each subject correspond to the grand average values of $x=1.0$ cm, $y=4.30$ cm, and $z=6.05$ cm reported in a previous study[15] using queues of Gaussian tone pulses with a repetition rate of 39 Hz and a carrier frequency of 250 Hz. This correspondence suggests that the SSR was generated from sources in the primary auditory cortex. Although individual MRI data were not available to corroborate this conclusion, coregistration of SSR sources was carried out in the previous investigation[15] and confirmed the presence of generators in the primary auditory cortex.

The goal of this study was to measure input–output characteristics (IOCs) relating SSR amplitude and phase to several parameters of steady-state stimulation in the same subjects. To achieve this goal, the SSR magnetic field was transformed by signal space projection into a time series representing the behavior of the cortical sources underlying SSR generation. Because the transformation was not affected by variations that occur when the MEG sensors are repositioned with respect to individual neuroanatomy during repeated sessions, the data from repeated sessions on the same subjects could be averaged. This allowed several IOCs to be investigated in these subjects with a very high signal-to-noise ratio. Signal space projection improved signal-to-noise ratios by a factor in the range of 2 to 4.

A. Modulation frequency

The data presented in this study demonstrate that the auditory cortex responds to amplitude fluctuations of SAM tones in the frequency range from 10 to 98 Hz.

The shape of the average MTF shown in Fig. 5, as well as the MTFs shown in Fig. 4 for single subjects, resemble those of a low-pass filter with an upper cutoff frequency near 50 Hz. Similar MTFs were obtained in psychophysical experiments[36] which measured the ability of subjects to resolve sinusoidal amplitude modulations at different modulation rates. Psychophysical MTFs describe perceptual sensitivity to AM sounds by determining just-noticeable-differences in modulation depth at different modulation frequencies. If the effects of pitch perception are eliminated either by using a wideband noise carrier[36] or instructing subjects to listen for ''roughness'' only,[37] the psychophysical MTF shows decreasing perceptual sensitivity by about 3 dB/octave with increasing modulation frequency above about 50 Hz. Correspondingly, an EEG study by Rees *et al.*[38] obtained a MTF with a slope of 3 to 6 dB/octave when modulation frequency was increased above 50 Hz. On the other hand, we found a steeper slope of 18 dB/octave above 50 Hz in the present study where MEG was measured. The difference between our findings and those of Rees *et al.* may be due to the fact that MEG measurements are sensitive only to cortical sources of brain activity. Animal data concur with this hypothesis by suggesting that the best modulation frequency for amplitude-modulated tones decreases along the auditory pathway. Best modulation frequencies of about 500 Hz have been reported for the cochlear nucleus of the gerbil,[39] of 100 to 120 Hz for the inferior colliculus of the rat,[40] and of more than 50 Hz in the anterior auditory field of the cat.[41] Rees and Møller[40] reported further that the frequency of the maximum response and the upper cutoff frequency of MTFs obtained from the inferior colliculus of the rat did not differ

between neurons with characteristic frequencies ranging between 1 and 40 kHz. Schreiner and Urbas[42] found, for neurons in the anterior auditory field of cat cortex, a small positive correlation between best modulation frequency and characteristic frequency with sinusoidal modulation, but no significant correlation when the modulating signal was a square wave. We did not measure MTFs at carrier frequencies above 250 Hz because diminishing response amplitudes necessitated an impractical measurement time. Nevertheless, animal data provide only little evidence for a general change of the upper slope of MTFs as carrier frequency changes.

The low-frequency slope of the MTF obtained in our study (6 dB/octave below 20 Hz; see Fig. 5) is flat compared to the high-frequency slope (18 dB/octave above 50 Hz). This shape is in contrast to bandpass-like MTFs found by Rees and Møller[40] in the inferior colliculus of the rat with a stimulus intensity of 60 dB. The MTFs obtained by Rees and Møller may have been influenced not only by the collicular origin of the auditory response, but also by the fact that MTFs were obtained from Fourier transform of correlograms between stimulus signal and period histograms of the response. For this calculation the assumption of a linear system is required which is not fulfilled, especially at low frequencies.

At low modulation frequencies the response waveform is periodic with the modulation frequency and consists of a number of distinguishable components. This results in an amplitude spectrum composed of harmonics of the modulation frequency, with maximum values in the 40-Hz range. It is rather unlikely that the response waveform to low modulation frequencies results from nonlinear distortions. More obviously it resembles an intrinsic waveform pattern that is triggered by the rising edge of the stimulus signal envelope. For modulation frequencies of 40 Hz and above, the SSR waveform is almost sinusoidal and follows the modulating waveform with some time delay. It may be possible that single periods of the modulated stimulus elicited transient responses. This would support the hypothesis of SSR generation by periodic superimposition of transient responses. This hypothesis was established first by Galambos,[2] who synthesized a 40-Hz response from repetitively superimposed middle latency responses that were evoked by short tone pulses presented at 10 per s. Azzena *et al.*[10] confirmed this finding for the modulation frequency of 40 Hz but not for frequencies below and above 40 Hz. Our findings regarding the frequency-group delay IOC [Fig. 6(b)], which showed local maxima at 20 Hz (72 ms), 40 Hz (48 ms), and 80 Hz (26 ms) provides an important contribution to this discussion. An explanation of the group delay is that the group of frequencies centered around ω_0 is delayed by the amount of group delay τ, in the sense that their contribution to the impulse response is maximum at that instant.[34] Therefore, the transient N_a-P_a waves (20 and 30 ms latency) could relate to the 80-Hz SSR, the N_b-P_b waves (latency of 38 and 52 ms) to the 40-Hz SSR, and the P_1-N_1 waves to the 20-Hz SSR. This interpretation is in line with the hypothesis of generation of SSR by linear superimposition of subsequent transient response components.

On the other hand, if it is assumed that an enhanced excitability for 40-Hz activity exists in some neural networks,[43] generation of SSR with a pronounced amplitude response around 40 Hz could be modeled by a damped oscillator, either linear or nonlinear. This model would necessarily contain one or more second-order components. A property of such a second-order linear system is that the phase response should show the steepest decrease, and therefore a maximum group delay, at the frequency of the response maximum. Indeed, the 20- and 40-Hz maxima of the group delay characteristic coincide with two peaks of the amplitude characteristic, and a third peak of the amplitude characteristic at 80 Hz might be blurred by its falling slope. Thus, the group delay characteristic provides data consistent with both hypotheses (linear superimposition and synchronization of intrinsic oscillations). It should be noted that these two hypotheses are not mutually exclusive.

The three distinct peaks of the modulation frequency-phase characteristic at 20, 40, and 80 Hz divide the frequency axis into three sections, a low-frequency section below 30 Hz, a mid range between 30 and 70 Hz, and an upper range above 70 Hz. At least for the carrier frequency of 250 Hz, this division is consistent with different perceptual categories within each frequency range. In the low-frequency range the amplitude-modulated tone is perceived as a beating sound, as roughness in the 30- to 70-Hz range, and as pitch in the upper-frequency range.

B. Stimulus intensity

The intensity-amplitude IOC obtained in this study suggests a stable causal system and is very similar to characteristics reported in previous studies.[18,44,33] The saturation level of 70- to 80-dB sensation level which we observed agrees with a previous EEG study with AM sounds[38] modulated to the level of 60 dB. However, in contrast to stimulus intensity IOCs which have been reported for SSRs evoked by short tone bursts,[18] we found evidence of saturation in the region of 90-dB sensation level. The reason could be that compared to short tone bursts, the power of the SAM tones is much higher due to the continuously presented carrier signal. This could cause habituation which prevents a further increase in SSR amplitude at stimulus intensities near 80–90 dB [Fig. 7(a)].

An extrapolation of the obtained intensity-amplitude IOC towards lower intensities crosses the zero-amplitude axis between 15- and 20-dB sensation level. This intensity can be interpreted as the threshold for elicitation of the SSR. This finding corresponds to thresholds previously reported by Stürzebecher[45] (10 dB above behavioral threshold), Szyter[46] (15 dB), Rees[38] (15 dB), and Rodriguez[33] (3–20 dB).

The intensity-phase IOC [Fig. 7(b)] can be explained by the threshold model of latency.[47] This model assumes that a neuron's onset response is triggered whenever the signal reaches a fixed threshold amplitude. If the stimulus signal has a rising edge with a constant form, as in the case of the cosine-shaped envelope of an SAM tone, the threshold will be reached earlier for a higher intensity than for a lower intensity. The argument of Heil[48,49] that the latency depends mostly on the acceleration of the envelope rather than on

amplitude, does not change the intensity-latency IOC expected in the case of an SAM signal. This is true because the second derivative of the envelope which represents its acceleration has the same shape, however with inverted sign. The observed latency shift of about 6 ms, which is the half of the rising edge of the envelope, fits very well to the model mentioned above.

C. Modulation depth

The modulation depth-amplitude IOC obtained in this study [Fig. 7(c)] agrees with results obtained by Rees *et al.*[38] in an EEG study of SAM sounds in which SSR amplitude increased linearly with modulation depth on a logarithmic scale. In contrast, Kuwada *et al.*[44] reported a linear relation between modulation depth and response amplitude from a single subject. The observed similarity between the modulation depth-amplitude IOC and the intensity-amplitude IOC shown in Fig. 7(a) may be explained by viewing the stimulus signal as the sum of two terms

$$y(t) = m \cdot a \cdot \sin(\omega_c t) \cdot (1 - \cos(\omega_m t)) + (1 - m) \cdot a \cdot \sin(\omega_c t). \quad (3.1)$$

The first term is an SAM signal with 100% modulation depth, whereas the second term is a sine wave signal with constant amplitude. Only the effect of the first term contributes to a SSR at the modulation frequency, and no response will be evoked by the second term. The amplitudes of both terms change with variation of *m*. When the stimulus signal is only the first term, variation of *m* is obviously equivalent to the investigation of the effect of overall intensity variation. Therefore, in Fig. 7(c), *m* is shown on a logarithmic scale. The resulting modulation depth-amplitude IOC is almost linear with a steeper slope than observed from the overall intensity-amplitude IOC. This steeper slope can be interpreted as the influence of the second term of the stimulus signal, which gives amplitude increases with decreasing modulation depth. The modulation depth-phase IOC is completely different in comparison to the intensity-phase IOC and shows a reversed relation between latency and intensity of the relevant part of the stimulus signal. These findings are completely in line with a phase characteristic presented by Kuwada *et al.*[44]

D. Carrier frequency

The SSR frequency-amplitude IOC [Fig. 7(e)], for which amplitude decreased by a factor of 2.8 with frequency increasing from 250 to 4000 Hz in this study, has been reported in the EEG literature in experiments using short tone-burst stimulation. Galambos[2] reported an amplitude decrease by a factor of about 2 to 3 for three subjects when carrier frequency increased from 250 to 5000 Hz. This result was attributed to activation of larger portions of the basilar membrane by low-frequency tones. Stapells[18] also obtained an amplitude decrease by a factor of 1.4 when increasing the carrier frequency from 500 to 4000 Hz. In an EEG study with short tone bursts (4-ms rise and fall time and a 2-ms plateau at a rate of 40/s), Rodriguez *et al.*[33] found the SSR amplitude decreased by a factor of 2.33 when frequency increased from 500 to 4000 Hz. Following the hypothesis of superimposition of middle latency responses (MLR), Rodriguez *et al.* explained the amplitude effect by suggesting that frequency-dependent latency variations of the different MLR components may have resulted in perfect aggregation of the SSR at low frequencies and less effective summation at higher frequencies. Rodriguez *et al.* noted especially the contribution of the frequency sensitive wave V of the brainstem responses to the MLR. However, the contribution of brainstem responses to the SSR can be excluded from the results of the present study, because the MEG recordings were not sensitive enough to see brainstem activity. The results of our study are in line with findings of a previous study by Pantev *et al.*[15] of the steady-state field (SSF). Pantev *et al.* reported a decrease by a factor of 2 for the global field power and the strength of the cortical source of the SSF in the frequency range from 250 to 4000 Hz. The same amplitude effect was reported by Kuwada *et al.*[44] for SSRs measured by EEG. Thus, the behavior of the frequency-amplitude IOC is consistent in both EEG and MEG recordings.

The variation in the amplitude of the SSR with carrier frequency certainly reflects the tonotopic organization of the SSR. In this respect, the cortical sources of steady-state responses recorded in this study were found to be tonotopically organized. This tonotopic organization was most pronounced in the mediolateral direction and spanned a distance averaging 0.5 cm over the eight subjects for a frequency variation of 250 to 4000 Hz. This result, which corrobates an earlier report by Pantev *et al.*,[15] means that the magnetic field power of a deeper source related to higher stimulus frequency is smaller than the magnetic field power of a laterally located source which is related to low stimulus frequency. The source space projections in this study, however, were calculated for a fixed estimation of dipole position which corresponded to the median of source parameters in the relevant experimental condition. The estimated coordinates matched exactly the 1000-Hz source of Pantev *et al.*,[15] and consequently the SSR amplitude for 1000 Hz was correctly estimated. Under the assumption of tonotopic organization, the source depth of the 250-Hz response was estimated to be too deep, which results in an overestimate of the 250-Hz SSR amplitude. In contrast, the 4000-Hz amplitude was underestimated. In order to evaluate this amplitude error, the source coordinates were systematically varied in all directions and the corresponding dipole moments were recalculated. This simulation showed that a 0.5 cm variation in the mediolateral direction would introduce a factor of no more than 1.3 in the amplitude of the equivalent source dipole. Thus, even accounting for the influence of tonotopy, the factor of greater than 2 obtained in this study for carrier frequency cannot be sufficiently explained.

In the frequency range from 250 to 4000 Hz, the SSR phase increases by $\pi/2$ as shown in Fig. 7(f). This corresponds to a 6.25-ms longer latency for the 250-Hz response compared to 4000 Hz. From EEG studies, Galambos[2] reported a latency shift of 6 ms, 5 ms of which he attributed to frequency-dependent traveling time along the basilar membrane. Rodriguez[33] as well reported a latency shift of 6 ms

but estimated that only 2–3 ms of the observed delay could be explained at the cochlear level.

IV. SUMMARY

Amplitude-modulated tones elicit cortical activity at the modulating frequency and its harmonics. Using MEG recordings, the SSR signal was detected with high precision and was separated from other brain sources. SSR amplitude decreased linearly when either stimulus intensity or modulation depth decreased on a logarithmic scale. In contrast, the phase lag between the SSR and the stimulus signal envelope decreased when the intensity increased but changed in the opposite direction when the modulation depth increased. A reduction of SSR amplitude was found for higher carrier frequencies, which is consistent with recently reported EEG recordings. A modulation transfer function was obtained from measurements of SSR amplitude when modulation frequency was incremented in fine steps between 10 and 98 Hz. The obtained MTF resembled psychophysical functions relating AM perception when pitch perception is excluded. From the SSR phase-modulation frequency characteristic, a group delay characteristic was derived which showed three distinct peaks with decreasing time delay as the modulation frequency increased.

ACKNOWLEDGMENTS

This work was supported by grants from the Deutsche Forschungsgemeinschaft (No. Pa392/7-1). The authors thank Karin Berning for assistance and Dr. Scott Makeig for helpful comments.

[1] J. S. Buchwald and C. M. Huang, ''Far-field acoustic response: origins in the cat,'' Science **189**, 382–384 (1975).
[2] R. Galambos, S. Makeig, and P. J. Talmachoff, ''A 40-Hz auditory potential recorded from the human scalp,'' Proc. Natl. Acad. Sci. USA **78**, 2643–2647 (1981).
[3] D. Regan, ''Comparison of transient and steady-state methods,'' Ann. (N.Y.) Acad. Sci. **388**, 45–71 (1982).
[4] D. Regan, *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine* (Elsevier, New York, 1989).
[5] D. S. Barth and S. Di, ''Three-dimensional analysis of auditory-evoked potentials in rat neocortex,'' J. Neurophysiol. **64**, 1527–1536 (1990).
[6] M. N. Franowicz and D. S. Barth, ''Comparison of evoked potentials and high-frequency (gamma-band) oscillating potentials in rat auditory cortex,'' J. Neurophysiol. **74**, 96–112 (1995).
[7] B. H. Gaese and J. Ostwald, ''Temporal coding of amplitude and frequency modulation in the rat auditory cortex,'' Eur. J. Neurosci. **7**, 438–450 (1995).
[8] J. P. Mäkelä, G. Karmos, M. Molnar, V. Csepe, and I. Winkler, ''Steady-state responses from the cat auditory cortex,'' Hear. Res. **45**, 41–50 (1990).
[9] M. Yoshida, L. D. Lowry, J. J. C. Liu, and K. Kaga, ''Auditory 40-Hz responses in the guinea pig,'' Am. J. Otolaryngol. **5**, 404–410 (1984).
[10] G. B. Azzena, C. Conti, R. Santarelli, F. Ottaviani, G. Paludetti, and M. Maurizi, ''Generation of human auditory steady-state responses (SSRs), I: Stimulus rate effects,'' Hear. Res. **83**, 1–8 (1995).
[11] N. Forss, J. P. Mäkelä, L. McEvoy, and R. Hari, ''Temporal integration and oscillatory responses of the human auditory cortex revealed by evoked magnetic fields to click trains,'' Hear. Res. **68**, 89–96 (1993).
[12] R. Hari, M. Hämäläinen, and S. L. Joutsiniemi, ''Neuromagnetic steady-state responses to auditory stimuli,'' J. Acoust. Soc. Am. **86**, 1033–1039 (1989).
[13] M. Joliot, U. Ribary, and R. Llinas, ''Human oscillatory brain activity near 40 Hz coexists with cognitive temporal binding,'' Proc. Natl. Acad. Sci. USA **91**, 11748–11751 (1994).
[14] C. Pantev, T. Elbert, S. Makeig, S. Hampson, C. Eulitz, and M. Hoke, ''Relationship of transient and steady-state auditory evoked fields,'' Electroencephalogr. Clin. Neurophysiol. **88**, 389–396 (1993).
[15] C. Pantev, L. E. Roberts, T. Elbert, B. Roß, and C. Wienbruch, ''Tonotopic organisation of the sources of human auditory steady-state responses,'' Hear. Res. **101**, 62–74 (1996).
[16] G. Plourde, D. R. Stapells, and T. W. Picton, ''The human auditory steady-state evoked potentials,'' Acta Oto-Laryngol. Suppl. **491**, 153–159 (1991).
[17] R. Santarelli, M. Maurizi, G. Conti, F. Ottaviani, G. Paludetti, and V. E. Pettorossi, ''Generation of human auditory steady-state responses (SSRs). II. Addition of responses to individual stimuli,'' Hear. Res. **83**, 9–18 (1995).
[18] D. R. Stapells, D. Linden, J. B. Suffield, G. Hamel, and T. W. Picton, ''Human auditory steady state potentials,'' Ear Hear. **5**, 105–113 (1984).
[19] A. A. Borbély, ''Changes in click evoked responses as a function of depth in auditory cortex of the rat,'' Brain Res. **21**, 217–247 (1970).
[20] G. Conti, R. Santarelli, C. Grassi, F. Ottaviani, and G. B. Azzena, ''Auditory steady-state responses to click trains from the rat temporal cortex,'' Clin. Neurophysiol. **110**, 62–70 (1999).
[21] G. V. Simpson and R. T. Knight, ''Multiple brain systems generating the rat auditory evoked potential. I. Characterization of the auditory cortex response,'' Brain Res. **602**, 240–250 (1993).
[22] L. S. Lee, H. Lueders, D. S. Dinner, R. P. Lesser, J. Hahn, and G. Klem, ''Recording of auditory evoked potentials in man using chronic subdural electrodes,'' Brain Res. **107**, 115–131 (1984).
[23] C. Liégeois-Chauvel, A. Musolino, J. M. Badier, P. Marquis, and P. Chauvel, ''Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components,'' Electroencephalogr. Clin. Neurophysiol. **92**, 204–214 (1994).
[24] B. W. Johnson, H. Weinberg, U. Ribary, D. O. Cheyne, and R. Ancill, ''Topographic distribution of the 40 Hz auditory evoked-related potential in normal and aged subjects,'' Brain Topogr. **1**, 117–121 (1988).
[25] J. P. Mäkelä and R. Hari, ''Evidence for cortical origin of the 40 Hz auditory evoked response in man,'' Electroencephalogr. Clin. Neurophysiol. **66**, 539–546 (1987).
[26] G. L. Romani, S. J. Williamson, L. Kaufman, and D. Brenner, ''Characterization of the Human Auditory Cortex by the Neuromagnetic Method,'' Exp. Brain Res. **47**, 381–393 (1982).
[27] R. C. Oldfield, ''The Assessment and analysis of handiness: The Edinburgh inventory,'' Neuropsychologica **9**, 97–113 (1971).
[28] J. Jerger, R. Chmiel, J. D. J. Frost, and N. Coker, ''Effect of sleep on the auditory steady state evoked potential,'' Ear Hear. **7**, 240–245 (1986).
[29] M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. Lounasmaa, ''Magnetoencephalography – theory, instrumentation, and applications to noninvasive studies of the working human brain,'' Rev. Mod. Phys. **65**, 413–497 (1993).
[30] R. J. Ilmoniemi, S. J. Williamson, and W. E. Hostetler, ''New method for the study of spontaneous brain activity,'' in *Biomagnetism '87*, edited by K. Atsumi, M. Kotani, S. Ueno, T. Katila, and S. J. Williamson (Tokyo Denki University Press, Tokyo, 1987), pp. 182–185.
[31] S. E. Robinson and D. F. Rose, ''Current source image estimation by spatially filtered MEG,'' in *Biomagnetism: Clinical Aspects*, edited by M. Hoke, S. Erné, Y. Okada, and G. Romani (Excerpta Medica, Amsterdam, 1992), pp. 761–765.
[32] S. E. Robinson, ''Theory and properties of lead field synthesis analysis,'' in *Advances in Biomagnetism*, edited by S. Williamson, M. Hoke, G. Stroink, and M. Kotani (Plenum, New York, 1989), pp. 599–602.
[33] R. Rodriguez, T. Picton, D. Linden, G. Hamel, and G. Laframboise, ''Human auditory steady state responses: Effects on intensity and frequency,'' Ear Hear. **7**, 300–313 (1986).
[34] L. Zadeh and C. A. Desoer, *Linear System Theory* (McGraw-Hill, New York, 1963).
[35] A. C. Davison and D. V. Hinkley, *Bootstrap Methods and their Application* (Cambridge University Press, Cambridge, England, 1997).
[36] N. F. Viemeister, ''Temporal modulation transfer functions based upon modulation thresholds,'' J. Acoust. Soc. Am. **66**, 1364–1380 (1979).
[37] E. Terhardt, ''On the perception of periodic sound fluctuations (roughness),'' Acustica **30**, 201–213 (1974).
[38] A. Rees, G. G. R. Green, and R. H. Kay, ''Steady-state evoked responses to sinusoidally amplitude-modulated sounds recorded in man,'' Hear. Res. **23**, 123–133 (1986).

[39] R. D. Frisina, R. L. Smith, and S. C. Chamberlain, ''Encoding of amplitude modulation in the gerbil cochlear nucleus. I. A hierarchy of enhancement,'' Hear. Res. **44**, 99–122 (1990).
[40] A. Rees and A. R. Møller, ''Stimulus properties influencing the responses of inferior colliculus neurons to amplitude-modulated sounds,'' Hear. Res. **27**, 129–143 (1987).
[41] C. E. Schreiner and J. V. Urbas, ''Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields,'' Hear. Res. **32**, 49–64 (1988).
[42] C. E. Schreiner and J. V. Urbas, ''Representation of amplitude modulation in the auditory cortex of the cat. I. The anterior auditory field (AAF),'' Hear. Res. **21**, 227–241 (1986).
[43] E. Basar, R. Rosen, C. Basar-Eroglu, and F. Greitschus, ''The associations between 40 Hz-EEG and the middle latency response of the auditory evoked potential,'' Int. J. Neurosci. **33**, 103–116 (1987).
[44] S. Kuwada, R. Batra, and V. L. Maher, ''Scalp potentials of normal and hearing-impaired subjects in response to sinusoidally amplitude-modulated tones,'' Hear. Res. **21**, 179–192 (1986).
[45] E. Stürzebecher, W. Kühne, and H. Berndt, ''Detectability of the acoustically evoked composite response (40 Hz potential) near threshold,'' Scand. Audiol. **14**, 23–25 (1985).
[46] W. Szyfter, R. Dauman, and R. C. de Sauvage, ''40 Hz middle latency responses to low frequency tone pips in normal hearing adults,'' J. Otolaryngol. **13**, 275–280 (1984).
[47] D. P. Phillips and S. E. Hall, ''Response timing constraints on the cortical representation of sound time structure,'' J. Acoust. Soc. Am. **88**, 1403–1411 (1990).
[48] P. Heil, ''Auditory cortical onset responses revisited. I. first-spike timing,'' J. Neurophysiol. **77**, 2616–2641 (1997).
[49] P. Heil, ''Auditory cortical onset responses revisited. II. response strength,'' J. Neurophysiol. **77**, 2642–2660 (1997).

The perceptual tone/noise ratio of merged, iterated rippled noises with octave, harmonic, and nonharmonic delay ratios

Stephen Handel
Department of Psychology, University of Tennessee, 307 Austin-Peay Building, Knoxville, Tennessee 37996-0900

Roy D. Patterson
Centre for the Neural Basis of Hearing, Physiology Department, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom

(Received 16 August 1999; revised 17 December 1999; accepted 15 May 2000)

The perceptual tone/noise ratio was measured for merged iterated ripple noise stimuli (IRNs) in which one of the individual IRNs always had a delay of 16 ms. The second IRN was chosen to create merged IRNs with single octave delay ratios (e.g., 16 ms:8 ms), double octave delay ratios (e.g., 16 ms:4 ms), harmonic delay ratios (e.g., 16 ms:12 ms), and nonharmonic delay ratios (e.g., 16 ms:3.9 ms). All stimuli were high-pass filtered at 800 Hz. The tone/noise ratio was significantly enhanced for the octave ratios, and there was a strong interaction between the single and double octave delay ratios and number of iterations. But, the perceptual tone/noise ratio for nonoctave ratios was determined solely by the number of iterations. The pattern of the results can be explained in terms of the height of the largest peak in the summary autocorrelogram [Meddis and Hewitt, J. Acoust. Soc. Am. **89**, 2866–2882 (1991)] provided the model is modified to improve the simulation of the loss of phase locking. [S0001-4966(00)04408-8]

PACS numbers: 43.66.Ba, 43.66.Hg, 43.66.Jh, 43.66.Mk [SPB]

INTRODUCTION

Rippled noise is constructed by delaying a random noise by d ms and adding it back to the same noise. If this process is repeated n times, the resulting iterated rippled noise (IRN) produces a two-component perception; there is a tonal component with a buzzy timbre sounding like an airplane propeller with a pitch at $1/d$, and a noise component sounding like a hiss. As the number of iterations increases, the resulting sound undergoes a continuous transition from a noisy *shshsh* in which the tonal component is hard to discern (one iteration), to a *tonal* noise in which the tone is muted (two to four iterations), to a noisy tone in which the tone is dominant (four to eight iterations), to a tonal rasp with only a hint of noise (16 iterations or more). This transformation illustrates that noise and tone exist on a continuum; to use Von Helmholtz's (1877/1954tr) words, ''they can pass insensibly into one another.'' Yet, the parsing of the sound into tone and noise is effortless and irresistible. It is one example of the general perceptual achievement of figure ground organization (Bregman, 1990).

To measure the perceptual tone/noise ratio of IRNs with 1, 2, 4, 8, and 16 iterations, Patterson *et al.* (1996) added broadband noise to a tonal standard IRN with 256 iterations until the perceptual ratios matched. An IRN with one iteration was matched by a standard with a perceptual tone/noise ratio of about -3 dB and each doubling of the number of iterations increased the perceptual tone/noise ratio by roughly 4 dB. Patterson *et al.* (1996), following Yost (1996), argued that at least as a first approximation, the perceptual tone/noise ratio is an exponential function of the height of the peak at $1/d$ in ms of the autocorrelation of the sound wave.

Subsequently, Patterson *et al.* (2000) used the same matching procedure to measure the perceptual tone/noise ratio of two IRNs with different delays played simultaneously. In this latter work, both IRNs had the same number of iterations while the delays differed by ± 0.1 ms or ± 1.1 ms (e.g., 16.0 ms and 16.1 ms for the smaller difference and 16 ms and 17.1 ms for the larger difference). In this case, the autocorrelation model predicts that the pitch of the merged IRN will split into two because there are two peaks in the autocorrelation, one at each delay. But, that is not what happened perceptually because the pitches are nearly identical. Listeners heard a single pitch and the perceptual tone/noise ratio decreased dramatically from that of a single IRN with the same number of iterations. Furthermore, the increase in the perceptual tone/noise ratio as a function of the number of iterations was much smaller for the merged IRNs than for a single IRN. Patterson *et al.* (2000) showed that a full auditory model incorporating limitations on phase locking provided a reasonably accurate quantitative simulation of these results.

The purpose of the present experiment was to measure the perceptual tone/noise ratios of merged IRNs across a wide range of delay ratios greater than 1.1 ms to determine if merged IRNs with octave and simple harmonic delay ratios produce higher perceptual tone/noise ratios than do merged IRNs with nonharmonic delay ratios.

I. METHOD

The current study was conceived as an extension to the previous studies so that both IRNs had the identical number of iterations and one of the two IRNs had a delay of 16 ms (as in previous work). The delay of the second IRN was

chosen to create several prototypical delay ratios. First, IRNs with delays of 8 and 4 ms were selected to create single and double octave delay ratios. Second, IRNs with delays of 6, 10, 12, and 14 ms were selected to create simple integer harmonic delay ratios, particularly for the 16:12 pair. Third, IRNs with delays of 3.9, 4.1, 7.8, 8.2, 11.7, and 12.3 ms were selected to create nonharmonic delay ratios that bounded the strongly harmonic ones at 4, 8, and 12 ms.

The stimuli were high-pass filtered at 800 Hz (as in previous work) to remove resolved harmonics using the criterion of Houtsma and Smurzynski (1990). Patterson *et al.* (1996) found no difference in the perceived tone/noise ratios between high-pass and unfiltered IRNs when both the test and standard stimuli were filtered in the same way. The pitch of IRNs with delays in the range of 12–16 ms is not as strong as that of IRNs with delays in the range of 4–6 ms, but when the energy of the stimuli is equated, the difference in pitch strength is not excessive. Moreover, recent measurements by Pressnitzer *et al.* (1999) indicate that IRNs with a 16-ms delay are within the musical pitch region for stimuli with many harmonics, as found for IRNs in the present study. The degree of resolution of harmonics is not of major importance here, but as a matter of interest when the second IRN has a delay greater than 10 ms, there are probably no truly resolved harmonics. However, when the delay of the second IRN is as small as 6 or 4 ms, and the delay ratio is harmonic, there may well be several local maxima in the auditory spectrum at harmonics of the fundamental of the second IRN. These peaks might enhance the perceived tone/noise ratio somewhat in this condition.

A. Stimuli

1. Standard IRNs

The tonal component for the standard sequence was based on an IRN with 256 iterations and a delay of 16 ms to maximize the similarity of the standards to the experimental merged stimuli, as in Patterson *et al.* (2000, 1996). We will use the same notation as before. An IRN with delay d ms and n iterations is designated IRN (d,n). Five samples of Gaussian noise were used to generate five 1.25-s samples of IRN (16, 256) using the add-same procedure (Yost *et al.*, 1996). The sequence of standards was produced by adding synchronously a fixed-level IRN (16, 256) to a 1.25-s sample of random noise. The noise varied in level from 3 dB above to 24 dB below the level of the IRN in 0.75-dB steps in the middle of the sequence and 1.5-dB steps at the extremes. Twenty samples of each standard were constructed by combining the five samples of IRN (16, 256) with four different random noises attenuated at the appropriate level.

2. Merged IRNs

As described above, the pair of IRNs both had the same number of iterations and one IRN always had a 16-ms delay to match the delay of the IRN in the standard sequence. The delays of the second IRN were: 3.9 or 4.1, 4, 6, 7.8 or 8.2, 8, 10, 11.7 or 12.3, 12, 14, 15.6 or 16.4 ms. (Each of the paired nonharmonic ratios will be treated as one condition.) We will use a similar notation to that for single IRNs: each merged IRN will be designated IRN $(16{:}d2,n)$ in which $d2$ is the delay in ms of the second IRN. Each of the merged IRNs was constructed at three different numbers of iterations: 4, 16, and 256. Patterson *et al.* (2000) found that four iterations was the minimum number to produce a strong tonal component and 256 iterations produced a nearly tonal percept.

3. Presentation

All stimuli were generated digitally and output by a 12-bit D/A converter, at a sampling rate of 10 kHz, and at the same level of 50 dB SPL. An antialiasing filter imposed an upper limit of 4500 Hz. The slope of the filter was 48 dB/octave below the cutoff frequency.

B. Procedure

On each trial, the subject was presented a standard followed by a merged IRN and was required to pick the sound with the greater tone/noise ratio. The duration of both the standard and merged IRN was 1.25 s and there was a 500-ms silent interval between the two sounds. Following the merged IRN, there was a 5-s interval for subjects to mark their choice on a scoring sheet.

The order of the trials was based on the method of limits. For each of the merged IRNs, preliminary listening identified a set of eight to ten standards that bounded the tone/noise ratio of the experimental merged IRN. Following this, two descending sequences (e.g., standards progressing from very tonal to very noisy) and two ascending sequences were constructed, each consisting of the identified standards. During the experiment, the four sequences for each merged IRN were presented together in a descending, ascending, descending, ascending order. The starting standard was varied so that the subjects could not simply count the number of trials to make a consistent choice of whether the standard or merged IRN had the higher tone/noise ratio.

For each of the three iteration levels, there was a total of ten conditions. Two conditions generated octave ratios: 16:8 and 16:4. Four conditions generated harmonic (integer) ratios: 16:14, 16:12, 16:10, and 16:6. Four conditions generated nonharmonic (noninteger) ratios: 16:16.4 or 16:15.6, 16:12.3 or 16:11.7, 16:8.2 or 16:7.8, and 16:4.1 or 16:3.9. In all sequences, the two nonharmonic ratios were alternated randomly. In addition to these 30 merged IRN conditions, there were two individual IRN conditions—IRN (16,4) and IRN (16,16)—to serve as baselines and to provide comparisons with conditions used previously.

The 32 conditions were presented in a different random order to each of the five subjects across four sessions on different days. Prior to beginning the experiment, each subject practiced on a variety of conditions for about 30 min. No feedback was given at any time. The practice session was merely for familiarity with the stimuli. The five subjects consisted of one of the authors (S. H.) and four undergraduates who were paid for their participation. All had normal hearing.

II. RESULTS

The matching point for each sequence was the midpoint between the standard judged as having a higher perceptual tone/noise ratio and the next standard judged as having a

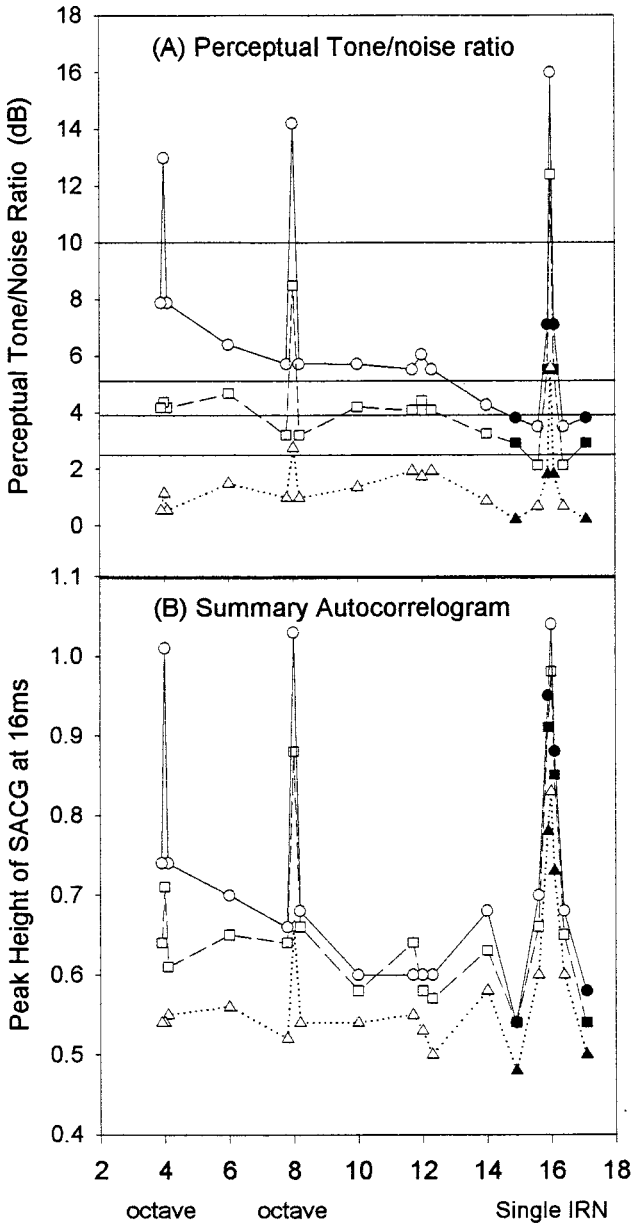


FIG. 1. (A) Perceptual tone/noise ratio; (B) height of summary autocorrelogram at base delay of 16 ms. The abscissa is the delay of the variable IRN of the merged IRN. Circles (○), squares (□), and triangles (△) represent merged IRNs with 256, 16, and 4 iterations, respectively. Filled symbols show data from Patterson *et al.* (2000). The perceptual tone/noise ratio for the single IRN with 256 iterations was set equal to 16 in (A).

lower ratio. The perceptual matching points found here and those found in Patterson *et al.* (2000) are shown in Fig. 1, panel (A).

The matching points for the two single IRNs with 4 and 16 iterations were compared to the results from Patterson *et al.* (2000, 1996) using independent *t*-tests. There were no differences between the results from all the experiments for either IRN.

The matching points for the merged IRNs were analyzed using a within-subject analysis of variance (ANOVA). The two descending and two ascending sequences were used as replications and the main effects were tested using the appropriate effect by subject interaction. The results were straightforward. There was no main effect due to direction of the sequences $\{F(1,4)=0.56\}$ and only a small nonsystematic IRN condition by direction effect $\{F(31,124)=1.92,\ p<0.05,\ \mathrm{MSE}=0.92\}$. The only significant effect was that due to the differences between the IRN conditions $\{F(31,124)=55.8,\ p<0.001,\ \mathrm{MSE}=4.4\}$. The differences between the merged conditions were analyzed further using Tukey's honestly significant difference procedure. Five groups of equivalent conditions were found and these groups are separated by the four horizontal lines in Fig. 1.

Consider first the octave delay ratios (i.e., 16:8, 16:4). For merged IRNs with four iterations, there was no significant increase in the perceptual tone/noise ratio for 16:8 or for 16:4 (in all cases, the increase is measured with respect to the surrounding nonharmonic ratios). For merged IRNs with 16 iterations, there was a significant increase in the perceptual tone/noise ratio for 16:8 (5.2 dB), but no increase for 16:4. For IRNs with 256 iterations, there was a significant increase in the perceptual tone/noise ratio for both 16:8 and 16:4 (8.6 and 5.2 dB, respectively). But, there was no significant difference (1.2 dB) between those two perceptual tone/noise ratios, which mirrors the lack of difference in the summary autocorrelograms described in the discussion. These results indicate that the delay ratio and the number of iterations interact to affect the perceptual tone noise ratio.

Consider second all the remaining harmonic and nonharmonic delay ratios. For 4, 16, and 256 iterations the perceptual tone/noise ratios did not differ significantly across the entire range of delay ratios. The average perceptual tone/noise ratio for 4, 16, and 256 iterations was 1.4, 4.1, and 5.6 dB, respectively, a significant but much smaller difference than found for individual IRNs. For each of the three iteration levels, the perceptual tone/noise ratio is lowest when the delay ratio is 16:15.6 or 16:16.4, gradually increases to a plateau at the perfect fourth delay ratio (16:12), and remains in the same region for the remaining delay ratios. There were no significant differences between any pair of harmonic to nonharmonic ratios. There were slight increases of about 2 dB at the 16:3.9 and 16:4.1 delay ratios for merged IRNs with 256 iterations. This increase may reflect the resolved harmonics and stronger pitch strength found for IRNs with shorter delays.

III. DISCUSSION

The basic question is why *only* octave ratios yield stronger perceptual tone/noise ratios. In this section we argue that the answer lies in the stimulus itself. The merged IRNs with octave ratios have a greater degree of periodicity at the base delay of 16 ms. Thus, the effect is peripheral rather than central. We calculate the tone noise ratio of the merged IRN as a unit peripherally and that increase in periodicity at the base delay creates the greater tone/noise ratio. We do not calculate the tone/noise ratios of the individual IRNs one at a time peripherally and then combine them centrally for the merged IRN.

In the autocorrelation function of a single IRN with delay d, the largest peak in the autocorrelation occurs at time d, but there are also smaller peaks at multiples of d. Their height decreases as the multiple increases, but the rate of

decrease is slower for IRNs with a greater number of iterations. In merged IRNs with octave ratios, a higher-order peak in the IRN with the shorter delay coincides with the first peak of the IRN with the longer delay and increases its relative height. This increases the perceived tone/noise ratio and the effect is bigger for IRNs with more iterations.

In time-domain auditory models, the equivalent of the autocorrelation function is the summary autocorrelogram (Meddis and Hewitt, 1991). We used the summary autocorrelogram (SACG) to measure the degree of periodicity in merged IRNs to see if the height of the peak at the base delay increased suddenly for octave delays as found in the perceptual data. Consider the SAGC produced by single IRNs with 8- and 16-ms delays and what happens when they are combined. The largest peak in the SACG of IRN $(8,n)$ occurs at 8 ms. There is, however, another peak that is not much smaller at 16 ms (particularly for 16 and 256 iterations), which is where the largest peak in the SACG of IRN $(16,n)$ occurs. When IRN $(8,n)$ and IRN $(16,n)$ are combined, the correlation at 16 ms in IRN $(8,n)$ reinforces the main peak in IRN $(16,n)$ with the result that the largest peak in the merged IRN $(16{:}8,n)$ occurs at 16 ms. This reinforcement of the peak at 16 ms only occurs for octave ratios and does not occur for any other delay combination.[1] Moreover, the reinforcement is strongest for 256 iterations, intermediate for 16 iterations, and weakest for four iterations. This suggests that the height of the largest peak in the SACG of the merged IRNs might explain the pattern of results observed in the current experiment.

Patterson *et al.* (2000) recently argued that the tone/noise ratio of merged IRNs indicated that current time-domain models do not accurately reflect the loss of phase locking during the early stages of neural processing. They increased the temporal smoothing in the auditory image model and with this modification showed that the height of the largest peak in the summary auditory image could be used to explain how the pitch strength of merged IRNs decreases as the difference in delay between the two component IRNs increases from 0.1 to 1.1 ms. We extended this approach to the merged IRNs of the current experiment using a slightly different auditory model.

We calculated SACGs for the merged IRN stimuli using a modified version of the Meddis and Hewitt model (1991) in which the minimum and maximum center frequencies in the gammatone filterbank were 400 and 4000 Hz. To limit phase locking, we introduced temporal smoothing between the hair-cell stage and the autocorrelation stage as recommended in Patterson *et al.* (2000). The smoothing was implemented by a three-stage low-pass filter with a time constant of 0.133 ms. For each merged IRN condition, 100 different samples were created and SACGs were calculated for each. The SACGs were then averaged and in each case, the maximum value of the SACG occurred at, or very close to, a 16-ms delay. The values are plotted in Fig. 1, panel (B). A comparison of panels (A) and (B) reveals that there is a close correspondence between the height of the largest peak in the SACG and the perceived tone/noise ratio, although the SACG values appear to predict slightly greater perceptual tone/noise ratios than found for the octave conditions. This correspondence between the perceptual judgments and the time intervals in the simulated neural activity suggests that the degree of phase locking in peripheral processing affects the perceptual tone/noise ratio, and that the simulation of the loss of phase locking is important for time-domain models of auditory processing.

ACKNOWLEDGMENTS

The stimuli were constructed at the MRC-APU while S. Handel was supported by professional development award from the Graduate School, University of Tennessee, Knoxville. The authors would like to thank Daniel Pressnitzer for his support in the calculation and analysis of the summary autocorrelograms and thank A. J. Datta for his ever-patient help in constructing the stimuli.

[1]It is possible for reinforcement to occur at nonoctave delay ratios for higher-order multiples of 16 ms, e.g. at 48 ms for an IRN $(16,n)$ merged with an IRN $(12,n)$. There was no evidence that this reinforcement increased the perceptual tone/noise ratio (see Fig. 1), perhaps because the common delay is associated with a repetition rate of 22 Hz which is below the lower limit of pitch.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).

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.

Meddis, R., and Hewitt, M. J. (**1991**). ''Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–82.

Patterson, R. D., Handel, S., Yost, W. A., and Datta, A. J. (**1996**). ''The relative strength of the tone and noise components in iterated rippled noise,'' J. Acoust. Soc. Am. **100**, 3286–3294.

Patterson, R. D., Yost, W. A., Handel, S., and Datta, A. J. (**2000**). ''The perceptual tone/noise ratio of merged iterated rippled noises,'' J. Acoust. Soc. Am. **107**, 1578–1588.

Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**1999**). ''The lower limit of melodic pitch with filtered harmonic complexes,'' J. Acoust. Soc. Am. **105**, 1152.

von Helmholtz, H. (**1877/1954tr**). *On the Sensations of Tone* (Dover, New York).

Yost, W. A. (**1996**). ''The pitch strength of iterated rippled noise,'' J. Acoust. Soc. Am. **100**, 3329–3335.

Yost, W. A., Patterson, R. D., and Sheft, S. (**1996**). ''A time domain description for the pitch strength of iterated rippled noise,'' J. Acoust. Soc. Am. **99**, 1066–1078.

Pitch matches between unresolved complex tones differing by a single interpulse interval

Christopher J. Plack
Department of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, England

Louise J. White
Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, England

(Received 31 August 1999; revised 13 January 2000; accepted 27 April 2000)

The experiment compared the pitches of complex tones consisting of unresolved harmonics. The fundamental frequency ($F0$) of the tones was 250 Hz and the harmonics were bandpass filtered between 5500 and 7500 Hz. Two 20-ms complex-tone bursts were presented, separated by a brief gap. The gap was an integer number of periods of the waveform: 0, 4, or 8 ms. The envelope phase of the second tone burst was shifted, such that the interpulse interval (IPI) across the gap was reduced or increased by 0.25 or 0.75 periods (1 or 3 ms). A ''no shift'' control was also included, where the IPI was held at an integer number of periods. Pitch matches were obtained by varying the $F0$ of a comparison tone with the same temporal parameters as the standard but without the shift. Relative to the no-shift control, the variations in IPI produced substantial pitch shifts when there was no gap between the bursts, but little effect was seen for gaps of 4 or 8 ms. However, for some conditions with the same IPI in the shifted interval, an increase in the IPI of the comparison interval from 4 to 8 ms (gap increased from 0 to 4 ms) changed the pitch match. The presence of a pitch shift suggests that the pitch mechanism is integrating information across the two tone bursts. It is argued that the results are consistent with a pitch mechanism employing a long integration time for continuous stimuli that is reset in response to temporal discontinuities. For a 250-Hz $F0$, an 8-ms IPI may be sufficient for resetting. Pitch models based on a spectral analysis of the simulated neural spike train, on an autocorrelation of the spike train, and on the mean rate of pitch pulses, all failed to account for the observed pitch matches.
[S0001-4966(00)01408-9]

PACS numbers: 43.66.Ba, 43.66.Hg, 43.66.Mk, 43.66.Nm [DWG]

INTRODUCTION

There is a large improvement in fundamental frequency ($F0$) discrimination with increasing duration for complex tones consisting of high harmonics that are *unresolved* by the peripheral auditory system, although a much smaller effect of duration is observed for *resolved* complex tones (Plack and Carlyon, 1995; White and Plack, 1998). The effect of duration suggests that the auditory system increases the accuracy of an $F0$ estimate by combining information over time.

White and Plack (1998) tried to determine the mechanism underlying the duration effect for unresolved complex tones. Specifically, they considered two rival hypotheses:
(i) The pitch mechanism uses a fixed, long-duration, sampling window: The more information within the integration window, the better the performance.
(ii) The pitch mechanism uses a short sampling window. The accuracy of the final pitch estimate is improved by combining several of these discrete pitch estimates, as in the ''multiple looks'' model (Viemeister and Plack, 1993; Viemeister and Wakefield, 1991). This hypothesis predicts that the detectability index, d', should increase by a factor of $\sqrt{n}$ for every n-fold increase in duration.

White and Plack showed that the improvement in $F0$ discrimination with duration was much greater than that predicted by the multiple looks model. Large performance improvements were found up to a duration of around 80 ms for an $F0$ of 250 Hz, and up to a duration of around 160 ms for an $F0$ of 62.5 Hz. The results suggest that the auditory system uses a long integration time for these stimuli. However, it was also demonstrated that the large improvement in performance for a 40-ms tone burst compared to a 20-ms burst depended on the 40-ms burst being continuous. When the 40-ms burst was split into two 20-ms bursts separated by a gap of only 5 ms, performance was much worse than when there was no gap between the tone bursts. In fact, performance for the gapped conditions was almost identical to the predictions of the multiple looks model.

As an explanation for these results, White and Plack proposed a third hypothesis that combines elements of the two outlined above:
(iii) The pitch mechanism for unresolved harmonics uses a long integration window when the tone is continuous, but the integration window is ''reset'' when there is a break in the tone, leading to two separate $F0$ estimates that are combined optimally to produce the discrimination performance predicted by multiple looks theory.

In other words, the third hypothesis proposes that a temporal discontinuity automatically causes the system to begin a separate pitch estimate. Nabalek and colleagues (Nabalek *et al.*, 1973; Nabelek, 1996) have made similar claims regarding the ''critical pause duration'' for the separate analysis of the frequency of pure tone bursts. Similarly, Bregman

et al. (1994a, b) suggested that the *onset* of a pure tone may cause the system to start a separate frequency estimate. Finally, Hafter and Buell (1990) showed that a type of resetting mechanism may operate in the binaural system. White and Plack argued that a flexible pitch resetting mechanism may be advantageous in the environment where a discontinuity often reflects the onset of a separate sound source, or of a feature within a sound, that requires independent analysis.

Although the resetting hypothesis is attractive, the results of White and Plack do not lead to this conclusion unequivocally. For example, it is conceivable that the gap interferes with performance in nonspecific ways, perhaps by causing a ''distraction,'' so that, although the auditory system may be integrating across the gap, performance is impaired compared to the continuous condition.

It can also be argued that a large effect of duration on performance does not imply a long integration process. In an analysis of absolute signal detection, Viemeister and Wakefield (1991) suggested that large improvements with duration can arise from a multiple-looks process if it is assumed that the looks do not carry equal information or do not receive equal weight. For example, if the pitch information is poor near stimulus onset, possibly because of a high variability in the neural activity at onset (Relkin and Pelli, 1987; Teich and Khanna, 1985; Young and Barta, 1986), then samples later in the stimulus may contain much more useful information than earlier samples. $F0$ discrimination for a 40-ms tone may be far superior to that for a 20-ms tone simply because the first 20-ms portion of the tone produces a very inaccurate pitch estimate (see White and Plack, 1998). Furthermore, in the two-burst condition the gap may effectively reset the onset response and so lead to two 20-ms segments with equally poor pitch information that are combined according to signal detection theory to produce a $\sqrt{2}$ improvement in d'.

Nabalek (1996) provided an experimental design that may be used to address both of these issues. If there is an instantaneous phase shift between two pure tone bursts, such that the second burst starts at a different phase than the first burst would have had if it had been continuous, then this can produce a shift in the position of the maximum peak in the Fourier spectrum. Specifically, a $\pi/2$ phase advance causes the spectral maximum to be located at a *higher* frequency than the carrier frequency, and a $3\pi/2$ phase advance causes the spectral maximum to be located at a *lower* frequency than the carrier frequency. Nabalek showed that these phase changes could produce shifts in the overall pitch of the tone pair, but only when the gap between the bursts was less than the critical pause duration of between 8 and 16 ms. The presence of a pitch shift suggests that the auditory system is performing the equivalent of a long-term Fourier transform that *includes both tone bursts*: If the two tone bursts are analyzed independently (for example, by a discrete-sampling, multiple-looks mechanism) then no pitch shift should be observed. However, a short time-constant, multiple-looks mechanism may predict a pitch shift if one of the looks includes the end portion of the first burst and the beginning portion of the second burst, bridging the phase discontinuity.

The main empirical contribution of the current article is

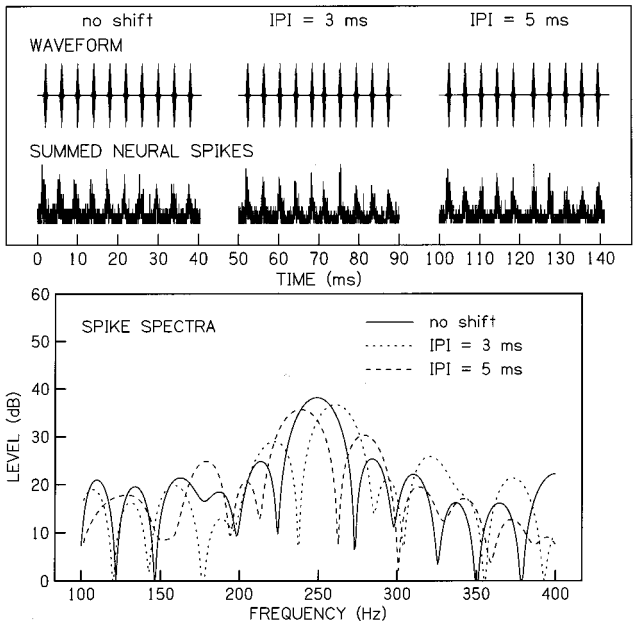


FIG. 1. The top panel shows three complex tone waveforms, two of which have shifts in the central IPI. Below each waveform is a simulation of the corresponding auditory nerve response. (The units are linear with respect to velocity and electrical potential, respectively.) The bottom panel shows the Fourier spectra of the neural activity patterns.

an adaptation of Nabalek's technique to unresolved complex tones. Harmonic complexes with an $F0$ of 250 Hz were passed through a single bandpass filter (5500–7500 Hz) so that there was no *spectral* information regarding $F0$ in terms of the excitation pattern on the basilar membrane. On this basis we might not expect to observe the pitch shifts predicted by Nabalek's experiment. However, for unresolved complexes there is information about $F0$ in the activity of auditory nerve fibers, which will tend to phase lock to the envelope produced by the interaction of the unresolved harmonics. Applying a phase shift to the $F0$ of the complex, while keeping all the harmonics in sine phase relative to $F0$, results in a shift in the envelope of the complex. Since complex tones tend to show discrete ''pitch pulses'' (i.e., envelope peaks), the envelope shift can be regarded as a shift in a single interpulse interval (IPI). The shift is reflected in the firing patterns in the auditory nerve, and in the *spectrum* of the firing patterns in the manner predicted by Nabalek.

As an illustration, Fig. 1 shows a simulation of auditory nerve activity provided by the LUTEar software[1] developed by Meddis and colleagues (Meddis and Hewitt, 1991a, b; Meddis *et al.*, 1990). The model comprises a simulation of the outer and middle ear transfer functions, a ''gammatone'' auditory filter (Patterson *et al.*, 1988) with a center frequency of 6.5 kHz, an inner-hair cell simulation (Meddis, 1986, 1988; Meddis *et al.*, 1990), and a stochastic auditory nerve spike generator. The simulation is described in more detail in White and Plack (1998).

The top panel of Fig. 1 shows three of the waveforms used in the experiment, together with the associated neural activity patterns, comprising the simulated neural spikes summed over 500 repetitions of the model. The firing pat-

terns exhibit strong phase locking to the envelope of the waveforms. The lower panel shows the *long-term* spectra of these activity patterns. The spectra were estimated by an FFT with the duration extended to 4 s by zero-padding (for illustrative purposes: The use of zero-padding produces a smooth spectrum that allows the maximum to be seen more clearly). Although the $F0$ component is not physically present in the stimuli, $F0$ is very strongly represented in the neural activity patterns. This occurs by virtue of the nonlinearities (mostly the half-wave rectification by the hair cells) in the transduction process; in effect, the spectra in Fig. 1 are approximations of the envelope spectra for the complex tones. Indeed, one definition of ''envelope'' is the low frequency distortion produced by half-wave rectification (Viemeister and Plack, 1993). The hair-cell section of the model performs rectification, and the spectral components illustrated in Fig. 1 reflect the low frequency components at the output of the nonlinearity.

It can be seen that a shift in a single IPI can produce a shift in the spectral maximum equivalent to the spectral shift seen for pure tones: A 1-ms reduction in the IPI (equivalent to a $\pi/2$ phase shift, or -0.25 periods) moves the spectral maximum upward, and a 1-ms increase in the IPI (equivalent to a $3\pi/2$ phase shift, or $+0.25$ periods) moves the spectral maximum downward. The present experiment uses these shifts to test hypotheses about the integration mechanisms involved in pitch perception.

I. METHOD

A. Rationale

Following the reasoning outlined above, if the pitch mechanism for unresolved harmonics is conducting the *equivalent* of long-term spectral analysis on the neural activity in order to extract $F0$, then manipulating the IPI between two 20-ms complex-tone bursts, so that the IPI is not an integer number of periods, should produce a shift in pitch. If the pitch shift is still heard when there is a gap between the tone bursts, then that will prove that the system is integrating information across the gap, and hence that the resetting hypothesis is incorrect. On the other hand, if the pitch shift is heard when the bursts are contiguous but is *not* heard when a gap is introduced, then that will be consistent with the hypothesis that the system is using a long-integration window for a continuous tone, but is analysing the two bursts independently when there is a discontinuity.

The use of unresolved harmonics ensures that there is no information about $F0$ in the distribution of excitation on the basilar membrane. In other words, the results can be interpreted solely in terms of auditory mechanisms acting upon the temporal firing patterns in the auditory nerve.

B. Stimuli

Stimuli consisted of two 20-ms complex-tone bursts (gated with no ramps), separated by silent intervals of 0, 4, and 8 ms. Each tone burst was composed of the harmonics of a 250-Hz $F0$, bandpass filtered between 5500 and 7500 Hz. Filtering occurred after gating. The harmonics were always added in sine phase, relative to the $F0$ component. In other words, the (nominal) waveform before filtering is given by

$$a(t)=A\sum_{n=1}^{\infty}[\sin(n(2\pi Ft+\theta))] \quad (0\leqslant t<0.02), \qquad (1)$$

$$a(t)=A\sum_{n=1}^{\infty}[\sin(n(2\pi Ft+\theta+\varphi))]$$

$$((G+0.02)\leqslant t<(G+0.04)), \qquad (2)$$

where t is time in seconds, $a(t)$ is the waveform of the complex, A is the peak amplitude of each harmonic, n is the harmonic number, F is $F0$ in Hz, θ is the starting phase of the $F0$ component ($n=1$) of the first tone burst, and φ is the starting phase of the $F0$ component of the second tone burst relative to the first. G is the gap duration, in seconds. Each harmonic component had a level of 50 dB SPL before filtering. The overall level of the complex after filtering was about 58 dB SPL.

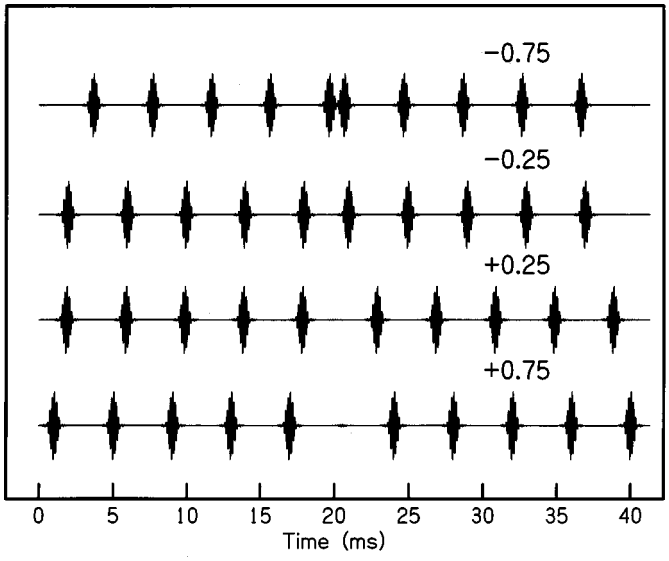


FIG. 2. The waveforms for four of the 0-ms gap conditions tested in the experiment. The labels give the IPIs in $F0$ periods relative to the unshifted comparison (see Table I).

The IPI between the two bursts was either an integer number of periods (no shift), advanced by 0.25 or 0.75 periods (1 or 3 ms) relative to this, or delayed by 0.25 or 0.75 periods relative to this (see Fig. 2). The no-shift reference was added to control for any bias in the matching procedure. Note that the actual silent interval between the bursts was always either 0, 4, or 8 ms: The shift in the IPI was produced by a phase change in the envelope of the second tone burst relative to the first (φ), and by varying the starting phase of the envelope of the first tone burst (θ), not by physically advancing or delaying the whole waveform. Similarly, the total stimulus duration was either 40, 44, or 48 ms (depending on the gap duration). The waveforms for these stimuli, for the 0-ms gap condition, are shown in Fig. 2. A full list of the conditions, detailing the IPIs across the gap, can be found in Table I. In addition, Table I shows the starting phase of the envelope of the first burst (θ), and the relative phase of

TABLE I. The interval between pitch pulses (''IPI'') across the gap between the tone bursts for the 15 conditions tested in experiment 1 (standard interval). Also shown is the starting envelope phase for the first burst (θ), and the envelope phase of the second burst relative to the first (φ).

Shift in periods	θ	φ	Gap=0 ms	Gap=4 ms	Gap=8 ms
−0.75	$\pi/4$	$3\pi/2$	1 ms	5 ms	9 ms
−0.25	π	$\pi/2$	3 ms	7 ms	11 ms
0 (no shift)	π	0	4 ms	8 ms	12 ms
+0.25	π	$3\pi/2$	5 ms	9 ms	13 ms
+0.75	$7\pi/4$	$\pi/2$	7 ms	11 ms	15 ms

the envelope of the second burst (ϕ), used to create each IPI. Notice that the same IPI for the standard interval can occur with different gap durations.

Stimuli were generated digitally with 32-bit resolution on a Silicon Graphics workstation and filtered after gating using a digital FIR filter with a nominal attenuation of 90 dB per octave. The sampling rate was 48 kHz. Stimuli were played out through a built-in 16-bit DAC and were presented monaurally to the right ear via Beyer Dynamic DT100 headphones (100 Ω) which were connected directly to the headphone output of the computer.

C. Procedure

Pitch matches were obtained using the adaptive procedure described by Jesteadt (1980). In each trial, two observation intervals (separated by 500 ms), each containing two tone bursts, were presented. The ''standard'' interval contained the shifted condition with the $F0$ fixed at 250 Hz. The ''comparison'' interval contained a no-shift reference with the $F0$ varied adaptively to obtain the pitch match.[2] The gap between the tone bursts in the comparison interval was the same as in the standard interval, but the IPI (in ms) varied depending on the $F0$. (For a comparison $F0$ of 250 Hz, the IPI was gap duration plus 4 ms. The IPI for the comparison interval was always an integer number of waveform periods.) On each trial the order of standard and comparison intervals was random.

Two interleaved 2AFC adaptive tracks were used to bracket the pitch match. In one track the $F0$ of the preserved-phase condition started 20% above 250 Hz (upper track) and in the other the $F0$ started 20% below 250 Hz (lower track). Each trial in a block was selected randomly from the two adaptive tracks. A ''two-down one-up'' rule was applied to the upper track and a ''two-up one-down'' rule to the lower track (Levitt, 1971). The step size was 4% for the first four reversals and 2% thereafter.

In other words, on each trial the listener was required to select which interval had the higher pitch. For the upper track, if the listener chose the comparison interval two times in a row, the $F0$ of the comparison interval was decreased (by the step size) for the next trial. Each time the listener chose the standard interval, the $F0$ of the comparison interval was increased (by the step size) for the next trial. Similarly for the lower track, if the listener chose the *standard* interval two times in a row, the $F0$ of the comparison interval was *increased* for the next trial. Each time the listener chose the *comparison* interval, the $F0$ of the comparison interval was *decreased* for the next trial.

Testing continued until 16 turnpoints (transitions from increasing to decreasing step size) had been completed for each track and the mean of the comparison $F0$s (in % relative to 250 Hz) for the last 12 turnpoints from each track was taken as the threshold for that track. The *mean* of the thresholds from the upper and lower tracks is an estimate of the pitch match. Half the *difference* between the thresholds from the upper and lower tracks is an estimate of the $F0$ discrimination threshold: The upper track converges, effectively, on the 70.7% point on the psychometric function for detecting an increase in $F0$, and the lower track converges on the 29.3% point for detecting an increase in $F0$ (or the 70.7% point for detecting a *decrease* in $F0$). The conditions were tested in a random order, within each experimental session. Each condition was repeated 12 times and the mean of the thresholds from the last 10 repetitions was used to calculate the pitch matches.

Listeners made their responses using the numeric keypad on the computer keyboard. ''Lights'' were presented in a graphical display on the computer monitor to delineate the observation intervals. No feedback was given.

D. Listeners

Four normally hearing listeners were tested. They were aged between 25 and 31 years. Listeners were given at least two hours training on the task before data collection began. There were no systematic improvements in performance with time. Listeners were tested individually in an Industrial Acoustics Company single-walled sound-attenuating booth.

II. RESULTS

The individual and mean pitch match results are shown in Figs. 3 and 4. In Fig. 3, the $F0$ of the comparison is plotted relative to the $F0$ of the standard (250 Hz) as a function of the nominal IPI in the comparison interval (the actual IPI varied slightly as the $F0$ of the comparison was varied). The IPI shift (in periods) of the standard stimulus relative to the comparison stimulus is the parameter. Figure 4 shows the mean data as a function of the IPI in the standard interval.

It is apparent from the results for listener SW that there was some bias in the pitch matching procedure. Even though the no-shift standard was nominally identical to the comparison, the measurement technique produced a consistent positive pitch shift for this listener. In some respects this is to be expected. The upper track is an estimate of the smallest detectable increment in $F0$ relative to 250 Hz and the lower track is an estimate of the smallest detectable decrement in

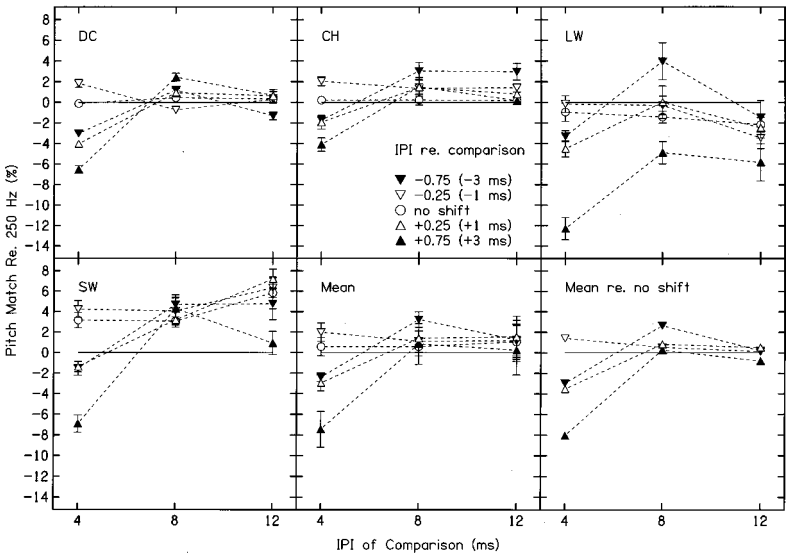


FIG. 3. The individual and mean pitch matches obtained in the experiment as a function of the IPI of the comparison stimulus (the actual gaps were 4 ms less than these values). The parameter is the IPI of the standard stimulus relative to the IPI of the comparison in $F0$ periods (and ms). The bottom right-hand panel shows mean pitch matches plotted relative to the ''no-shift'' conditions. Error bars show standard errors.

$F0$ relative to 250 Hz. Because the $F0$ discrimination threshold (in Hz) increases as $F0$ is increased (Shackleton and Carlyon, 1994), the threshold for the upper track will be larger than the threshold for the lower track, so that the mean of the two will be higher than the ''true'' pitch match.

The bottom right-hand panel of Fig. 3 shows the mean pitch matches relative to the mean matches for the no-shift conditions (for the same gap duration). Only for the 0-ms gap conditions (4-ms comparison IPI) do the pitch matches for the IPI shift conditions differ substantially from the no-shift reference. The $+0.75$ condition produced the largest mean shift of around -8%. The pitch shifts for the -0.75 and $+0.25$ conditions were similar (-3%), while the -0.25 condition produced a small positive pitch shift. The basic pattern of results for the 4-ms IPI was similar across the four listeners, although the magnitude of the pitch shifts varied. For the other comparison IPIs, three of the listeners showed no substantial effect of the IPI shift relative to the no-shift reference. Listener LW did show a negative pitch shift for the -0.25 conditions across all the comparison IPIs, although her data were highly variable. Even for this listener, the largest effects of the IPI manipulations were observed for the 0-ms gap (4-ms comparison IPI).

A repeated-measures ANOVA (listener×IPI shift×gap/IPI duration) was performed on the pitch match data. There was a significant effect of IPI shift [$F(4,12)=4.7$, $p<0.05$] and of gap/IPI duration [$F(2,6)=8.9$, $p<0.005$]. There was also a significant interaction between IPI shift and gap/IPI duration [$F(8,24)=22.1$, $p<0.01$]. Planned pairwise comparisons for the 0- and 4-ms gap conditions (4- and 8-ms comparison IPIs) were conducted. When the gap was 0 ms, the pitch matches for all the IPI shift conditions were significantly different from the pitch match for the no-shift reference. For the 4-ms gap (8-ms comparison IPI), only the -0.75 condition was significantly different from the reference.

Figure 4 demonstrates that the interpretation of the data is not as straightforward as might be hoped from the analysis purely in terms of gap duration. It is clear from this figure that conditions with the same IPI in the standard interval can produce different pitch matches. The $+0.25$ condition with a 0-ms gap, has the same absolute IPI (5 ms) as the -0.75 condition with a 4-ms gap, but a different pitch match is observed. Similarly, for the $+0.75$ condition with a 0-ms gap and the -0.25 condition with a 4-ms gap (IPI=7 ms). Of course, the IPI for the comparison stimulus is different for conditions with the same IPI in the standard interval, and it is likely that this is the cause of the differences in the pitch matches. This issue will be addressed in Sec. III B.

No additional statistics were conducted on IPI *per se*, since the relevant planned comparisons had already been tested. In short, with the exception of the -0.25 condition with a 7-ms absolute IPI, the pitch matches for all IPIs of 7 ms or less were significantly different from the no-shift control. No other pitch matches differed from the control.

The $F0$ discrimination thresholds, derived from the data using the procedure described in Sec. I C, are shown in Fig. 5. As expected from the results of White and Plack (1998), overall performance for the continuous conditions (0-ms gap/4-ms comparison IPI) was better than that for the gap conditions, consistent with the resetting hypothesis.

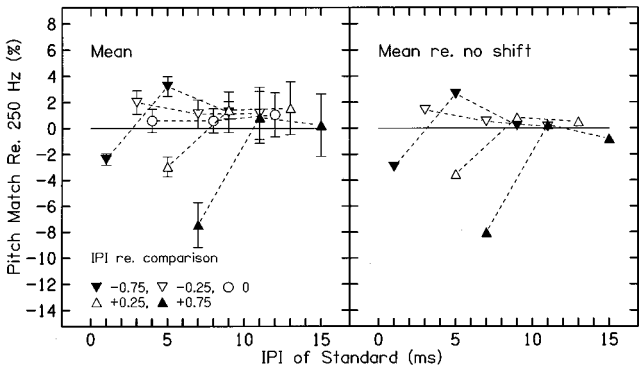


FIG. 4. The mean pitch matches obtained in the experiment as a function of the IPI of the standard stimulus. The parameter is the IPI of the standard stimulus relative to the IPI of the comparison in $F0$ periods. The right-hand panel shows mean pitch matches plotted relative to the ''no-shift'' conditions. Error bars show standard errors.

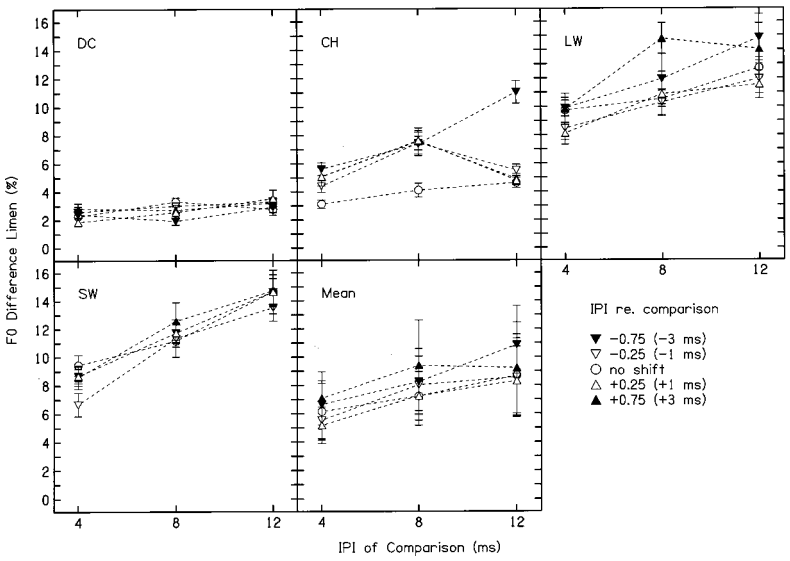


FIG. 5. The individual and mean $F0$ difference limens, derived from the tracking procedure, as a function of the IPI of the comparison stimulus. Error bars show standard errors.

III. DISCUSSION

A. Models of *F*0 extraction

So far the article has avoided being specific about the actual type of analysis that the auditory system uses to extract $F0$ from the phase-locked neural response. It is important to consider the types of analysis that may be involved in order to evaluate the resetting hypothesis in the light of the present results. Specifically, the aim is to determine whether a mechanism for $F0$ extraction using a long-integration time can account for the results *without the need for a resetting process*.

The spectral model (Fig. 1) was introduced primarily to demonstrate that information about $F0$ is encoded in the neural activity patterns: The model is equivalent to an analysis of the envelope spectrum of the complex. It would seem unlikely, however, that the auditory system implements an explicit Fourier transform in this manner. Two alternative methods of extracting $F0$ from phase-locked neural activity are described below.

1. Autocorrelation

Most temporal models of pitch perception make use of the fact that the neural activity in response to a periodic waveform is phase locked, i.e., that a repetitive waveform will produce a regular pattern of neural activity. In the models of Goldstein and Srulovicz (1977), Srulovicz and Goldstein (1983), van Noorden (1982), and Moore (1997), $F0$ is derived by analyzing the time intervals between *consecutive* spikes in the auditory nerve. More generally, however, it is possible to extract $F0$ from the time intervals between consecutive and nonconsecutive spikes using autocorrelation (Licklider, 1951).

Autocorrelation is a process whereby a signal is correlated with a delayed version of itself. For a periodic signal, the normalized correlation will be equal to one (maximum) when the delay is equal to an integer multiple of the period. Meddis and colleagues (Meddis and Hewitt, 1991a, b; Meddis and O'Mard, 1997) have shown how a model based on the autocorrelation function (ACF) of simulated neural activity can provide a very good account of the pitch of a variety of resolved and unresolved complex tones.

The version of the Meddis and Hewitt model used here is the same as that described in the Introduction, with the exception that a simulation of the actual neural spikes is not generated. For each center frequency, the *spike probability* function is generated and autocorrelated. (The spikes themselves are derived from this function using a simple random number generator, and if this process is repeated a large number of times, the summed spikes approximate the form of the original probability function.) In their simulation, Meddis and Hewitt calculated ACFs for a number of auditory filter center frequencies and averaged these patterns to produce a ''summary'' ACF (SACF). The highest point in the SACF was used as an estimate of the perceived pitch (pitch being equal to the inverse of the delay). In the event of a number of equally high peaks, the peak with the shortest delay (i.e., highest frequency) was used. In a further development of the decision stage of the model, Meddis and O'Mard (1997) compared the SACF of the test stimulus with the best fitting template to obtain a pitch estimate. As in the Meddis and Hewitt version, however, they only considered delay values close to the first main peak in the SACF.

Following Licklider (1959), Meddis and Hewitt used an exponential temporal window, with a time constant of 2.5 ms, to sum the ACF over time. For the stimuli from the present experiments such a short integration time is not appropriate since the waveforms are not steady state. Specifically, a sample that includes the phase shift will produce a very different pitch estimate compared to one that does not. For some of their simulations, Meddis and Hewitt assumed that several SACF samples were averaged to produce the final SACF. Since autocorrelation is a linear process, combining short SACF samples across the stimulus is equivalent to calculating the SACF from ACFs of the whole stimulus (i.e., using a long-integration time), and this is the approach that will be adopted here. A rectangular window, with a du-

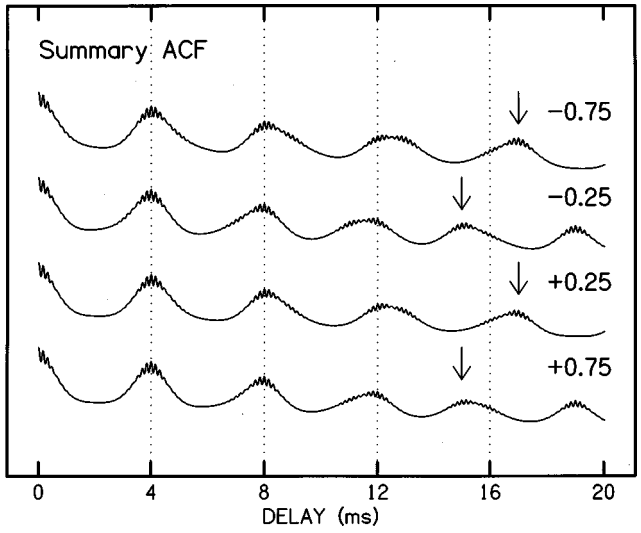


FIG. 6. Summary autocorrelation functions for the four waveforms from Fig. 2. ''Delay'' refers to the autocorrelation delay. Arrows indicate the location of the fourth peak in each case.

ration equal to that of the stimulus, was used to sample the spike probabilities.

For the present simulation, the SACF was computed using 11 filter channels with center frequencies, at 0.25 ERB spacing, between 5500 and 7500 Hz. Figure 6 shows the SACFs for the four IPI shift conditions from the experiment with a 0-ms gap. As expected, there are peaks in the function at integer multiples of the period (4 ms). It is interesting that the effect of the phase shift only becomes apparent after the second main peak in the SACF. This is because for delays of 4 and 8 ms, only the pitch pulses either side of the IPI shift contribute to the decorrelation, whereas for a delay of say 16 ms, most of the pitch pulses in the first burst contribute by virtue of being decorrelated with those from the second burst. If the SACF approach is correct, then the results of the experiment suggest that the auditory system may use higher-order peaks in the SACF than is often assumed in these models.

The predictions of the SACF model were generated by taking the inverse of the delay that produced the fourth peak in the SACF, and multiplying this value by four (the location of the fourth peak is indicated by the arrows in Fig. 6). The fourth peak was used to give the autocorrelation model a reasonable chance of predicting the pitch shifts for the 0-ms gap. The choice of which peak to choose is arbitrary, but the overall pattern of pitch matches for the higher-order peaks was similar to that for the fourth. Another method of estimating the pitch match is to find the $F0$ of the comparison stimulus whose SACF most closely matches that of the standard, as defined by the Euclidean distance between the two normalized SACFs (Meddis and Hewitt, 1991a; Meddis and O'Mard, 1997). Although a detailed set of predictions was not generated using this technique, simulations were conducted using SACFs with delays up to 20 ms, and with the $F0$ of the comparison stimulus varied in 1% increments. The pattern of results was similar to that based on the fourth peak, but, as expected, the actual pitch matches were much less ($\pm$2%–3%).

It should be noted that the complete ACF of a stimulus actually contains the same information as the magnitude spectrum. The different predictions of the SACF and spectral models are dependent on which aspects of the SACF are used to determine $F0$.

2. Mean rate

Carlyon (1997) distinguished between two cues that may be used to derive $F0$ for complex tones consisting of unresolved harmonics. The ''common interval'' cue refers to the fact that the time intervals between peaks in the waveform will be integer multiples of $1/F0$. It is this information that can be extracted using the ACF or by taking the Fourier spectrum. The ''mean rate'' cue refers to the fact that the number of waveform peaks in a given period of time is proportional to $F0$. The mean rate cue does not depend on regular interspike intervals. All that is necessary is that the auditory system has some representation of the number of waveform peaks in a given time.

For stable, steady state complex tones the common interval and mean rate cues produce identical pitch estimates. However, Carlyon manipulated the cues independently by randomly deleting pulses in the waveform to decrease the mean rate cue without affecting the common interval cue. He found that pitch judgements were affected by the mean rate cue, although the common interval cue could be used when mean rate was not informative.

The predictions of the mean rate model were generated by taking the inverse of the mean interval between pitch pulses (in seconds) across the whole stimulus. The approach allows for the variation in the time between the first and last pitch pulse produced by manipulating the IPI.

B. Comparing the models

To provide an evaluation of the three different pitch models with respect to the present data, the models were used to generate pitch matches for each of the conditions from the experiment. It was assumed that the analysis was performed over *the whole stimulus*, i.e., with a long integration time and *no resetting*. As described above, the main aim of the simulation was to determine whether any of the models could explain the results without requiring a resetting mechanism for the gap conditions. The predictions of the spectral model were based on the position of the maximum peak in the long-term Fourier spectra of the simulated neural activity patterns, generated using the procedure described in the Introduction.

Figure 7 shows the mean data from the experiment together with the pitch matches predicted by the three models. The results are plotted as a function of IPI in the standard interval. For the simulated results, the predicted pitch of the standard is plotted *relative* to the predicted pitch of the comparison. The majority of the predictions are surprisingly wide of the mark. For example, the spectral and SACF models predict large upward pitch shifts for the -0.25 and $+0.75$ conditions. These conditions correspond to an envelope phase shift of $\pi/2$. As described in the Introduction, a $\pi/2$ phase shift applied to pure tones produces an upward shift in

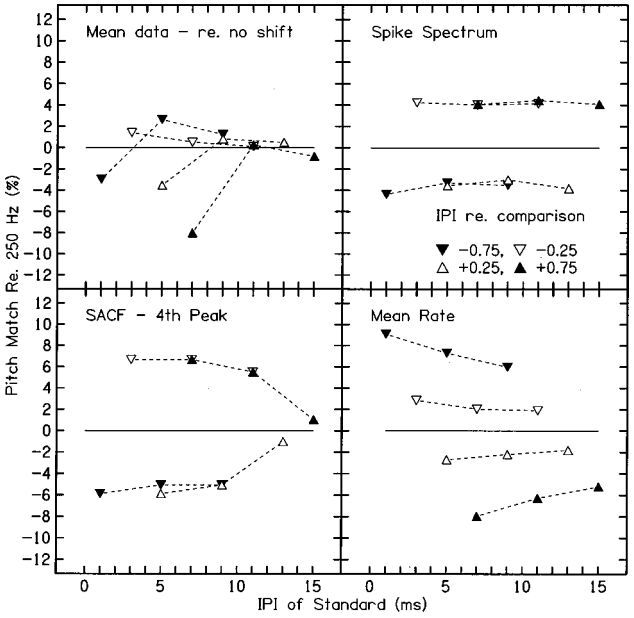


FIG. 7. The mean pitch-match data from the experiment together with the values predicted by the three long-integration models described in the text.

the spectral maximum. The SACF model does predict that the two longest IPI conditions for the 8-ms gap should produce little pitch shift. This can be understood in terms of the actual size of the IPIs (13 ms and 15 ms) for these conditions. These intervals are only slightly less than the autocorrelation delay of around 16 ms (fourth peak), so that effectively there is little interaction between the two tone bursts at the output of the autocorrelator. Moving on to the mean rate simulation, the model predicts correctly the large downward shift for the +0.75, 7-ms IPI, condition, but predicts an equally large upward shift for the −0.75, 1-ms IPI, condition. As with the other models, there was little effect of IPI on the predicted pitch shifts.

To summarize, not one of three models can account for all the data. Crucially, all the models predict substantial pitch shifts for the longer IPIs, whereas no such shifts were seen in the mean results. On this basis, the hypothesis of a fixed long-integration time for the pitch mechanism is not supported.

One simple modification of the mean rate model could allow it to account qualitatively for the data for the short IPIs. A 1-ms IPI is close to the refractory period of the auditory nerve (Kiang *et al.*, 1965) so that the second of the two closely spaced pitch pulses in this condition may illicit no response at all in many fibers. It is unlikely that *every* fiber would fire on the first pulse, and so only a proportion of the total population would be expected to be in refractory when the second pulse arrived. Nevertheless, it is possible that the overall neural response is strongly biased toward the first pulse in the pair. Effectively, this would mean that the −0.75, 1-ms IPI condition and the +0.25, 5-ms IPI condition (both 0-ms gap) in the experiment may be similar with respect to the rate of phase-locking peaks in the auditory nerve. The modified model can account for the similar downward shift for these two conditions.

One of the concerns in the experimental results was that conditions with the same standard IPI sometimes produced different pitch matches. Unlike the SACF and spectral models, the mean rate model does predict these differences, qualitatively at least. For example, the (long-integration) mean rate model illustrated in Fig. 7 predicts a higher pitch match for the 7-ms standard IPI with the −0.25 shift than for the 7-ms standard IPI with the +0.75 shift (relative to the comparison IPI). The higher pitch match occurs because the IPI in the comparison stimulus is 8 ms in the former case (gap=4 ms) and 4 ms in the latter (gap=0 ms). In other words, the mean rate model predicts that the comparison stimulus has a lower pitch in the former case so that the $F0$ of the comparison has to be increased for the comparison to match the pitch of the standard. Notice that this assumes that the pitch mechanism does *not* reset in response to the comparison stimulus.

Another possible explanation for the differences in pitch matches for the same standard IPI is that the auditory system always uses the same integration strategy for the two intervals on each trial. If the IPIs in both intervals are relatively long (e.g., 7-ms standard and 8-ms comparison) then resetting might occur in both cases, resulting in a pitch match of zero. If the IPI in one interval is short (e.g., 7-ms standard and 4-ms comparison) then long integration may be used in both cases, resulting in a negative pitch match. There is some evidence that the auditory system can select an integration time depending on the expectations of the temporal characteristics of the stimulus, as opposed to the selection of integration time being a simple automatic process (Wright and Dai, 1994).

C. Introducing resetting

Having determined that the three models fail badly at the longer IPIs when a long integration time is assumed, it can now be asked whether the models would do any better with a resetting mechanism. One way of introducing a resetting mechanism is to assume that any IPI greater than a certain value will trigger separate $F0$ estimates for the two tone bursts. In this case, all of the models predict that the reset interval has a pitch corresponding to an $F0$ of 250 Hz. As an illustration, Fig. 8 shows the predictions of the three models on the assumption that the integration window resets after an IPI of 8 ms. Although the models now account for the lack of a pitch shift at the long IPIs, the predictions for the short IPIs are still wide of the mark, even for the mean rate model.

It is clear that even if the models are allowed a resetting mechanism to deal with the long IPI conditions, *post hoc* modifications are needed to fit the present results for the short IPIs alone. The failure of the SACF model in particular is of theoretical importance since this model has proved very successful for accounting for the pitch of other complex-tone stimuli (Meddis and Hewitt, 1991b; Meddis and O'Mard, 1997).

D. Long integration and resetting

The results of the experiment show that shifting the IPI in the center of an unresolved complex tone can produce a

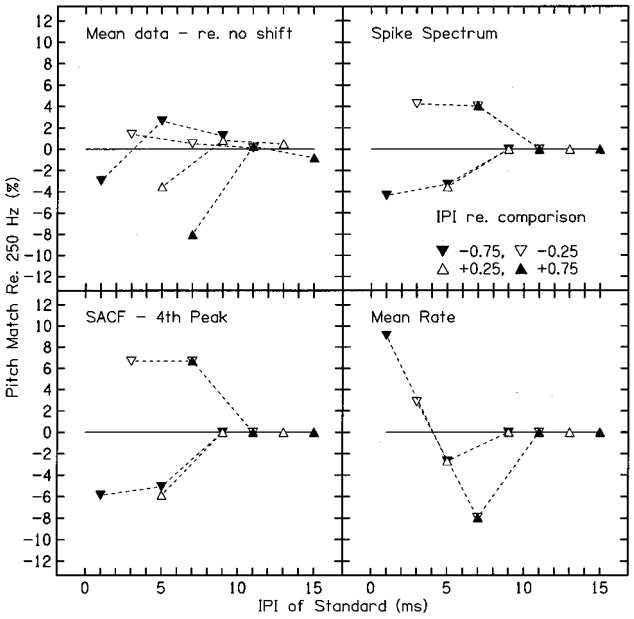


FIG. 8. As Fig. 7 except showing the predictions of the models using integration windows that are reset after an IPI of 8 ms or more (see text for details).

pitch shift. The fact that a pitch shift is seen at all implies that the auditory system is sensitive to the temporal relationship of the pitch pulses either side of the discontinuity, and this in turn implies that the pitch mechanism is integrating across the discontinuity. As noted in the Introduction, on this basis alone it cannot be inferred that the integration time extends over the whole stimulus, but this seems likely when the pitch match results are considered along with those showing a large improvement in $F0$ discrimination as duration is increased (White and Plack, 1998).

In contrast, for the conditions where the IPI shift was preceded by a gap, the pitch was almost identical to that of a gapped complex with no shift. The finding suggests that the auditory system is not sensitive to the temporal relationship of the pitch pulses either side of the gap and hence that the system is not integrating across the gap. It has to be said, however, that the evidence for resetting in the gap conditions is weaker than that for integration in the ungapped conditions. The absence of a pitch shift falls short of *proving* that the system is not integrating across the discontinuity. On the other hand, all the long-integration models evaluated in Sec. III B predict substantial pitch shifts for the conditions with long IPIs. Unless an alternative pitch model can be proposed, the evidence seems to be in favor of hypothesis (iii) described in the Introduction: The auditory system uses a long-integration time for determining the pitch of continuous unresolved complex tones, but the integration time is reset in response to temporal discontinuities.

IV. CONCLUSIONS

(i) Applying an increase or decrease to a single IPI in the center of an otherwise continuous 250-Hz unresolved complex tone can produce a substantial shift in the pitch of the tone.

(ii) Little or no pitch shift is seen for an IPI variation (relative to an integer number of periods) when the shifted IPI is greater than about 7 ms. It is not clear whether the size of the critical interval is an absolute value or whether it may be dependent upon $F0$.

(iii) The results support the hypothesis that the auditory system uses a long-integration time for continuous unresolved complex tones.

(iv) The lack of a pitch shift in the conditions with large IPIs suggests that a temporal discontinuity may cause the auditory system to reset the integration time.

ACKNOWLEDGMENTS

The research was supported by a grant from the Medical Research Council (UK). The first author was supported by a Royal Society University Research Fellowship. The authors are very grateful for the effort put in by Trevor Shackleton and two anonymous reviewers in improving the clarity of the manuscript and raising important theoretical issues. In addition, the authors would like to thank Peter Cariani for some illuminating discussions on the nature of temporal processing in the auditory pathway.

[1]The LUTEar/DSAM software is available on the internet: http://www.essex.ac.uk/psychology/hearinglab.

[2]A comparison stimulus with the same basic temporal parameters (stimulus gating) as the standard was used to encourage listeners to concentrate on the effects of IPI alone, and not on any other aspect of the stimuli. Because pitch is comparatively weak for these stimuli, it was found helpful for the listeners to simplify the task as much as possible. The test of the hypotheses depended on whether a shift in the IPI of the standard could be heard as a pitch shift relative to the comparison (where the IPI was an integer number of periods), not on the absolute pitch of either. The use of the same basic temporal parameters for the two intervals also simplifies the interpretation of the $F0$ discrimination data that were derived from the pitch match procedure.

Bregman, A. S., Ahad, P., Kim, J., and Melnerich, L. (**1994a**). ''Resetting the pitch-analysis system: 1. Effects of rise times of tones in noise backgrounds or of harmonics in a complex tone,'' Percept. Psychophys. **56**, 155–162.

Bregman, A. S., Ahad, P. A., and Kim, J. (**1994b**). ''Resetting the pitch-analysis system. 2. Role of sudden onsets and offsets in the perception of individual components in a cluster of overlapping tones,'' J. Acoust. Soc. Am. **96**, 2694–2703.

Carlyon, R. P. (**1997**). ''The effects of two temporal cues on pitch judgements,'' J. Acoust. Soc. Am. **102**, 1097–1105.

Goldstein, J. L., and Srulovicz, P. (**1977**). ''Auditory-nerve spike intervals as an adequate basis for aural frequency measurement,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London).

Hafter, E. R., and Buell, T. N. (**1990**). ''Restarting the adapted binaural system,'' J. Acoust. Soc. Am. **88**, 806–812.

Jesteadt, W. (**1980**). ''An adaptive procedure for subjective judgements,'' Percept. Psychophys. **28**, 85–88.

Kiang, N. Y.-S., Wantanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). ''Discharge patterns of single fibers in the cat's auditory nerve,'' Res. Monogr. No. 35 (MIT, Cambridge, MA).

Levitt, H. (**1971**). ''Transformed up-down methods in psychophysics,'' J. Acoust. Soc. Am. **49**, 467–477.

Licklider, J. C. R. (**1951**). ''A duplex theory of pitch perception,'' Experientia **7**, 128–133.

Licklider, J. C. R. (**1959**). ''Three auditory theories,'' in *Psychology: A Study of Science*, edited by S. Koch (McGraw-Hill, New York).

Meddis, R. (**1986**). ''Simulation of mechanical to neural transduction in the auditory receptor,'' J. Acoust. Soc. Am. **79**, 702–711.

Meddis, R. (**1988**). ''Simulation of auditory-nerve transduction: Further studies,'' J. Acoust. Soc. Am. **83**, 1056–1063.
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.
Meddis, R., Hewitt, M. J., and Shackleton, T. M. (**1990**). ''Implementation details of a computational model of the inner hair-cell/auditory-nerve synapse,'' J. Acoust. Soc. Am. **87**, 1813–1818.
Meddis, R., and O'Mard, L. (**1997**). ''A unitary model of pitch perception,'' J. Acoust. Soc. Am. **102**, 1811–1820.
Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, New York).
Nabelek, I. V. (**1996**). ''Pitch of a sequence of two short tones and the critical pause duration,'' Acustica **82**, 531–539.
Nabalek, I. V., Nabalek, A. K., and Hirsh, I. J. (**1973**). ''Pitch of sound bursts with continuous or discontinuous changes in frequency,'' J. Acoust. Soc. Am. **53**, 1305–1312.
Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1988**). Spiral Vos Final Report, Part A: The Auditory Filterbank. Applied Psychology Unit, Cambridge, England.
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.
Relkin, E. M., and Pelli, D. G. (**1987**). ''Probe tone thresholds in the auditory nerve measured by two-interval forced choice procedures,'' J. Acoust. Soc. Am. **82**, 1679–1691.
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 communication of frequency spectrum,'' J. Acoust. Soc. Am. **73**, 1266–1276.
Teich, M. C., and Khanna, S. M. (**1985**). ''Pulse-number distribution for the neural spike train in the cat's auditory nerve,'' J. Acoust. Soc. Am. **77**, 1110–1128.
van Noorden, L. (**1982**). ''Two channel pitch perception,'' in *Music, Mind, and Brain*, edited by M. Clynes (Plenum, New York).
Viemeister, N. F., and Plack, C. J. (**1993**). ''Time analysis,'' in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer-Verlag, New York).
Viemeister, N. F., and Wakefield, G. H. (**1991**). ''Temporal integration and multiple looks,'' J. Acoust. Soc. Am. **90**, 858–865.
White, L. J., and Plack, C. J. (**1998**). ''Temporal processing of the pitch of complex tones,'' J. Acoust. Soc. Am. **103**, 2051–2063.
Wright, B. A., and Dai, H. P. (**1994**). ''Detection of unexpected tones with short and long durations,'' J. Acoust. Soc. Am. **95**, 931–938.
Young, E. D., and Barta, P. E. (**1986**). ''Rate responses of auditory-nerve fibers to tones in noise near masked threshold,'' J. Acoust. Soc. Am. **79**, 426–442.

Effect of masker harmonicity on informational masking

Eunmi L. Oh
Waisman Center, University of Wisconsin, Madison, Wisconsin 53705

Robert A. Lutfi
Department of Communicative Disorders, University of Wisconsin, Madison, Wisconsin 53705

(Received 5 November 1998; revised 26 October 1999; accepted 14 April 2000)

Detection thresholds for a tone in an unfamiliar tonal pattern can be greatly elevated under conditions of masker uncertainty [Neff and Green, Percept. Psychophys. **41**, 409–415 (1987); Oh and Lutfi, J. Acoust. Soc. Am. **101**, 3148 (1997)]. The present experiment was undertaken to determine whether harmonicity of masker tones can reduce the detrimental effect of masker uncertainty. Inharmonic maskers were comprised of $m=2-49$ frequency components selected at random on each presentation within 100–10 000 Hz, excluding frequencies between 920–1080. Harmonic maskers were comprised of frequency components selected at random within this same range, but constrained to have a fundamental frequency of 200 Hz. For inharmonic maskers the signal was a 1000-Hz tone. For harmonic-maskers the signal was a tone whose frequency was either harmonically (1000 Hz) or inharmonically (1047 Hz) related to the masker. In all conditions the amount of masking was greatest for $m=20-40$ components. At this point, harmonic maskers with harmonic signal produced an average of 9–12 dB less masking than inharmonic maskers. Harmonic maskers with inharmonic signal produced an average of 16–20 dB less masking. © *2000 Acoustical Society of America.* [S0001-4966(00)00608-1]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Fe [JWH]

INTRODUCTION

Masked thresholds for a tonal signal are often greatly elevated when maskers are made uncertain by varying the frequencies of masker components at random from trial to trial (Neff and Callaghan, 1988; Neff and Green, 1987). Masking produced by stimulus uncertainty is referred to as *informational masking* (Pollack, 1975), and is distinguished from *energy-based masking* in minimal uncertainty conditions that are caused by masker energy falling in close temporal–spectral proximity to the signal (Fletcher, 1940; Patterson *et al.*, 1982). Informational masking can amount to as much as 30 dB for some listeners when maskers are comprised of 10–20 spectral components (Neff and Dethlefs, 1995; Oh and Lutfi, 1998). The large elevations in threshold are often attributed to the lack of any predictable structure in the masker that would allow listeners to segregate the single spectral component belonging to the signal from the collection of components belonging to the masker. In this study we investigate the effect of masker harmonicity as one type of structure that might be expected to aid detection under conditions of high masker uncertainty. Our conjecture is that a masker comprised of harmonics of a single fundamental frequency will produce less masking of a signal that is inharmonically related to the masker than one that is harmonically related. This conjecture is based on the expectation that the inharmonicity of the signal will provide a cue for segregating the signal from the masker. We also test a specific prediction regarding how harmonicity of masker components can be expected to reduce informational masking by reducing uncertainty associated with the possible combinations of masker components that are presented from trial to trial.

I. METHOD

A. Stimuli

Maskers were constructed from 50 Gaussian noise samples, bandpass filtered from 0.1 to 10 kHz. The magnitude and phase spectra of each Gaussian noise sample were analyzed into individual spectral components, with 11-Hz spacing between components. The amplitude distribution of these samples is Rayleigh, and the phase distribution is rectangular. On each trial, two of the noise samples for a two-interval, forced-choice procedure were selected at random. For each sample, a fixed number of frequency components was drawn randomly from a rectangular distribution ranging from 0.1 to 10 kHz on a linear scale. The frequencies were constrained to be harmonics of a 200 Hz fundamental from 200 to 10 000 Hz not including 1000 Hz. The phases and amplitudes of the selected frequency components were then used to synthesize the multitone maskers played on each trial. In different blocks of trials, the number of frequency components m was varied from 2 to 49. The rms values of the maskers were adjusted to be the same regardless of the number of masker components. Average total power of maskers was kept constant at 60 dB SPL.

The two major conditions of the experiment are illustrated in panels (a) and (b) of Fig. 1. In both conditions the maskers are the same; the only difference is that the signal is a 1000-Hz tone that is harmonically related to the masker tones, or a 1047-Hz tone that is inharmonically related to the masker tones. The 200-Hz fundamental frequency was chosen in this case so that the 1000-Hz signal would be in the dominance region for pitch (Ritsma, 1967; Moore *et al.*, 1985). The frequency of the inharmonic signal was chosen to produce a clear percept of inharmonicity without introducing

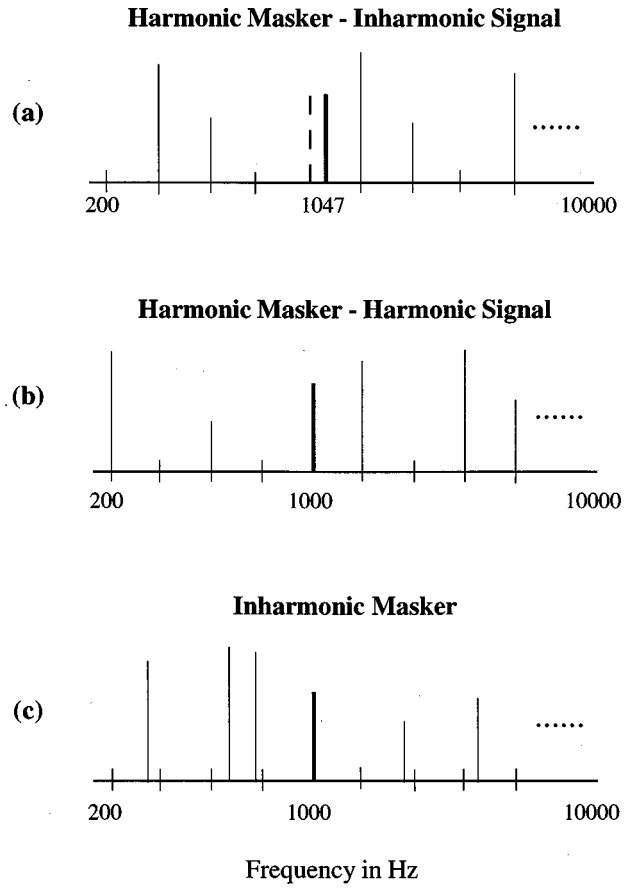


FIG. 1. Schematic illustration of the stimuli used in the three experimental conditions of this study. (a) The signal is a 1047-Hz tone inharmonically related to randomly selected masker frequencies with 200-Hz fundamental. (b) The masker is the same as in (a), the signal is a 1000-Hz tone harmonically related to the masker. (c) the masker is an inharmonic complex with frequencies chosen at random and without constraint.

significant differences in the amount of energy-based masking for the two signals. An inharmonic masker condition was also included, panel (c) of Fig. 1, to assess the effect of imposing harmonic structure on maskers independent of the effect of the relation of the signal to the masker. The components of the inharmonic maskers were simply chosen at random on each presentation without any constraints. The data for this condition are from a previous study by the authors involving identical procedures, identical methods for synthesizing stimuli, and the same group of listeners (see Oh and Lutfi, 1998). Differences in energy-based masking for the three conditions is estimated to be 2 dB or less based on roex-filter predictions with parameters representing the average normal-hearing adult (see Patterson *et al.*, 1982).[1]

In all conditions signals and maskers were gated on and off together with 10-ms, cosine-squared onset/offset ramps for a total duration of 370 ms. The signal and maskers were computer generated and played over a 16-bit, digital-to-analog converter (Tucker Davis Technologies DD1) at a sampling rate of 44.1 kHz. All stimuli were presented monaurally through Sennheiser model HD-520 headphones. The Sennheiser headphones were calibrated using a binaural loudness balancing procedure using a TDH-49 earphone with known transfer function.

B. Procedure

An adaptive, two-interval, forced-choice procedure was used to measure signal threshold in quiet and in the presence of the multitone maskers. The decision rule estimated the 70.7%-point on the psychometric function (Levitt, 1971). The level of signal was varied by a programmable attenuator (Tucker Davis Technologies PA4) with an initial step size of 4 dB reduced to 2 dB after the third reversal. A trial block consisted of a total of 12 reversals, and threshold for that block was designated as the average of the last 8 reversal levels. Within an experimental condition, the number of masker components was fixed. Threshold for each condition was the average of five thresholds obtained from five blocks of trials. Total masking (dB difference between masked threshold and quiet threshold) was reported as a function of the number of masker components. Blocks of trials were run consecutively to facilitate practice within each condition. However, the order in which experimental conditions were run was randomized. Ten blocks of trials were collected for an experimental session lasting approximately 1 h with breaks.

Individual listeners were seated in a double-walled, IAC sound-attenuation chamber. At the beginning of each block, the signal was presented several times in isolation to familiarize the listener with the tone to be detected. On each trial, two multitone complexes were presented monaurally through the headphones. The interval between two stimuli was 500 ms. Listeners were asked to indicate by key press which one of the two intervals contained the signal on each trial. The listener's response was followed by visual feedback on a CRT indicating whether the response was correct or incorrect. Listeners were instructed to make use of the feedback to improve their performance.

C. Listeners

Eight normal-hearing listeners, three males and five females, were paid at an hourly rate for their participation. The ages of the listeners ranged from 20 to 33 years. The listeners' pure-tone air-conduction thresholds were less than 15 dB HL (ANSI, 1989) between 150 and 8000 Hz in both ears. All listeners had participated in the earlier study by Oh and Lutfi (1998).

II. RESULTS

Figure 2 shows total masking for individual listeners when the signal was harmonically (open triangles) and inharmonically (open circles) related to the masker. Filled symbols give the data from the nonharmonic masker condition of Oh and Lutfi (1998). Each circle represents the total masking averaged over five blocks of trials for a fixed number of masker components. The effect of masker and signal harmonicity varies from listener to listener. For the one listener showing the least-amount of masking overall (LOE), it matters little whether the masker is harmonic or inharmonic, or whether the signal is harmonically or inharmonically related to the masker. For the six remaining listeners, the harmonic maskers produces significantly less masking than the inharmonic masker, as much as 25 dB less for listener LCM.

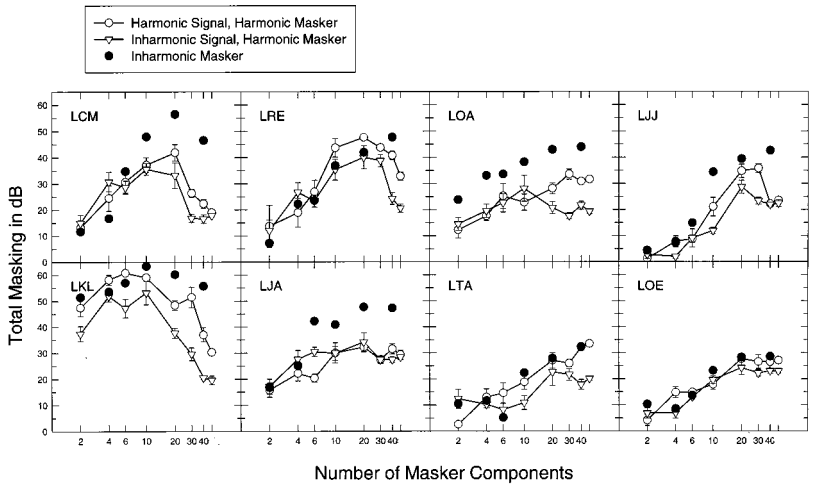


FIG. 2. Total masking for individual listeners for the harmonic signal (open circles), inharmonic signal (open triangles), and the inharmonic masker (filled symbols) conditions. Each circle represents the total masking averaged over 5 blocks of trials with standard errors.

Among these six listeners, five also show significantly less masking for the inharmonic signal than the harmonic signal, as much as 20 dB less masking for listener LKL, and an average a 5–10 dB less masking for the other listeners. Generally, the difference in masked thresholds across conditions is greatest where the amount of informational masking is also the greatest, for $m>6$.

Figure 3 shows the predictions of a model that has been successfully applied to describe the results of previous studies of informational masking. The component-relative entropy (CoRE) model assumes that listeners integrate information in a maximum-likelihood statistic within an ''attentional band'' far exceeding the bandwidth of the signal. Details regarding these predictions can be found in Oh and Lutfi (1998) and in Lutfi (1993). The important point for the present application is that the components of the maskers and signal are assumed to be statistically independent. Hence, the predictions allows us to evaluate whether the signal's specific *relation* to masker components (harmonic or inharmonic) has an effect above and beyond that due to the reduction in uncertainty associated with imposing harmonic structure on maskers. The symbols give the total masking averaged over eight listeners for the three masker conditions, the dashed curve gives predictions of the model for the inharmonic masker, and the continuous curve gives predictions for the two harmonic masker conditions. The agreement of the model predictions with the data for the inharmonic masker and harmonic signal condition is good, but the predictions significantly overestimate the amount of masking for the inharmonic signal condition. This outcome supports the conclusion that only the inharmonic relation of the signal to the masker has an effect beyond that resulting from the reduction in uncertainty associated with constraining maskers to be harmonic.

III. DISCUSSION

The present results are noteworthy with regard to past studies of the effects of harmonicity on detection which have shown inconsistent results. Studies on pitch and sound quality, for example, tend to support the idea that an inharmonic tone would be segregated from a harmonic tone complex. Generally, when a low harmonic in a complex comprised of the first ten harmonics is mistuned by more than 3% of the harmonic frequency, it is heard as a separate tone, not belonging to the complex (Moore *et al.*, 1986). Martens (1984) also reports that an inharmonic tone becomes easier to hear out from a complex comprised of the first ten harmonics of a 200-Hz fundamental frequency. According to Martens, the level of a 1000-Hz tone should be 6 dB higher than the other harmonics to be audible as a separate spectral pitch. For a 1040-Hz tone that is not a multiple of 200 Hz, the audibility threshold for pitch can be reduced as much as 9 dB. Darwin and Gardner (1986) examined the effect of harmonicity on vowel perception. A harmonic in the first formant region of vowels at 500 Hz was mistuned from the other harmonics of a 125-Hz fundamental. For a large amount of mistuning (about 8%), the inharmonic component did not affect a listener's judgment on the category of the vowel. Bregman and

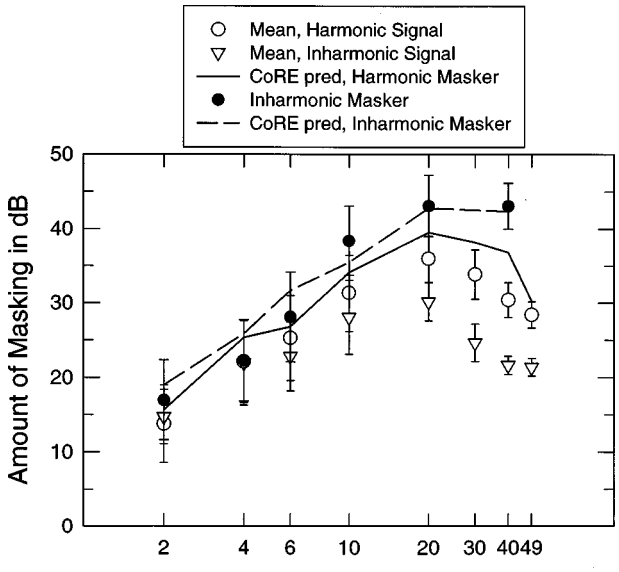


FIG. 3. Predictions of the CoRE model for total masking in nonharmonic (dashed curve) and harmonic (continuous curve) conditions. Values of the two free parameters of the model are those previously used in the study by Oh and Lutfi (1998). Symbols give total masking averaged over the 8 listeners for harmonic signal (open circles), inharmonic signal (open triangles), and inharmonic masker (filled symbols) conditions.

Doehring (1984) demonstrated that harmonicity could play a role in auditory streaming. They alternated a pure tone with a three-tone complex having the same frequency as the middle tone of the triad. When the middle tone was inharmonic in relation to the other two tones, it was more easily grouped with the pure tone as a sequential stream.

Other studies employing forced-choice detection and discrimination of signals generally fail to reveal a significant effect of harmonicity. In the study of Stellmack *et al.* (1997), for example, listeners were asked to discriminate the level of a 1000-Hz tone embedded in a masker that was comprised of two components whose amplitudes varied from trial to trial. The masker components were either harmonically related or inharmonically related to the signal. There was little difference in percent-correct discrimination scores for the harmonic and inharmonic maskers. The effect of harmonicity on comodulation masking release (CMR) was investigated by Grose and Hall (1993). A tonal complex whose frequencies were harmonically or inharmonically related to the signal was added to one ear, and a tonal complex whose frequencies were harmonically or inharmonically related to the comodulated noise bands was added to the other ear. Grose and Hall hypothesized that CMR would be reduced more by harmonic complexes than inharmonic ones because harmonic complexes might prevent comodulated noise bands from forming a separate auditory object. However, the same amount of disruption of CMR was obtained for both the harmonic and inharmonic complexes.

The present results suggest a possible reason for the failure to find a significant effect of harmonicity in these past detection studies. Unlike the present study, past studies were conducted under conditions of minimal uncertainty, that is, conditions where the masker did not vary considerably from one trial to the next. Under such conditions one would expect little if any informational masking that would elevate thresholds above that of energy-based masking. Since energy-based masking is often taken to represent fixed limitations in the resolving power of the peripheral ear, it may, in past studies, have established a lower bound on threshold beyond which any significant improvement due to harmonicity was not possible. Our data seem to support this explanation inasmuch as the reduction in masking for the inharmonic signal tended to be the greatest at the point where informational masking was also the greatest.

ACKNOWLEDGMENTS

This research was supported by a grant from the NIDCD (Grant No. R01 CD01262-09). We wish to thank two anonymous reviewers for comments on an early version of this manuscript.

[1]Empirical estimates of energy-based masking in these conditions is impractical as it would require the measurement of masked threshold for each of the masker samples in fixed masker conditions. The problem and the justification for using theoretical estimates based on the roex filter is discussed in Oh and Lutfi (1998).

ANSI (**1989**). ANSI S3.6-1989, ''Specifications for audiometers'' (American National Standards Institute, New York).

Bregman, A. S., and Doehring, P. (**1984**). ''Fusion of simultaneous tonal glides: The role of parallelness and simple frequency relations,'' Percept. Psychophys. **36**(3), 251–256.

Darwin, C. J., and Gardner, R. B. (**1986**). ''Mistuning a harmonic of vowel: Grouping and phase effects on vowel quality,'' J. Acoust. Soc. Am. **79**, 838–845.

Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–56.

Grose, J. H., and Hall, J. W. (**1993**). ''Comodulation masking release: Is comodulation sufficient?'' J. Acoust. Soc. Am. **93**, 2896–2902.

Levitt, H. (**1971**). ''Transformed up-down methods in psychacostics,'' J. Acoust. Soc. Am. **49**, 467–477.

Lutfi, R. A. (**1993**). ''A model of auditory pattern analysis based on component-relative-entropy,'' J. Acoust. Soc. Am. **94**, 748–758.

Martens, J. P. (**1984**). ''Comment on Algorithm for extraction of pitch and pitch salience from complex tonal signals,'' J. Acoust. Soc. Am. **75**, 626–628.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1985**). ''Relative dominance of individual partials in determining the pitch of complex tones,'' J. Acoust. Soc. Am. **77**, 1853–1860.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1986**). ''Thresholds for hearing mistuned partials as separate tones in harmonic complexes,'' J. Acoust. Soc. Am. **80**, 479–483.

Neff, D. L., and Dethlefs, T. M. (**1995**). ''Individual differences in simultaneous masking with random-frequency, multicomponent maskers,'' J. Acoust. Soc. Am. **98**, 125–134.

Neff, D. L., and Callaghan, B. P. (**1988**). ''Effective properties of multicomponent simultaneous maskers under conditions of uncertainty,'' J. Acoust. Soc. Am. **83**, 1833–1838.

Neff, D. L., and Green, D. M. (**1987**). ''Masking produced by spectral uncertainty with multicomponent maskers,'' Percept. Psychophys. **41**, 409–415.

Oh, E. L., and Lutfi, R. A. (**1997**). ''Effect of number and frequency-spacing of masker components on multitone masking,'' J. Acoust. Soc. Am. **101**, 3148.

Oh, E. L., and Lutfi, R. A. (**1998**). ''Nonmonotonicity of informational masking,'' J. Acoust. Soc. Am. **104**, 3489–3499.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). ''The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold,'' J. Acoust. Soc. Am. **72**, 1788–1803.

Pollack, I. (**1975**). ''Auditory informational masking,'' J. Acoust. Soc. Am. **57**, S5.

Ritsma, R. J. (**1967**). ''Frequencies dominant in the perception of the pitch of complex sounds,'' J. Acoust. Soc. Am. **42**, 191–198.

Stellmack, M. A., Willihnganz, M. S., Wightman, F. L., and Lutfi, R. A. (**1997**). ''Spectral weights in level discrimination by preschool children: Analytic listening conditions,'' J. Acoust. Soc. Am. **101**, 2811–2821.

Neighboring spectral content influences vowel identification

Lori L. Holt[a)]
Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213

Andrew J. Lotto
Department of Psychology & Parmly Hearing Institute, Loyola University Chicago, 6525 North Sheridan Road, Chicago, Illinois 60626

Keith R. Kluender
Department of Psychology, University of Wisconsin-Madison, 1202 West Johnson Street, Madison, Wisconsin 53706

(Received 9 December 1999; accepted for publication 27 April 2000)

Four experiments explored the relative contributions of spectral content and phonetic labeling in effects of context on vowel perception. Two 10-step series of CVC syllables ([bVb] and [dVd]) varying acoustically in *F*2 midpoint frequency and varying perceptually in vowel height from [ʌ] to [ɛ] were synthesized. In a forced-choice identification task, listeners more often labeled vowels as [ʌ] in [dVd] context than in [bVb] context. To examine whether spectral content predicts this effect, nonspeech–speech hybrid series were created by appending 70-ms sine-wave glides following the trajectory of CVC *F*2's to 60-ms members of a steady-state vowel series varying in *F*2 frequency. In addition, a second hybrid series was created by appending constant-frequency sine-wave tones equivalent in frequency to CVC *F*2 onset/offset frequencies. Vowels flanked by frequency-modulated glides or steady-state tones modeling [dVd] were more often labeled as [ʌ] than were the same vowels surrounded by nonspeech modeling [bVb]. These results suggest that spectral content is important in understanding vowel context effects. A final experiment tested whether spectral content can modulate vowel perception when phonetic labeling remains intact. Voiceless consonants, with lower-amplitude more-diffuse spectra, were found to exert less of an influence on vowel perception than do their voiced counterparts. The data are discussed in terms of a general perceptual account of context effects in speech perception. © *2000 Acoustical Society of America.* [S0001-4966(00)01908-1]

PACS numbers: 43.66.Lj, 43.71.An, 43.71.Es [CWT]

INTRODUCTION

One of the earliest and most influential demonstrations of effects of context in speech perception focused on the influence adjacent consonants exert on vowel identification in consonant–vowel–consonant (CVC) syllables. For this study, Lindblom and Studdert-Kennedy (1967) synthesized three vowel series. The first consisted of steady-state series varying perceptually from [ʊ]–[ɪ] via manipulation of *F*2 frequency. The second and third series were comprised of these same vowels embedded in time-varying [wVw] and [jVj] contexts. Critically, vowel formant frequencies at the midpoint of the synthetic stimuli were identical across series. Thus if [wVw] and [jVj] contexts affect vowel perception, then listeners' vowel identification functions for these series should be shifted relative to those of the isolated [ʊ]–[ɪ] series.

Lindblom and Studdert-Kennedy's (1967) observations supported this prediction. Listeners more often identified vowels in [wVw] context as [ɪ]. In [jVj] context, listeners labeled the same vowels more often as [ʊ]. These significant findings have since been supported by Nearey (1989), who extended this perceptual evidence to stop-consonant CVC syllables ([bVb] and [dVd]) with vowel sounds ranging from [ɔ]–[ʌ] and [ʌ]–[ɛ].

By what means might adjacent consonants influence vowel perception? A later study, building upon the results of Lindblom and Studdert-Kennedy (1967), suggests one possibility. Using stimuli very similar to those employed by Lindblom and Studdert-Kennedy, Williams (1986) examined listeners' identification responses across three conditions. Subjects in the first condition identified vowels in series varying perceptually from [wʊw]–[wɪw] and [ʊ]–[ɪ] in an AXB task where they judged whether the random stimulus (*X*) was more similar to the first or third member of the triad (*A* or *B*). These comparison stimuli were always series endpoints ([wʊw]/[wɪw] or [ʊ]/[ɪ]). This condition replicated, in part, the task used by Lindblom and Studdert-Kennedy.[1] In the remaining conditions, Williams capitalized on the observation that nonspeech sine-wave stimuli following the frequency trajectories of speech formants may be perceived as either speech or nonspeech depending upon instructions given to listeners (e.g., Bailey *et al.*, 1977; Remez *et al.*, 1981; Best *et al.*, 1981). Both of the remaining conditions used sine-wave tone complexes varying in frequency trajectory to mimic *F*1, *F*2, and *F*3 formant paths of stimuli from the [wʊw]–[wɪw] and the [ʊ]–[ɪ] series. In the second condition, subjects were informed that these stimuli were speech

a)Electronic mail: lholt@andrew.cmu.edu

and were instructed to identify them as [ʊ] or [ɪ]. In the third condition, the same stimuli were described as nonspeech (i.e., as a ''chord'') and a separate set of subjects identified stimuli as having ''low'' or ''high'' pitch.

Replicating Lindblom and Studdert-Kennedy (1967), Williams (1986) observed a shift in vowel identification for speech stimuli; subjects more often labeled vowels as [ɪ] in the context of [wVw]. Interestingly, Williams also found an analogous shift in identification of sine-wave complexes that listeners labeled as speech. Mirroring the results for full-spectrum speech stimuli, time-varying sine-wave triads modeling [wʊw]–[wɪw] formant trajectories were more often labeled [ɪ] than steady-state complexes patterned after [ʊ]–[ɪ]. Importantly, though, there was no evidence of an identification shift among subjects who made pitch judgments of the same sine-wave stimuli. This pattern of results, with identification shifts for stimuli identified as speech, but not for acoustically identical stimuli identified on pitch, suggests that phonetic labeling may be critical in understanding effects of context.

There are several reasons to interpret the conclusions arising from these results with reservation, however. For one, it is important to acknowledge that Williams' (1986) conclusions rely on a null result—namely, the failure to demonstrate a context-dependent shift in pitch identification of sine-wave triads. At the very least, conclusions drawn from null hypotheses need be interpreted with care. However, further evidence also advises caution. In an attempt to replicate Williams' results, Mullennix *et al.* (1988) adopted a stimulus set and an identification task identical to those used by Williams. Whereas Williams reported no context-dependent shift in sine-wave pitch identification, Mullennix *et al.* found that listeners' identification of [wʊw]–[wɪw] tone analogues was shifted toward lower frequencies than was their identification of tone complexes modeled after [ʊ]–[ɪ]. Subjects were significantly more likely to label the pitch of the triad as ''high'' for stimuli mimicking [wʊw]–[wɪw] than for those imitating [ʊ]–[ɪ]. Mullennix *et al.* were therefore able to induce a shift in pitch labeling dependent upon context. Moreover, the observed shift for the sine-wave complexes was in the same direction as that elicited by the speech stimuli after which the tone triads were modeled.

Considering the stimuli and methodology of this experiment were replicas of those used by Williams (1986), these results make it difficult to assess whether phonetic labeling indeed plays a role in the influence of consonant context on vowel perception. Furthermore, the observations of Mullennix *et al.* suggest an alternative hypothesis. Listeners' vowel identification functions shifted regardless of whether they labeled the sine-wave triads phonetically or not. Considering that the sine waves modeled some of the putatively important spectral energy in the speech formants, it may be that the spectral content of adjacent stimuli is the key variable in effects of context on vowel perception.

The present experiments investigate the relative contributions of phonetic labeling and spectral content to context effects in vowel perception. If phonetic labeling is fundamental, then nonspeech contexts that mimic spectral characteristics of consonant context (and are not perceived as speech) should not produce shifts in vowel identification. Alternatively, if spectral content is important, nonspeech stimuli should influence vowel identification in a manner that mimics the influence of the consonants they model. Experiments 1–3 examine these competing hypotheses. Experiment 4 addresses the issue in another way. If spectral content is the important variable in understanding effects of context on vowel perception, then there may be cases in which phonetically labeled consonants' influence on vowel identification can be modulated by manipulations of spectral content. Experiment 4 examines this using voicing as a manipulation of spectral content.

I. EXPERIMENT 1

Pilot study replications of Lindblom and Studdert-Kennedy (1967) and Nearey (1989) demonstrated that native American–English subjects find the stop-consonant series ([bVb]) and [dVd] employed by Nearey to be more compelling instances of CVC syllables than the semi-vowel series ([wVw] and [jVj]) of Lindblom and Studdert-Kennedy.[2] Thus experiment 1 utilizes stop-consonant contexts similar to those of Nearey. However, not all aspects of Nearey's stimuli were modeled in the present studies. In the same pilot experiments, American–English speaking subjects were not adept at labeling the vowel [o] used by Nearey. Even for the best exemplars, subjects labeled stimuli as [o] only approximately 85% of the time. This is likely due to the fact that nondiphthongized [o] (familiar in Western Canadian dialects) is not common to Midwest–American speech. Consequently, experiment 1 introduces vowel stimuli spanning a perceptual range from [ʌ] to [ɛ] via manipulation of $F2$ frequency in a test of effects of [bVb] and [dVd] contexts on vowel identification.

A. Method

1. Subjects

Twenty-eight native English-speaking undergraduates at the University of Wisconsin–Madison participated in return for course credit in Introductory Psychology. All subjects reported normal hearing.

2. Stimuli

Two 10-step series of 200-ms stimuli with constant fundamental frequencies of 120 Hz were synthesized with 12-bit resolution using the cascade branch of the speech synthesizer developed by Klatt (1980). Series varied perceptually from [bʌb]–[bɛb] and [dʌd]–[dɛd]. Synthesis was implemented on a Pentium microcomputer at a sampling rate of 10 kHz. Figure 1 provides schematic spectrograms of endpoint stimuli for each series.

As is illustrated in the figure, formant frequencies are symmetric around the stimulus midpoint (100 ms) and formants $F1-F3$ have a curvilinear trajectory. The trajectory from stimulus onset to midpoint was specified by an equation introduced by Lindblom and Studdert-Kennedy (1967) and used by Nearey (1989),

$$F(t)=F_v+(F_i-F_v)[(t-t_v)^p/t_v^p], \quad (1)$$

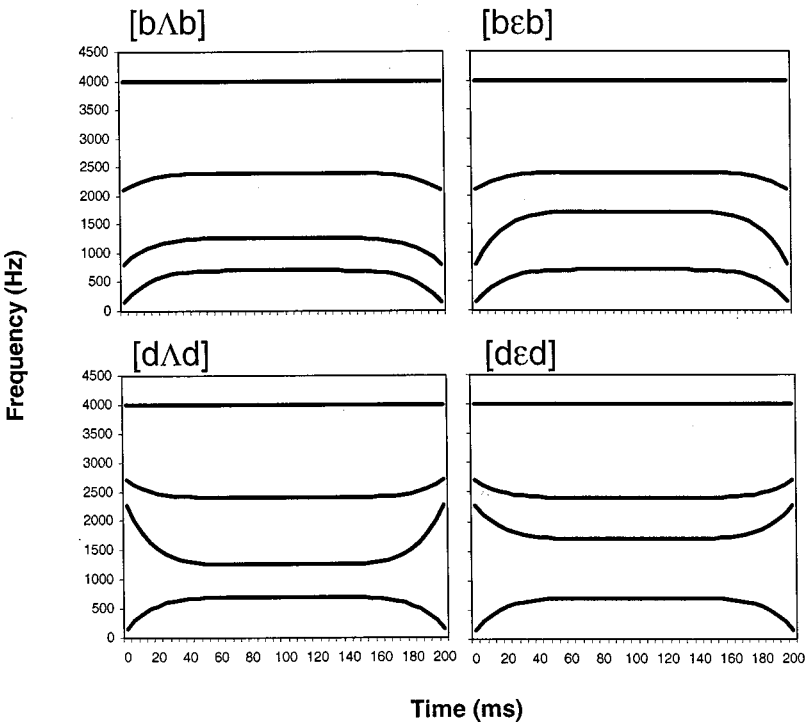


FIG. 1. CVC Series Endpoints. Representative pseudo-spectrograms of experiment 1 stimulus-series endpoints. The top row represents [bVb] stimuli. The bottom row shows [dVd] endpoints. The left column corresponds to low-$F2$ ([Λ]) stimulus endpoints and the right column depicts high-$F2$ ([ɛ]) endpoints.

where $F(t)$ is the frequency of the formant at time t, t_v is the midpoint of the stimulus, F_v is the formant frequency at stimulus midpoint (time t_v), F_i is the formant frequency at stimulus onset, and p is an exponent determining the shape of the formant-frequency trajectory. Here $p=6.0$.[3] The second half of each stimulus was a mirror image of the first.

Onset, offset, and midpoint nominal frequency values for each formant of the endpoint stimuli are displayed in Table I. Within each series, onset and offset frequencies were equal across series members; only $F2$ midpoint frequency varied.

Stimulus presentation and response collection were under the control of an 80486-25 microcomputer. Following D/A conversion (Ariel DSP-16), stimuli were low-pass filtered (4.8 cutoff frequency, Frequency Devices, #677), rms matched in amplitude, amplified (Stewart HDA4), and presented to subjects via headphones (Beyer DT-100) at a level of 70 dB SPL(A).

B. Procedure

Listeners participated in a 2AFC identification task. One to three subjects were tested concurrently in individual sound-attenuated chambers during a single experimental session. In segments mixed across series ([bVb] and [dVd]), participants first identified the vowel as [ʌ] or [ɛ] by pressing either of two buttons on a handheld electronic response box with buttons labeled ''PUTT'' and ''PET'' for reference. Next, listeners identified the consonant as either ''B'' or ''D.'' Although vowel identification was the focus of the experiment, consonant identification responses were elicited to assure that listeners perceived synthetic renditions of CVCs as reasonably good examples of the intended syllables. This also ensured that listeners explicitly phonetically labeled the consonant contexts as well as the vowels.

Participants responded to each stimulus 20 times; order of stimulus presentation was randomized. In all, subjects identified 400 stimuli (10 stim/series×10 repetitions/stim×2 series×2 blocks). The entire session lasted approximately 45 min.[4]

C. Results and discussion

Figure 2 displays mean vowel identification curves for experiment 1.[5] Data were submitted to a 2×10 (consonant context×stimulus-step) within-subjects analysis of variance (ANOVA). As predicted from earlier reports (Lindblom and Studdert-Kennedy, 1967; Nearey, 1989), there was a significant effect of consonant context on vowel identification ($F_{(1,27)}=43.09$, $p<0.0001$). Subjects differentially labeled vowels dependent on consonant context; vowels in [dVd] context were identified as [ʌ] more often than vowels with

TABLE I. Nominal formant frequency values used in synthesis of experiment 1 CVC stimuli.

	Context	Onset	Midpoint	Offset
$F1$	All CVC	150	700	150
$F2$	[bΛb]	800	1260	800
	[bɛb]	800	1710	800
	[dΛd]	2270	1260	2270
	[dɛd]	2270	1710	2270
$F3$	[bVb]	2100	2400	2100
	[dVd]	2700	2400	2700
$F4$	All CVC	4000	4000	4000

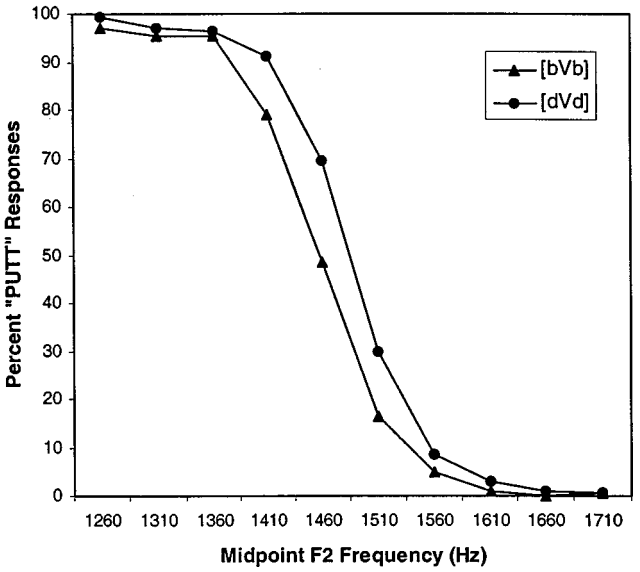


FIG. 2. Identification of Vowels in CVC Context. Mean identification functions for experiment 1. Percent ''PUTT'' responses as a function of midpoint $F2$ frequency are plotted by consonant context.

identical $F2$ midpoint frequencies in [bVb] context. Flanked by high-frequency $F2$ onset–offset consonants ([dVd]), vowels were significantly more likely to be identified as the low $F2$ frequency vowel [ʌ]. In the context of lower $F2$ frequency onset/offset consonants ([bVb]), vowels were more likely to be identified as those with higher $F2$ frequencies (i.e., as [ɛ]). These data thus support findings of earlier studies (Lindblom and Studdert-Kennedy, 1967; Nearey, 1989) and provide a foundation from which to examine the potential influence of nonspeech context upon perceived vowel identity.

II. EXPERIMENT 2

Experiment 2 explores the relative influence of phonetic labeling and spectral content on the effect observed in experiment 1. If spectral content of adjacent context is significant, then any salient acoustic energy, whether speech or nonspeech, should elicit shifts in vowel identification. Experiment 2 stimuli consist of synthetic vowels flanked by nonspeech stimuli that mimic some of the spectral characteristics of the CVC stimuli from experiment 1. In this way, phonetic labeling of context stimuli is eliminated, but some elements of spectral content of CVC context is preserved.

The primary acoustic dimension distinguishing consonant contexts in experiment 1 was $F2$ onset/offset frequency and $F2$ transition trajectories. Experiment 2 models this energy with nonspeech sine-wave glides that mimic the $F2$ frequency trajectories of the CVC stimuli. If spectral content is important, vowels presented in nonspeech contexts with higher-frequency acoustic energy (those modeled after [dVd]) should more often be labeled as [ʌ] than the same vowels flanked by lower-frequency nonspeech contexts modeled after [bVb]. However, nonspeech sounds should not be phonetically labeled. If phonetic labeling is important in effects of context on vowel perception, nonspeech contexts should not shift vowel identification.

A. Method

1. Subjects

Thirty native English-speaking University of Wisconsin undergraduates with normal hearing participated in return for course credit.

2. Stimuli

Experiment 2 utilized a 10-step series of 60-ms steady-state vowels synthesized according to the methods described for experiment 1.[6] As in experiment 1, series members differed only in $F2$ frequency (1260–1710 Hz, in 50-Hz steps) and varied perceptually from [ʌ] to [ɛ]. First, third, and fourth formant frequencies were constant across series members ($F1=700$ Hz, $F3=2400$ Hz, and $F4=4000$ Hz). All formants were steady state over the entire 60-ms duration.

Members of the synthetic vowel series were the basis for two series of nonspeech–speech hybrid stimuli. Each hybrid stimulus was made up of a 70-ms frequency-modulated (FM) sine-wave glide abutted in time to a member of the synthetic vowel series that, in turn, abutted a second 70-ms FM glide that was a mirror image of the first. Frequency trajectories of the sine-wave glides mimicked the $F2$-frequency trajectories of experiment 1 stimuli. For example, the hybrid series modeled after experiment 1 [bVb] series was composed of a nonlinear FM glide from 800 Hz to the $F2$ midpoint frequency (for the vowel) followed by a member of the synthetic vowel series and a final nonlinear FM glide from $F2$-midpoint frequency to 800 Hz. Stimuli mimicking [dVd] series were similar except FM glide onset/offset frequency was 2270 Hz. Equation (1) determined the precise trajectory of each FM glide from midpoint to steady-state values. Glides and vowels were appended online to create a nonspeech–speech hybrid stimulus to model each of [bVb] and [dVd] stimuli of experiment 1.

Pseudo-spectrograms of representative hybrid stimuli are illustrated in Fig. 3. It is important to note that Fig. 3 conveys only the frequency *trajectory* of the stimuli. Although the frequency trajectory of nonspeech precursors is identical to the $F2$ transitions of CVC stimuli from experiment 1, these stimuli should not be considered to be perceptually or acoustically equivalent to the speech that they model. The nonspeech precursors, though modeling putatively important characteristics of CVC context, differ considerably in overall spectral content and are not perceived as speechlike.[7]

The CVC stimuli possess a rich harmonic structure, with energy at each multiple of the fundamental frequency ($f0=120$ Hz). The FM glide, in contrast, possesses energy across only a limited band of frequencies, with no fine harmonic structure and no energy mimicking $F1$ or $F3$. Moreover, formant transitions like those of the CVC stimuli of experiment 1 are not truly frequency-modulated (as are the glide stimuli) because the actual component frequencies do not vary; only the amplitude peak of the spectral envelope varies for formant transitions. Despite these differences, the very simple FM glides capture some of the putatively important spectral energy in the region of CVC $F2$ frequencies. If spectral content accounts for context effects in vowel percep-

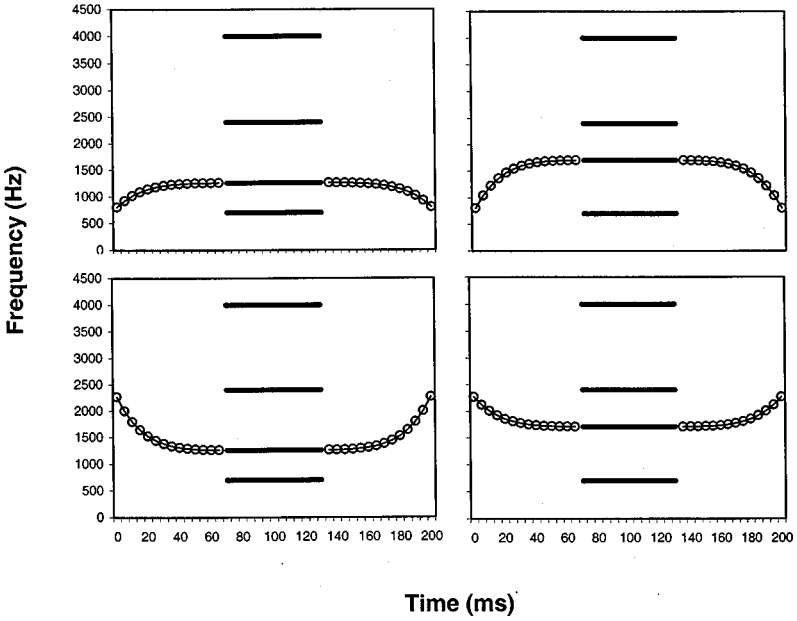


FIG. 3. Glide-Vowel-Glide Series Endpoints. Representative pseudo-spectrograms of experiment 2 stimulus-series endpoints. The format follows that of Fig. 1. The top row represents nonspeech–speech hybrid stimuli modeling the [bVb] series of experiment 1 whereas the bottom row shows endpoints mimicking the [dVd] series. The left column corresponds to low-$F2$ ([ʌ]) stimulus endpoints. The right column depicts high-$F2$ ([ɛ]) endpoints. In each frame, full-formant synthetic speech stimuli are illustrated as solid lines and nonspeech FM glides are shown as stippled lines.

tion, then these very simple nonspeech caricatures, by virtue of their spectral similarity to experiment 1 consonant contexts, may be sufficient to produce effects of context on vowel identification. If phonetic labeling is responsible, then these nonspeech glides should exert no influence on vowel identification.

B. Procedure

Stimulus presentation and response collection were identical to experiment 1. In a 2AFC task, subjects identified the vowel of each stimulus as [ʌ] or [ɛ] by pressing buttons labeled ''PUTT'' or ''PET'' on a handheld electronic response box. After categorizing the vowel, listeners classified the nonspeech segment of the stimulus as ''LOW'' or ''HIGH'' in a task that mirrored the consonant identification task subjects performed in experiment 1. This task also helped to ensure that listeners were not encouraged (perhaps implicitly, by mere participation in a vowel identification task) to attempt to phonetically label the FM glides. Nonspeech–speech hybrid stimuli modeling [bVb] and [dVd] were mixed in stimulus presentation. Each stimulus was presented 20 times in a random order for a total of 400 presentations. Experimental apparatus and procedure were otherwise like that of experiment 1.

C. Results and discussion

Mean identification functions are presented in Fig. 4. A 2×10 (nonspeech context×stimulus-step) ANOVA revealed a significant effect of sine-wave glide context ($F_{(1,29)} = 20.96$, $p < 0.0001$). These results demonstrate that a nonspeech analogue modeling $F2$ frequency trajectory is sufficient to elicit a context effect similar to that obtained with full-formant synthetic speech contexts. In the context of FM glides with higher-frequency onsets and offsets (modeling [dVd]), subjects more often identified vowels as [ʌ], the vowel with a lower $F2$ frequency.

Spectral content thus appears to be an important variable in explaining context effects on vowel perception. Shifts in vowel identification endure even when phonetic labeling is eliminated. Even a very simple sine-wave caricature of a portion of the energy present in rich full-spectrum CVC stimuli is sufficient to influence vowel identification.

One concern that might be raised with respect to these data arises from consideration of auditory grouping principles. Darwin and his colleagues (Darwin, 1984; Darwin and Sutherland, 1984; Darwin *et al.*, 1989), for example, have demonstrated that adjacent nonspeech tones and glides can ''capture'' harmonics of a vowel. Listeners ''attribute'' the energy of the vowel's harmonic to the nonspeech rather than vowel. In their studies, this resulted in an apparent change in $F1$ frequency and a concomitant shift in vowel identification. Considering that Dannenbring (1976) has

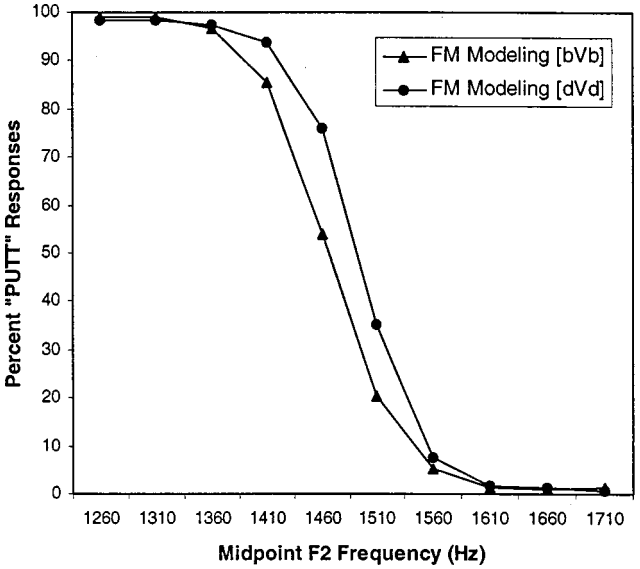


FIG. 4. Identification of Vowels in Glide Context. Mean identification functions for experiment 2. Percent ''PUTT'' responses as a function of midpoint $F2$ frequency are plotted by FM glide context.

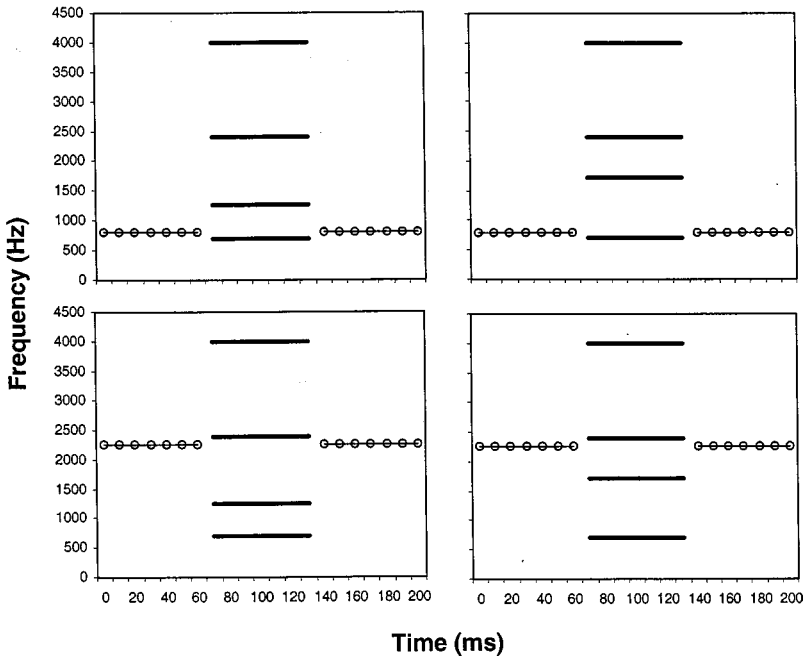


FIG. 5. Tone-Vowel-Tone stimulus series endpoints. Representative pseudo-spectrograms of experiment 3 stimulus-series endpoints. The format follows that of Figs. 1 and 3. The top row represents stimuli that mimic characteristics of [bVb] stimuli; the bottom depicts those that model [dVd] context. The left column illustrates low-$F2$ ([Λ]) vowels whereas the right column depicts high-$F2$ ([ɛ]) vowels. In each frame, full-formant synthetic speech stimuli are illustrated as solid lines and nonspeech FM glides are shown as stippled lines.

demonstrated that perceived continuity of an intermittent sound is greatest when rising or falling transitions have their discontinuity filled with another sound at the maximum or minimum of their transition trajectory, the stimuli of experiment 2 may have been susceptible to effects of auditory grouping. That is, subjects may have grouped several harmonics around $F2$ with the rising and falling sinusoidal context, thus leading to a perceived shift in $F2$.

III. EXPERIMENT 3

It is worth noting that experiment 2 stimuli were constructed to minimize effects of grouping inasmuch as was possible within this particular stimulus paradigm. Both the nominal vowel $F2$ frequency and the maximum/minimum frequencies of the FM glides that were temporally closest to the vowel were positioned such that they did not fall upon harmonics of the fundamental frequency. To even further reduce potential influences of auditory grouping, experiment 3 used much simpler nonspeech context stimuli. For this new set of nonspeech–speech hybrid stimuli, the nonspeech context was a simple steady-state sine-wave tone situated at the onset/offset frequency of either [bVb] or [dVd] stimuli from experiment 1 (i.e., 800 or 2270 Hz). The tone–vowel–tone stimuli of experiment 3 mitigate grouping of nonspeech energy with harmonics critical to vowel perception because tone context stimuli are closer in frequency to $F1$ and $F3$ (for contexts modeling [bVb] and [dVd], respectively) than $F2$. Thus ''harmonic capture,'' should it be a factor, is much less likely to influence the $F2$ region of the vowel spectra that distinguishes stimuli. Further, because these steady-state tones model even more limited characteristics of CVC spectral characteristics than the FM glides of experiment 2, this stimulus paradigm allows an examination of how closely spectral characteristics of nonspeech analogues must model CVCs to produce shifts in vowel identification.

A. Method

1. Subjects

Thirty-one native English-speaking undergraduate students from the University of Wisconsin–Madison participated in partial fulfillment of the requirements of Introductory Psychology. All listeners reported normal hearing.

2. Stimuli

The 10-step series of steady-state vowel stimuli from experiment 2 were utilized again in experiment 3. For this experiment, steady-state-sine-wave tones, rather than FM glides, served as context stimuli. Each stimulus was comprised of a 70-ms constant-frequency sine-wave tone immediately followed by a member of the 60-ms synthetic vowel series in turn followed by a second 70-ms tone identical to the first. Tones and vowels were appended online to create a model of each of the experiment 1 stimuli. For stimuli modeling [bVb] series members, tone frequency was equivalent to the $F2$ onset/offset frequency of [bVb] stimuli, 800 Hz. Tones with frequencies of 2270 Hz modeled [dVd] series stimuli. Endpoint stimuli are illustrated in Fig. 5.

B. Procedure

In a 2AFC task, listeners identified the vowel of each stimulus as [ʌ] or [ɛ] by pressing buttons labeled ''PUTT'' or ''PET'' on a handheld electronic response box. After categorizing the vowel, subjects classified the nonspeech segment as ''LOW'' or ''HIGH'' in accord with the tasks of experiments 1 and 2. Stimulus presentation was mixed across the two series and each stimulus was presented 20 times in a random order for a sum of 400 total stimuli. Experimental apparatus and procedure were otherwise like that of experiments 1 and 2.

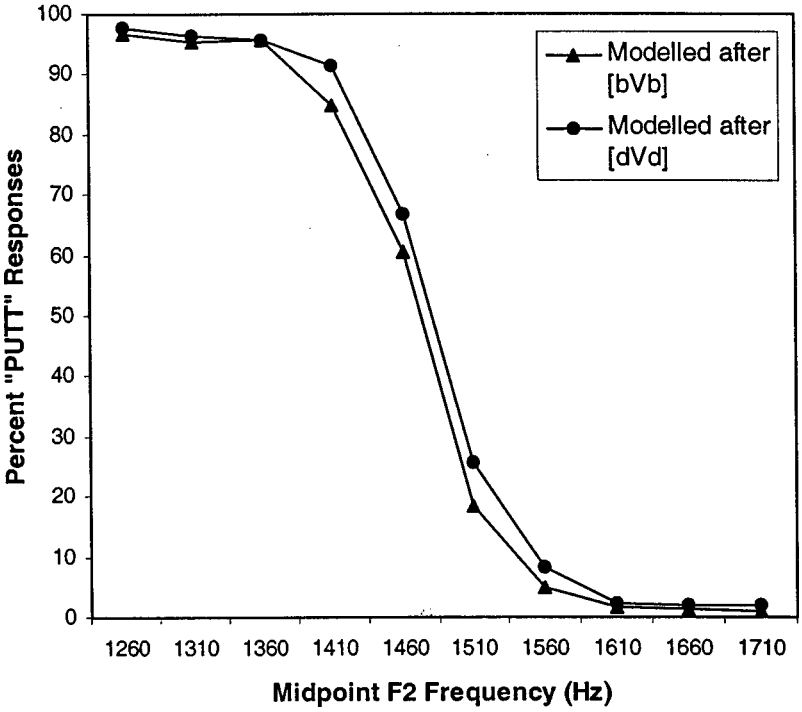


FIG. 6. Identification of Vowels in Tone-Vowel-Tone Context. Mean identification functions for experiment 3. Percent ''PUTT'' responses as a function of midpoint $F2$ frequency are plotted by tone context.

C. Results and discussion

Mean identification functions for experiment 3 are illustrated in Fig. 6. The results of experiment 3 extend the generality of the perceptual effects observed in experiments 1 and 2 to include constant-frequency tones. A static pure tone appended to the onset and offset of a vowel was sufficient to produce a small, but quite reliable, shift in vowel identification ($F_{(1,30)}=8.15$, $p<0.05$). Vowels were more likely to be identified as [ʌ] (having low-frequency $F2$) when flanked by high-frequency tones equivalent to the onset/offset frequency of [dVd] series stimuli in experiment 1.

Although steady-state sine-wave stimuli exert an influence on vowel identification, their effect is smaller than the effects of context observed for consonant and FM glides. This may be because these stimulus hybrids were subject less to auditory grouping, but there are also at least several other explanations for why tones may have relatively less influence on vowel identification. For experiment 3, frequency of the tones was equivalent to onset and offset frequencies of experiment 1 CVCs. Although this seems to be a reasonable choice, it may not have been optimal for producing the largest effects of context. Evidence from psychophysics, for example, suggests listeners perceive brief glides to be a sort of average of frequency across glide trajectory (Brady *et al.*, 1961; Nábělek *et al.*, 1970, 1973). Given this work, it could be argued that tone analogues at mid-trajectory may have more closely modeled context stimuli of experiments 1 and 2. Future psychophysical work that maps the extent of context effects with hybrid speech–nonspeech sounds ultimately will demonstrate whether effect size is sensitive to precise placement of context stimuli and whether these functions can be predicted from psychoacoustics. A second hypothesis is that the dynamic nature of the FM glides from experiment 2 contributes to the influence upon context effects. Given the results of experiment 3, kinematic information offered by dynamic stimuli cannot explain these context effects entirely, but it is possible that this information is a factor. Another difference between the glide and tone stimuli was the overall spectral content. Owing to frequency modulation, glide stimuli possess a richer spectrum than do their tone counterparts. It is conceivable that this spectral distinction influenced the magnitude of the context effect produced by tone stimuli. The results of Shigeno and Fugisaki (1979) suggest that at least one class of context effect becomes more intense as stimulus complexity increases.

IV. EXPERIMENT 4

Time-varying sinusoids and steady-state tones, neither labeled phonetically, were sufficient to shift listeners' identification of vowel stimuli. Taken together, results of experiments 2 and 3 suggest that when phonetic labeling is absent, spectral content is an important variable in understanding how context affects vowel perception. However, it remains at issue whether spectral content is as important when phonetic labeling is intact.

Experiment 4 addresses this question using voicing as a means of manipulating spectral content, while maintaining phonetic labeling. Spectrally, voiced and voiceless consonants vary along many acoustic dimensions. Several distinctions are convenient for the present hypothesis. Namely, voiceless consonants tend to be lower in amplitude than voiced consonants and, owing to the aspiration that often accompanies voicelessness, voiceless consonants tend to have more diffuse spectra than their voiced counterparts. The results of experiments 2 and 3 suggested that acoustic energy in the region of $F2$ frequency trajectories of CVC syllables is important in predicting effects of context. Given this significance, one might predict smaller effects of context from

stimuli with less prominent acoustic energy. By this view, voiceless consonants, with lower-amplitude harmonics and more diffuse spectra, should have relatively less influence on vowel identification than their voiced consonant complements. Experiment 4 examines this hypothesis.

A. Method

1. Subjects

Fourteen undergraduate Introductory Psychology students with normal hearing participated in return for course credit. All listeners learned English as a first language.

2. Stimuli

Two pairs of CVC stimulus series were used in experiment 4. One pair was identical to the [bʌb]–[bɛb] and [dʌd]–[dɛd] series used in experiment 1, except that their duration was 100 ms. This shorter duration was used because previous research (Lindblom and Studdert-Kennedy, 1967) has demonstrated that CVC context effects are smaller for longer stimuli. The present experiment investigates whether a spectral change introduced by manipulating voicing changes the degree to which consonants influence vowel identification. The experimental manipulation is predicted to produce smaller context effects. Therefore, it is desirable to elicit the largest effect possible while maintaining adequate phonetic identification. Shortening the overall stimulus duration to 100 ms thus allowed a more sensitive test of the prediction. In addition, this change allowed experiment 4 to test whether the effect of stimulus duration reported for [jVj] and [wVw] contexts (Lindblom and Studdert-Kennedy, 1967) extends to stop-consonant CVCs.[8]

Duration and formant frequency trajectories for the second pair of series were identical to those of the first. The critical distinction between pairs was that, for the second pair, consonant context was synthesized to mimic voiceless stop consonants rather than voiced consonants as in experiment 1. Thus these stimuli varied perceptually from [pʌp]–[pɛp] and [tʌt]–[tɛt]. It is important to note that the formant frequency trajectories for [bVb] and [pVp] series were identical. Likewise, [dVd] and [tVt] series shared equivalent formant frequency trajectories (as in Fig. 1, except stimuli are 100 ms). Voicing source was the key difference between pairs. Synthesis of this distinction was accomplished by changing the synthesis parameters (Klatt, 1980; parallel branch) of the voiceless series to model the aperiodic aspirated source typical of voiceless consonants. Specifically, the voicing source (Klatt parameter AV) was substituted with an aperiodic source (AH). This change had the effect of lowering the amplitude of the formants during the period of voicelessness. Klatt (1980) synthesizer parameters AV and AH were manipulated carefully so that the transitions of voiceless stimuli were approximately 6 dB lower in amplitude than the same segments of their voiced counterparts. To improve perceptual salience of the voiceless stops, $F1$ cutback (typical of voiceless stop consonants) was also modeled by reducing the amplitude of $F1$ to zero the voiceless portion of the token. No consonant bursts were added to these stimuli. VOT for all stimuli was 40 ms, a value intermediate that of typical natural English [p] and [t] productions (Lisker and Abramson, 1964). Thus voiced and voiceless series shared formant frequency trajectories, but differed in amplitude, periodicity, and low-frequency ($F1$) energy during the voiceless portion of the stimulus.

A final note is in order regarding the voiceless stimulus series used in this experiment. In synthesizing speech for use in controlled laboratory experiments, there is always a compromise between stimulus naturalness and experimental control. In the present case, it was very important to be meticulous with regard to stimulus characteristics because the hypothesis under investigation is coupled closely with the spectral content of the stimuli. With regard to stimulus naturalness, the present investigation demands only that the syllables be perceived phonetically and that listeners' ability to label voiced versus voiceless stimuli is not disproportionate. Thus although the voiceless stimuli employed here do not precisely model natural productions of voiceless CVC syllables (e.g., natural voiceless CVCs tend not to be symmetric), the question under investigation dictated that voiced and voiceless stimuli be as spectrally similar as possible, manipulating only very specific characteristics of the consonant context.

To be certain that listeners do, in fact, perceive these voiceless syllables in the intended manner, experiment 4 incorporates a consonant identification pre-test. In comparing listeners' identification of vowels in the context of voiced and voiceless consonants, it is important to ensure that any observed differences do not stem merely from a difference in the ability of listeners to label voiced versus voiceless consonants; differences in phonetic labeling for voiced versus voiceless consonants may influence the degree to which spectral content can be held responsible for effects of context on vowel identification.

B. Procedure

Therefore, each listener first participated in a preliminary task in which consonant identification was assessed. During this task, listeners heard each stimulus from each 10-member stimulus series ([bVb], [dVd], [pVp], [tVt]) five times for a total of 200 stimuli. On each trial, subjects labeled the consonants of the syllables as ''B,'' ''D,'' ''P,'' or ''T.''

Following this task, each listener completed two blocks of trials to assess vowel identification in the context of adjacent consonants. The trials were blocked by voicing, such that listeners identified vowels in the context of [bVb] and [dVd] separate from vowels in [pVp] and [tVt] contexts. The order of blocks was counterbalanced across subjects. Within a block, subjects heard 20 repetitions of each member of each stimulus series. On every trial, subjects indicated perceived vowel identity as [ʌ] or [ɛ] by pressing one of two buttons labeled ''HUD'' and ''HEAD,'' then responded to the consonant as ''P/T'' or ''B/D.''[9]

C. Results and discussion

Subjects performed very well on the consonant identification pretest. On average, subjects correctly labeled conso-

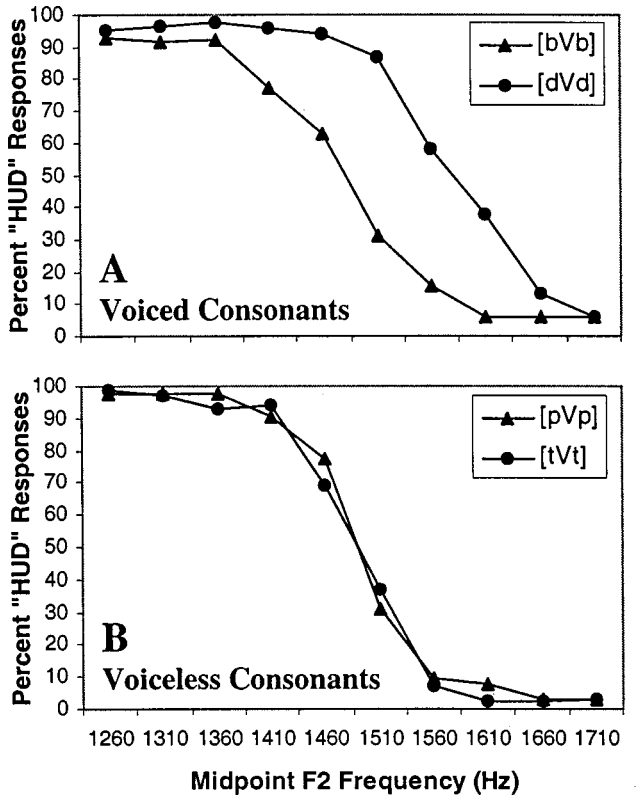


FIG. 7. Identification of Vowels in ±Voice CVC Contexts. Panel A illustrates listeners' identification of vowels in the context of voiced consonants. Panel B depicts the same listeners' identification responses for vowels in the context of voiceless consonants.

nants from [bVb], [dVd], [pVp], and [tVt] series 95.3%, 91.2%, 93.3%, and 90.7% of the time, respectively. Thus despite strict constraints on stimulus construction, all consonant contexts were readily identifiable.

The influence of voiced consonants on vowel identification is shown in the top panel of Fig. 7. As in experiment 1, voiced consonant context elicited a substantial effect on vowel identification ($F_{(1,13)}=73.18$, $p<0.0001$). In the context of [dVd], vowels were significantly more likely to be identified as [ʌ]. Consistent with earlier observations (Lindblom and Studdert-Kennedy, 1967), the magnitude of the context effect produced by these shorter stimuli (100 ms versus 200 ms) was greater than that observed for experiment 1, thus extending effects of duration upon phonetic context effects to stop-consonant CVCs.

The significant question of experiment 4 was whether voiceless consonant contexts, though easily labeled and possessing the same formant trajectories as their voiced counterparts, would exhibit relatively less influence on vowel identification due to the differences in spectral content between series. The lower panel of Fig. 7 depicts the data pertinent to this question. As is clear from the graph, voiceless consonants were substantially less effective in shifting vowel identification than their voiced counterparts. In fact, there is no significant influence of neighboring voiceless consonants upon vowel identification at all ($F_{(1,13)}=0.66$, $p=0.435$). These data thus strongly implicate spectral content as an important factor in context effects in vowel perception, even in stimulus paradigms were phonetic labeling is intact.

These results are especially interesting because it is unlikely that extant accounts of context effects would forecast these outcomes. Most theoretical accounts of context effects in speech perception, whether embodied by modular processes (Liberman *et al.*, 1957; Liberman and Mattingly, 1985; Mann, 1980, 1986), Direct Realist accentuation of distal events (Fowler *et al.*, 1990; Fowler, 1986, 1996), or reference to tacit knowledge (Repp, 1982, 1983) suggest that mechanisms of context effects in speech perception are intimately linked with speech production generally, and to effects of coarticulation more specifically. However, the dynamics of articulation for voiced/voiceless consonants at the same place of articulation should be very similar because the voicing distinction is related primarily to a change in the timing relationship between release of oral closure and onset of vocal fold vibration, not movement of supra-laryngeal articulators (Lisker and Abramson, 1964, 1971). Thus articulatory patterns for voiced versus voiceless stimuli should be quite similar and their coarticulatory influences on neighboring sounds should be analogous, too. Given this similarity, it is unclear how accounts that rely on knowledge of or recovery of coarticulatory aspects of speech production to aid perception would account for the results of experiment 4.

V. GENERAL DISCUSSION

The goal of the present work was to assess the relative contributions of phonetic labeling and spectral content on context effects in vowel categorization. Williams (1986) reported data suggesting that effects of context on vowel identification are dependent upon phonetic identification. Listeners exhibited context effects for sine-wave speech in conditions where sine-wave triads were phonetically labeled, but not in conditions where the same sounds were labeled on pitch. Mullennix *et al.* (1988) later called the necessity of phonetic labeling into question by reporting a context effect for sine wave triads labeled either phonetically or with pitch labels, thus suggesting a possible influence of spectral content.

The present data support this possibility. Spectral content, rather than phonetic labeling, seems to be the chief characteristic contributing to context effects in vowel perception. The results of experiments 2 and 3 indicate that nonspeech acoustic contexts with spectral characteristics modeled after consonant contexts of experiment 1 are sufficient to produce shifts in vowel identification. Frequency-modulated sine-wave glides and, to a lesser extent, steady-state sine-wave tones, elicit shifts in listeners' vowel identification responses. Moreover, these shifts are in accord with patterns of perception produced by consonant neighbors. These nonspeech contexts were not phonetically labeled, so it appears that the limited spectral characteristics of the consonant acoustics that they modeled are adequate to predict effects of context on vowel identification. Furthermore, because these nonspeech stimuli elicited context effects, the data suggest that close correspondence to productions of the human vocal tract is not a prerequisite for context effects in speech perception.

Experiment 4 linked results of experiments 2 and 3 back to more ordinary speech perception in CVC stimuli. Voiced,

but not voiceless, consonant context modulated listeners' vowel identity responses. This is a curious finding in that extant theoretical accounts of speech perception that rely upon links to speech production to account for context effects would not predict such an interaction. Several new empirical investigations, however, suggest a manner by which to interpret these results. Recently, there have been a number of reports investigating explicitly the possibility that the influence of neighboring context on speech identification may be a consequence of general auditory processes (Holt, 1999; Lotto *et al.*, 1997; Lotto and Kluender, 1998; for an overview see Holt and Kluender, in press). These studies have provided two primary lines of evidence to suggest effects of adjacent neighbors on speech perception may have origins in very general auditory mechanisms.

The first line of evidence ties closely with results of experiments 2 and 3 and involves correspondence between the effects of neighboring speech and nonspeech sounds upon perception of speech. Lotto and Kluender (1998) demonstrated that sine-wave FM glides modeling $F3$ transitions of [al] or [ar] and steady-state tones situated at the $F3$ offset frequencies of [al] or [ar] induce the same pattern of [ga]–[da] identification responses as do natural and synthetic speech tokens of [al] and [ar]. Likewise, Holt (1999) has reported that listeners exhibit shifts in identification of CV syllables when the CVs are preceded by [i], [u], or by nonspeech precursors modeling $F2$ characteristics of these vowels. The correspondence of speech and nonspeech–speech hybrid context effects suggests that general auditory alternatives to speech-specific explanations may be tenable and that these mechanisms are related to spectral content of the acoustic signal.

A second line of evidence bolsters this account. Lotto *et al.* (1997) trained Japanese quail (*Coturnix coturnix japonica*) to peck to a lighted key in response to endpoints of the same [da]–[ga] series for which humans had exhibited context-dependent shifts in identification (Lotto and Kluender, 1998). Birds trained to peak to [ga] pecked most vigorously to novel intermediate members of the [da]–[ga] series that were preceded by [al]. Correspondingly, [da] positive quail pecked most robustly when novel stimuli were preceded by [ar]. Japanese quail thus exhibited a context effect analogous to human shifts in CV identification. Quail are unlikely to have relied upon recovery or representation of human vocal tract dynamics. Furthermore, the quail had no experience with coarticulated speech, so their behavior cannot be explained based on learned covariance of the acoustic attributes of coarticulated speech.

These twin lines of evidence favor a general auditory account of context effects in speech perception rather than explanations hinging on correspondences between speech perception and speech production or speech-specific mechanisms. Taken together, these data suggest spectral content and its influence upon the auditory system may be critical variables in explaining context effects. Thus these data join previous reports of speech perception phenomena that appear to arise from general perceptual processes and are not dependent upon phonetic labeling (e.g., Diehl and Walsh, 1989; Lotto *et al.*, 1996; Sinnott *et al.*, 1998).

Discussion of previous results (Holt, 1999; Lotto and Kluender, 1998; Lotto *et al.*, 1997) has cast these findings as cases of spectral contrast. Considered in this way, the results of experiments 1–3 could be interpreted in the following manner: In the context of higher-frequency spectral energy (i.e., [dVd] stimuli and the nonspeech analogues that model them), vowels are perceived as the alternative with lower-frequency spectra (i.e., as [Λ]). This pattern of perception is typical of phonetic context effects in speech and has been noted by other authors as well (e.g., Lindblom and Studdert-Kennedy, 1967; Repp, 1983; Fowler *et al.*, 1990). However, the present results implicate general auditory mechanisms in explanations of context effects and, therefore, suggest that contrast may serve as more than a mere description. Mechanisms of contrast that enhance change in the acoustic signal potentially could produce perceptual results like those reported here.

This hypothesis is all the more feasible considering that contrast is an important mechanism for exaggerating differences between neighboring objects and events across perceptual modalities. The best-known examples are in the visual domain (enhancement of edges produced by lateral inhibition, Hartline and Ratliff, 1957; lightness judgments, Koffka, 1935; judgment of line orientation, Gibson, 1933), but context effects in behavior are observed for all sensory modalities (von Bekesy, 1967; Warren, 1985). Across domains, contrast is a central characteristic of mechanisms that serve to exaggerate change in the physical stimulus and to maintain an optimal dynamic range. Perceptual contrast, in this case spectral contrast, may play an important role in perception of speech, too.

This perspective also contributes an explanation for the curious results of experiment 4. Indulging the hypothesis that general perceptual mechanisms of spectral contrast may play a role in vowel context effects, it is useful to consider how the auditory system might implement spectral contrast. Delgutte (1996; Delgutte *et al.*, 1996), for example, has suggested spectral contrast might be realized by the auditory system via neural adaptation. Adaptation can be described very generally as a decrease in neurons' discharge rate following an initial response, such that responses to subsequent stimuli are depressed (Harris and Dallos, 1979; Smith, 1979). Though neural adaptation is perhaps best known for its influence in the auditory system at the level of the auditory (VIIth) nerve, adaptation exists at every level of the auditory system and occurs on many time scales, from a few milliseconds to several seconds or even minutes (Kiang *et al.*, 1965; Smith, 1979).

Neurons in the auditory system have a preferred, or characteristic, frequency to which they are most likely respond with an action potential (Brugge and Reale, 1985; Irvine, 1992; Ruggero, 1992). Adaptation may produce contrast because neurons excited by stimulus components spectrally close to their preferred frequency fire and subsequently become adapted. Consider what happens when another stimulus follows. Neurons most sensitive to frequencies present in the preceding stimulus will be likely to have fired in response to it, leaving them relatively less responsive due to adaptation. However, frequencies absent in the pre-

ceding stimulus will tend to be encoded by more responsive *unadapted* neurons. Thus on a population level, there is a shift in neural response toward frequencies absent in the preceding stimulus.

If we entertain neural adaptation as a candidate mechanism for spectral contrast, a prediction arises. All other things being equal, less spectrally distinct or less intense precursors should tend to result in fewer adapted neurons. Thus when a subsequent stimulus follows, less adaptation should result in less of a population shift in neural response. That is, there should be less neural contrast. If this change influences perception, such stimuli should exert a smaller effect of context on their neighbors. For the differences in spectra between voiced and voiceless consonants in experiment 4, the account outlined above predicts that lower-amplitude, more diffuse formant energy of voiceless consonants such as [p] and [t] ought to exert relatively less influence on neighboring vowels than should their voiced counterparts, [b] and [d]. In previous discussions (Lotto *et al.*, 1997; Lotto and Kluender, 1998; Holt and Kluender, in press), spectral contrast has served as a useful descriptive convenience in discussing general auditory mechanisms by which context effects might occur. Experiment 4 provides the first perceptual evidence to advise that spectral contrast may provide more than descriptive convenience in explaining context effects in speech perception. Rather, it serves as a framework from which to make novel predictions about perception. Much work remains to be done before the precise means by which context effects in speech perception arise is known. Nonetheless, a number of things are clear, even at this point.

Often, arguments that context effects in speech perception arise from ''general auditory'' processes are taken as suggesting that the mechanisms must be *peripheral*. If, by peripheral, authors mean mechanisms that exert their influence in the cochlea or at the level of the auditory nerve, then this is almost certainly false. Evidence in other stimulus paradigms (e.g., Holt, 1999; Lotto, 1996) suggests that more central regions of the auditory system are likely involved in context effects. In addition, neurophysiological investigation of these effects has borne little evidence of robust neural encoding of context effects at the level of the auditory nerve (Holt and Rhode, 2000). This need not rule out ''general auditory'' accounts. Consider, again, adaptation as a candidate mechanism. As Delgutte (1996) has pointed out, adaptation occurs at all levels of the auditory system and possesses a host of time courses.

It is very important to note that the results of experiment 4 do not demonstrate that neural adaptation serves as a mechanism for context effects in speech perception. As Summerfield and others have argued within other theoretical domains (e.g., Summerfield *et al.*, 1984, 1987), effects similar to those putatively accomplished by neural adaptation may be accomplished by adaptation of suppression (a mechanism whereby inhibitory inputs are suppressed, thus enhancing the response of some population of neurons by virtue of decreasing inhibition) or other temporal interaction mechanisms such as disinhibition, long-lasting inhibition, or facilitation. Temporal interactions abound within the auditory system (Delgutte, 1996). Neural adaptation, even if it does play a role, is unlikely to account entirely for such effects. Especially at higher levels of the auditory system, mechanisms of temporal interaction are likely to be more complex than simple adaptation. Even at the cochlear nucleus, responses of single neurons to brief vowel stimuli presented in rapid succession are suppressed when a particular vowel is preceded by another vowel that, by itself, produces no response (Caspary *et al.*, 1977). These data suggest that there exist mechanisms distinct from adaptation that are involved in temporal interactions among neighboring speech sounds. The present results, coupled with these neurophysiological clues, suggest that although neural adaptation may not provide a complete account of effects of context in speech perception, spectral contrast has a useful role in predicting perceptual results.

In explaining their original finding that CVC context influences listeners' identification of vowels, Lindblom and Studdert-Kennedy (1967) offered several possible theoretical interpretations. They noted that their data were agreeable to articulation-based theoretical perspectives like Motor Theory (Liberman *et al.*, 1957) and Analysis-by-Synthesis (Stevens and House, 1963; Stevens and Halle, 1967). However, Lindblom and Studdert-Kennedy also examined the possibility that more general perceptual mechanisms may be responsible. Indeed, they devoted most of their discussion to potential explanations in terms of general auditory processes. In most all examinations of phonetic context effects since, authors have interpreted their data more categorically as evidence that context effects in speech perception originate in properties of speech perception distinct from its auditory characteristics (e.g., Mann, 1980; Repp, 1982; Williams, 1986; Fowler *et al.*, 1990). The present data, in combination with other recent results (Lotto *et al.*, 1997; Lotto and Kluender, 1998; Holt, 1999) suggest that general perceptual mechanisms sensitive to similarities in spectral content play an important role in context effects in speech perception.

ACKNOWLEDGMENTS

This work was supported in part by a National Science Foundation Predoctoral Fellowship to the first author. Additional support was provided by NSF Young investigator Award DBS-9258482 to the third author. Some of the data were presented at the 131st Meeting of the Acoustical Society of America in Indianapolis, IN. The authors thank Bjorn Lindblom and Terry Nearey for their helpful comments as this work progressed. Correspondence and requests for reprints should be addressed to Lori L. Holt, Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 (email: lholt@andrew.cmu.edu).

[1]One distinction between these studies is that Williams (1986) did not include a [jVj] series. In addition, Lindblom and Studdert-Kennedy employed a two-alternative forced-choice (2AFC) identification procedure. Williams also collected 2AFC data for two of his three conditions. Results were comparable across methods, but the overall effects observed for 2AFC identification were smaller. The ABX results are described here because they provide the most complete account of Williams' findings.

[2]In these pilot studies, we replicated Lindblom and Studdert-Kennedy (1967) and Nearey (1989). However, we were interested in the putative role of phonetic identity in phonetic context effects, so we assessed listeners' ability to identify consonants as well as vowels in the CVC syllables. Though most subjects converged upon consistent labeling of [wVw] and

[jVj] stimuli, some listeners reversed their identifications. The same listeners had little trouble identifying stop-consonant CVCs. Therefore, we chose to use stop-consonant CVCs for the remainder of the study.

[3]Lindblom and Studdert-Kennedy (1967) created CVCs perceptually corresponding to [wVw] and [jVj] using $p=2.0$, a parabolic function. Like Nearey (1989), the present experiment employed $p=6.0$ to produce percepts of CVCs conforming to [bVb] and [dVd]. Acoustically, use of a larger exponent creates formant transitions that traverse frequency more quickly, creating a longer duration at relative steady-state midpoint frequency relative to smaller exponents.

[4]In studies of context effects in speech perception, the method of comparing identification of isolated exemplars to identification of the same exemplars in the context of adjacent stimuli has been used extensively (e.g., Lindblom and Studdert-Kermedy, 1967; Mann, 1980; Mann and Repp, 1980). Despite this common practice, this means may not be the most appropriate manner of assessing ''baseline'' perception in the case of vowel identification. This is to say that, considered in terms of adjacent spectral energy and its influence on perception of vowel stimuli, silence may not be the best control ''stimulus'' for comparison with prominent low- or high-frequency energy. For this reason, context effects are assessed here in terms of the extent to which identification of a vowel stimulus is shifted relative to adjacent context rather than the extent to which it is shifted from listeners' identification of the vowel in isolation.

[5]Listeners were extremely accurate in labeling consonant contexts as ''B'' or ''D'' (96.7% and 97.2% accuracy, respectively), indicating that CVC syllables were perceived as reasonable exemplars of the intended speech sounds. To remain in accord with this first experiment, the remaining studies (experiments 2–4) use similar methods for labeling context. In all cases, listeners performed very well on the labeling tasks. Listeners' responses will be reported in detail only when they are central to the hypotheses of interest.

[6]The value 60 ms was chosen as a compromise between preserving spectral information inherent to the formant transition of CVC syllables and preserving the spectral information of the relatively steady-state vowel portion of the syllables. Inspection of the Klatt synthesis parameters derived from Eq. (1) (updated every 5 ms) led us to choose 60 ms as a compromise between these competing demands. It was also desirable to have the overall stimulus duration of experiment 2 stimuli be equivalent to that of the CVC stimuli of experiment 1 (200 ms). Each nonspeech context stimulus was thus 70 ms, for a total stimulus duration of 200 ms ($70+60+70$ ms $=200$ ms).

[7]Independent of experiment 2, a small group of naïve listeners described these stimuli. No listener described the nonspeech context stimuli as speechlike. The most common responses were ''computer game sounds,'' ''beeps,'' ''electronic music,'' and ''chirps.''

[8]Nearey (1989) used 100-ms stop-consonant CVCs in an analysis of the effect of consonant context on vowel identification. However, this study did not include a direct comparison across syllable duration.

[9]Different response labels were used in this experiment to avoid confounding the labels with the voiceless series.

Bailey, P. J., Summerfield, A. Q., and Dorman, M. (**1977**). ''On the identification of sine wave analogues of certain speech sounds,'' Haskins Lab. Status Report on Speech Res., **SR-51/52**, 1–25.

von Bekesy, G. (**1967**). *Sensory Inhibition* (Princeton University Press, Princeton, NJ).

Best, C. T., Morrongiello, B., and Robson, R. (**1981**). ''Perceptual equivalence of acoustic cues in speech and nonspeech perception,'' Percept. Psychophys. **29**, 191–211.

Brady, P. T., House, A. S., and Stevens, K. N. (**1961**). ''Perception of sounds characterized by a rapidly changing resonant frequency,'' J. Acoust. Soc. Am. **33**, 1357–1362.

Brugge, J. F., and Reale, R. A. (**1985**). ''Auditory cortex,'' in *Cerebral Cortex*, Vol. 4, edited by A. Peters and E. G. Jones (Plenum, New York), pp. 229–271.

Caspary, D. M., Rupert, A. L., and Moushegian, G. (**1977**). ''Neuronal coding of vowel sounds in the cochlear nuclei,'' Exp. Neurol. **54**, 414–431.

Dannenbring, G. L. (**1976**). ''Perceived auditory continuity with alternately rising and falling frequency transitions,'' Can. J. Psychol. **30**, 99–114.

Darwin, C. J. (**1984**). ''Perceiving vowels in the presence of another sound: Constraints on formant perception,'' J. Acoust. Soc. Am. **76**, 1636–1647.

Darwin, C. J., Pattison, H., and Gardner, R. B. (**1989**). ''Vowel quality changes produced by surrounding tone sequences,'' Percept. Psychophys. **45**, 333–342.

Darwin, C. J., and Sutherland, N. S. (**1984**). ''Grouping frequency components of vowels: When is a harmonic not a harmonic?'' Q. J. Exp. Psychol. **36A**, 193–208.

Delgutte, B. (**1996**). ''Auditory neural processing of speech,'' in *The Handbook of Phonetic Sciences*, edited by W. J. Hardcastle and J. Laver (Blackwell, Oxford), pp. 507–538.

Delgutte, B., Hammond, B. M., Kalluri, S., Litvak, L. M., and Carian, P. A. (**1996**). ''Neural encoding of temporal envelope and temporal interactions in speech,'' in *Proceedings of Auditory Basis of Speech Perception*, edited by W. Ainsworth and S. Greenberg, European Speech Communication Association.

Diehl, R. L., and Walsh, M. A. (**1989**). ''Effects of syllable duration on stop-glide identification in syllable-initial and syllable-final position by humans and monkeys,'' J. Acoust. Soc. Am. **85**, 2154–2164.

Fowler, C. A. (**1986**). ''An event approach to the study of speech perception from a direct-realist perspective,'' J. Phonetics **14**, 3–28.

Fowler, C. A. (**1996**). ''Listeners do hear sounds, not tongues,'' J. Acoust. Soc. Am. **99**, 1730–1741.

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.

Gibson, J. J. (**1933**). ''Adaptation, after-effect and contrast in the perception of curved lines,'' J. Exp. Psychol. **16**, 1–31.

Harris, D. M., and Dallos, P. (**1979**). ''Forward masking of auditory-nerve fiber responses,'' J. Neurophysiol. **42**, 1083–1107.

Hartline, H. K., and Ratliff, F. (**1957**). ''Inhibitory interaction of receptor units in the eye of *Limulus*,'' J. Gen. Physiol. **40**, 1357–1376.

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 Kluender, K. R. (in press). ''General auditory processes contribute to perceptual accommodation of coarticulation,'' Phonetica.

Holt, L. L., and Rhode, W. R. (**2000**). ''Examining context-dependent speech perception in the chinchilla cochlear nucleus,'' Midwinter Meeting of the Association for Research in Otolaryngology.

Irvine, D. R. F. (**1992**). ''Physiology of the auditory brainstem,'' in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay (Springer Verlag, New York), pp. 153–231.

Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). ''Discharge patterns of single fibers in the cat's auditory nerve,'' Res. Monogr., 35 (MIT Press, Cambridge, MA).

Klatt, D. K. (**1980**). ''Software for a cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 971–995.

Koffka, K. (**1935**). *Principles of Gestalt Psychology* (Harcourt Brace, New York).

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1957**). ''Perception of the speech code,'' Psychol. Rev. **74**, 431–461.

Liberman, A. M., and Mattingly, I. G. (**1985**). ''The motor theory of speech perception revised,'' Cognition **21**, 1–36.

Lindblom, B., and Studdert-Kennedy, M. (**1967**). ''On the role of formant transitions in vowel recognition,'' J. Acoust. Soc. Am. **42**, 830–843.

Lisker, L., and Abramson, A. S. (**1964**). ''A cross-language study of voicing in initial stops: Acoustical measurements,'' Word **20**, 384–422.

Lisker, L., and Abramson, A. S. (**1971**). ''Distinctive features and laryngeal control,'' Language **47**, 767–785.

Lotto, A. J. (**1996**). ''General auditory constraints in speech perception,'' unpublished doctoral dissertation, University of Wisconsin–Madison.

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 Green, K. P. (**1996**). ''Spectral discontinuities and the vowel length effect,'' Percept. Psychophys. **58**, 1005–1014.

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.

Mann, V. A. (**1980**). ''Influence of preceding liquid on stop-consonant perception,'' Percept. Psychophys. **28**, 407–412.

Mann, V. A. (**1986**). ''Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English ''l'' and ''r'','' Cognition **24**, 169–196.

Mann, V. A., and Repp, B. H. (**1980**). ''Influence of vocalic context on perception of the [sh]–[s] distinction,'' Percept. Psychophys. **28**, 213–228.

Mullennix, J. W., Pisoni, D. B., and Goldinger, S. D. (**1988**). ''Some effects of time-varying context on the perception of speech and nonspeech sounds,'' Res. on Speech Percept.: Prog. Report No. 14, Indiana University.

Náběelek, I. V., Náběelek, A. K., and Hirsh, I. J. (**1970**). ''Pitch of tone bursts of changing frequency,'' J. Acoust. Soc. Am. **48**, 536–553.

Náběelek, I. V., Náběelek, A. K., and Hirsh, I. J. (**1973**). ''Pitch of sound bursts with continuous or discontinuous change of frequency,'' J. Acoust. Soc. Am. **53**, 1305–1312.

Nearey, T. (**1989**). ''Static, dynamic, and relational properties in vowel perception,'' J. Acoust. Soc. Am. **85**, 2088–2113.

Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (**1981**). ''Speech perception without traditional speech cues,'' Science **212**, 947–950.

Repp, B. H. (**1982**). ''Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception,'' Psychol. Bull. **92**, 81–110.

Repp, B. H. (**1983**). ''Bidirectional contrast effects in the perception of VC–CV sequences,'' Percept. Psychophys. **33**, 147–155.

Ruggero, M. A. (**1992**). ''Physiology and coding of sound in the auditory nerve,'' in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay (Springer Verlag, New York), pp. 34–93.

Shigeno, S., and Fugisaki, H. (**1979**). ''Effect of a preceding anchor upon the categorical judgment of speech and nonspeech stimuli,'' Japanese Psych. Res. **21**, 165–173.

Sinnott, J. M., Brown, C. H., and Borneman, M. A. (**1998**). ''Effects of syllable duration on stop-glide identification in syllable-initial and syllable-final position by humans and monkeys,'' Percept. Psychophys. **60**, 1032–1043.

Smith, R. L. (**1979**). ''Adaptation, saturation, and physiological masking in single auditory-nerve fibers,'' J. Acoust. Soc. Am. **65**, 166–178.

Stevens, K. N., and Halle, M. (**1967**). ''Remarks on analysis by synthesis and distinctive features,'' in *Models for the Perception of Speech and Visual Form*, edited by W. Wathem-Dunn (MIT Press, Cambridge, MA).

Stevens, K. N., and House, A. S. (**1963**). ''Perturbations of vowel articulations by consonantal context: An acoustical study,'' J. Speech Hear. Res. **6**, 111–128.

Summerfield, Q., Haggard, M. P., Foster, J., and Gray, S. (**1984**). ''Perceiving vowels from uniform spectra: Phonetic exploration of an auditory aftereffect,'' Percept. Psychophys. **35**, 203–213.

Summerfield, Q., Sidwell, A., and Nelson, T. (**1987**). ''Auditory enhancement of changes in spectral amplitude,'' J. Acoust. Soc. Am. **81**, 700–707.

Warren, R. M. (**1985**). ''Criterion shift rule and perceptual homeostasis,'' Psychol. Rev. **92**, 574–584.

Williams, D. R. (**1986**). ''Role of dynamic information in the perception of coarticulated vowels,'' unpublished doctoral dissertation, University of Connecticut.

The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers[a)]

Armin Kohlrausch[b)] and Ralf Fassel[c)]
IPO-Center for User-System Interaction, P.O. Box 513, NL-5600 MB Eindhoven and Philips Research Laboratories Eindhoven, Prof. Holstlaan 4, NL-5656 AA Eindhoven, The Netherlands

Torsten Dau
Carl von Ossietzky Universität Oldenburg, AG Medizinische Physik, P.O. Box 2503, D-26111 Oldenburg, Germany

(Received 29 September 1997; revised 2 August 1999; accepted 31 March 2000)

This paper is concerned with modulation and beat detection for sinusoidal carriers. In the first experiment, temporal modulation transfer functions (TMTFs) were measured for carrier frequencies between 1 and 10 kHz. Modulation rates covered the range from 10 Hz to about the rate equaling the critical bandwidth at the carrier frequency. In experiment 2, TMTFs for three carrier frequencies were obtained as a function of the carrier level. In the final experiment, thresholds for the detection of either the lower or the upper modulation sideband (beat detection) were measured for ''carrier'' frequencies of 5 and 10 kHz, using the same range of modulation rates as in experiment 1. The TMTFs for carrier frequencies of 2 kHz and higher remained flat up to a modulation rate of about 100–130 Hz and had similar values across carrier frequencies. For higher rates, modulation thresholds initially increased and then decreased rapidly, reflecting the subjects' ability to resolve the sidebands spectrally. Detection thresholds generally improved with increasing carrier level, but large variations in the exact level dependence were observed, across subjects as well as across carrier frequencies. For beat rates up to about 70 Hz (at 5 kHz) and 100 Hz (at 10 kHz), beat detection thresholds were the same for the upper and the lower sidebands and were about 6 dB higher than the level per sideband at the modulation-detection threshold. At higher rates the threshold for both sidebands increased, but the increase was larger for the *lower* sideband. This reflects an asymmetry in masking with *more* masking towards lower frequencies. Only at rates well beyond the maximum of the TMTF did detection for the lower sideband start to be better than that for the upper sideband. The asymmetry at intermediate frequency separations can be explained by assuming that detection always takes place in filters centered above the stimulus spectrum. The shape of the TMTF and the beat-detection data reflects a limitation in resolving fast amplitude variations, which must occur central to the inner-ear filtering. Its characteristic resembles that of a first-order low-pass filter with a cutoff frequency of about 150 Hz. © *2000 Acoustical Society of America.* [S0001-4966(00)03407-X]

PACS numbers: 43.66.Mk, 43.66.Ba [JWH]

INTRODUCTION

This paper studies the influence of the modulation rate on the sensitivity for detecting amplitude modulation of sinusoidal carriers. Previous research with sinusoidal and with noise carriers has suggested that the dependence of modulation detection thresholds on modulation rate, the temporal modulation transfer function or TMTF, increases with increasing rate in a similar way. Both choices for the carrier type have certain advantages and disadvantages. A sinusoidal carrier allows the influence of carrier frequency, and thus the possible contribution of peripheral filtering on the shape of the TMTF, to be studied. On the other hand, it is not always possible to control which detection cues are used by the subjects in discriminating a modulated from an unmodulated sinusoid, especially at high modulation rates (see Strickland and Viemeister, 1997). Besides using the temporal cue associated with the periodic modulation, subjects may also perform the discrimination task on the basis of spectral cues. The spectrum of a sinusoidally amplitude-modulated sinusoid consists of the carrier and two sidebands with a spectral distance from the carrier that equals the modulation frequency. Above a certain modulation rate, the sidebands become audible as separate tones and the impression of modulation or roughness gradually vanishes.

The possible spectral cues associated with detecting modulation in sinusoids have led many investigators to use broadband noise as the carrier. For such a carrier signal the modulation sidebands cannot be perceived separately because they generally fall within the spectrum of the carrier and are therefore masked. A disadvantage of a broadband carrier is that the contribution of the individual auditory filter outputs to the TMTF cannot be inferred from such experiments. An additional problem lies in the fact that a *stochastic*

[a)]Part of the data in this paper has been presented at the 131st meeting of the Acoustical Society of America, Indianapolis, 1996.
[b)]Author to whom correspondence should be addressed; electronic mail: armin.kohlrausch@philips.com
[c)]Present address: Akustik Technologie Göttingen, Theaterstr. 10, D-37073 Göttingen, Germany.

carrier has intrinsic modulations, perhaps changing the task of modulation detection to a task of modulation masking, with the intrinsic fluctuations functioning as masker and the test modulation as signal. These intrinsic fluctuations become particularly relevant for narrow-band-noise carriers, for which not only the modulation detection thresholds increase, but also the shape of the TMTF changes from low pass to bandpass or even high pass (Fleischer, 1982; Dau *et al.*, 1997a, 1999).

Therefore, in order to obtain a good insight into modulation perception in the auditory system, both experiments with noise carriers and with sinusoidal carriers should be performed. A look into the literature reveals, however, that the vast majority of studies concerned with the TMTF used noise carriers. This is true for research with normal-hearing and in particular, for research with hearing-impaired listeners (e.g., Rodenburg, 1972, 1977; Viemeister, 1977, 1979; van Zanten, 1980; Bacon and Viemeister, 1985; Formby, 1986, 1987; Takahashi and Bacon, 1992; Bacon and Gleitman, 1992). In addition to the above-mentioned disadvantages of sinusoidal carriers, there might be other factors involved in this choice. Probably most relevant is the idea that both carrier types reveal basically the same temporal process, which is usually described by a low-pass function with a cutoff frequency around 50 Hz. This notion is supported by the seemingly quite similar shape of data obtained with sinusoids (e.g., Zwicker, 1952) and with broadband noise (e.g., Viemeister, 1979); it can be found in many places in the original literature and in review articles, sometimes more explicit, sometimes more implicit (e.g., Viemeister, 1979, p. 1369; Langner, 1992; Shannon, 1992). In some cases, however, differences in the thresholds for these two conditions are also acknowledged (e.g., Viemeister, 1979, p. 1369; Zwicker and Fastl, 1990, p. 157).

We argue that such a view does not consider several differences between noise and sinusoidal carriers. First of all, the limitations of detecting amplitude modulation are different in the two cases: For a broadband noise, it is the amount and the spectral distribution of intrinsic envelope fluctuations in the various auditory filters that probably limit detection (see, e.g., Dau *et al.*, 1997a,b; Kohlrausch *et al.*, 1997; Dau *et al.*, 1999). For a sinusoidal carrier, the limitation has to be internal noise. In addition, TMTFs for these two carrier types show a very different level dependence: While for broadband noise, modulation detection thresholds are level independent for carrier levels above about 20-dB sensation level, they decrease continuously with increasing levels for sinusoidal carriers (for a review on this topic, see Kohlrausch, 1993). This difference cannot be attributed solely to the fact that, for the narrow-band sinusoid, spread of excitation plays a role, because thresholds are also nearly level independent for narrow-band-noise carriers (see Maiwald, 1967; Dau *et al.*, 1997a).

The one-TMTF view is also not in line with more recent data on modulation detection for sinusoidal carriers. Most of these data show a shape quite different from the most often cited study by Zwicker (1952), as can be seen in Fig. 1. Fleischer (1982, 1983) published TMTFs for a 5-kHz gated sinusoid, which had a constant threshold for modulation rates

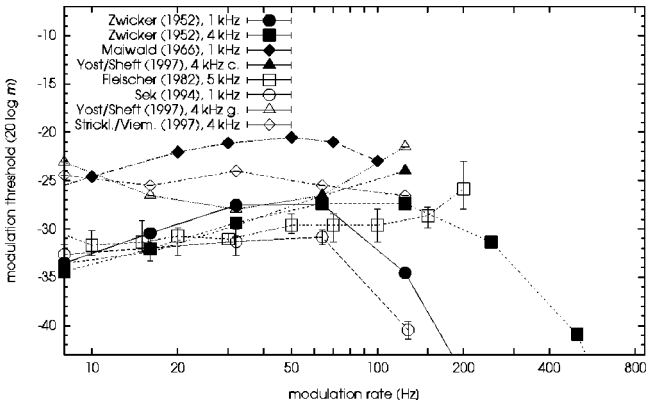


FIG. 1. Compilation of various modulation detection data for sinusoidal carriers at different carrier frequencies (see legend). The data from Strickland and Viemeister (Strickl./Viem.) are for the discrimination between AM and QFM. The carrier level for the data sets from Zwicker was 70 phon, for the data from Fleischer it was 80 dB, for Maiwald's data it was 60 dB, and for all other data it was 70 dB. Filled symbols indicate conditions with a continuous carrier, open symbols those with a gated carrier.

between 5 and about 100 Hz. Similarly, Sek (1994) found no increase in modulation thresholds between 4 and 64 Hz for a gated sinusoidal carrier at 1 kHz. In contrast, Zwicker's data increased with a slope of about 2 dB/doubling of the modulation rate at rates above 4 Hz. While the present study was in progress, additional data for a 4-kHz gated sinusoidal carrier were published by Strickland and Viemeister (1997) and Yost and Sheft (1997). Strickland and Viemeister measured the discriminability of amplitude modulation and quasi-frequency modulation (QFM) and found constant thresholds up to a modulation rate of 128 Hz. From additional data on QFM detection, the authors concluded that the flatness in thresholds between 64 and 128 Hz may have been caused by the use of nontemporal cues (see Dau *et al.*, 1997a, for a comment on this argument). But even the flatness of the TMTF up to a modulation rate of 64 Hz clearly disagrees with Zwicker's data. The same is true for the gated carrier data from Yost and Sheft (1997) (open triangles in Fig. 1). The only results in line with Zwicker's data are from Maiwald's thesis (Maiwald, 1966) which used a continuous carrier (filled diamonds in Fig. 1), and the continuous carrier data from the recent study by Yost and Sheft (1997) (filled triangles in Fig. 1).

Summarizing these data, it appears that gating of the carrier is one of the parameters influencing the shape of tonal TMTFs.[1] Imposing a gated modulation on a continuous carrier leads to a decrease in modulation detection thresholds towards low modulation rates, while for a gated carrier, modulation thresholds stay much more constant at rates between 8 and 64 Hz.

Based on this analysis of the literature and the problems in interpreting TMTFs for noise carriers, and given the sparse data for tonal TMTFs (at the point in time when we started this research), we see a clear need for more modulation experiments using sinusoidal carriers. We report here on two measurements of modulation-detection thresholds for such carriers. In the first experiment, TMTFs were measured as a function of the carrier frequency, and in the second experiment they were measured as a function of the carrier

level. These parameter variations were also studied in Zwicker's paper. In a third experiment, beat-detection (or single-sideband-detection) thresholds were measured for the lower and the upper sideband separately. By comparing single-sideband with modulation-detection data, we hoped to understand better the contribution of each sideband to the modulation thresholds.

I. METHODS

The time signal of a sinusoidally amplitude-modulated sinusoid is defined as follows:

$$s(t) = A(1 + m\cos(2\pi f_m t))\sin(2\pi f_c t). \quad (1)$$

Here, A indicates the overall amplitude, m is the degree of modulation taking values between 0 and 1, and f_m and f_c are the modulation rate and the carrier frequency, respectively. Modulation detection thresholds indicate the just noticeable value of m [often expressed as $20\log(m)$]. The spectrum of such a stimulus consists of three components: the carrier frequency and the two sidebands which are spectrally separated from the carrier by the modulation frequency. The level of the sidebands relative to the carrier level can be derived by subtracting 6 dB from $20\log(m)$. Thus, for 100% amplitude modulation ($m = 0$ dB), the sideband level is 6 dB lower than the carrier level.

Data were collected at IPO in Eindhoven and at the university of Oldenburg with slightly different setups, which are described in the following.

A. IPO setup

The thresholds were measured using an adaptive 3-IFC 2-down/1-up procedure which tracks the 71%-correct point of the psychometric function. The subject's task was to identify the one interval out of three in which the sinusoidal carrier was modulated. After two correct responses the modulation index (or the level of the sideband in the experiments described in Sec. II C) was decreased; after each incorrect response it was increased. At the start of the experiment the modulation index (or the level of the sideband) was varied in relatively large steps. After every two reversals this step size was halved until a final step size was reached which was then held constant. The threshold was calculated as the median of the following ten reversals. Begin and end step sizes were 0.04 and 0.005, respectively, for the modulation detection experiments, and 8 and 1 dB, respectively, for the sideband-detection experiment.[2]

In each interval the carrier had a duration of 800 ms including Hanning on- and off-ramps of 150 ms each. In the test interval the modulation was applied over the whole length of the carrier including the on/off ramps. The carrier frequency was randomly varied over a range of 3% in each interval. Carrier levels were in the range of 20 to 75 dB SPL. Between intervals a random level variation of ± 1.5 dB was applied in order to minimize cues that might arise from loudness changes due to the modulation. Both aspects of roving were applied in a corresponding way to the stimuli in the sideband-detection experiment. Due to the long ramps, the three intervals were clearly perceived separately and were presented without any additional intervening silent periods. No precautions were taken to mask possible distortion products. Measurements were repeated four times, and the individual results are given as median value and interquartile ranges of these four threshold values. Averages across subjects are represented as means and standard deviations across these medians.

All signals were generated digitally and played via the DA converter and antialiasing filter of the computer. Analog attenuators were used to set the overall levels. The stimuli were presented diotically via Beyer DT 880 M or DT 990 headphones. Three subjects participated in the experiments at IPO; all of them were experienced listeners in psychoacoustic tasks.

B. Oldenburg setup

The experiments in Oldenburg were performed with basically the same experimental procedure. We do not think that the slight differences between the two setups, which are listed below, have any systematic influence on the results:

(i) The modulation was applied only to the central 500 ms of the carrier and it was gated with a Hanning window over the total length of the modulation.
(ii) The pause between the intervals was 300 ms.
(iii) The modulation index m was varied in dB steps rather than in linear steps with begin and end values for the step size of 4 and 1 dB, respectively.
(iv) Stimuli were presented at 65 dB SPL carrier level via Sennheiser HDA 200 headphones.
(v) Carrier frequency and level were kept constant (no roving was applied).
(vi) Averages across subjects were computed as medians and interquartile ranges across all 12 individual threshold values.

Three subjects took part in the Oldenburg experiments. All subjects had experience in psychoacoustic measurements.

II. EXPERIMENTAL RESULTS

A. Modulation thresholds as a function of carrier frequency

The results of the modulation-detection experiments, obtained with the two experimental setups, are shown in Figs. 2 and 3. The modulation index m at threshold, expressed as $20\log m$, is plotted as a function of the modulation rate. All data are based on the results of three observers, with the exception of the 3-kHz IPO data (two subjects) and the 8-kHz IPO data (one subject).

The following common trends can be seen:

(i) For low modulation rates up to about 100 to 130 Hz and carrier frequencies above 1 kHz the modulation detection thresholds do not depend systematically on the modulation rate. Unsystematic variations of the thresholds are not greater than 2–3 dB at each individual carrier frequency.
(ii) With the exception of the data for the 1- and 2-kHz carriers, an initial *increase* of the thresholds at modu-

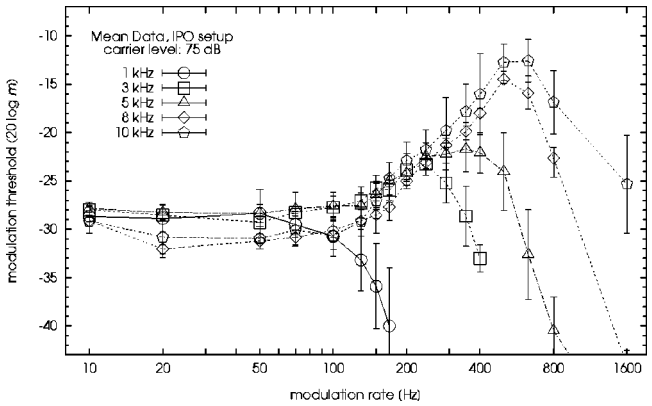


FIG. 2. Mean modulation-detection thresholds as a function of the modulation rate for different carrier frequencies (see legend, IPO setup).

lation rates above 100 to 130 Hz is observed. This increase is seen for those modulation rates for which sideband detection is, probably, not the prominent cue. In the region of increasing thresholds the curves follow a common slope, which varies across subjects from 5 to 8 dB/oct.

(iii) Threshold curves for the different carrier frequencies reach their maximum at different modulation rates. The modulation rate of the maximum shifts with increasing carrier frequency towards higher modulation rates. Also the maximum threshold value increases with increasing carrier frequency. These relations are shown in Fig. 4. The left panel shows the modulation rate derived from the individual threshold maxima as a function of the carrier frequency, while the corresponding threshold values are plotted in the right panel. Also included in the figure are data from the study by Zwicker (1952) for a carrier level of 70 phon.

(iv) For the highest modulation rates thresholds decrease with increasing rate. In this region it is likely that the lower sideband of the modulation becomes audible as separate tone as a result of the greater spectral separation between carrier and sidebands. The audibility of this tone determines thresholds and turns the modulation-detection task into a tone-on-tone masking experiment.

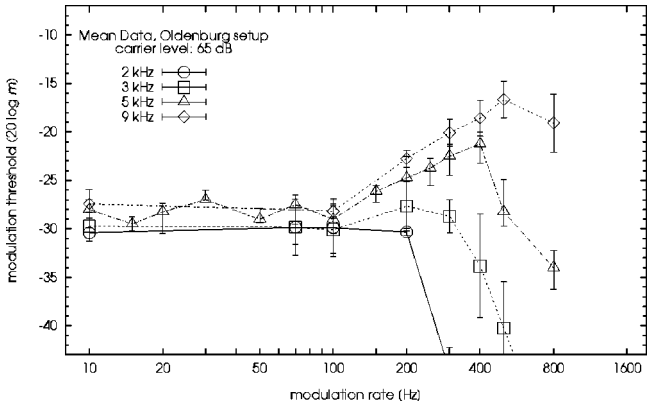


FIG. 3. Mean modulation-detection thresholds as a function of the modulation rate for different carrier frequencies (see legend, Oldenburg setup).

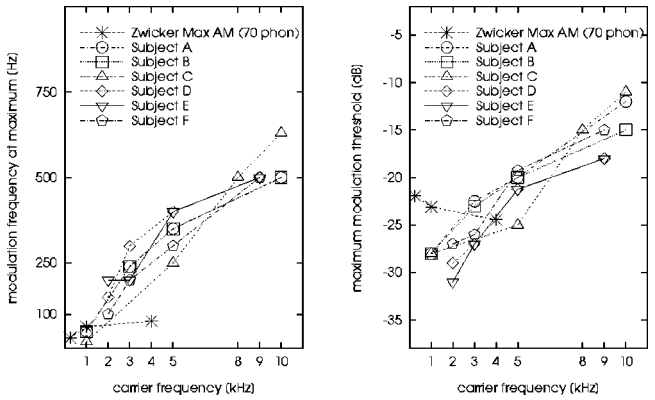


FIG. 4. The left panel shows the modulation rate for which the individual TMTFs at a certain carrier frequency reach their maximum. The curve labeled ''Zwicker'' (∗) indicates the maximum of the TMTF for a carrier level of 70 phon. The right panel shows for the same TMTF data the degree of modulation at the maximum in the TMTF.

B. Modulation thresholds as a function of the carrier level

In the following experiment, modulation-detection thresholds were determined as a function of the carrier level at carrier frequencies of 1, 5, and 10 kHz. At each carrier frequency, two of the three subjects from the IPO group participated. Since the applied carrier levels differed between subjects and due to differences in their absolute threshold values, we prefer to show here individual data instead of averages across subjects.

Figure 5 shows the results for a carrier frequency of 1 kHz. In general, thresholds decrease with increasing carrier level. In both data sets in Fig. 5, a transition of the shape of the threshold curves can be seen as the carrier level changes. For carrier level curves of 45 dB and above in the lower panel and 60 dB and above in the upper panel, thresholds decrease for modulation rates above 70 Hz. At lower levels an initial *increase* in thresholds for these modulation rates is observed (similar to that seen at higher carrier frequencies; see Figs. 2 and 3), and a local maximum of thresholds occurs around a modulation rate of 150 Hz. For the highest modulation rates thresholds decrease again, and detection probably is limited by the absolute thresholds of one or both resolved sidebands. As an example, the modulation thresholds in the lower panel of Fig. 5 at a rate of 400 Hz increase by 25 dB for a decrease in carrier level from 45 to 20 dB. Thus, the absolute level of the sidebands at modulation threshold is the same at both carrier levels.

At a carrier frequency of 5 kHz (Fig. 6) the shapes of the threshold curves are roughly the same at all levels. For modulation rates greater than 100 Hz thresholds increase towards a maximum at about 300–400 Hz and decrease again for higher modulation rates. For a carrier level of 30 dB (upper panel) thresholds increase constantly by ca. 6 dB between 20- and 130-Hz modulation rate while they remain flat within 2 dB in this region for the higher carrier levels. Thresholds decrease with increasing carrier level. For subject A (upper panel), this decrease is strongest at *low* carrier levels and vanishes at high carrier levels (approximately 5 dB from 30 to 45 dB, but similar thresholds at 60 and 75 dB).

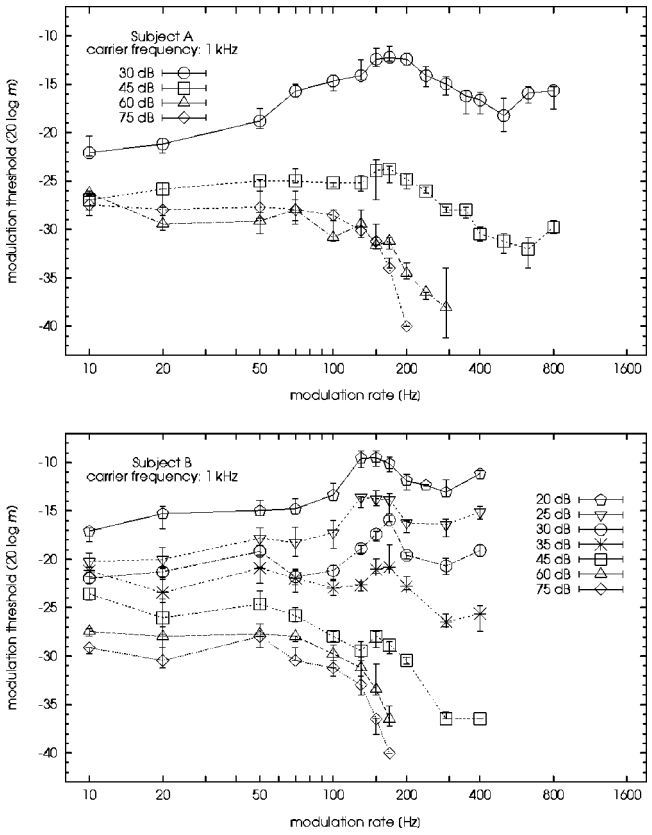


FIG. 5. Modulation-detection thresholds at different carrier levels (see legend) at a carrier frequency of 1 kHz. The two panels show results for two different subjects. The absolute thresholds at the carrier frequency were 14 dB for subject A and 4 dB for subject B.

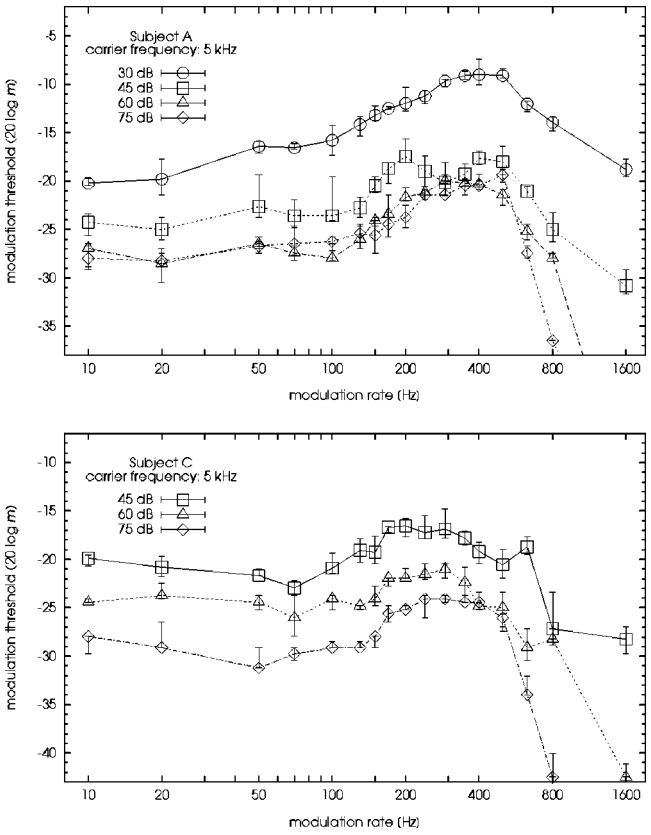


FIG. 6. Modulation-detection thresholds at different carrier levels (see legend) at a carrier frequency of 5 kHz. The two panels show results for two different subjects. The absolute thresholds at the carrier frequency were 15 dB for subject A and 17 dB for subject C.

For the other subject (lower panel), who was tested over a smaller level range, the decrease in threshold amounts to about 4 dB for each 15-dB increase in level.

At a carrier frequency of 10 kHz (Fig. 7) the overall shape of the TMTF is similar at all levels. An exception is again the lowest carrier level, where thresholds tend to increase with a constant rate already below 100-Hz modulation rate. The decrease in thresholds with increasing carrier level is largest at *high* carrier levels for both subjects (upper panel: approximately 8 dB, lower panel: approximately 4 dB for a 15-dB change in carrier level from 60 to 75 dB for modulation rates less than 100 Hz). The change of thresholds with carrier level is greatest at modulation rates less than 130 Hz and least at the maximum of the threshold curves around 500–600 Hz.

C. Detection of single sidebands

In order to investigate the extent to which each sideband of the modulated signal contributes to the modulation-detection threshold, a tone-on-tone masking experiment was performed. The thresholds of the lower and the upper sidebands were measured separately in the presence of the carrier as masker. Measurements for a carrier frequency of 10 kHz and a carrier level of 75 dB SPL were performed by two subjects (IPO setup). In addition, data for a carrier frequency of 5 kHz and 65-dB carrier level were obtained for three subjects using the Oldenburg setup. As with AM detection, detection in the tone-on-tone masking experiment can be based on either envelope modulation or spectral resolution, depending on the spectral distance Δf.

1. Experimental results

Mean experimental results are shown in Figs. 8 and 9. Sideband level at threshold, relative to the carrier level, is shown as a function of the frequency difference Δf. In addition, the modulation detection thresholds are included in each panel. Modulation depth at threshold was transformed into signal level *per sideband* relative to the carrier level, which implies a -6-dB vertical shift of the data relative to the values shown in the TMTF figures. In the case of the modulation-detection condition, the abscissa indicates the modulation rate.

At 10 kHz (Fig. 8), thresholds for the lower and upper sidebands are similar up to about 100 Hz. They are approximately 6 dB higher than the corresponding modulation detection thresholds, indicating that the modulation depths of the AM and the beat conditions are approximately equal at detection threshold. Thresholds are flat in this region of frequency separation. At medium spectral distances (between 100 Hz and 400 or 500 Hz), all threshold values show an increase with increasing spectral distance which is steepest for the lower sideband and shallowest for the upper sideband. Thus, thresholds for the *upper* sideband are substantially lower than those of the lower sideband, with differences being as large as 8 dB in this region. At the largest

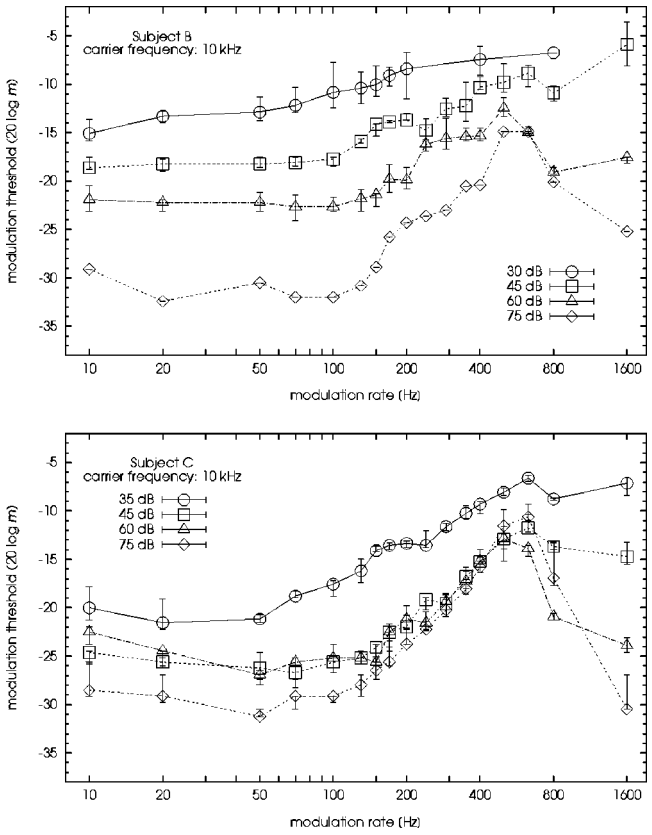


FIG. 7. Modulation-detection thresholds at different carrier levels (see legend) at a carrier frequency of 10 kHz. The two panels show results for two different subjects. The absolute thresholds at the carrier frequency were 15 dB for subject B and 23 dB for subject C.

spectral distance of 1600 Hz the situation is reversed and the *lower* sideband threshold is lower than that for the upper sideband. We had expected that the lower sideband threshold is very close to the modulation detection threshold, as it is the case at 5 kHz (see Fig. 9). While this is indeed true for one of the two subjects, the other measured a rather high modulation-detection threshold at 1600 Hz, so that in the average data, the modulation threshold at 1600 Hz lies in between the thresholds for the upper and the lower sideband.

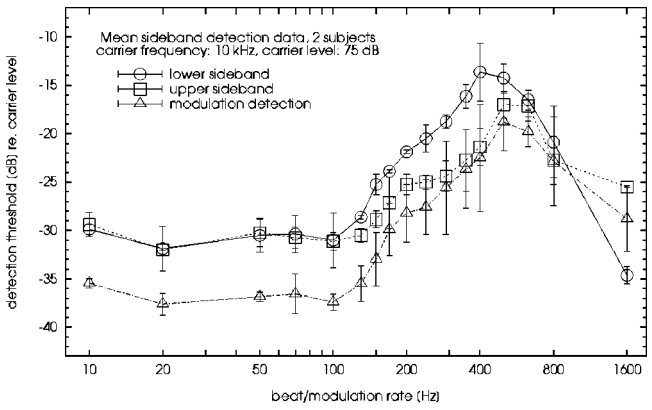


FIG. 8. Thresholds for detecting the lower (○) or the upper (□) modulation sideband in the presence of a 10-kHz carrier with a level of 75 dB. Thresholds are indicated relative to the carrier level. For comparison, the TMTF for the same carrier is indicated by the triangles (△). Modulation-detection thresholds are expressed as level per sideband.

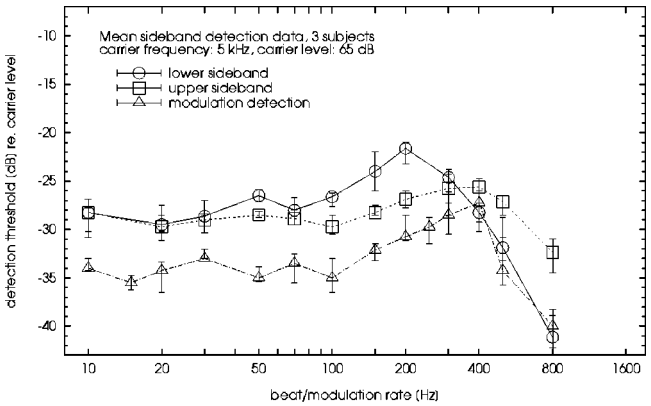


FIG. 9. Thresholds for detecting the lower (○) or the upper (□) modulation sideband in the presence of a 5-kHz carrier with a level of 65 dB. Thresholds are indicated relative to the carrier level. For comparison, the TMTF for the same carrier is indicated by the triangles (△). Modulation-detection thresholds are expressed as level per sideband.

At 5 kHz (Fig. 9) thresholds for the upper and lower sideband are very similar for Δf values up to about 70 Hz. The offset between the modulation detection data and the sideband detection data in this range is approximately 6 dB. For frequency differences between 100 and 300 Hz the upper sideband thresholds are lower than those for the lower sideband. For the highest frequency differences, the modulation detection curve coincides with the lower sideband threshold.

2. Conclusions

The single-sideband detection thresholds are qualitatively similar for the 5- and 10-kHz carriers. The curves differ in the height and the beat frequency of the threshold maximum which again seems to be related to the characteristics of the peripheral filter at the carrier frequency. They also differ in the modulation rate up to which upper and lower sideband are equally detectable.

In order to reach the same modulation depth in an AM and a single-sideband stimulus, the level-per-sideband for the latter has to be 6 dB higher than for the AM stimulus. The observed threshold difference of 6 dB between these two conditions at small spectral distances can therefore be taken as strong evidence that envelope modulation was the detection cue.

An unexpected finding which has so far only been mentioned explicitly by Hartmann and Hnath (1982), is that, at intermediate spectral distances, thresholds for the upper sideband are lower than for the lower sideband. A careful inspection of related studies revealed, however, that this asymmetry has been measured in other data sets as well. Examples are the study by Zwicker and Jaroszewski (1982), where for masker frequencies of 4 and 6.03 kHz higher thresholds for the lower sideband at frequency separations of a few hundred Hz are shown (Figs. 4 and 5 in their paper). In the study by Moore *et al.* (1998) with a masker frequency of 1 kHz, one subject (Fig. 1 in their paper) showed the same asymmetry for a spectral separation between masker and signal of 50 Hz. Hartmann and Hnath (1982), who also measured at 1 kHz, stated that all three subjects showed a downward spread

of masking (i.e., higher thresholds for the lower sideband) for a frequency separation of 25 Hz.

The coincidence of the lower-sideband thresholds and the modulation-detection thresholds at high modulation frequencies indicates that, here, modulation detection thresholds are determined by the masked threshold of the spectrally resolved lower sideband. This observation is in line with the literature (see, e.g., Hartmann and Hnath, 1982).

III. DISCUSSION

A. Modulation detection

Comparing our results with the most systematic other study available (Zwicker, 1952), we see a major difference in the shape of the TMTF: while the curves presented by Zwicker show a monotonic increase of thresholds between the 8- and 64-Hz modulation rates by approximately 6 dB, we observed a *flat* threshold function between the 10-Hz and 100- to 130-Hz modulation rates, with unsystematic variations of typically less than 2.5 dB. In addition, the rate ($f_{\max}$) and threshold values ($m_{\max}$) at the maximum of the TMTF show a frequency dependence which differs from that in Zwicker's data (cf. Fig. 4). In our data, $f_{\max}$ increases with the center frequency and has a value of approximately 5% of the carrier frequency. In Zwicker's data, $f_{\max}$ is nearly constant between 1 and 4 kHz. Even stronger is the discrepancy for $m_{\max}$, which, in Zwicker's data, is slightly lower at 4 kHz than at 1 kHz. In our data, $m_{\max}$ has the same value at 1 and 2 kHz and increases above 2 kHz with a slope of nearly 6 dB per octave. In Zwicker's data [as in the data from Yost and Sheft (1997) for a continuous carrier] there is some trend for decreasing modulation thresholds with increasing carrier frequency. This effect is, however, limited to a few dB and we did not observe such a systematic trend in our data.

One obvious difference between the two studies is the threshold estimation procedure, which in our case was an adaptive 3IFC procedure. Zwicker obtained his data with a kind of Békésy tracking procedure. This, however, does not seem a good candidate to explain the differences in the shape of the TMTF. Fleischer (1982) used a similar procedure as Zwicker, in which thresholds were defined as the degree of modulation, at which the subjects reported in 50% of the presentations to hear a difference between a modulated and an unmodulated signal. Nevertheless, his TMTF for a 5-kHz carrier was flat up to 100 Hz and thus resembles the slope we see in our data. Other recent data by Sek (1994) for 1-kHz carriers, by Strickland and Viemeister (1997) for the discrimination between AM and QFM for a 4-kHz carrier, and the data by Yost and Sheft (1997) for a gated carrier also did not replicate the increase seen in Zwicker's data.

As mentioned in the Introduction, another difference between the various sets of tonal TMTFs is the gating of the carrier. It is known that for noise carriers the shape of the low-frequency part of the TMTF differs for continuous and for gated carriers (e.g., Viemeister, 1979; Sheft and Yost, 1990), but no satisfactory explanation has been found so far. All studies which showed a rather flat tonal TMTF between about 8 and 64 Hz impressed the modulation over the full duration of a gated carrier. The only exceptions to this procedure are the data from Zwicker (1952) and Maiwald (1966) and the data for a continuous carrier published by Yost and Sheft (1997). All these data show a continuous increase of modulation thresholds at low modulation rates (see Fig. 1). It is still unclear to what extent these differences are caused by gating the carrier (which has been discussed in relation to adaptation effects) or whether thresholds are influenced by the audibility of the modulation *onset*. Modulation is typically gated rectangularly (Yost and Sheft, 1997) and this introduces spectral splatter in the envelope-frequency domain. This envelope splatter might be much easier to detect on a continuous carrier than on a carrier gated synchronuously with the modulation. If this idea was correct, measurements for gated and continuous carriers should not be expected to show the same shape of the TMTF and one would predict that TMTFs for continuous carriers are influenced by the duration of the modulator onsets and offsets.

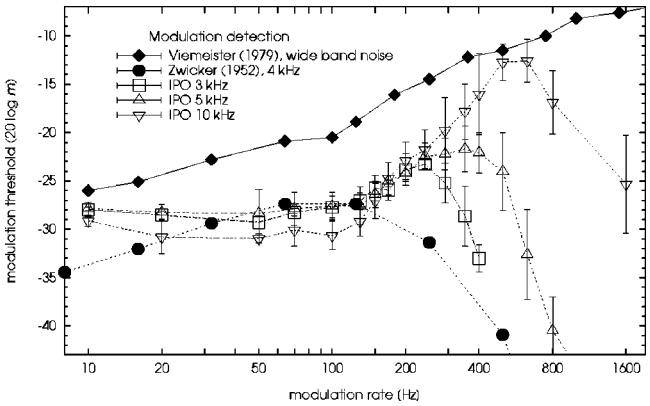


FIG. 10. Comparison of TMTFs obtained with broadband noise (filled diamonds, Viemeister, 1979) and for a 4-kHz sinusoidal carrier at 70 phon (filled circles, Zwicker, 1952) with data from the present study for sinusoidal carriers of 3, 5, and 10 kHz (open symbols).

The results further show that the initial increase in modulation thresholds for sinusoidal carriers and wideband noise carriers differs. Figure 10 compares data for broadband noise (Viemeister, 1979) with data for a 4-kHz sinusoidal carrier from Zwicker (1952) and for carriers at 3, 5, and 10 kHz from the present study. As mentioned before, the tonal data from Zwicker increase up to about 64 Hz with the same slope as the noise data. Compared with the 10-kHz data from the present study, the noise TMTF has a much steeper slope below about 100 Hz and a clearly shallower slope above 100 Hz.

In the following we discuss which part of the tonal TMTF can be attributed to effects of peripheral filtering. Because peripheral filters have a bandpass characteristic, their modulation transfer function is low pass. One might therefore assume that the initial increase in the measured modulation thresholds is caused by the attenuation of high modulation frequencies in the peripheral filters. However, from the results of the modulation detection experiments at high carrier frequencies we conclude that peripheral filtering alone cannot explain the initial increase in modulation thresholds, a conclusion which was also reached earlier by Zwicker (1956) and Viemeister (1979). The bandwidths of peripheral filters

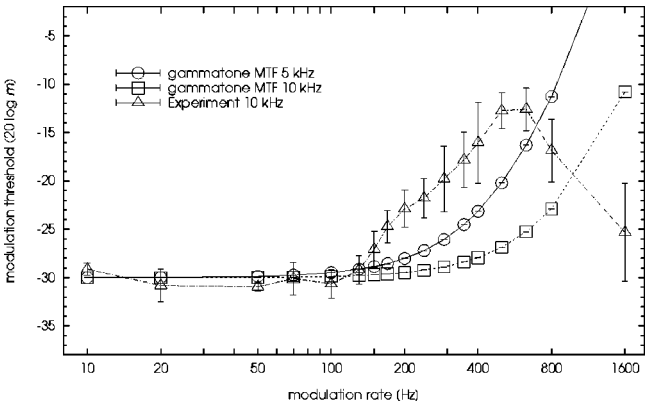


FIG. 11. Comparison of the TMTF obtained at 10 kHz (triangles) and modulation transfer functions of Gammatone filters at 5 (circles) and 10 kHz (squares).

generally increase with increasing carrier frequency. If the observed increase in modulation detection thresholds was due to peripheral filtering, the increasing part of the curves for different carrier frequencies should not coincide but rather should show a horizontal shift on the modulation rate scale. This can be seen by calculating the TMTF for auditory filters using gammatone filters. Figure 11 shows the TMTFs for the 10-kHz carrier averaged across the three subjects together with the modulation transfer functions of the gammatone filters at 5 and 10 kHz center frequency. These transfer functions were computed by filtering a sinusoidally amplitude-modulated 5- or 10-kHz sinusoid by the corresponding gammatone filter. Using a modulation depth of -30 dB for the input stimulus, the modulation depth of the filtered signal (ignoring the filter response to the carrier onset) was extracted and converted to a dB value. It can be seen that the filters allow a much better transformation of high modulation rates than is observed psychophysically.

Filtering does, however, contribute to the shape of the TMTFs in the region of the maximum and beyond. Here, subjects have not only temporal cues, but also spectral cues available for discriminating a modulated from an unmodulated sinusoid. The usability of spectral cues depends on the width (and shape) of the auditory filters at and around the carrier frequency. We can generally state that, for a fixed carrier frequency, the role of spectral cues will increase with the modulation rate, while for a fixed modulation rate, it will decrease with increasing carrier frequency.

In summary, this implies the following characterization of the underlying temporal processes. The increase in modulation thresholds above 100 to 130 Hz reflects a limitation in resolving fast amplitude variations, which must occur central to the inner-ear filtering. This limitation is independent of the carrier frequency and its characteristic resembles that of a first-order low-pass filter with a cutoff frequency of about 150 Hz. The characteristics of the temporal limitation can be seen best by using high-frequency sinusoidal carriers.

At the lowest carrier frequency included in our measurements (1 kHz), spectral cues start to dominate detection already at modulation rates below 100 Hz and therefore one cannot see the initial increase in modulation thresholds. But when the usability of spectral cues is reduced due to a low carrier level, the effect of the proposed temporal process can be seen even at this low carrier frequency (see Fig. 5).

This description of a low-pass characteristic of tonal TMTFs is not meant to be in contrast with the modulation filterbank model derived in Dau *et al.* (1997a,b), but it rather forms an extension. For sinusoidal carriers, this model does not predict any change in modulation detection thresholds with modulation rate, except for very low and very high rates (Dau, 1996). At low rates below about 10 Hz, the adaptation stage causes modulation thresholds to increase with decreasing rate. At high rates, the modulation transfer function of the peripheral auditory filters comes into play (see Fig. 11). A possible extension to explain the tonal TMTFs could comprise a low-pass filter in combination with the filterbank. Such a low-pass filter would not affect the modulation threshold for noise carriers, because they depend on the relation between test modulation and *intrinsic* fluctuations of the noise carrier, and both would be attenuated in the same way by the low-pass filter. For sinusoidal carriers, on the other hand, the modulation depth would be attenuated *relative to the level of the internal noise* added after the modulation filters and the predicted modulation thresholds would increase with a slope given by the low-pass filter. Such an extension of the filterbank model has been investigated by Ewert and Dau (2000).

B. Carrier level

The results obtained by varying the carrier level agree with many earlier studies on this issue which have revealed a continuous decrease in modulation detection thresholds with increasing carrier level (for an overview, see Kohlrausch, 1993). For instance, the data published by Zwicker (1952) show an average decrease in modulation threshold of 6 dB for a 30-dB increase in the carrier level. The TMTFs obtained at different carrier levels run parallel over a wide range of modulation rates. Usually, two effects are mentioned that can contribute to the level dependence. First of all, with increasing level, there is spread of excitation across an increasing number of auditory filters. Assuming that the limitation for detecting modulation is internal noise, and assuming uncorrelated noises in the various auditory filters, detection thresholds will decrease with increasing level. The second argument is the nonlinear growth of the upward spread of excitation. This argument is usually explained in the following way: at medium and high levels, the excitation of a sinusoid grows expansively in the spectral region above its frequency. Therefore, for a modulated sinusoid, the depth of modulation will here be larger than in the acoustic input signal. By analyzing auditory filters in the upward-spread region, lower detection thresholds are possible than would be possible without the nonlinear upward spread.[3]

One interesting and new aspect in our data can be observed at the lowest levels. Here, the internal noise producing absolute threshold influences the shape of the TMTF in a characteristic way. At 1 kHz, it hampers the detection of the sidebands as separate tones. As long as the sidebands are processed within the same auditory filter as the carrier, they lead to a certain amount of envelope fluctuations within this filter. They do this, although their individual levels are

clearly below their own absolute threshold. For subject A in Fig. 5, the absolute threshold at 1 kHz is 14 dB SPL and we can assume similar values for nearby frequencies. At a modulation rate of 10 Hz and a carrier level of 30 dB, the level per sideband at modulation threshold ($m = -22$ dB, which corresponds to a level per sideband relative to the carrier level of -28 dB) corresponds to only 2 dB SPL, about 10 dB below the assumed absolute threshold. Above a certain modulation rate, the sidebands can be resolved spectrally. At high carrier levels, this effect leads to a decrease in modulation thresholds above about 70 Hz, because spectral cues dominate over temporal modulation cues. At low levels, this is no longer true. Temporal cues seem to influence the TMTF at 1 kHz up to a much higher modulation rate, because the spectral cues are not effective. Only when the temporal modulation cues have become sufficiently weak, the detection of the resolved sidebands determines the modulation threshold. In the above example, this occurs at modulation rates above 150 Hz.

The increase in modulation thresholds between 10 and about 100 Hz at the lowest levels might be caused by the following effect. The auditory filters at 1 kHz are so narrow that they attenuate higher modulation rates. Assuming an ERB value of 100 Hz at low levels, the modulation transfer function of the 1-kHz filter will already be 3 dB down at 44 Hz. Since, as described above, spectral cues do not contribute to detection at the lowest level for modulation rates below about 130 to 150 Hz, thresholds have to increase with increasing modulation rate.

C. Single sideband detection

The influence of the assumed central limitation of temporal resolution also agrees with the results for single sideband detection. For small frequency differences between ''carrier'' and sideband, detection is based on the detection of envelope fluctuations: both the lower and the upper sidebands are equally detectable and the degree of modulation at threshold is the same as for AM (Figs. 8 and 9). This equality holds at 5 kHz up to about 70 Hz and at 10 kHz up to about 100 Hz.

For higher modulation rates, thresholds first increase as they do for AM, before they finally drop, because the sidebands are resolved. The increase for the *lower* sideband has a slope similar to, but somewhat steeper than, that for AM, while the slope for the upper sideband is shallower. Thus in this range of increasing beat-detection thresholds, the upper sideband is *easier* to detect than the lower. In the following we will discuss possible mechanisms for this asymmetry.

The difference could be caused by using the increased effective modulation depth in a filter placed away from the carrier frequency (cf. Goldstein, 1967; Hartmann and Hnath, 1982; Dau *et al.*, 1997a). In order for such a mechanism to lead to threshold differences between lower and upper sidebands, the auditory filters have to be asymmetric. It is generally assumed that, at higher levels such as those used for the single-sideband measurements, the auditory filter has a steep high-frequency and a shallow low-frequency slope. A graphical example of how to apply such asymmetric filters to single-sideband detection and to AM detection is provided by Hartmann and Hnath (1982) in the top panels of their Fig. 3. With such an asymmetry, a stronger increase in the effective modulation depth will be reached for the lower sideband. In the filter centered on the lower sideband, the carrier will be more strongly attenuated—and thus, the modulation depth will be increased by a greater amount—than in a filter centered on the upper sideband. That is, if detection were influenced by the gain in effective modulation depth at a filter placed away from the carrier, thresholds should be lower for the lower sideband, in contrast to the data.

If detection was based on spectral resolvability, this cue should become available already at a smaller frequency separation for the lower than for the upper sideband. After all, this is the accepted source for the asymmetry in tonal masking patterns, which show more masking (higher thresholds) towards high than towards low frequencies. This asymmetry, however, is only seen in our data at the largest spectral separation of 800 Hz for a 5-kHz carrier and 1600 Hz for a 10-kHz carrier.

An explanation that is in line with the experimentally observed spectral asymmetry assumes that detection of single sidebands always takes place in filters centered at frequencies *above* the stimulus spectrum. This view takes into account that the number of auditory filters in which the modulation can be detected is larger above the stimulus spectrum than below. As discussed earlier in Dau *et al.* (1997a), the limitation for detecting amplitude modulation of a sinusoidal carrier must be sought in ''internal noise.'' If this internal noise is uncorrelated in the various auditory filters, its influence will decrease as more auditory filters are analyzed for the presence of the modulation which occurs coherently in all excited filters.

For the following argument, we assume linear auditory filters with an asymmetric shape as used by Goldstein (1967) and Hartmann and Hnath (1982). For mathematical simplicity, we also assume a constant filter slope on a linear frequency scale.[4] For detection of the upper sideband, the effective degree of modulation in a filter centered on the sideband will be higher than the degree of the externally applied modulation, because the carrier is somewhat attenuated relative to the sideband. The same degree of modulation is present in all filters above the upper sideband. Since these filters have a shallow low-frequency slope, many of these filters will be excited and can contribute to detection of the modulation. It should be noted that the increase in the effective degree of modulation will be larger for larger spectral separations between carrier and upper sideband.

The situation for the lower sideband is as follows: Filters centered at and below the lower sideband will have a higher degree of modulation than the acoustic stimulus. However, since the high-frequency slope of the filters below the stimulus spectrum is very steep, there are only a few filters that are excited and can contribute to detection of the modulation. In filters *above* the stimulus spectrum, the effective degree of modulation will be smaller than in the acoustic stimulus, due to the stronger attenuation of the sideband relative to the carrier. This decrease in the effective degree of modulation will be stronger for larger separations of sideband and carrier.

TABLE I. Effect of a filter slope of 1 dB/100 Hz on the effective degree of modulation for detection of the lower (low) or upper (up) sideband. The second and third columns indicate the change in degree of modulation relative to the input modulation. The fourth column gives the predicted difference in threshold for the two conditions and the last column gives the experimentally observed difference at 10 kHz. All entries are in dB.

Δf (Hz)	Lower	Upper	Low−Up (predicted)	Low−Up (observed)
100	−1.0	1.0	2.0	0.1
130	−1.3	1.3	2.6	1.9
150	−1.5	1.5	3.0	3.6
170	−1.7	1.7	3.4	3.2
200	−2.0	2.0	4.0	3.4
240	−2.4	2.4	4.8	4.5
290	−2.9	2.9	5.8	5.6
350	−3.5	3.5	7.0	6.7
400	−4.0	4.0	8.0	7.8

How strong this effective degree of modulation differs between lower and upper sideband depends on the slope of the auditory filter as well as on the frequency separation in the two-tone stimulus. The following example assumes a lower slope of the auditory filter of 1 dB/100 Hz. That is, for a frequency separation of 100 Hz between carrier and sideband, the effective degree of modulation in filters above the center frequency will be 1 dB decreased for the lower sideband and 1 dB increased for the upper sideband, both relative to the input degree of modulation. Under these assumptions the asymmetry (in dB) increases proportional to the spectral separation between carrier and sidebands. For the same spectral separation, the asymmetry will be smaller at a frequency of 10 kHz than at 5 kHz, because the filter slopes are shallower (on a linear frequency axis) at the higher frequency.

In Table I, we give some quantitative results based on this simple scheme. We show how strongly the degree of modulation is changed in a filter above the carrier. The assumed filter slope was 1 dB/100 Hz. The first column indicates the frequency separation between the carrier and the sideband. The next two columns give, for the lower and the upper sideband, the change in modulation depth relative to the input, where positive numbers indicate an increase in modulation depth. The next column computes the difference between the values from the previous two columns which indicates the *predicted* threshold difference for the two sidebands. The final column gives the *observed* threshold difference for a carrier frequency of 10 kHz. As can be seen, this extremely simplified scheme is able to describe the observed threshold asymmetry with an accuracy of 1 dB (except for $\Delta f = 10$ Hz).

In the following we will show that this scheme also predicts that the threshold increase for AM is different from that for the two sidebands. In Table II we show how strongly the degree of modulation is changed for AM in filters centered above the AM spectrum. These values were computed on the basis of the sideband amplitude values, ignoring possible phase changes between the components. The net effect of the filter asymmetry on modulation depth is smaller for AM than for either sideband. An increase in relative level for the upper sideband is partly compensated by a decrease in relative level for the lower sideband, so that even at 400-Hz

TABLE II. Effect of a filter slope of 1 dB/100 Hz on the effective degree of modulation for AM detection. The second column indicates the change in degree of modulation relative to the input modulation. The third column gives the predicted difference in threshold for the detection of the upper sideband and for detection of AM, and the last column gives the experimentally observed difference between these two conditions at 10 kHz. All entries are in dB.

AM rate (Hz)	Change in m	Up−AM (predicted)	Up−AM (observed)
10	0.00	5.9	6.0
50	0.01	5.5	6.6
100	0.06	5.0	6.3
130	0.10	4.8	5.0
150	0.13	4.6	4.1
170	0.17	4.5	2.8
200	0.23	4.2	2.9
240	0.33	3.9	2.6
290	0.48	3.6	1.2
350	0.7	3.2	0.9
400	0.9	2.9	1.1

modulation rate, the effective degree of modulation is changed by less than 1 dB. In the third column of this table, we have computed the predicted difference between upper-sideband and AM detection thresholds (expressed as level per sideband), taking only the attenuation by the filter slope into account. At very low rates the difference approaches 6 dB, because both sidebands contribute equally to AM detection. With increasing separation, the contribution of the upper sideband to AM detection will become more and more important, which shows up in an approach of the two thresholds. The final column in the table gives the observed difference for the two conditions. The trend is as predicted, but the decrease in the data is somewhat stronger than in the prediction. This could be caused by an additional effect of phase changes in the AM complex at the largest modulation rates. Phase shifts will change AM into a mixture of AM and FM and further reduce the modulation depth at the output of the auditory filter, but they do not affect the modulation depth for a two-tone stimulus.

In summary, this simple scheme is qualitatively in line with the following experimental observations:

(i) Thresholds for the lower sideband are higher than those for the upper sideband for frequency separations above 70 to 100 Hz.
(ii) The threshold difference increases with increasing frequency separation between carrier and sideband.
(iii) The asymmetry becomes visible at a smaller frequency separation at 5 kHz compared to 10 kHz. This is expected from the decrease in auditory filter slopes (on a linear frequency scale) with increasing center frequency.
(iv) The increase in threshold for the lower and the upper sideband differs from the slope observed in the TMTF.

In addition, this scheme makes predictions that can be tested in future experiments. For instance, if the region of upward spread was masked by a high-pass noise, this should not only lead to increased single sideband thresholds in a similar way as observed for modulation detection thresholds

(e.g., Zwicker, 1956), but it should also reduce the asymmetry. In a similar way, the asymmetry should be reduced if the carrier level was reduced, because auditory filters are thought to be more symmetric at low levels. We hope that such future experiments will help to further clarify our understanding of modulation and beat detection thresholds for sinusoidal carriers.

D. The possible role of distortion products

As briefly mentioned in Sec. I, we did not add any noise to mask possible distortion products. We think that the main results summarized below are not influenced by detection cues based on distortion products. The condition where distortion products could contribute strongest to detectability is the single-sideband detection experiment for the upper sideband. In a recent paper, Moore *et al.* (1998) investigated this condition for a carrier frequency of 1 kHz. Based on a variety of experimental conditions, they concluded that temporal beats between masker and signal were a stronger detection cue than distortion products for frequency differences up to 250 Hz. In our experiments, we used masker frequencies of 5 and 10 kHz and here, this statement should apply up to even higher frequency separations, because the optimal separation for generating a cubic difference tone is a constant proportion of the center frequency. Also the initial increase in the single-sideband thresholds up to 200 (at 5 kHz) or 400 Hz (at 10 kHz) spectral separation is not in line with the idea that thresholds are *predominantly* determined by distortion products. We conclude the same for the amplitude-modulation condition in the region of increasing modulation thresholds, because the thresholds for this condition (if expressed as level per sideband) approach the upper-sideband thresholds, but are never clearly lower.

IV. SUMMARY

The results of this study can be summarized as follows:

(i) The TMTFs for gated sinusoidal carriers show a different dependence on modulation rate than TMTFs for broadband noise carriers.

(ii) With the exception of very low carrier levels, TMTFs for carrier frequencies at and above 2 kHz are flat for rates between 10 and about 100 to 130 Hz and have similar values across carrier frequencies. For these carrier levels and carrier frequencies above 2 kHz, modulation thresholds first *increase* with a slope of 5 to 8 dB/oct above 100 to 130 Hz, before decreasing due to the detection of resolved sidebands.

(iii) Modulation-detection thresholds generally decrease with increasing carrier level. The details of the level dependence differ across subjects and carrier frequencies. At very low levels, where spectral cues cannot contribute to detection, the initial increase in modulation thresholds with modulation rate can also be observed for a carrier frequency of 1 kHz.

(iv) When presented separately, both the lower and the upper sidebands are equally detectable for frequency separations up to about 70 Hz at 5 kHz and up to about 100 Hz at 10 kHz. For these ranges, single-sideband thresholds are 6 dB higher than the level per sideband in modulation detection. At larger frequency separations, single-sideband thresholds at 5 and 10 kHz first increase, before decreasing due to the detection of a resolved sideband. In the range of increasing thresholds, the upper sideband has a lower threshold than the lower sideband.

(v) The different slopes in the range of increasing detection thresholds for the single-sideband conditions and the AM condition are in line with a processing scheme, in which detection for all three conditions is based on the temporal fluctuations in filters centered above the stimulus spectrum. If this influence of peripheral filtering is taken into account, TMTFs and beat-detection thresholds reflect the same central limitation in resolving fast envelope fluctuations. Its characteristic resembles that of a first-order low-pass filter with a cutoff frequency of about 150 Hz.

ACKNOWLEDGMENTS

We thank Andrew Oxenham and the two anonymous reviewers for helpful comments on earlier versions of this manuscript.

[1]Zwicker does not give details of the duration of carrier and modulation. He mentions that, with a temporal distance of about 3 s, the modulated and the unmodulated signal were presented (Zwicker, 1952, p. AB128). The German text is not unambiguous, but we tend to interpret this description to indicate the use of a continuous carrier and a gated modulation.

[2]These linear and dB step sizes are approximately equal for a degree of modulation of −30 dB.

[3]In terms of basilar membrane mechanics, it might be more appropriate to speak about linear growth in the region of upward spread and compressed growth at the center frequency (Oxenham and Plack, 1997; Ruggero *et al.*, 1997).

[4]Our scheme would also work if we did assume nonlinear auditory filters, the equivalent of assuming nonlinear upward spread of excitation in Zwicker's excitation pattern scheme for AM detection. However, the filter nonlinearity is not the critical aspect in our scheme.

Bacon, S. P., and Gleitman, R. M. (**1992**). ''Modulation detection in subjects with relatively flat hearing losses,'' J. Speech Hear. Res. **35**, 642–653.

Bacon, S. P., and Viemeister, N. F. (**1985**). ''Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners,'' Audiology **24**, 117–134.

Dau, T. (**1996**). ''Modeling auditory processing of amplitude modulation,'' Ph.D. thesis, Universität Oldenburg, Germany.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). ''Modeling auditory processing of amplitude modulation: I. Detection and masking with narrowband carriers,'' J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). ''Modeling auditory processing of amplitude modulation: II. Spectral and temporal integration,'' J. Acoust. Soc. Am. **102**, 2906–2919.

Dau, T., Verhey, J., and Kohlrausch, A. (**1999**). ''Intrinsic envelope fluctuations and modulation-detection thresholds for narrowband noise carriers,'' J. Acoust. Soc. Am. **106**, 2752–2760.

Ewert, S., and Dau, T. (**2000**). ''Characterizing frequency selectivity for envelope fluctuations,'' submitted for publication.

Fleischer, H. (**1982**). ''Modulationsschwellen von Schmalbandrauschen,'' Acustica **51**, 154–161.

Fleischer, H. (**1983**). ''Modulation thresholds of narrow noisebands,'' in *11th International Congress on Acoustics, ICA*, Paris, Lyon-Toulouse, pp. 99–102.

Formby, C. (**1986**). ''Modulation detection by patients with eighth-nerve tumors,'' J. Speech Hear. Res. **29**, 413–419.

Formby, C. (**1987**). ''Modulation threshold functions for chronically impaired meniere patients,'' Audiology **26**, 89–102.
Goldstein, J. L. (**1967**). ''Auditory spectral filtering and monaural phase perception,'' J. Acoust. Soc. Am. **41**, 458–479.
Hartmann, W. M., and Hnath, G. M. (**1982**). ''Detection of mixed modulation,'' Acustica **50**, 297–312.
Kohlrausch, A. (**1993**). ''Comment on 'Temporal modulation transfer functions in patients with cochlear implants' [J. Acoust. Soc. Am. **91**, 2156–2164 (1992)],'' J. Acoust. Soc. Am. **93**, 1649–1650.
Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J., and Püschel, D. (**1997**). ''Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations,'' Acust. Acta Acust. **83**, 659–669.
Langner, G. (**1992**). ''Periodicity coding in the auditory system,'' Hear. Res. **60**, 115–142.
Maiwald, D. (**1966**). ''Zusammenhang zwischen Mithörschwellen und Modulationsschwellen,'' Ph.D. thesis, TH München, Institut für Elektrotechnik, München.
Maiwald, D. (**1967**). ''Die Berechnung von Modulationsschwellen mit Hilfe eines Funktionsschemas,'' Acustica **18**, 193–207.
Moore, B. C., Alcántara, J. I., and Dau, T. (**1998**). ''Masking patterns for sinusoidal and narrow-band noise maskers,'' J. Acoust. Soc. Am. **104**, 1023–1038.
Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.
Rodenburg, M. (**1972**). ''Intensity discrimination of noise bands as a function of band-width and duration,'' in *Hearing Theory*, edited by B. L. Cardozo, E. de Boer, and R. Plomp (Institute for Perception Research, Eindhoven), pp. 429–439.
Rodenburg, M. (**1977**). ''Investigation of temporal effects with amplitude modulated signals,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 429–439.
Ruggero, M. A., Rich, N. C., Recia, A., Narayan, S. S., and Roble, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Sek, A. (**1994**). ''Modulation thresholds and critical modulation frequency based on random amplitude and frequency changes,'' J. Acoust. Soc. Jpn. (E) **15**, 67–75.
Shannon, R. V. (**1992**). ''Temporal modulation transfer functions in patients with cochlear implants,'' J. Acoust. Soc. Am. **91**, 2156–2164.
Sheft, S., and Yost, W. A. (**1990**). ''Temporal integration in amplitude modulation detection,'' J. Acoust. Soc. Am. **88**, 796–805.
Strickland, E. A., and Viemeister, N. F. (**1997**). ''The effects of frequency region and bandwidth on the temporal modulation transfer function,'' J. Acoust. Soc. Am. **102**, 1799–1810.
Takahashi, G. A., and Bacon, S. P. (**1992**). ''Modulation detection, modulation masking, and speech understanding in noise in the elderly,'' J. Speech Hear. Res. **35**, 1410–1421.
van Zanten, G. A. (**1980**). ''Temporal modulation transfer functions for intensity modulated noise bands,'' in *Psychoacoustical, Physiological and Behavioural Studies in Hearing*, edited by G. van den Brink and F. Bilsen (Delft U.P., Delft), pp. 206–209.
Viemeister, N. F. (**1977**). ''Temporal factors in audition: A system analysis approach,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 419–428.
Viemeister, N. F. (**1979**). ''Temporal modulation transfer function based upon modulation thresholds,'' J. Acoust. Soc. Am. **66**, 1364–1380.
Yost, W. A., and Sheft, S. (**1997**). ''Temporal modulation transfer functions for tonal stimuli: Gated versus continuous conditions,'' Aud. Neurosci. **3**, 401–414.
Zwicker, E. (**1952**). ''Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones,'' Acustica, Akust. Beih. **2**, AB125–AB133.
Zwicker, E. (**1956**). ''Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs,'' Acustica **6**, 365–381.
Zwicker, E., and Fastl, H. (**1990**). *Psychoacoustics: Facts and Models* (Springer, Berlin).
Zwicker, E., and Jaroszewski, A. (**1982**). ''Inverse frequency dependence of simultaneous tone-on-tone masking patterns at low levels,'' J. Acoust. Soc. Am. **71**, 1508–1512.

An analysis of quasi-frequency-modulated noise and random-sideband noise as comparisons for amplitude-modulated noise

Elizabeth A. Strickland and Sumit Dhar
Department of Audiology and Speech Sciences, Purdue University, West Lafayette, Indiana 47907

(Received 31 August 1999; revised 5 January 2000; accepted 23 April 2000)

Experiments were performed to determine under what conditions quasi-frequency-modulated (QFM) noise and random-sideband noise are suitable comparisons for AM noise in measuring a temporal modulation transfer function (TMTF). Thresholds were measured for discrimination of QFM from random-sideband noise and AM from QFM noise as a function of sideband separation. In the first experiment, the upper spectral edge of the noise stimuli was at 2400 Hz and the bandwidth was 1600 Hz. For sideband separations up to 256 Hz, at threshold sideband levels for discriminating AM from QFM noise, QFM was indiscriminable from random-sideband noise. For the largest sideband separation used (512 Hz), listeners may have used within-stimulus envelope correlation in the QFM noise to discriminate it from the random-sideband noise. Results when stimulus bandwidth was varied suggest that listeners were able to use this cue when the carrier was wider than a critical band, and the sideband separation approached the carrier bandwidth. Within-stimulus envelope correlation was also present in AM noise, and thus QFM noise was a suitable comparison because it made this cue unusable and forced listeners to use across-stimulus envelope differences. When the carrier bandwidth was less than a critical band or was wideband, QFM noise and random-sideband noise were equally suitable comparisons for AM noise. When discrimination thresholds for QFM and random-sideband noise were converted to modulation depth and modulation frequency, they were nearly identical to those for discrimination of AM from QFM noise, suggesting that listeners were using amplitude modulation cues in both cases. [S0001-4966(00)01208-X]

PACS numbers: 43.66.Mk, 43.66.Nm [DWG]

INTRODUCTION

Most of the complex sounds of interest to us, such as speech, music, and environmental sounds, are identifiable by patterns of amplitude change over time. A systematic method of measuring the perception of dynamic amplitude changes has been to measure threshold modulation depth for detection of sinusoidal amplitude modulation of a carrier. Modulation thresholds as a function of modulation frequency have been called the temporal modulation transfer function (TMTF). For noise carriers with bandwidths of at least twice the highest modulation frequency, the resulting function has the form of a low-pass filter (Viemeister, 1979; Dau *et al.*, 1997). The TMTF has the advantage of separating the effects of temporal resolution from intensity resolution, which may be confounded in other temporal measures such as gap detection. At low modulation frequencies, the system can easily follow the temporal fluctuations, and thus these thresholds reflect the intensity resolution of the system. As the modulation frequency increases, the system cannot follow the amplitude changes faithfully, and the frequency at which the function begins to roll off reflects the temporal resolution of the system.

The bandwidth of the carrier must be restricted when the goal is to examine the effects of frequency region or bandwidth on temporal processing. It is critical to find stimuli that are discriminable only on the basis of envelope cues. In the case of tonal carriers, sinusoidal amplitude modulation adds sideband components at plus and minus the modulation frequency, as shown in Eq. (1):

$$\mathrm{AM}(t)=\cos(2\pi f_c t)+\frac{m}{2}(\cos(2\pi(f_c-f_s)t)+\cos(2\pi(f_c+f_s)t)), \quad (1)$$

where f_c is the carrier frequency, and f_s is the modulation frequency. If this is used as the signal stimulus, and a tone at the carrier frequency is used as the comparison, listeners may be able to discriminate them either on the basis of the envelope modulation or on the basis of the long-term spectrum. In order to eliminate the spectral cue, a quasi-frequency-modulated (QFM) tone may be used as the comparison, as shown in Eq. (2):

$$\mathrm{QFM}(t)=\cos(2\pi f_c t)+\frac{m}{2}\left(\cos\left(2\pi(f_c-f_s)t+\frac{\pi}{2}\right)+\cos\left(2\pi(f_c+f_s)t+\frac{\pi}{2}\right)\right). \quad (2)$$

The long-term amplitude spectrum is identical to that of the AM tone, and thus this is no longer a cue. However, changing the phase of the sidebands introduces other differences from the AM tone. The instantaneous frequency of a QFM tone is modulated at a rate of f_s. A QFM tone is amplitude modulated at a frequency of $2*f_s$. For the same value of *m*, the modulation depth is smaller for the QFM stimulus than that of an AM tone. In addition, the difference in phases of the components may introduce differences in the long-term spectra within the ear, due to distortion products. In a previous paper (Strickland and Viemeister, 1997), it was argued

that the presence of these multiple cues makes it unclear whether thresholds for discriminating an AM from a QFM tone are based solely on envelope cues when f_s is high.

In that paper, the approach taken was to use narrowband noise carriers instead of tones. It was argued that this might eliminate the above-mentioned frequency modulation cue in the QFM stimulus, because the pitch of a noise is less salient than that of a tone. Differences in internal long-term spectra might also be unusable, due to the nondeterministic nature of the stimuli. Indeed, TMTFs with a low-pass filter shape can be obtained when QFM noise is used as the comparison, but the detectability of other cues in QFM noise was not measured. In addition, once the carrier is broadened to a noise band, it would be possible to simply add sidebands in random phase. The goal of this paper is to determine what noise is the most suitable comparison for AM noise in measuring a TMTF.

An AM noise may be created by using the Rice model (Rice, 1944) to extend Eq. (1) for a noise. A carrier band is created, and sidebands are added at plus and minus f_s from each component, as shown in Eq. (3):

$$\mathrm{AM}(t)=\sum_{f_c=f_l+f_s}^{f_h-f_s} A_c\Big[\cos(2\pi f_c t+\phi_c) + \frac{m}{2}(\cos(2\pi(f_c-f_s)t+\pi+\phi_c) + \cos(2\pi(f_c+f_s)t-\pi+\phi_c))\Big], \tag{3}$$

where A_c is a Rayleigh-distributed amplitude, ϕ_c is a random phase rectangularly distributed from 0 to 2π radians, and m is the modulation index (0 to 1). For the frequency terms, f_c is the frequency of each carrier component, f_l and f_h are the lower and upper spectral edges of the whole noise, including the sidebands, and f_s is the sideband separation from the carrier frequency in Hz. AM can be created by any phase relationship between the sidebands such that one is in positive phase and the other is in negative phase by the same amount relative to the carrier phase. In this experiment, the phases were chosen so that the modulation would start and end in a valley. AM noise is sinusoidally amplitude modulated at frequency f_s, at a modulation depth of m, from 0 to 1.

QFM noise may be created by extending Eq. (2) in the same manner, as shown in Eq. (4):

$$\mathrm{QFM}(t)=\sum_{f_c=f_l+f_s}^{f_h-f_s} A_c\Big[\cos(2\pi f_c t+\phi_c) + \frac{m}{2}\Big(\cos\Big(2\pi(f_c-f_s)t-\frac{\pi}{2}+\phi_c\Big) + \cos\Big(2\pi(f_c+f_s)t-\frac{\pi}{2}+\phi_c\Big)\Big)\Big], \tag{4}$$

which differs from Eq. (3) in that the sideband phases are shifted by minus $\pi/2$ radians from the carrier phase rather than plus and minus π radians. QFM noise has the same frequency and amplitude modulation characteristics as a QFM tone. The amplitude modulation comes from interaction between the two sidebands and the carrier, and has the form of a fully rectified sine wave at high modulation depths, and is nearly sinusoidal at lower modulation depths.

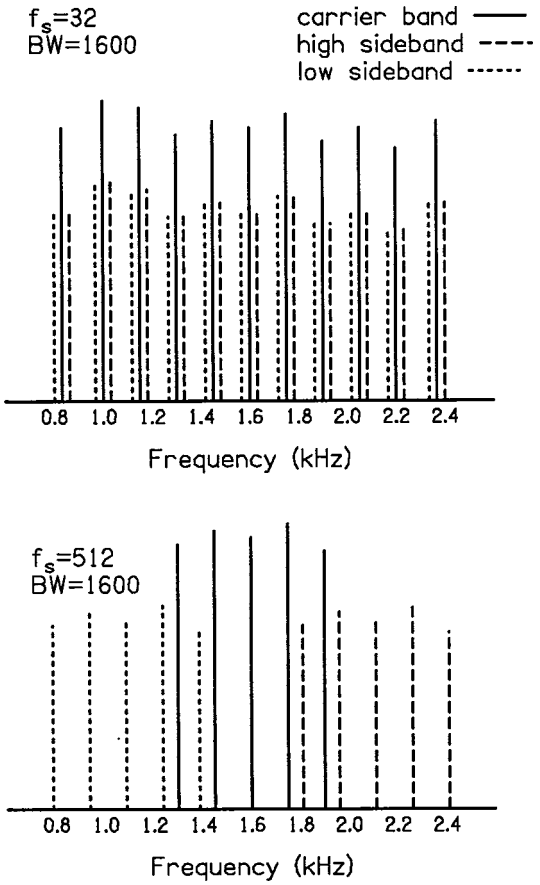


FIG. 1. Schematic depiction of the carrier band and sidebands for $f_s=32$ (top panel) and 512 Hz (bottom panel), for $m=1.0$. Component separation is 150 Hz, for clarity. In the experiments, the component separation was 2 Hz. The bandwidth is 1600 Hz in both panels.

As noted above, it would also be possible to create a noise with the same long-term average power spectrum as AM noise simply by adding the sidebands in random phase, as shown in Eq. (5):

$$\text{Random-sideband}(t)=\sum_{f_c=f_l+f_s}^{f_h-f_s} A_c\Big[\cos(2\pi f_c t+\phi_c) + \frac{m}{2}(\cos(2\pi(f_c-f_s)t+\phi_{c1}) + \cos(2\pi(f_c+f_s)t+\phi_{c2}))\Big], \tag{5}$$

where ϕ_{c1} and ϕ_{c2} are phases randomly selected for each sideband. Thus, the phase of these sidebands is random relative to the carrier phase, in contrast to the AM and QFM noise, where there is a specified phase relationship between the sidebands and the carriers.

These three types of noises have identical long-term average amplitude spectra. As shown schematically in Fig. 1, the amplitude spectra of these stimuli have a central flat area,

composed of carriers and sidebands, and a lower amplitude region flanking the central area, composed only of sidebands. Figure 1 is structured to make the relationship between each carrier and its sidebands clear, so the frequency spacing of the components is 150 Hz, in contrast to the spacing of 2 Hz in the actual stimuli. Of course, sidebands would add with carrier components in the central area, so they would not be separately visible as they are in the figure. In this experiment, the total bandwidth of the stimuli stays constant at $f_h - f_l$, and thus the width of the central area decreases and the width of the sidebands increases with f_s. Thus, in the top panel the central area is 1536 Hz wide, and the sidebands extend out 32 Hz to either side, while in the bottom panel the central area is only 576 Hz wide, and the sidebands extend out 512 Hz. The phases in which the sidebands are added relative to the carrier components create different dynamic characteristics in each type of noise.

Both QFM and random-sideband noise have average long-term amplitude spectra that are identical to AM noise. Random-sideband noise might appear to be a better comparison for AM noise, because it does not have other cues that could make it discriminable from AM noise, while QFM noise may have cues such as frequency modulation. Calculations done using an envelope detector model for a previous paper (Strickland and Viemeister, 1997), however, made it apparent that QFM noise might be a more suitable comparison for AM noise under certain conditions. For AM noise, the orderly relationship between modulation depth and sideband level relative to the carrier only holds true if the carrier and both sidebands interact. In the auditory system, the stimuli are filtered by the cochlea. When the modulation frequency is high, as shown in the bottom panel of Fig. 1, the sidebands overlap very little with the carrier. If the listener monitored a filter which passed the carrier band and one sideband, the output of the filter would have the form of a fully rectified sine wave (known as ''beats'') with a frequency equal to f_s. Goldstein (1967) has shown that for a sinusoidally AM signal, the modulation depth will be maximal when the carrier and one sideband are equal in level. If random-sideband noise were the comparison, a listener could attend to the filter that maximized beats in the AM noise. If the comparison were QFM noise, monitoring the filter that produced beats in the AM noise would also produce beats in the QFM noise, and therefore this would no longer be a usable cue for discrimination.

From the above discussion it is apparent that the structure of QFM noise may make it a better comparison for AM noise than random-sideband noise under some conditions. In previous studies, the detection of a specific cue in AM noise, amplitude modulation, has been measured by adapting on modulation index *m*, that is, the level of the sidebands relative to the carrier band (Strickland and Viemeister, 1997; Strickland, 2000; Eddins, 1999). The comparison used has been QFM noise. As the sideband level increases, the modulation depth of QFM noise also increases and the frequency sweeps in the QFM become less sinusoidal. The aim of this experiment is to determine under what conditions QFM noise and random-sideband noise are suitable comparisons for AM noise, in that the sinusoidal AM in the AM noise is the only cue for discrimination. The first experiment was conducted to determine at what values of *m* QFM noise is discriminable from random-sideband noise, as a function of f_s. These threshold values of *m* will be compared to those for discriminating AM noise from QFM noise. If threshold values of *m* for discriminating QFM from random-sideband noise are reliably well above those for discriminating AM from QFM noise, then QFM noise is a suitable comparison for AM noise.

I. METHODS

A. Stimuli

AM, QFM, and random-sideband noises were generated according to the above equations. The test stimuli had an upper spectral edge of 2400 Hz and a bandwidth of 1600 Hz. This condition was chosen because it was the largest bandwidth used in a previous paper (Strickland and Viemeister, 1997). The rate and extent of the frequency sweeps in the QFM noise increase with f_s, but f_s cannot be more than half of the bandwidth of the test stimulus. Using the widest bandwidth from the previous study maximizes f_s, and thus the chances that QFM might be discriminable from random-sideband noise on the basis of frequency sweeps.

The sideband frequency separation, f_s, was set from 4 to 512 Hz in equal log steps. Test stimuli were presented at a spectrum level of 40 dB. A low-frequency band of masking noise extending from 100 to 600 Hz was presented at a spectrum level of 30 dB, in order to mask detection of low-frequency distortion products that have been noted in other studies (Strickland and Viemeister, 1997; Eddins, 1999). Discrimination functions for QFM and random-sideband noise were measured with and without a high-frequency band of masking noise. This noise had a bandwidth of 1600 Hz and had a lower spectral edge of 2450 Hz. A previous study (Strickland and Viemeister, 1997) showed that adding such a masking noise above the test stimulus increased the modulation detection threshold for modulation frequencies of 64 Hz and below. These results suggested that when no masking noise was present, listeners might attend to the frequency region above the test stimulus, where the modulation depth might be magnified by nonlinear spread of excitation. It was of interest to determine the effects of such a noise on discrimination of QFM from random-sideband noise, where different cues may be used. Discrimination of AM from QFM noise was only measured with the high-frequency masking noise present.

The test stimuli and masking noises were 500 ms in duration, including 10-ms raised-cosine ramps. The test stimuli and masking noises were digitally generated by a computer and output through separate digital-to-analog channels (TDT DA1) at a rate of 16 384 Hz. They were low-pass filtered at 6 kHz (Kemo VBF/23). The levels were adjusted by programmable attenuators (TDT PA4). Stimuli were presented through one earphone of Sennheiser HD450 headphones to a listener seated in a double-walled sound-attenuating booth.

B. Procedure

Sideband separation, f_s, was fixed within a run, and $20\log(m)$ was adjusted in an adaptive two-interval forced-choice (2IFC) procedure with a two-down, one-up stepping rule, to track 71% correct (Levitt, 1971). Fifty trials were presented in each block, and feedback was given after each trial. The initial step size was 3, and was reduced to 1 after the first two reversals. The threshold estimate for each block was calculated from the average of the last even number of reversals at the smaller step size. When measuring discrimination of AM from QFM, initial values of $20\log(m)$ ranged from -9 at low modulation frequencies to -6 at high modulation frequencies, and if $20\log(m)$ would have exceeded zero the track was terminated and was not included in threshold estimates. This was done because for AM, when $20\log(m)$ exceeds zero, the modulation frequency doubles and the modulation is no longer sinusoidal. When the discrimination was between QFM and random-sideband noise, $20\log(m)$ was allowed to exceed zero, and initial values ranged from 5 at low modulation frequencies to 15 at high modulation frequencies. Threshold values are based on the average of at least three blocks.

C. Subjects

Three females and one male were tested. The age range was from 23 to 41 years, with a mean of 29 years. All listeners had thresholds within laboratory norms for pure tones at octave frequencies from 250–8000 Hz in the ear tested. S1, S2, and S3 had at least 10 h of experience with psychoacoustic tasks before testing, while S4 was a naive listener. S1 was not tested in the AM vs QFM condition due to time constraints.

II. RESULTS

A. Thresholds for QFM vs random-sideband noise

Thresholds in units of $20\log(m)$ (left axis) or m (right axis) as a function of f_s are shown in Fig. 2. Thresholds increase *downward* on these figures, as that has been the convention often used in plotting AM thresholds. Note also that in contrast to AM thresholds, where m is always less than 1, m in this task generally exceeded 1. Thresholds with no high-frequency masking noise are shown by the filled circles, while those with high-frequency masking noise are shown by the open triangles. Error bars are one standard deviation about the mean. With no masking noise, thresholds were fairly constant and variability was low at low f_s. At higher f_s, thresholds increased and variability increased markedly. This variability was seen even if tracks with high standard deviations were excluded. For two subjects (S2 and S4) threshold decreased again for $f_s=512$. The addition of the high-frequency masking noise had a similar effect across subjects. Thresholds increased by about 2 units for f_s of 64 Hz and below. At higher f_s, thresholds increased markedly and were quite variable within and across subjects.

In Fig. 3, the average thresholds for discrimination of QFM from random-sideband noise are shown compared to thresholds for discrimination of AM from QFM noise, using

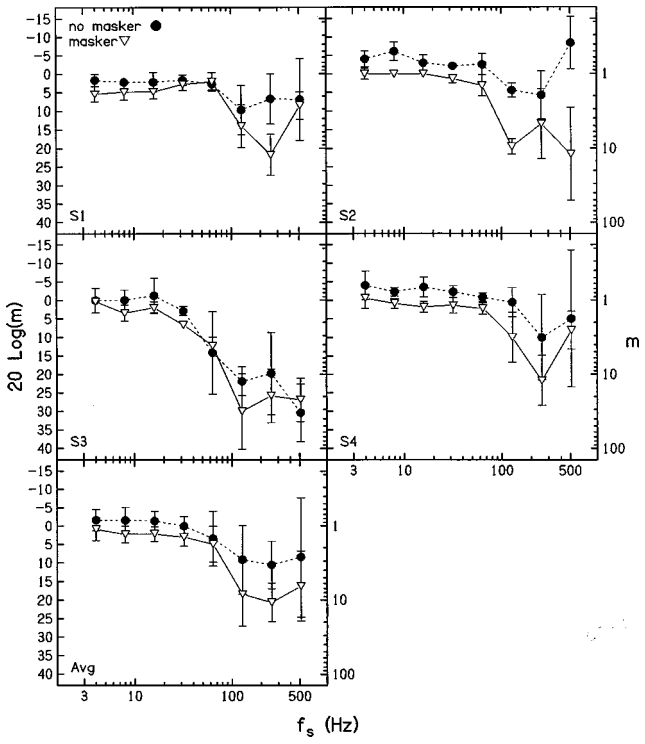


FIG. 2. Thresholds for discriminating QFM from random-sideband noise, either with (open triangles) or without (filled circles) a high-frequency masker. Error bars in all figures are one standard deviation about the mean.

test stimuli with the same bandwidth, upper cutoff frequency, and level as the QFM and random-sideband noises. For $f_s=512$ Hz, when AM was being compared to QFM noise, only one of the three subjects tested in this condition could achieve a threshold without terminating the track. Thresholds for the other two subjects were averaged in as $20\log(m)=0$. Thresholds for discrimination of AM from QFM noise were well below those (higher on the figure) for discrimination of QFM from random-sideband noise when $f_s=256$ Hz and below. This indicates that QFM noise is a suitable comparison for AM noise when $f_s=256$ Hz or less, for the particular stimulus conditions used here. However, it is not clear that this is true when $f_s=512$ Hz, due to the variability of the thresholds for discriminating QFM from

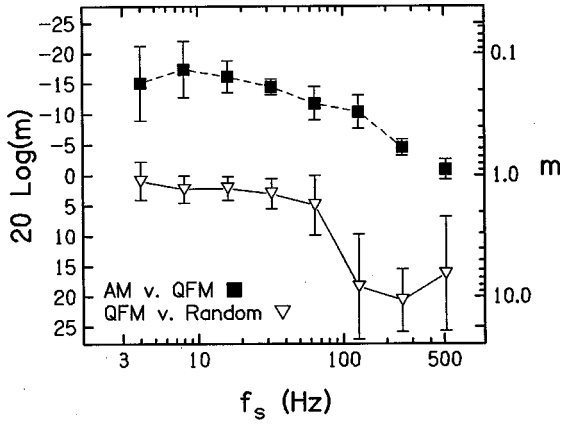


FIG. 3. Average thresholds for discriminating QFM from random-sideband noise (open triangles, replotted from Fig. 2) and AM from QFM noise (filled squares).

random-sideband noise, and the fact that no threshold could be obtained for discriminating AM from QFM noise for two of the subjects. Even though the average thresholds were clearly separated, the variability within subjects made it impossible to be certain that other cues in the QFM noise were unusable in discriminating AM from QFM.

B. Psychometric functions

Thresholds for discrimination of QFM from random-sideband noise for $f_s = 128$ Hz and above were quite variable both within and across subjects, and were more affected by the addition of the high-frequency masking noise than thresholds for lower f_s. These thresholds were also at higher values of *m*. The gross similarity between the two functions in Fig. 3 suggested that listeners might be using modulation depth as a cue in discriminating QFM noise from random-sideband noise. For AM noise, *m* corresponds directly to amplitude modulation depth. The envelope is shown in Eq. (6):

$$E(t) = 1 + m\cos(f_s t). \tag{6}$$

For QFM noise, the envelope is related to *m* in a more complex manner, as shown by Eq. (7):

$$E(t) = \sqrt{1 + m^2\cos^2(f_s t)}, \tag{7}$$

which was modified from the relationship derived by Edwards and Viemeister [1994a, Eq. (2)]. If listeners were using modulation depth as a cue, this complex relationship between *m* and modulation depth for QFM noise might account for the change in variability at different values of *m*.

For the same four subjects, psychometric functions were measured for high f_s (64 Hz and above) to investigate the higher variability at these frequencies. A two-cue 2IFC procedure (Bernstein and Trahiotis, 1982) was used. Intervals one and four were random samples of the random-sideband noise, and the QFM noise could occur with equal probability in interval two or three. This was done so that subjects only had to discriminate between the two types of noise, not identify them. Subjects also began each run by listening to a sample interval in which $20\log(m)$ was fixed at 27, and the QFM noise was always in the first interval. The value of $20\log(m)$ was fixed for a block of 50 trials, and randomized across blocks. Thresholds for $f_s = 64$ Hz were also measured, using a 2IFC procedure with no cues. The high-frequency masking noise was present in all conditions. The results are shown in Fig. 4. For $f_s = 64$ Hz, performance increased with $20\log(m)$ and plateaued at approximately 100% correct. For higher f_s, however, performance plateaued at a lower percent correct or was even nonmonotonic. This explains the extreme variability, both within and between subjects, seen in the thresholds for discrimination of QFM from random-sideband noise at high f_s in Figs. 2 and 3. Because percent correct may change little with large changes in $20\log(m)$, the adaptive tracking procedure does not converge on a consistent value.

To see what the psychometric functions would look like if subjects were using amplitude modulation depth as a cue, modulation depth (mod) was calculated for each value of *m* by creating a QFM tone in MATLAB and calculating (max −min)/(max+min). These functions are shown in Fig. 5. The psychometric functions for $f_s = 64$ and 128 Hz were monotonic, and the form was similar to that found by Eddins (1993) for AM noise for similar modulation frequencies (125 and 250 Hz, which would be close to the modulation frequencies of 128 and 256 Hz for the QFM noise), shown as asterisks in Fig. 5. Subject behavior was ambiguous when $f_s = 256$ Hz. One of the four subjects was at chance (50%); two others were above chance but showed almost no change in performance as a function of modulation depth. The fourth subject showed a monotonically increasing psychometric function. For $f_s = 512$ Hz, however, the functions clearly plateaued below 100%, and some were nonmonotonic. This suggests that up to $f_s = 128$ Hz, for most subjects, modulation depth was the cue used to discriminate QFM noise from random-sideband noise. For $f_s = 512$ Hz, however, Fig. 4

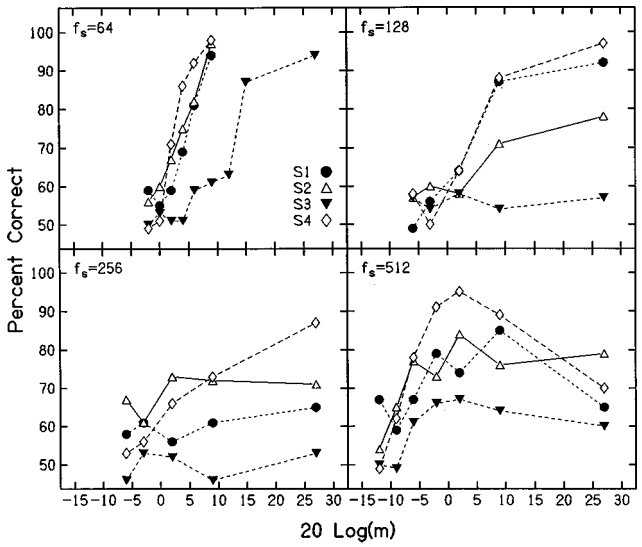


FIG. 4. Psychometric functions for discriminating QFM from random-sideband noise for four subjects.

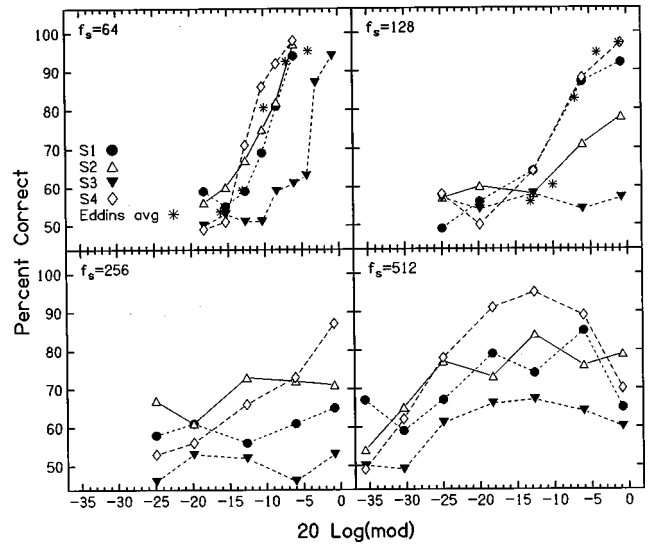


FIG. 5. The data from Fig. 4 plotted in terms of calculated modulation depth. The asterisks are average data from Eddins (1993), for a modulation frequency of 125 Hz in the upper-left panel, and 250 Hz in the upper-right panel.

shows that there was a different cue that became more salient with increases in $20\log(m)$ up to approximately -5 or 0, and then plateaued or decreased again. The value of $20\log(m)$ is 6 dB higher than the level of the sidebands relative to the carrier. In Fig. 4, whatever cue the subjects were using was most salient when $20\log(m)$ was 2, or the sidebands were just 4 dB below the level of the carrier.

C. Possible cues at $f_s = 512$ Hz

The psychometric function for $f_s = 512$ Hz suggests that subjects are able to detect other cues that are available within these noises. To understand one possible cue, it is necessary to think about the structure of QFM noise, which was shown schematically in Fig. 1. It consists of a carrier band of noise, and sidebands that are the same bandwidth, whose outer spectral edges extend out from the carrier band by f_s. The total bandwidth of the test stimulus is held constant. When f_s is low, the carrier band and the sidebands overlap almost entirely, as shown schematically in the top panel for $f_s = 32$ Hz. At high f_s, however, the sidebands overlap with the carrier band less and less. This leads to three nearly separate bands of noise that have the same power spectrum, except that the sidebands may be shifted in level by a constant, as shown in the bottom panel of Fig. 1 for $f_s = 512$ Hz. If each of these bands of noise were passed through a separate filter, the envelopes would be exactly correlated. This is true because the sideband components have exactly the same phases, and the phases of the carrier components are all shifted by a constant phase. In reality some of the carrier components overlap the sideband components, but the envelopes would still be highly correlated. Random-sideband noise would have the same amplitude spectrum, but because the phases of the components are randomized, the envelopes of the carrier and the two sidebands would be less correlated. Richards (1988) has shown that subjects can discriminate with a high degree of accuracy between the equivalent of the random carrier and sidebands, which have identical power spectra but different envelopes, and the QFM carrier and sidebands, which have identical power spectra and envelopes.

Although listeners could conceivably compare any of the bands, data from Richards (1987) suggest it is likely that they are using the carrier band and the upper sideband. In that study, performance improved as the center frequencies of the noise bands were increased. The fact that some subjects show a decline in performance as the sideband level is increased relative to the carrier (see Fig. 4) is probably due to masking of one band by another. When the carriers are more separated in frequency, it is possible for subjects to make envelope judgments even when carrier levels are unequal (Goldstein, 1965; Strickland *et al.*, 1989).

In order to test the theory that listeners were detecting envelope correlation when $f_s = 512$, QFM versus random-sideband discrimination was tested when the total extent of the test stimulus was 10 to 6000 Hz. In this case the situation would be closer to the top panel of Fig. 1, and the carrier band and sidebands would always overlap to a large extent. Even when the sidebands extend out by 512 Hz, the carrier is approximately 5000-Hz wide, and the upper sideband would be related to the high-frequency end of the carrier while the lower sideband would be related to the low-frequency end of the carrier. Thus, the sidebands would not be highly correlated with each other. Thresholds measured using the broadband stimulus are shown in Fig. 6. It can be seen that thresholds now fell off monotonically as f_s increased. This suggests that listeners were forced to use the temporal cue at the modulation frequency, $2*f_s$, and now have no envelope correlation cue.

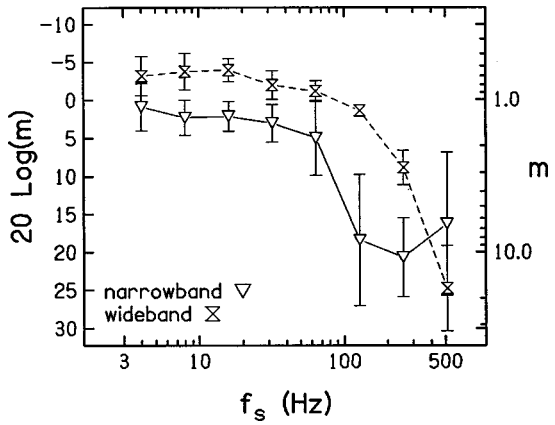


FIG. 6. Average thresholds for discrimination of QFM from random-sideband noise with a test stimulus bandwidth of 1600 Hz (open triangles, replotted from Fig. 2) or 6000 Hz (hourglasses).

D. Comparison to AM

The psychometric functions in Fig. 5 suggest that listeners are discriminating QFM from random-sideband noise on the basis of amplitude modulation depth for $f_s = 128$ Hz and below. To examine this, the thresholds from Fig. 6 were converted to modulation depths and plotted at their amplitude modulation frequencies, $2*f_s$. These are shown in Fig. 7 along with AM vs QFM noise thresholds, also plotted in terms of modulation depth of the AM signal. In the top panel, the bandwidth of the test stimuli was 1600 Hz. In the bottom panel, it was 6000 Hz, and the AM vs QFM data are from a previous study (Strickland, 2000). The large error bars are due to variability between subjects. Although not significant, there was a tendency for thresholds to be higher for QFM vs random-sideband noise than for AM vs QFM noise at the lowest modulation frequencies. This was also noted by Edwards and Viemeister (1994b) when comparing thresholds for tonal carriers for AM versus tone and for QFM vs FM. Following their suggestion, there may have been enough frequency modulation in the QFM noise to make detection of the amplitude modulation slightly more difficult. Above a modulation frequency of 16 Hz, thresholds were nearly identical up to a modulation frequency of 256 Hz in the top panel and 1024 Hz in the bottom panel, supporting the idea that listeners were basing their judgments on modulation depth in both cases.

E. AM vs QFM compared to AM vs random sideband

The data suggest that for low f_s, QFM noise is indiscriminable from random-sideband noise at values of m at

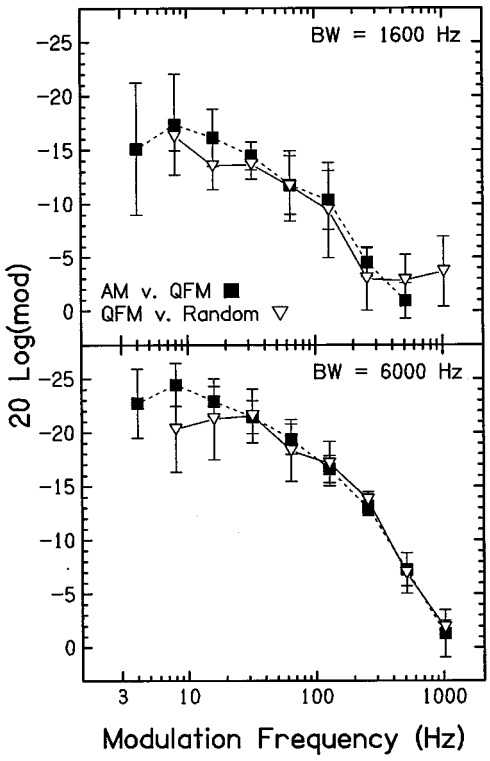


FIG. 7. Average thresholds for discrimination of QFM from random-sideband noise (open triangles) and AM from QFM noise (filled squares) with a test stimulus bandwidth of 1600 Hz (top panel) or 6000 Hz (bottom panel). The data are plotted in terms of modulation frequency and calculated modulation depth.

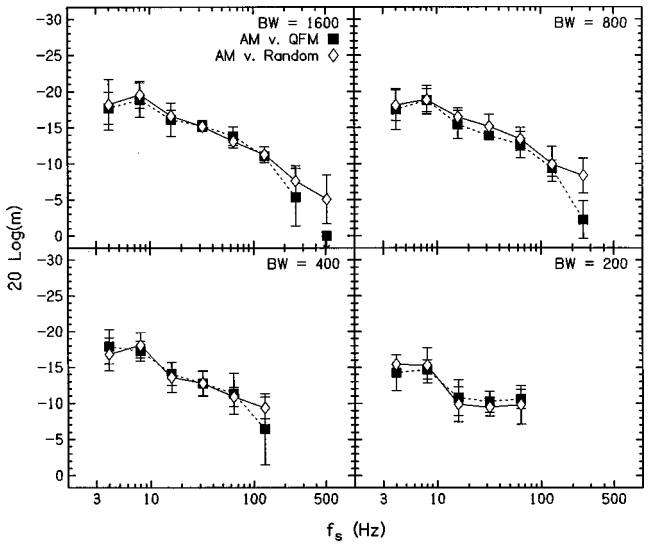


FIG. 8. Average thresholds for four subjects for discrimination of AM from QFM noise (filled squares) and AM from random-sideband noise (open diamonds) as a function of test stimulus bandwidth.

which AM is just discriminable from QFM noise. It would be expected then that thresholds for discriminating AM noise from QFM noise should be the same as those for discriminating AM noise from random-sideband noise for these sideband separations. For $f_s = 512$ Hz, envelope correlation cues should be present for the AM noise just as they are for QFM noise, and these may be usable when the comparison is random-sideband noise. As part of a previous study, thresholds for these two cases were measured for S2 and three additional subjects. Subjects were two males and two females, with an average age of 24.8 years, and a range of 18 to 36 years. Each subject had thresholds for pure tones within laboratory norms for the ear tested, which was the left ear for all subjects. The methods and stimuli were the same as those used above. In addition to a test stimulus bandwidth of 1600 Hz, bandwidths of 800, 400, and 200 Hz were used, with the upper spectral edge fixed at 2400 Hz. Figure 8 shows threshold modulation depths for discrimination of AM from QFM noise, and AM from random-sideband noise. Data points are the average of the four subjects. For bandwidths of 800 and 1600 Hz, thresholds were clearly lower (higher on the figure) for the highest f_s when random-sideband noise was the comparison. The same trend was seen when the bandwidth was 400 Hz. For the 200-Hz bandwidth, thresholds were nearly identical for all f_s. A filter centered on the 200-Hz bandwidth stimulus would have a center frequency of 2.3 kHz. According to the formula of Glasberg and Moore (1990), the equivalent rectangular bandwidth would be 273 Hz. This suggests that when the carrier bandwidth is larger than a critical band, an envelope correlation cue may be usable when the sidebands only minimally overlap the carrier band and the comparison is random-sideband noise. When QFM noise is the comparison, both of the test stimuli have correlation between the carrier and the sideband envelopes. Thus, this is not a usable cue and thresholds are higher.

III. DISCUSSION

The results indicate that if the goal is to measure detection of the temporal fluctuations in the AM noise, QFM is a suitable comparison in all conditions, while random-sideband noise is not suitable in some conditions. When the carrier is wider than a critical band, and the sideband separation approaches the carrier bandwidth, it appears that listeners can discriminate QFM noise from random-sideband noise using envelope correlation between sidebands and the carrier. This would also be true for AM noise, and thus in these conditions QFM noise is a better comparison than random-sideband noise, because it eliminates this as a cue for discrimination. Both QFM and random-sideband noise are good comparisons in all other conditions.

As noted in the Introduction, the first author originally chose QFM over random-sideband noise as the comparison noise for AM noise due to concern about detection of beats between the sidebands and the carrier (see also Strickland and Viemeister, 1997, footnote 1). Thus, that study and others that have used the technique for creating noises used in this paper have used QFM noise as the comparison (Strickland, 2000; Eddins, 1999). Other studies using bandlimited stimuli have filtered after modulation (Formby and Muir, 1988; Rodenburg, 1977; van Zanten, 1980; Viemeister, 1979; Bacon and Viemeister, 1985; Dau *et al.*, 1997; Eddins, 1993) or restricted the frequency region with notched noise (Patterson *et al.*, 1978), so envelope correlation cues would not be available. Detection of envelope correlation should only be a possibility when the sideband separation ap-

proaches or exceeds the carrier bandwidth, and the carrier is wider than a critical band. Under these conditions, it is important that QFM noise be used as the comparison.

When a QFM tone is being discriminated from a pure tone, sideband level at threshold decreases with sideband separation for small sideband separations (Edwards and Viemeister, 1994b). This is consistent with the idea that listeners use the size of the instantaneous frequency sweeps, which would increase with sideband separation. Figure 2 shows that thresholds do not decrease with sideband separation when QFM noise is compared to random-sideband noise, and thus that frequency sweeps are not a usable cue. This is further supported by the fact that the data in Fig. 2 are very similar to data from Edwards and Viemeister (1994b, Fig. 2) for discrimination of QFM and FM tones, when the sideband separation is 64 Hz or less. This is consistent with the idea that temporal envelope cues are used in both cases. When the sideband separation is greater than 64 Hz, thresholds for QFM vs FM become lower than those for QFM versus random-sideband noise, due to detection of sideband differences. The use of noise allows comparison of the QFM tone to be extended to higher sideband separations, and shows that envelope fluctuations are the only cue under these conditions.

IV. CONCLUSIONS

(1) When QFM noise is compared to random-sideband noise, listeners use amplitude modulation cues when sideband separations are well below the carrier bandwidth. As sideband separation approaches the carrier bandwidth, they may use correlation between carrier and sideband envelopes as a cue.
(2) Frequency modulation is not a usable cue in QFM noise with a bandwidth of 1600 Hz.
(3) When measuring modulation depth thresholds with narrow-band noise carriers, QFM noise is a suitable comparison for AM noise. When sideband separations are well below the carrier bandwidth, the amplitude modulation cues in QFM noise are below threshold at threshold sideband levels for discriminating AM from QFM. When the sideband separation approaches the carrier bandwidth, and the carrier bandwidth is greater than a critical band, AM and QFM noise have correlations between the carrier and sideband envelopes, and thus listeners are forced to use the amplitude modulation cue.

ACKNOWLEDGMENTS

The authors thank Tamara Scott for assistance in data collection, and John Madden and two anonymous reviewers for constructive comments on an earlier version of the paper.

Bacon, S. P., and Viemeister, N. F. (**1985**). ''Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners,'' Audiology **24**, 117–134.

Bernstein, L. R., and Trahiotis, C. (**1982**). ''Detection of interaural delay in high-frequency noise,'' J. Acoust. Soc. Am. **71**, 147–152.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). ''Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers,'' J. Acoust. Soc. Am. **102**, 2892–2905.

Eddins, D. A. (**1999**). ''Amplitude-modulation detection at low- and high-audio frequencies,'' J. Acoust. Soc. Am. **105**, 829–837.

Eddins, D. A. (**1993**). ''Amplitude modulation detection of narrow-band noise: Effects of absolute bandwidth and frequency region,'' J. Acoust. Soc. Am. **93**, 470–479.

Edwards, B. W., and Viemeister, N. F. (**1994a**). ''Modulation detection and discrimination with three-component signals,'' J. Acoust. Soc. Am. **95**, 2202–2212.

Edwards, B. W., and Viemeister, N. F. (**1994b**). ''Psychoacoustic equivalence of frequency modulation and quasi-frequency modulation,'' J. Acoust. Soc. Am. **95**, 1510–1513.

Formby, C., and Muir, K. (**1988**). ''Modulation and gap detection for broadband and filtered noise signals,'' J. Acoust. Soc. Am. **84**, 545–550.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

Goldstein, J. L. (**1965**). *An Investigation of Monaural Phase Perception*, Doctoral dissertation, The University of Rochester, Rochester, NY (University Microfilms, Ann Arbor, MI, Publ. No. 66-6852).

Goldstein, J. L. (**1967**). ''Auditory spectral filtering and monaural phase perception,'' J. Acoust. Soc. Am. **41**, 458–479.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Patterson, R. D., Johnson-Davies, D., and Milroy, R. (**1978**). ''Amplitude-modulated noise: The detection of modulation versus the detection of modulation rate,'' J. Acoust. Soc. Am. **63**, 1904–1911.

Rice, S. O. (**1944**). ''Mathematical analysis of random noise,'' Bell Syst. Tech. J. **23**, 238–332; **24**, 46–156 (**1945**). Also available in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York, 1954).

Richards, V. M. (**1988**). ''Components of monaural envelope correlation perception,'' Hear. Res. **35**, 47–58.

Richards, V. M. (**1987**). ''Monaural envelope correlation perception,'' J. Acoust. Soc. Am. **82**, 1621–1630.

Rodenburg, M. (**1977**). ''Sensitivity of the auditory system to differences in intensity,'' Ph.D. dissertation, Erasmus University of Rotterdam, Netherlands.

Strickland, E. A. (**2000**). ''The effects of frequency region and level on the temporal modulation transfer function,'' J. Acoust. Soc. Am. **107**, 942–952.

Strickland, E. A., and Viemeister, N. F. (**1997**). ''The effects of frequency region and bandwidth on the temporal modulation transfer function,'' J. Acoust. Soc. Am. **102**, 1799–1810.

Strickland, E. A., Viemeister, N. F., and Fantini, D. A. (**1989**). ''Within- versus cross-channel mechanisms in detection of envelope phase disparity,'' J. Acoust. Soc. Am. **86**, 2160–2166.

van Zanten, G. A. (**1980**). ''Temporal modulation transfer functions for intensity modulated noise bands,'' in *Psychophysical, Physiological and Behavioural Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (Delft University Press, Delft, The Netherlands), pp. 206–209.

Viemeister, N. F. (**1979**). ''Temporal modulation transfer functions based upon modulation thresholds,'' J. Acoust. Soc. Am. **66**, 1364–1380.

The effect of aging on horizontal plane sound localization

Sharon M. Abel
The Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada

Christian Giguère
Audiology and Speech-Language Pathology Program, University of Ottawa, 545 King Edward Avenue, Ottawa, Ontario K1N 6N5, Canada

Angela Consoli
The Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada

Blake C. Papsin
Department of Otolaryngology, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada

(Received 11 February 1999; revised 28 June 1999; accepted 10 May 2000)

An experiment was conducted to determine the effect of aging on sound localization. Seven groups of 16 subjects, aged 10–81 years, were tested. Sound localization was assessed using six different arrays of four or eight loudspeakers that surrounded the subject in the horizontal plane, at a distance of 1 m. For two 4-speaker arrays, one loudspeaker was positioned in each spatial quadrant, on either side of the midline or the interaural axis, respectively. For four 8-speaker arrays, two loudspeakers were positioned in each quadrant, one close to the midline and the second separated from the first by 15°, 30°, 45°, or 60°. Three different 300-ms stimuli were localized: two one-third-octave noise bands, centered at 0.5 and 4 kHz, and broadband noise. The stimulus level (75 dB SPL) was well above hearing threshold for all subjects tested. Over the age range studied, percent-correct sound-source identification judgments decreased by 12%–15%. Performance decrements were apparent as early as the third decade of life. Broadband noise was easiest to localize (both binaural and spectral cues were available), and the 0.5-kHz noise band, the most difficult to localize (primarily interaural temporal difference cue available). Accuracy was relatively higher in front of than behind the head, and errors were largely front/back mirror image reversals. A left-sided superiority was evident until the fifth decade of life. The results support the conclusions that the processing of spectral information becomes progressively less efficient with aging, and is generally worse for sources on the right side of space.
[S0001-4966(00)03408-1]

PACS numbers: 43.66.Qp, 43.66.Ba [DWG]

INTRODUCTION

Several experimental paradigms have been used to assess directional hearing ability (Blauert, 1997). In lateralization experiments, the sound is presented to the two ears by means of earphones. If there are no interaural temporal or level differences, the subject will perceive a single acoustic image within the head, in the midline. In young, normal-hearing adults, introduction of an interaural delay of 15 μs or a difference in level of 1 dB will shift the image toward the ear leading in time or receiving the higher level (e.g., Klumpp and Eady, 1956; Abel and Kunov, 1983; Grantham, 1984). Another paradigm, free-field sound localization, tests the listener's ability to identify the direction of a sound source. For continuous signals, the primary cue to azimuthal perception for stimulus frequencies below 1.5 kHz is the temporal difference in the sound waveform at the two ears, and the primary cues for higher frequencies are spectral cues and interaural differences in level (Stevens and Newman, 1936; Mills, 1972; Wightman and Kistler, 1992). Front/back discrimination and elevation perception for frequencies beyond 3 kHz are aided by spectro–temporal information from the filtering effect of the pinna of the ear (Musicant and Butler, 1984).

Accuracy in sound-source identification is affected by characteristics of the test environment, the loudspeaker array, the gating of the stimulus to be localized, and noise conditions in the room. It has been shown, for example, that accuracy in localizing sounds with slow onsets can be significantly higher in absorbent than in reverberant test rooms (Hartmann, 1983; Giguère and Abel, 1993). For horizontal speaker arrays, errors increase with an increase in both the number of possible alternatives and their proximity to the interaural axis (Abel and Banerjee, 1996). While left/right errors are rare, there is a relatively high degree of confusion between sources in front of and behind the ear (Butler, 1994; Burke *et al.*, 1994; Abel *et al.*, 1999). In rooms with reflecting surfaces, pure tones with abrupt onsets or offsets are more easily localized than those with longer rise/decay times (Rakerd and Hartmann, 1986). Performance may also be affected by the presence of background noise, depending on

the signal-to-noise ratio and noise location relative to the speaker (Abel and Hay, 1996; Good and Gilkey, 1996; Lorenzi *et al.*, 1999).

The present experiment was conducted to explore life-cycle changes in the utilization of cues underlying directional hearing, from adolescence to senescence. Sound localization ability is available at birth. Infants, 2 to 4 days of age, can correctly turn their heads toward the sound of a rattle presented 90° to the right or left of midline (Muir and Field, 1979). They also have limited ability to discriminate among positions within the right and left hemifields (Morrongiello *et al.*, 1994). Acuity improves with age, a finding ascribed to maturation of the inner ear and brainstem, growth of the head, and shift of locus of control from subcortical to cortical structures (Clifton, 1992; Muir *et al.*, 1989). Over the first year and a half of life, the minimum audible change in the straight-ahead position decreases from 12° to 4° (Morrongiello, 1988), reaching a minimum of 1°–2° in adulthood (e.g., Mills, 1958). Spatial resolution is better the closer the sound source is to the midline, a finding consistent with adult performance patterns (Morrongiello and Rocca, 1990).

Aging in the later years has a deleterious effect. Herman *et al.* (1977) found that men aged 60 to 72 years required a significantly greater interaural time delay than those aged 22 to 32 years to lateralize a 2-s click train. The interaural level difference limen was the same, pointing to an age-specific decrement in the ability to process interaural temporal disparity. A recent study by Abel and Hay (1996) showed that subjects aged 40–60 years had more difficulty with front/back discrimination than did younger subjects aged 20–38 years, suggesting a progressive failure in the utilization of spectro–temporal information. Hearing loss, if symmetrical bilaterally, did not appreciably affect performance, as long as the stimulus was clearly audible. Age-related central degenerative changes have been identified throughout the auditory system. These may possibly underlie observed sound localization deficits with aging (e.g., Kirikae *et al.*, 1964; Johnsson and Hawkins, 1972; Casey and Feldman, 1982).

Sound localization studies conducted to date have focused almost exclusively on either developmental changes or performance patterns in young adulthood. There are, to date, no studies of how the utilization of binaural and spectral cues evolve with age. The recent literature suggests that age-related hearing loss will not be a critical factor, as long as the sounds to be localized are clearly audible and hearing is symmetrical in the two ears (Noble *et al.*, 1994; Abel and Hay, 1996). In the present study of aging, sound localization was assessed in an environment which modeled real-world listening. The task involved identification of the direction of a sound emanating from a set of loudspeakers surrounding the subject in the horizontal plane. Three parameters were manipulated: (1) the number of alternative sound sources, (2) the separation between speakers, and (3) the bandwidth and frequency of the stimulus, in order to evaluate contextual effects, directional acuity, and the use of binaural and spectral cues.

TABLE I. The six speaker arrays.

Array	Configuration	Azimuth angles
Four-speaker	midline	±15° ±165°
	lateral	±75° ±105°
Eight-speaker	15° separation	±15° ±30° ±150° ±165°
	30° separation	±15° ±45° ±135° ±165°
	45° separation	±15° ±60° ±120° ±165°
	60° separation	±15° ±75° ±105° ±165°

I. METHODS AND MATERIALS

A. Experimental design

The sound localization experiment was conducted in a semireverberant chamber that modeled listening in a small office environment (Giguère and Abel, 1990; Abel *et al.*, 1999). Subjects were required to judge the direction of three different stimuli: broadband noise and one-third-octave noise bands centered at 0.5 and 4 kHz, presented at a comfortable listening level of 75 dB SPL measured at the test location. The duration of each stimulus was 300 ms, including a linear rise/decay of 50 ms. The broadband stimulus assessed the utilization of binaural and spectral cues, in combination. The low-frequency band was intended to assess primarily the utilization of interaural temporal difference cues, and the high-frequency band, spectral and interaural level difference cues (Wightman and Kistler, 1992). Noise bands were chosen in preference to pure tones to avoid local minima or maxima in sound level in the test chamber due to the small degree of room reverberation. The present study primarily assessed the utilization of the ongoing cues in the stimulus waveform rather than the onset/offset cues. A previous study in the same environment and using the same stimuli showed that a change in rise/decay time from 5 to 200 ms had only a very small effect on localization (Giguère and Abel, 1993).

Sound localization ability was tested using two arrays of four loudspeakers and four arrays of eight loudspeakers, surrounding the subject at ear level in the horizontal plane, at a distance of 1 m. For the four-speaker arrays, speakers were placed on either side of the midline, front (±15°) and back (±165°), or on either side of the interaural axis, right (+75° and +105°) and left (−75° and −105°), relative to the straight-ahead position, defined as 0°. For the eight-speaker arrays, two speakers were positioned in each of the four spatial quadrants, one close to the midline (±15° and ±165°) and the second separated from the first by 15°, 30° or 45° and 60° (see Table I and Abel *et al.*, 1999). The separations chosen for the present study exceed the minimum audible angles for frontal positions in the horizontal plane (Mills, 1972). However, as demonstrated by Abel and Banerjee (1996), with a speaker identification localization paradigm, accuracy decreases as the number of loudspeaker alternatives increases or as the separation between speakers decreases.

To ensure that subjects understood the task, they were given 20 random presentations of the broadband noise from an array of four loudspeakers, positioned at ±45° and ±135°, i.e., at the center of each spatial quadrant. Previous research has shown that infants and normal-hearing adults can discriminate between right and left sound sources with-

out difficulty (Morrongiello, 1988; Abel and Hay, 1996). Thus, any subject who did not achieve a score of at least 70% correct during this practice block was discontinued. One subject in each of the 30–39 and 70–81-year-old groups did not meet this criterion and was replaced.

B. Subjects

Seven groups of 16 subjects (40 males and 72 females), aged 10–19, 20–29, 30–39, 40–49, 50–59, 60–69, and 70–81 years, respectively, participated. The male/female ratio was similar across groups. All but six individuals were right-handed. Subjects were recruited by posting advertisements at various University of Toronto campus sites. Prospective candidates were screened by telephone for a history of birth defects, head injury, systemic disease, neurological disorders, and the use of medications that might affect performance. Those who met these criteria came to the laboratory for a hearing screening test.

Hearing thresholds were measured in each ear by means of a variation of Bekesy tracking, using Telephonics TDH 49P matched earphones, calibrated with a Bruel & Kjaer 4153 coupler (see Abel and Hay, 1996). The four pure-tone frequencies tested, 0.5, 1, 2, and 4 kHz, covered the range of the stimuli that would be localized. The admission guidelines allowed for mild hearing loss due to presbycusis: less than 25 dB SPL for those aged 10–29 years, no greater than 35 dB SPL for those 30–59 years, and no greater than 50 dB SPL for those 60–81 years (Brant and Fozard, 1990), with interaural differences at any frequency no greater than 10 dB. Based on ANSI S3.6-1996, the reference equivalent threshold levels for 0 dB HL at the four test frequencies are 13.5, 7.5, 9, and 12 dB SPL, respectively. Thus, a threshold of 50 dB SPL at 4 kHz corresponds to 38 dB HL.

C. Apparatus

Subjects were tested individually while seated at the center of a 3.5 by 2.7 by 2.3 m semireverberant chamber. Reverberation times were 0.6 s at 0.125 and 0.25 kHz, 0.4 s from 0.5 to 4 kHz, and 0.3 s at 6.3 and 8 kHz (see Giguère and Abel, 1990). The stimuli to be localized were generated by a noise generator (Bruel & Kjaer, type 1405). Spectral shaping of the stimuli into one-third-octave bands was realized by means of a matched pair of variable-frequency six-pole Butterworth filter sets (Bruel & Kjaer, type 1617). The level of the stimulus was controlled by means of a programmable attenuator (Coulbourn S85-08), and a set of integrated stereo amplifiers (Realistic SA-150) connected with the loudspeakers (Realistic Minimus 3.5). Measurements recently conducted in an anechoic room (by C.G.) on a set of 22 identical Minimus 3.5 speakers from the University of Ottawa showed a close match in frequency response (within ± 2.5 dB) from 0.125–12 kHz. The amplifiers were balanced so that the sound levels emanating from the speakers at the different azimuths were within 1.5 dB at the subject's position, precluding the use of a change in loudness as a cue to direction. Specification of the stimulus bandwidth and frequency, as well as fine adjustment of the stimulus duration and envelope shaping, was accomplished through a Coulbourn Instruments modular system. The subject responded using a specially designed laptop response box with a circular array of microswitches that was configured exactly like the speaker array under test, in number of elements and azimuth angles. Thus, subjects had access to a visual representation of both frontal and rearward speakers at all times.

TABLE II. Hearing thresholds (dB HL) for the seven age groups.

	Frequency (kHz)			
Age (yrs)	0.5	1	2	4
10–19	2.9(3.3)[a]	1.4(3.7)	0.8(3.8)	−1.3(5.4)
20–29	−1.3(2.8)	0.4(2.2)	2.0(4.7)	−1.2(5.4)
30–39	1.3(6.0)	2.7(6.3)	2.6(6.3)	0.5(6.1)
40–49	0.0(3.6)	3.4(3.6)	5.8(5.8)	6.8(10.4)
50–59	5.7(4.2)	5.4(4.6)	6.8(4.4)	11.5(7.4)
60–69	7.4(7.3)	8.7(8.6)	11.9(7.7)	18.6(9.2)
70–81	11.3(8.3)	11.6(7.8)	18.7(8.8)	24.4(11.3)

[a]Mean (s.d.).

D. Procedure

One block of trials was given for each of the 18 array-by-speaker configuration-by-stimulus frequency listening conditions. The four-speaker arrays preceded the eight-speaker arrays. Order of loudspeaker configuration, and frequency within configuration, were counterbalanced across the 16 subjects in the group. Within a block, the stimulus was presented randomly from each of the four or eight speakers on 16 trials/speaker. Each trial began with a $\frac{1}{2}$-s warning light on the subject's response box, followed by a $\frac{1}{2}$ -s delay, and then the presentation of the 300-ms stimulus. Subjects were instructed to focus on a straight-ahead visual target mounted on the wall of the booth and to keep the head steady during the presentation of the stimulus. This procedure ensured that the speaker array and head-related coordinate system were aligned, and also minimized the use of head movements to resolve possible front/back confusions (Wightman and Kistler, 1999).

A forced-choice speaker identification paradigm was used, in which the response azimuths corresponded to the stimulus azimuths (see Green and Swets, 1966). The subject's task was to indicate the direction of the sound by depressing one of the alternative response buttons. A maximum of 7 s was given for the response. Guessing was encouraged and no feedback was given about the correctness of the judgment. The loudspeaker that emitted the stimulus and the response button selected were automatically stored on each trial.

II. RESULTS

Table II shows the mean hearing thresholds (converted to dB HL) from 0.5–4 kHz, averaged across left and right ears, for each of the seven groups of 16 subjects. Within subject, between-ear differences were no greater than 10 dB at any frequency, except in seven individuals aged 70–81 years. In these, 11–12-dB differences were observed at one frequency (in five at 4 kHz). All subjects met the admission guidelines, except for four cases. Two subjects aged 47 and

TABLE III. Percent-correct sound localization scores for the seven age groups. Effects of number of speakers, speaker configurations and stimulus.

			Age group (yrs)						
Array	Config.	Stimulus	10–19	20–29	30–39	40–49	50–59	60–69	70–81
4-spk	midline	0.5 kHz	57(11)[a]	55(15)	55(7)	55(7)	51(7)	50(11)	42(8)
		4.0 kHz	79(14)	76(19)	62(17)	66(17)	67(18)	66(12)	56(13)
		BB	99(1)	100(1)	99(3)	98(6)	94(13)	95(10)	88(18)
	lateral	0.5 kHz	56(14)	55(10)	55(9)	54(7)	53(10)	49(5)	48(7)
		4.0 kHz	74(17)	70(13)	60(14)	65(16)	63(15)	64(14)	55(12)
		BB	90(8)	93(9)	91(10)	92(8)	85(13)	82(12)	76(15)
8-spk	15° sep	0.5 kHz	43(10)	46(12)	45(9)	41(8)	41(9)	36(9)	33(8)
		4.0 kHz	60(11)	61(14)	60(16)	53(14)	50(13)	50(15)	51(11)
		BB	85(7)	89(4)	92(3)	84(7)	84(9)	84(7)	78(11)
	30° sep	0.5 kHz	50(9)	51(11)	52(7)	45(8)	45(10)	39(8)	34(9)
		4.0 kHz	72(13)	69(21)	65(18)	58(15)	58(16)	57(16)	52(8)
		BB	92(5)	94(4)	97(3)	92(6)	91(8)	89(10)	85(9)
	45° sep	0.5 kHz	49(9)	53(11)	54(8)	47(9)	46(6)	44(9)	36(8)
		4.0 kHz	72(13)	70(20)	66(17)	61(16)	61(17)	62(16)	50(12)
		BB	93(6)	96(3)	97(5)	94(7)	91(9)	89(10)	84(9)
	60° sep	0.5 kHz	46(10)	51(9)	53(8)	47(8)	49(10)	44(9)	39(8)
		4.0 kHz	67(17)	65(18)	65(12)	56(14)	60(15)	57(15)	53(10)
		BB	90(7)	92(6)	95(4)	91(9)	87(13)	86(8)	80(12)

[a]Mean (s.d.).

53 years exceeded the 35 dB SPL (23 dB HL) guideline by 19 and 13 dB (worse ear), respectively, at 4 kHz. Two subjects aged 70 years exceeded the 50 dB SPL (38 dB HL) guideline by 7 and 12 dB (worse ear), respectively, at 4 kHz. However, based on the reference equivalent threshold sound-pressure levels for binaural listening in free field given in ANSI S3.6-1996, the sound localization stimuli were sufficiently loud (75 dB SPL) to be well above threshold for all subjects.

A. Accuracy in sound-source identification

Table III shows the mean percent correct, $P(C)$, sound-source identification judgments for the 16 subjects in each group for each of the 18 array-by-configuration-by-stimulus listening conditions. Over the age range studied, $P(C)$ decreased by 7%–23% depending on the combination of number of speakers, speaker configuration, and stimulus to be localized. Standard deviations were similar for the seven groups, ranging nonsystematically from 1%–21% across conditions. In the case of the four-speaker arrays, two binary decisions were required for accurate sound-source identification, left vs right, and front vs back. By chance subjects could achieve a $P(C)$ of 25%. Inspection of the data revealed that there was little difficulty in distinguishing left from right. Thus, accuracy scores reflect the degree of difficulty with front/back discrimination. A score of 50% signifies chance performance in this dimension. Averaged across age groups and speaker configurations, the $P(C)$s observed for broadband noise, 4 kHz, and 0.5 kHz were 92%, 66%, and 53%, respectively. Thus, subjects were most likely to confuse front and back with the 0.5-kHz stimulus and least likely with the broadband noise. For the eight-speaker arrays, three binary decisions were necessary for accurate sound-source identification, left versus right, front versus back, and midline versus lateral speaker within quadrant. By chance subjects could achieve a $P(C)$ of 12.5%. Averaged across groups and speaker configurations, the observed $P(C)$s were 89%, 60%, and 45% for broadband noise, 4 kHz, and 0.5 kHz, respectively. The difference between the outcomes for the four-and eight-speaker arrays, 3%–8%, largely reflects confusion of the two speakers within each quadrant.

To determine whether an age-related change in hearing threshold might account for the decrease in accuracy in sound-source identification with aging, multiple linear regression analyses were applied to the data for the 112 subjects for each of the four- and eight-speaker arrays (averaged across speaker configurations) for each of the three stimulus frequencies (Howell, 1997). Table IV shows the proportion of variance (R^2) in $P(C)$ accounted for by changes in age, hearing threshold (HTH), and age in combination with HTH. For the 0.5-kHz and 4-kHz stimuli, HTH was calculated as the average of the hearing thresholds (in dB HL) measured for the two ears at 0.5 and 4 kHz, respectively. For the broadband noise, HTH was the average of the thresholds obtained at 0.5, 1, 2, and 4 kHz, averaged for the two ears.

In order to examine the possibility that the effect of age on sound localization was mediated by hearing sensitivity, a series of multiple regression analyses was run. The point was to compare the proportion of variance in $P(C)$ that could be accounted for by hearing sensitivity with the proportion of variance that could be accounted for by also considering age. If age adds to the ''prediction'' of sound localization performance when hearing has already been taken into account, then it is reasonable to assume that sound localization is affected by age, independent of hearing sensitivity. For the simple regression model with only hearing threshold (HTH) as the predictor, the R^2 values shown for HTH in Table IV

TABLE IV. Outcomes of the multiple regression analyses. Proportion of variance accounted for (R^2) is shown for the factors age, hearing threshold (HTH), and age+HTH for the different array and stimulus combinations employed.

Array	Stimulus	Age R^2	HTH R^2	Age+HTH R^2	$F(df)$[a]	p
4-spk	0.5 kHz	0.20	0.10	0.21	15.00(1,109)	0.001
	4.0 kHz	0.13	0.03	0.15	15.37(1,109)	0.001
	BB	0.26	0.24	0.29	7.20(1,109)	0.01
8-spk	0.5 kHz	0.22	0.13	0.24	16.16(1,109)	0.001
	4.0 kHz	0.12	0.02	0.15	17.22(1,109)	0.001
	BB	0.20	0.21	0.23	3.26(1,109)	0.10

[a] $F_{(f-r,N-f-1)}=[(N-f-1)(R_f^2-R_r^2)]/[(f-r)(1-R_f^2)]$, where $f=2$ (HTH, Age), $r=1$ (HTH), $N=112$.

were significant for the 0.5-kHz and broadband stimuli but not the 4-kHz stimulus, for the four- and eight-speaker arrays. When age was included in the model, the R^2 values increased. The difference in R^2 was significant at the 0.01 level or better for all conditions except for the eight-speaker array with broadband noise. In the latter case the F statistic was borderline significant ($0.05<p<0.10$). These findings indicate an effect of age on localization even when hearing sensitivity was taken into account.

Nested analyses of variance (ANOVAs) (Daniel, 1983) were applied to the $P(C)$s observed for the four- and eight-speaker arrays to evaluate the effects of age (seven groups), speaker configuration (midline versus interaural) or separation (15°, 30°, 45°, and 60°), and stimulus frequency (0.5, 4, and broadband noise). The ANOVA for the four-speaker arrays showed that there were significant effects of age group [$F(6,105)=11.30$, $p<0.001$], speaker configuration [$F(1,105)=18.74$, $p<0.001$], stimulus frequency [$F(2,210)=524.55$, $p<0.001$], and speaker configuration by frequency [$F(2,210)=15.64$, $p<0.001$]. $P(C)$ decreased by 15% over the age range studied. Averaged across the seven age groups, $P(C)$ was highest for the broadband noise and lowest for the one-third-octave noise band centered at 0.5 kHz, with a difference of 40%. *Post hoc* pairwise comparisons using Fisher's LSD test (Daniel, 1983) indicated that, averaged across groups, there were significant differences among the $P(C)$s for the three stimuli ($p<0.05$). A significant advantage (9%) for the midline compared with the lateral configuration of speakers was apparent only for the broadband noise ($p<0.05$).

In the case of the eight-speaker arrays, the data were partitioned, and a series of nested ANOVAs was carried out. First, the effects of age (seven groups) and speaker separation (15°, 30°, 45°, and 60°) were analyzed for each of the three stimulus frequencies. The trends were similar: for 0.5 kHz, significant effects of age group [$F(6,105)=7.86$, $p<0.001$] and separation [$F(3,315)=32.39$, $p<0.001$]; for 4 kHz, significant effects of age group [$F(6,105)=3.12$, $p<0.001$], separation [$F(3,315)=18.15$, $p<0.001$] and age by separation [$F(18,315)=1.74$, $p<0.025$]; and for broadband noise, significant effects of age group [$F(6,105)=7.61$, $p<0.001$] and separation [$F(3,315)=44.32$, $p<0.01$]. Decrements in $P(C)$ with aging were 15%, 17%, and 13% for 0.5, 4, and broadband noise, respectively. *Post hoc* analyses indicated that $P(C)$ did not decrease systematically as age increased from the first to the third decade. Significant decrements ($p<0.05$) were found between the third and seventh decade. Differences due to speaker separation, averaged across age groups, ranged from 6%–8% across the three stimuli. Regardless of stimulus frequency, $P(C)$ for the 15° separation was significantly less than $P(C)$s for the other three separations ($p<0.05$), which were not different. Only for the 4-kHz stimulus was there a significant interaction of age by speaker separation, with $P(C)$ decreasing significantly ($p<0.05$) by 9%, 20%, 22%, and 14% for the 15°, 30°, 45°, and 60° separations, respectively, over the age range studied.

The final nested ANOVA compared differences due to stimulus frequency and separation for subjects in the youngest, middle, and oldest age groups, i.e., those in the the first, fourth, and seventh decades. There were significant effects of age group [$F(2,45)=10.95$, $p<0.001$], speaker separation [$F(3,135)=27.46$, $p<0.001$], stimulus frequency [$F(2,90)=662.93$, $p<0.001$], age group by stimulus frequency [$F(4,90)=3.87$, $p<0.01$], age group by separation [$F(6,135)=2.61$, $p<0.025$], and age by separation by frequency [$F(12,270)=1.99$, $p<0.025$]. $P(C)$ was highest for broadband noise and lowest for 0.5 kHz, the difference ranging from 43%–46% across the three groups. *Post hoc* pairwise comparisons indicated that there were significant differences ($p<0.05$) among the three frequencies for each of the age groups examined.

B. Erroneous responses in sound-source identification

Detailed analyses of erroneous responses were confined to the data for the eight-speaker array. As indicated above, differences in $P(C)$ for the four- and eight-speaker arrays were largely the result of confusion between speakers within quadrant for the latter of these two conditions. Figures 1 and 2 show age-related breakdowns of the types of erroneous response observed for the four quadrants for each of the 0.5-kHz and 4-kHz stimuli, for the eight-speaker array. The results have been averaged for the two speakers within each quadrant and the four speaker separations. Each bar in the histogram contains the mean percent-correct speaker identifications, $P(C)$, together with the percentages of four types of error. $P(RE)$ is the mean percentage of front/back mirror image reversal errors (e.g., 15° vs 165°), $P(QE)$ is the mean percentage of within-quadrant errors (e.g., 15° vs 45°), $P(SE)$ is the remaining error on the correct side of midline

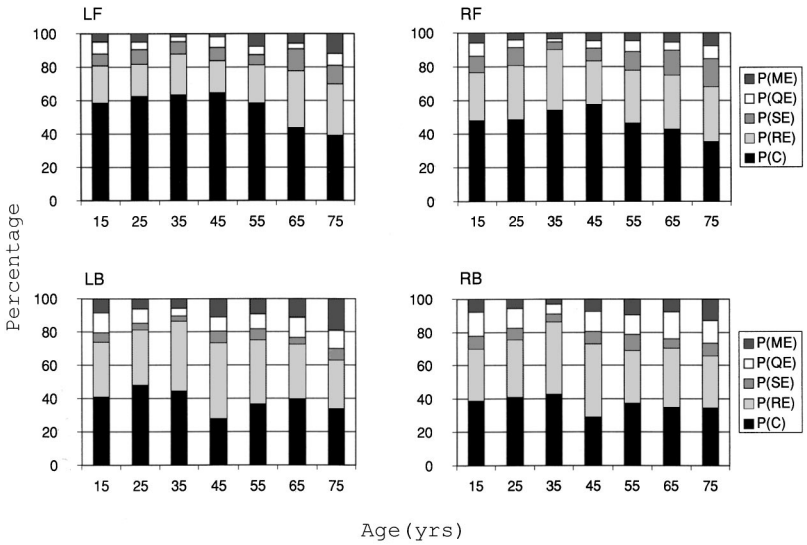


FIG. 1. Distribution of responses for 0.5 kHz for the eight-speaker arrays for the four quadrants (LF—left front, RF—right front, LB—left back, and RB—right back) as a function of age. $P(C)$—percent correct, $P(RE)$—percent front/back mirror image reversal error, $P(QE)$—percent within-quadrant error, $P(SE)$—percent within-side errors other than *RE* and *QE*, and $P(ME)$—percent opposite side errors are compared.

(e.g., 15° vs 135°), and finally, $P(ME)$ is any error involving the wrong side of midline (e.g., 15° vs −45°). The five components sum to 100%.

At 0.5 kHz (Fig. 1), $P(C)$ was relatively higher for the frontal than the rearward quadrants by 14%, on average. This was mainly due to an 8% greater $P(RE)$ for the rearward quadrants, and to a lesser extent by increases in $P(QE)$ and $P(ME)$. In front, $P(C)$ was about 60% over the first five age decades, declining to 40% in the last two decades, on the left side of space. This decline was mainly the result of a relative increase in $P(RE)$ with age. $P(C)$ was relatively lower on the right than on the left over the first five decades, again mostly due to a greater $P(RE)$ on the right. In back, distributions of responses on the left and right were similar and did not vary systematically with age.

At 4 kHz (Fig. 2), $P(C)$ was relatively higher for the frontal quadrants than the rearward quadrants by 22%, on average. This was due to higher values for $P(ME)$, $P(QE)$, and $P(RE)$ of about 10%, 7%, and 5%, respectively, for the rearward quadrants. In particular, while $P(ME)$ was slightly less than 2% in the front, it accounted for almost 12% of responses in back. Older groups were especially prone to this type of error. In front, $P(C)$ did not show systematic changes with aging but was somewhat higher on the left side than the right by 6%. As for the 0.5-kHz stimulus, this was due primarily to a higher $P(RE)$ on the right. In back, $P(C)$ decreased progressively from the third decade to reach a minimum of about 35% by the seventh decade, as a result of increases in $P(RE)$, $P(SE)$, and $P(ME)$. The distributions of responses on the left and right sides were fairly similar.

For the broadband noise, a similar error analysis revealed that $P(C)$ was relatively higher in the frontal than the rearward quadrants by 12%, on average. This was mainly due to an 8% greater $P(QE)$ for the rearward quadrants. In front, scores almost reached the ceiling value of 100%. $P(C)$ was about 96% over the first six age decades, declining to 90% in the seventh decade, on the left and right sides of space. This was mainly due to a greater value for $P(RE)$. In back, $P(C)$ increased slightly from the first to the third decade where it reached 93%, then decreased progressively to 74% in the seventh decade due to increases in $P(QE)$, $P(RE)$, and $P(ME)$. $P(C)$ was relatively lower on the right than on the left side of space in the back, by 4% on average, but this asymmetry did not vary systematically with age.

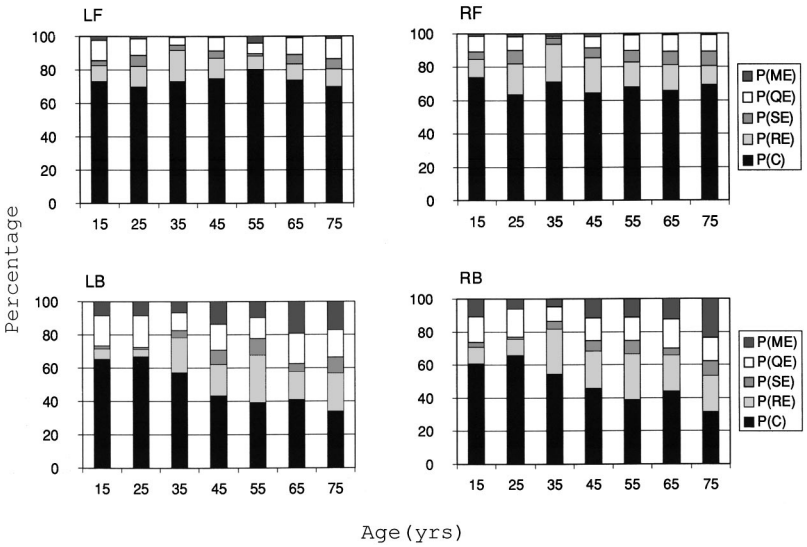


FIG. 2. Distribution of responses for 4 kHz. As in Fig. 1.

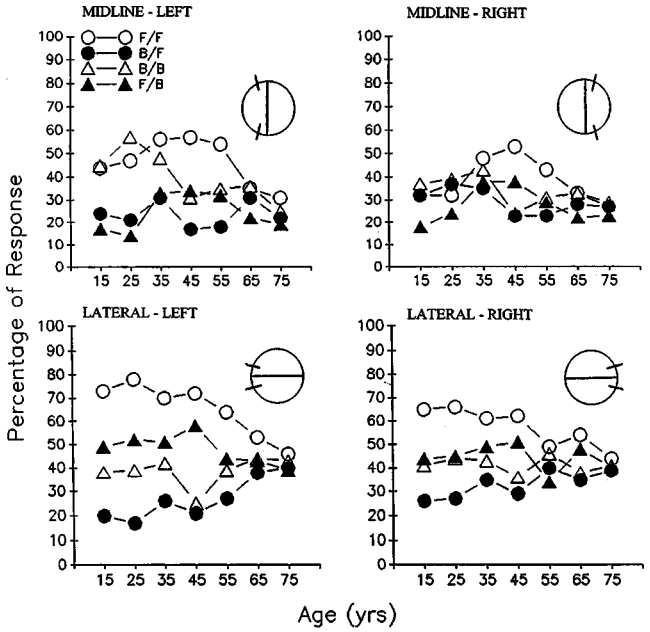


FIG. 3. The effect of age on front/back discrimination of speakers located near the midline and interaural axes, on the left and right, in the eight-speaker arrays. Results are presented for 0.5 kHz. Percentages of correct sound-source identification (F/F—front given front and B/B—back given back) and front/back mirror image reversals (F/B—front given back and B/F—back given front) are compared.

The data shown in Figs. 1 and 2 suggested that front/back mirror image reversal errors, $P(RE)$, were likely responsible for the quadrant differences in accuracy. Figures 3 and 4 show just the $P(C)$s and $P(RE)$s for speakers located close to the midline (midline speakers) and interaural axes (lateral speakers) for 0.5 and 4 kHz, respectively, for the eight-speaker array. The results have been averaged for the four separations. Across separations, the midline speakers did not change position (top panels). The position of the lateral speakers (bottom panels) was linked to separation. A nested ANOVA on the two types of responses for the midline speakers with the 0.5-kHz stimulus (Fig. 3, upper panels) for subjects in the youngest, middle, and oldest groups, i.e., those in the first, fourth, and seventh age decades, confirmed that there were statistically significant effects of age [$F(2,45)=4.38$, $p<0.025$], age by side [$F(2,45)=4.10$, $p<0.025$], response type, i.e., correct response or mirror image reversal error [$F(3,135)=4.32$, $p<0.01$], and side by response type [$F(3,135)=7.70$, $p<0.001$]. As indicated in the figure, on the left, subjects showed some ability to distinguish the midline speaker in front (F/F) until the fifth decade and the midline speaker in back (B/B) until the third decade, after which they were equally likely to respond front and back (upper-left panel). On the right, front and back were confused by all age groups (upper right panel). An ANOVA on the data for the lateral speakers produced significant effects of response type [$F(3,135)=11.88$, $p<0.001$], age by response type [$F(6,135)=2.49$, $p<0.025$], and side by response type [$F(3,135)=3.60$, $p<0.025$]. Inspection of the figure suggests that for the speaker in front of the ear, there was an age-related decline in accuracy (F/F), accompanied by an increase in front-to-back (B/F) mirror image reversal errors, on both the left and right (lower panels). Generally, sounds coming from the speaker behind the ear were equally likely to be called front (F/B) and back (B/B).

ANOVAs applied to the $P(C)$s and $P(RE)$s for the 4-kHz stimulus (Fig. 4, upper panels) for the midline speakers indicated that there were statistically significant effects of response type [$F(3,135)=58.16$, $p<0.001$], age by response type [$F(6,135)=2.13$, $p<0.05$], and side by response type [$F(3,135)=2.85$, $p<0.05$]. For the lateral speakers, there were significant effects of age [$F(2,45)=4.24$, $p<0.025$], side [$F(1,45)=5.87$, $p<0.025$], response type [$F(3,135)=96.21$, $p<0.001$], and age by response type [$F(6,135)=4.60$, $p<0.001$]. As shown in the figure, the ability to localize speakers in front (F/F), whether proximal to the midline (upper panels) or the interaural axis (lower panels), did not change systematically with age. Progressive age-related decrements, accompanied by increases in back-to-front (F/B) mirror image reversal errors, in localizing the four rearward speakers, were evident by the third decade.

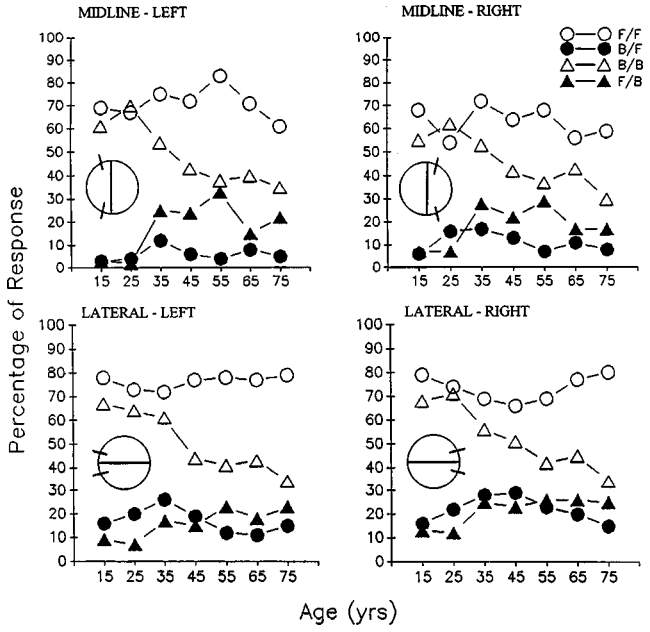


FIG. 4. The effect of age on front/back discrimination of speakers located near the midline and interaural axes on the left and right. Results are presented for the 4-kHz stimulus. As in Fig. 3.

III. DISCUSSION

There are few studies in the literature on the effects of aging on sound localization. Cranford *et al.* (1990) studied the precedence effect in young and elderly subjects with normal hearing or bilateral sensorineural hearing loss. Older subjects with normal hearing performed significantly more poorly, i.e., the percentage of correct identifications of the leading speaker was significantly less, for interspeaker delays in the range of 0.1–0.5 ms. Still greater deficits were observed with hearing loss. Abel and Hay (1996) tested normal-hearing subjects under and over the age of 40 years in a six-speaker sound-source identification experiment. Aging resulted in a decrement in front/back discrimination of about 10% for both high- and low-frequency test sounds. In

contrast, left/right discrimination was close to 100%. The results of the present experiment clearly demonstrated that sound localization ability deteriorated from adolescence through senescence. The effect was relatively small, with $P(C)$ decreasing on average by 15% and 12% over the age range studied for the four- and eight-speaker arrays, respectively. Decrements were observed for all three stimuli, suggesting that the processing of both binaural and spectral cues could be involved.

In any study of auditory perception in which age effects are identified, the question of hearing loss due to presbycusis as a contributor to the deficit must be considered (Marshall, 1981). In the present study, subjects were screened for moderate to severe hearing loss. However, mild age-related changes were acceptable and noted at both 0.5 and 4 kHz (Yantis, 1985). Multiple regression analyses confirmed that there was a significant effect of age even when hearing sensitivity was taken into account, supporting the conclusion that sound localization was affected by central deficits with aging. This conclusion is in line with previously reported findings of central age-related deficits in such tasks as duration discrimination and sound lateralization (Abel *et al.*, 1990; Schneider *et al.*, 1994; Strouse *et al.*, 1998). The present results extend previous findings of aging effects to binaural processing and sound localization in rooms.

The age-related deficits associated with the broadband noise and 4-kHz stimulus were largely the result of progressive difficulty in identifying sound sources located behind. A surprising outcome was that performance decrements began as early as the third decade of life. For the 0.5-kHz stimulus, accuracy scores for the speakers in back remained fairly constant at a relatively low level of 40% across age. Accuracy in identifying the frontal speakers began to decline by the fifth decade from 60% to 40%. Previous studies have suggested that errors might be the result of front/back mirror image reversal errors which differ on the left and right sides of space (Burke *et al.*, 1994; Butler, 1994; Abel *et al.*, 1999). In the case of the 0.5-kHz stimulus, which provided primarily the interaural temporal difference cue, left/right differences in accuracy favoring the left were clearly apparent. On the right, speakers close to the midline, front and back, were not discriminable, whatever the age; on the left, front and back were differentiable until the fourth decade. The ability to localize the lateral speaker in front of the ear decreased with aging starting at the fourth decade but was better on the left. With aging, there was an increase in front/back mirror image reversal errors. For the 4-kHz stimulus, which provided primarily the interaural level difference cue, errors in identifying speakers close to both the midline and interaural axes, in back, were apparent by the third decade. Back-to-front errors increased in prevalence only for the speakers close to the midline axis. Back-to-front and front-to-back errors were equally likely at 10%–20% for the lateral speakers. Reversal errors were rare in the case of the broadband stimulus. Errors in back, mainly associated with the speaker close to the midline, were predominately within-quadrant confusions.

To further characterize the size of left/right and front/back differences in sound localization and the effect of aging on these differences, the data were rescored according to a head-related coordinate system consistent with the cone of confusion (Morimoto *et al.*, 1983; Giguère and Abel, 1993; Abel *et al.*, 1999). Each speaker was assigned two coordinates: (1) a positive angle α between the interaural axis and a straight line connecting the speaker with the center head position, and (2) a binary hemifield classification of ''front'' or ''back.'' In this scheme, mirror image sound sources with respect to the interaural axis differ only in terms of the front/back classification. This provides a framework for assessing azimuthal performance (angle α) independently of front/back decisions. Figure 5 shows the percentage of correct judgments of angle α (upper panels) and the percentage of correct hemifield classifications (lower panels). For the eight-speaker array there were four possible values of angle α and two values for the hemifield classification (front or back). The results were averaged across the two speakers in each quadrant and the four separations.

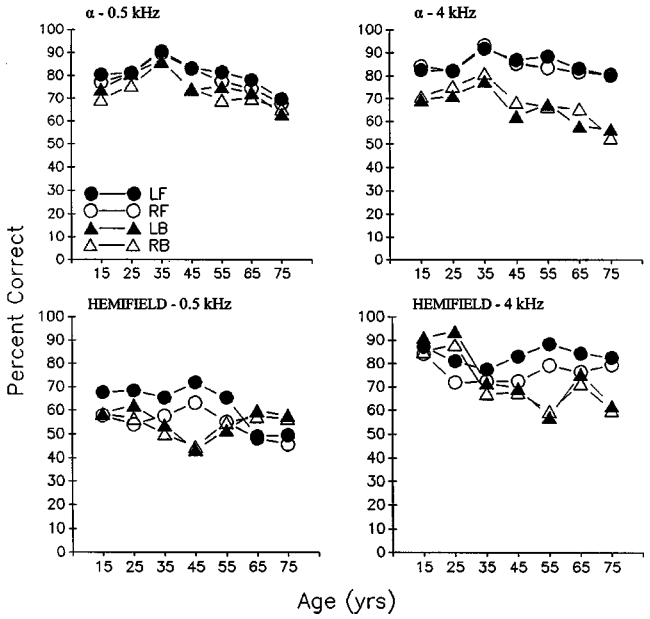


FIG. 5. Percent correct judgments of angle α and and front/back hemifield at 0.5 and 4 kHz for the eight-speaker arrays. Within each panel, the results are presented for each of the four quadrants. Chance performance is 25% correct for the perception of angle α and 50% correct for the hemifield classification. Note that the percentage correct for angle α corresponds to the sum of the $P(C)$ and $P(RE)$ bars in Figs. 1 and 2.

As indicated in Fig. 5, there is no consistent left/right difference in the perception of angle α for either 0.5 or 4 kHz. There is, however, a relatively large difference between the front and rearward quadrants of 7%–27% at 4 kHz, favoring the front. This may indicate poorer encoding of interaural level differences in back. At 0.5 kHz, the difference between front and back is smaller, but again favors the front by up to 9%. At both frequencies, the correct perception of angle α peaks in the third decade and then decreases. This may indicate a progressive inability to process interaural temporal and level differences later in life.

In contrast, a left/right difference in the front/back hemifield classification is clearly evident at both frequencies for the frontal quadrants. At 4 kHz, there is a 3%–11% left-sided advantage at 4 kHz across age, and at 0.5 kHz, an 8%–15% left-sided advantage over the first five decades in

front. Thus, the main source of left/right asymmetry in the sound localization data is associated with front/back discrimination. This outcome confirms and extends the findings of Burke *et al.* (1994) and Abel *et al.* (1999) over a wide age range. Since front/back discrimination depends on the extraction of spectral cues from the stimulus (Blauert, 1997), the left-sided advantage points to a right cerebral hemisphere advantage for processing spectral information (Butler, 1994). The decline in the discrimination of front from back with age supports the conclusion that the processing of spectral information becomes progressively less efficient with aging, beginning as early as the third decade.

IV. CONCLUSIONS

(1) Sound-source identification in the horizontal plane decreased by 12%–15% from adolescence and young adulthood to old age, and was clearly apparent by the third decade of life. This finding derives from a forced-choice experiment with a fixed response set that was chosen to explore left/right and front/back mirror image confusions. Both the paradigm and semireverberant listening environment were typical of real-world scenarios.

(2) Age-related decrements were found for all three stimuli tested: one-third-octave noise bands centered at 0.5 kHz (primarily interaural temporal difference cues available) and 4 kHz (primarily spectral and interaural level difference cues available), and broadband noise (both binaural and spectral cues available). Performance was best with broadband noise and worst with the 0.5-kHz noise band.

(3) Deficits in sound localization with aging were greater on the right side of space and were largely the result of front/back mirror image reversal errors. The patterns observed support the conclusion that the processing of spectral information is a primary factor underlying the aging effect and the left/right asymmetry in sound localization.

ACKNOWLEDGMENTS

This research was supported by a grant from the Physicians' Services Incorporated Foundation of Ontario, Canada, and in part by a senior scientist award (SMA) from The Saul A. Silverman Family Foundation. The authors are grateful to Dr. Ian Spence and Dr. Daniel Ashmead for statistical consultation and to Mr. Adam Trottola for his help in carrying out the analyses. They also thank Dr. Wesley Grantham, Dr. Daniel Ashmead, and two anonymous reviewers for their helpful comments on earlier drafts of the manuscript.

Abel, S. M., Krever, E. M., and Alberti, P. W. (**1990**). ''Auditory detection, discrimination and speech processing in ageing, noise-sensitive and hearing-impaired listeners,'' Scand. Audiol. **19**, 43–54.

Abel, S. M., and Banerjee, J. (**1996**). ''Accuracy vs choice response time in sound localization,'' Appl. Acoust. **49(4)**, 405–417.

Abel, S. M., Giguère, C., Consoli, A., and Papsin, B. C. (**1999**). ''Front/back mirror image reversal errors and left/right asymmetry in sound localization,'' Acustica **85(3)**, 378–386.

Abel, S. M., and Hay, V. H. (**1996**). ''Sound localization: The interaction of HPDs, aging and hearing loss,'' Scand. Audiol. **25**, 3–12.

Abel, S. M., and Kunov, H. (**1983**). ''Lateralization based on interaural phase differences: Effects of frequency, amplitude, duration and rise/decay,'' J. Acoust. Soc. Am. **73**, 955–960.

ANSI (**1996**). ANSI S3.6-1996, ''Specification for Audiometers'' (American National Standards Institute, New York).

Blauert, J. (**1997**). *Spatial Hearing* (MIT Press, Cambridge, MA).

Brant, L. J., and Fozard, J. L. (**1990**). ''Age changes in pure-tone hearing thresholds in a longitudinal study of normal human aging,'' J. Acoust. Soc. Am. **88**, 813–820.

Burke, K. A., Letsos, A., and Butler, R. A. (**1994**). ''Asymmetric performance in binaural localization of sound in space,'' Neuropsychol. **32**, 1409–1417.

Butler, R. A. (**1994**). ''Asymmetric performances in monaural localization of sound in space,'' Neuropsychol. **32**, 221–229.

Casey, M. A., and Feldman, M. L. (**1982**). ''Aging in the rat medial nucleus of the trapezoid body. I. Light microscopy,'' Neurobiol. Ag. **3**, 187–195.

Clifton, R. K. (**1992**). ''The development of spatial hearing in human infants,'' in *Developmental Psychoacoustics*, edited by L. A. Werner and E. W. Rubel (APA, Washington), pp. 135–157.

Cranford, J. L., Boose, M., and Moore, C. A. (**1990**). ''Effects of aging on the precedence effect in sound localization,'' J. Speech Hear. Res. **33**, 654–659.

Daniel, W. W. (**1983**). *Biostatistics: A Foundation for Analysis in the Health Sciences*, 3rd ed. (Wiley, New York).

Giguère, C., and Abel, S. M. (**1990**). ''A multi-purpose facility for research on hearing protection,'' Appl. Acoust. **31**, 295–311.

Giguère, C., and Abel, S. M. (**1993**). ''Sound localization: Effects of room reverberation, speaker array, stimulus frequency and stimulus rise/decay,'' J. Acoust. Soc. Am. **94**, 769–776.

Good, M., and Gilkey, R. H. (**1996**). ''Sound localization in noise: The effect of signal-to-noise ratio,'' J. Acoust. Soc. Am. **99**, 1108–1117.

Grantham, D. W. (**1984**). ''Interaural intensity discrimination: Insensitivity at 1000 Hz,'' J. Acoust. Soc. Am. **75**, 1191–1194.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Hartmann, W. M. (**1983**). ''Localization of sounds in rooms,'' J. Acoust. Soc. Am. **74**, 1380–1391.

Herman, G. E., Warren, L. R., and Wagener, J. W. (**1977**). ''Auditory lateralization: Age differences in sensitivity to dichotic time and amplitude cues,'' J. Gerontol. **32**, 187–191.

Howell, D. C. (**1997**). *Statistical Methods for Psychology*, 4th ed. (Duxbury, Belmont, CA).

Johnsson, L. G., and Hawkins, Jr., J. E. (**1972**). ''Sensory neural degeneration with aging, as seen in microdissections of the human ear,'' Ann. Otol. **81**, 179–193.

Kirikae, I., Sato, T., and Shitara, T. (**1964**). ''A study of hearing in advanced age,'' Laryngoscope **74**, 205–220.

Klumpp, R. G., and Eady, H. R. (**1956**). ''Some measurements of interaural time difference thresholds,'' J. Acoust. Soc. Am. **28**, 859–860.

Lorenzi, C., Gatehouse, S., and Lever, C. (**1999**). ''Sound localization in noise in normal-hearing listeners,'' J. Acoust. Soc. Am. **105**, 1810–1820.

Marshall, L. (**1981**). ''Auditory processing aging listeners,'' J. Speech Hear. Dis. **8**, 226–240.

Mills, A. W. (**1958**). ''On the minimum audible angle,'' J. Acoust. Soc. Am. **30**, 237–246.

Mills, A. W. (**1972**). ''Auditory localization,'' in *Foundations of Modern Auditory Theory*, edited by J. W. Tobias (Academic, New York), Vol. 2, pp. 301–348.

Morimoto, M., Aokata, H., and Maekawa, Z. (**1983**). ''Sound localization on the upper hemisphere,'' in Proceedings of 11th International Congress on Acoustics, Paris, France, Vol. 3, pp. 143–146 (unpublished).

Morrongiello, B. A. (**1988**). ''Infants' localization of sounds along the horizontal axis: Estimates of minimum audible angle,'' Develop. Psychol. **24**, 8–13.

Morrongiello, B. A., Fenwick, K. D., Hillier, L., and Chance, G. (**1994**). ''Sound localization in newborn human infants,'' Dev. Psychobiol. **27(8)**, 519–538.

Morrongiello, B. A., and Rocca, P. T. (**1990**). ''Infants' localization of sounds within hemifields: Estimates of minimum audible angle,'' Child Dev. **61**, 1258–1270.

Muir, D., Clifton, R. K., and Clarkson, M. G. (**1989**). ''The development of a human auditory localization response: A U-shaped function,'' Can. J. Psychol. **43**, 199–216.

Muir, D., and Field, J. (**1979**). ''Newborn infants orient to sounds,'' Child Dev. **50**, 431–436.

Musicant, A. D., and Butler, R. A. (**1984**). ''The influence of pinnae-based spectral cues on sound localization,'' J. Acoust. Soc. Am. **75**, 1195–1200.

Noble, W., Byrne, D., and Lepage, B. (**1994**). ''Effects on sound localization of configuration and type of hearing impairment,'' J. Acoust. Soc. Am. **95**, 992–1005.
Rakerd, B., and Hartmann, W. M. (**1986**). ''Localization of sound in rooms. III. Onset and duration effects,'' J. Acoust. Soc. Am. **80**, 1695–1706.
Schneider, B. A., Pichora-Fuller, M. K., Kowalchuk, D., and Lamb, M. (**1994**). ''Gap detection and the precedence effect in young and old adults,'' J. Acoust. Soc. Am. **95**, 980–991.
Stevens, S. S., and Newman, E. B. (**1936**). ''The localization of actual sources of sound,'' Am. J. Psychol. **48**, 297–306.
Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (**1998**). ''Temporal processing in the aging auditory system,'' J. Acoust. Soc. Am. **104**, 2385–2399.
Wightman, F. L., and Kistler, D. J. (**1992**). ''The dominant role of low-frequency interaural time differences in sound localization,'' J. Acoust. Soc. Am. **91**, 1648–1661.
Wightman, F. L., and Kistler, D. J. (**1999**). ''Resolution of front–back ambiguity in spatial hearing by listener and source movement,'' J. Acoust. Soc. Am. **105**, 2841–2853.
Yantis, P. A. (**1985**). ''Puretone air-conduction testing,'' in *Handbook of Clinical Audiology*, 3rd ed., edited by J. Katz (Williams & Wilkins, Baltimore), pp. 153–169.

Binaural effects in center-frequency modulation detection interference for vowel formants

J. Lyzenga[a] and R. P. Carlyon
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, United Kingdom

(Received 21 April 2000; accepted for publication 8 May 2000)

The detection of slow (5 Hz) center-frequency modulations of formants (signals) can be impaired by the simultaneous presentation of off-frequency modulated formants (maskers) to the same ear [J. Lyzenga and R. P. Carlyon, J. Acoust. Soc. Am. **105**, 2792–2806 (1999)]. In the present study we examine this ''formant-frequency modulation detection interference (FMDI)'' for various binaural masker presentation schemes. Signals and maskers were formantlike complex tones, centered around 1500 and 3000 Hz, respectively. Fundamentals of 80 and 240 Hz were used. The signals were presented to the right ear. The maskers were presented either to the right, the left, or to both ears, and they were either unmodulated or modulated at a slow rate (10 Hz). They had the same fundamental as the signals. Hardly any interference was found for the unmodulated maskers. For modulated maskers, the amount of FMDI depended strongly on the binaural masker presentation scheme. Substantial interference was found for the ipsilateral maskers. Interference was smaller for the contralateral maskers. In both cases the FMDI increased with increasing masker level. Substantial interference was also found for the binaural maskers. Imposing different interaural time and level differences (ITDs and ILDs) on maskers and signals did not affect FMDI. The same was true for the ITD condition when the maskers had different fundamentals than the signals, though FMDI was slightly smaller here. The amount of interference for the binaural maskers was roughly equal to that of the corresponding monaural masker with the largest effect. The data could not be described accurately using a model based on the loudness of the maskers. On the other hand, they were well described by a model in which the amount of FMDI was predicted from a ''weighted combination'' of the monaural masker levels.
[S0001-4966(00)03508-6]

PACS numbers: 43.66.Pn, 43.66.Mk, 43.66.Fe, 43.71.Es [RVS]

INTRODUCTION

In a previous study (Lyzenga and Carlyon, 1999) we reported center-frequency modulation detection interference (FMDI) for formants for slow modulation rates (5 to 10 Hz). We presented normal-hearing listeners with formantlike complex tones that were modulated in center (formant) frequency. The modulation was created by imposing a sinusoidal translation along the frequency axis on the spectral envelopes of the formants, while holding the fundamental frequency fixed. The task of the subjects was to detect these modulations. The modulated complexes were presented simultaneously with modulated or unmodulated formantlike off-frequency maskers. We found no effect of the unmodulated interferers on the detection of the slow formant-frequency modulations. For modulated interferers we found FMDI under a number of stimulus conditions. It was found that this FMDI depended on the fundamental (F_0) of the masker relative to that of the signal (with the larger FMDI for equal fundamentals). In the present study we attempt to characterize the interference effects further by introducing interaural differences in signals and maskers. In this way we may be able to determine the influences of the ear of presentation and of the perceived lateral location of the interferer relative to that of the signal on FMDI, as well as searching for (central) side-independent interference effects. For purely peripheral interference processes we would expect only effects of ear of entry with no FMDI produced by contralateral presentation. In contrast, central FMDI effects would be driven either by perceived location, or by a weighted combination of inputs to the two ears. An important aim of the experiments reported here was to differentiate between these two latter possibilities.

The literature offers little work on binaural effects in the interference of modulation detection. To our best knowledge there is a single paper, Sheft and Yost (1997), that addresses this issue directly. Sheft and Yost studied binaural effects on MDI for amplitude-modulated pure tones. Though the nature of the modulation, and accordingly the low-level detection mechanism involved, differ from the present ones, correspondence between their and the present data can be expected when higher (central) processes are involved in effects of perceived lateral location (or perceived location, for short) on MDI. Sheft and Yost used phase differences in the carriers to generated interaural time differences (ITDs). With these ITDs a separation between the perceived location of signals and maskers was created. Sheft and Yost used tones with near-harmonic relationships, but no harmonic complexes. They found small effects of perceived location on MDI for unmodulated maskers, but none for modulated maskers.

[a]Current address: Department of ENT/Audiology, University Hospital, VU, De Boelelaan 1117, 1081 HV, Amsterdam, The Netherlands.

It is hard to predict whether effects of perceived location will occur in modulation detection interference when harmonic complexes are used, and, if so, how this may be affected by the fundamentals of these complexes. In experiments on grouping, different interactions have been reported between ear of entry and perceived location on the one side, and signal fundamentals on the other side. When presenting the formants of harmonic vowels to different ears, Broadbent and Ladefoged (1957) found no effect of ear of entry on the fusion of vowels into a single perceived location (i.e., grouping of formants) when these formants had identical fundamentals. In contrast, for a task incorporating processing of the fundamental, the detection of a mistuning in one of the harmonics of a complex sound, Gockel and Carlyon (1998) found effects of ear of entry (to which ear the mistuned harmonic and the remainder of the complex were presented), but not of perceived location of these signals (set using ILDs). In these experiments, mistuning was detected by virtue of the signal ''standing out'' from the complex, i.e., when the stimuli did not group together and were perceived as two separate sounds. Likewise, Culling and Summerfield (1995) found effects of ear of entry (but not of lateralization by ITD) on the perceptual grouping of noise bands that formed synthetic whispered vowels (i.e., sounds without a specific F_0). Grouping of stimuli can affect across-channel interference (Hall and Grose, 1991), and because of the different results from the mentioned studies it seems desirable to examine the effects of binaural presentation and fundamental on FMDI for complex tones more closely. In the present study, we use detection of formant modulations to further examine the nature of binaural effects in FMDI. The ITDs are used to change the perceived locations of maskers that have either the same or a different fundamental than the signals.

When utilizing interaural level differences (ILDs) to manipulate the perceived location of the tones, Sheft and Yost (1997) found larger MDI for modulated maskers when signals and maskers shared the same rather than different locations. However, most of the variability in these data appeared to be consistent with within-ear level effects, so it may not be an effect of the perceived location *per se*. To take a closer look at these level effects in the present study, we use a set of ipsilateral and contralateral maskers with different levels, as well as a dichotic masker with an ILD that caused it to be perceived contralateral to the signal.

Under all conditions with modulated interferers, Sheft and Yost (1997) observed a substantial amount of MDI, even when signals and maskers were presented to, or perceived in, opposite ears. Under a dichotic condition where signal and masker were presented to opposite ears, they found 10 dB MDI for modulated maskers, which, in a diotic listening condition, was increased by a further 6 dB. So, it seems that the largest effects were due to a side-independent mechanism, and that a further increase in MDI could be achieved by having signals and maskers in the same ears. However, it is not yet clear whether there are separate central and peripheral processes involved, or whether they are central processes with different input weights for each ear.

In the present experiments, we measured listeners' ability to detect a slow modulation in the center frequency of formants (with a fixed fundamental) presented to the right ear. These stimuli were presented in the absence and presence of simultaneous competing formants that were either unmodulated or modulated at a 10-Hz rate. The maskers were presented under different diotic and dichotic conditions. The effects of ear of entry were examined by presenting the maskers either to the right or to the left ear. The effects of perceived location were examined with binaural maskers, utilizing both interaural time and level differences to separate the perceived locations of signals and maskers. If the perceived location affects FMDI, smaller FMDI is expected when the perceived locations of signals and maskers are further apart. In addition, we examined for the binaural maskers how the relative contributions of the ipsilateral and contralateral parts of these maskers combined to produce the total interference effect.

I. STIMULI AND METHODS

A. Stimuli

The signals and maskers were band-limited harmonic complexes closely resembling vowel formants. All stimuli were harmonic tones which were band-pass filtered with a triangular (with 50 dB/oct) spectral envelope, on a log–log scale. Except where otherwise stated, the signals and maskers had the same fundamental, which was either 80 or 240 Hz. The center frequencies of the unmodulated stimuli were located at a harmonic. For the fundamental of 80 Hz, the center frequencies of signal and masker were 1520 and 2960 Hz, respectively. For the 240-Hz fundamental, these center frequencies were 1440 and 2880 Hz, respectively. To modulate the center frequencies of the stimuli, the center frequencies of the triangular filter were varied sinusoidally over time, thus resulting in a sinusoidal movement of the center frequency of the spectral envelopes of the stimuli.

The subjects' task was detection of a modulation of the center frequency of the signals. These signals were either presented in isolation or together with maskers that were either unmodulated or had a modulated center frequency. The stimuli used in this study were a subset from the ones used by Lyzenga and Carlyon (1999). From those stimuli we chose two signal/masker combinations for which substantial FMDI had been found, while the cues used by the listeners for detecting the modulations in both signals and maskers were found to be based predominantly on detection of shifts in excitation patterns. The formant-frequency modulation detection thresholds (FMTs) found for these stimuli were roughly two times larger than the formant-frequency discrimination thresholds found for comparable stimuli by Lyzenga and Horst (1997). The excitation-pattern differences corresponding to the peak-to-peak center-frequency excursion at threshold were about three times larger than those found for the discrimination thresholds of Lyzenga and Horst (1997). A fuller description of stimuli, thresholds, and models is given by Lyzenga and Carlyon (1999).

During the stimulus presentations, a continuous pink noise was present at a spectrum level of 0 dB SPL/Hz at 1 kHz. For each subject we measured the detection thresholds

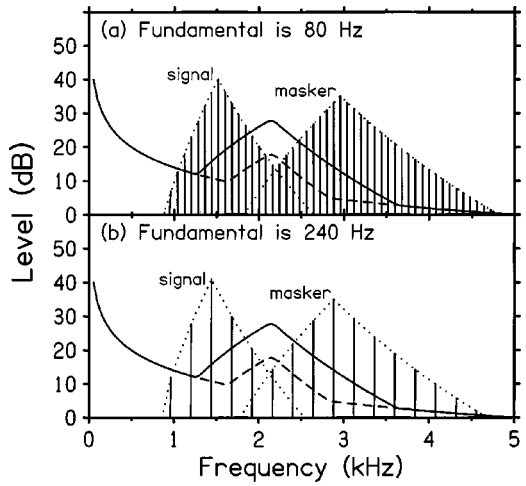


FIG. 1. Spectra of signal/masker combinations including background noise. The bandwidth of this spectral analysis is 47.8 Hz. Between 1 and 4 kHz, the width of the auditory filter is larger than 48 Hz. Therefore, after auditory filtering the levels of the noise band (here shown at either 20 or 10 dB above the pink noise) will have been slightly larger in respect to the complexes than is depicted.

of all stimuli both in quiet and in the pink noise background. Relative to the threshold in the pink noise, the signals were presented at a level of +30 dB, and the maskers were presented either +10, +30, or +50 dB, depending on the condition. Averaged over all five subjects, the 30-dB stimulus-presentation level corresponded to a sensation level of 47±5 dB SL relative to the thresholds in quiet. In the conditions with maskers, we inserted a narrow-band noise between the signals and the maskers. The center frequency of this noise band was 2150 Hz, and its spectrum level was set 10 dB (for two subjects) or 20 dB (for three subjects) above the pink noise background. The 10-dB bandwidth of the noise band was 200 Hz, which is roughly half the critical bandwidth (400 Hz) at this center frequency (Scharf, 1970). It was added to mask any interference effects in the slopes of the signals and the maskers, such as beating of a signal and a masker component that lie very close in frequency. A second strategy to prevent such interference effects from simultaneous presentation of adjacent formants was to have their fundamentals at integer multiples (either 1 or 3) of each other. In this way the occurrence of beating between very closely placed signal and masker components could be avoided. The frequency spectra of the two signal/masker combinations with identical fundamentals, including the background noise spectrum, are shown in Fig. 1.

The stimulus waveforms were computed digitally off-line with 16-bit resolution on a Sun workstation and stored on disk. Each stimulus was generated by calculating its (modulated) spectral envelope as a function of time, after which all sinusoidal frequency components (with fixed frequency) were added to the stimulus with their thus calculated time-varying amplitudes. In this way the formant-frequency modulations resulted in combined amplitude modulations of the components, where these modulations were in anti-phase for components to the lower- and the higher-frequency sides of the formant. All frequency components of the complexes were in sine phase. In the spectral slopes, components were included down to a level of at least 50 dB below the largest one. The steady-state duration of the stimuli was 200 ms, with 25-ms raised-cosine on and off ramps. A set of signals, needed for the determination of one formant-frequency modulation detection threshold, consisted of a reference sound with center frequency F_C and 23 targets with an increasingly strong center frequency modulation. The target modulation depth ranged from 0.0125% to 25.6% in 22 steps of a factor $\sqrt{2}$, giving a range with a small step size from clearly audible modulations to well below detection threshold.

In the experiments, the stimulus waveforms were retrieved by a PC and converted to analog signals using a 16-bit DAC (CED 1401 plus) at a sample rate of 40 kHz. The waveforms were low-pass filtered at 17.2 kHz (3-dB-down point, VBF/25.01, slope=100 dB/oct). Attenuation and mixing were performed by computer-controlled attenuators and mixers (TDT PA4 & SM3). The noise band interspersed between signals and maskers was generated using white noise (TDT WG1), which was band-pass filtered to the desired bandwidth (Kemo VBF/25.03, slope=48 dB/oct). All stimuli were presented through Sennheiser HD414 headphones. They were checked using an HP 3561A spectrum analyzer.

For the signals, the center frequency of the triangular filter was modulated at a 5-Hz rate, so, their steady-state portion contained a single modulation cycle. Two different types of maskers were used: unmodulated, and modulated. They could be presented to either ear or to both. They had the same fundamental as the signal, except for the binaural conditions with equal masker levels in both ears (the ITD conditions), where their fundamentals could be either equal to or different from those of the signals. When modulated, the modulation rate of the maskers was 10 Hz (Lyzenga and Carlyon, 1999) and, therefore, these stimuli contained two modulation cycles. We gave the maskers roughly equally detectable modulations, corresponding to four times the detection threshold (at $d'=1$). The FM applied to all stimuli (both signals and maskers) had a starting phase of 0 degrees.

The binaural masker conditions used in the present experiment are given in Table I. The signals were always presented to the right ear. Examples of the relations between signals and maskers over time are given in Fig. 2. Lyzenga and Carlyon (1999) have already reported FMDI for these stimuli when the masker is ipsilateral (condition I), compared to when the masker is either absent (''None'') or unmodulated (''SI''). We now test its level dependency for ipsilateral maskers (conditions I−, I, and I+) and whether it also occurs contralaterally (conditions C−, C, and C+ vs SC and SC+). In addition, we will compare the FMDI for the maskers that are presented diotically and heard near the middle of the head (condition M) to those where the perceived location of the masker is shifted. Moving the perceived location of the maskers towards the ipsilateral side is achieved by advancing the onset of both the (monaural) signal and the ipsilateral channel of the masker (ITDI). Moving it towards the contralateral side is achieved by advancing the onset of either the contralateral channel of the masker alone ($ITDC_1$) or of both the signal and the contralateral channel

TABLE I. Binaural masker conditions. A level of 0 dB signifies that the masker had the same level as the signal. In conditions ITDI and DTDI both the signal and the ipsilateral channel of the masker were advanced. In conditions $ITDC_1$ and DTDC, only the contralateral masker channel was advanced, and in condition $ITDC_2$ both the signal and the contralateral masker channel were advanced.

Code	Binaural Y/N	Modulated Y/N	Ipsilateral level (dB)	Contralateral level (dB)	Masker presentation condition
None		...	$-\infty$	$-\infty$	...
SI	N	N	0	$-\infty$	Ipsilateral
SC	N	N	$-\infty$	0	Contralateral
SC+	N	N	$-\infty$	+20	Contralateral
I−	N	Y	−20	$-\infty$	Ipsilateral
I	N	Y	0	$-\infty$	Ipsilateral
I+	N	Y	+20	$-\infty$	Ipsilateral
C−	N	Y	$-\infty$	−20	Contralateral
C	N	Y	$-\infty$	0	Contralateral
C+	N	Y	$-\infty$	+20	Contralateral
M	Y	Y	0	0	Central
ITDI	Y	Y	0	0	ITD advance ipsilateral
$ITDC_1$	Y	Y	0	0	ITD advance contralateral (1)
$ITDC_2$	Y	Y	0	0	ITD advance contralateral (2)
ILDC	Y	Y	0	+20	ILD increase contralateral
DM	Y	Y	0	0	Central, different F_0
DTDI	Y	Y	0	0	ITD advance ipsilateral, different F_0
DTDC	Y	Y	0	0	ITD advance contralateral, different F_0

of the masker ($ITDC_2$), or by raising the level of the contralateral masker (ILDC). Finally, and unlike all other cases reported here, in conditions DM, DTDI, and DTDC the maskers had a different fundamental than the signals; in these conditions the effects of perceived masker location were assessed by varying their ITDs.

To lateralize the maskers, a 1-ms onset difference was used in the ITD conditions; the presentation of the masker was either advanced or delayed relative to the signal. When presented in isolation (but not necessarily when presented together with the signals), these ITDs lateralized the perceived location of the maskers (Mossop and Culling, 1998; Bernstein and Trahiotis, 1985). However, adding more components to a sound can affect sensitivity for ITDs (Dye, 1990; Buell and Hafter, 1991). Furthermore, for the conditions with equal fundamentals for signals and maskers, they tend to fuse into a single perceived location (Broadbent and Ladefoged, 1957). Therefore, the presence of a signal with the same fundamental may affect the perceived location of the maskers compared to when these maskers are presented in isolation. This is in agreement with the reports of our subjects. For the binaural masker conditions we performed a separate experiment, in which we asked subjects to indicate the perceived location of the sound or sounds on a scale ranging from −1 for the contralateral ear to +1 for the ipsilateral ear (where 0 is center). Subjects could replay each stimulus as often as they liked before giving an estimate of its perceived location. Estimates were collected for targets at the starting modulation depth (25.6%) of the adaptive procedure and at threshold (FMT). The estimates for these two conditions were very similar. For signals and maskers with equal fundamentals, subjects usually indicated positions shifting from ipsilateral to central rather than to contralateral, as the isolated-masker conditions changed from ITDI, via M, to $ITDC_1$. The average positions (over 14 judgments) for these three conditions were found to be 0.43, 0.26, and 0.12, respectively. For comparison, the ILDC condition produced −0.51. For the conditions with different fundamentals, fusion of maskers and signals did not occur, and subjects reported the expected perceived locations for the maskers under the conditions DTDI, DM, and DTDC (here the average positions were 0.46, −0.04, and −0.52, respectively; ten judgments). For these three conditions, the signals and maskers were always perceived as two different sounds, with the signal located on the far right.

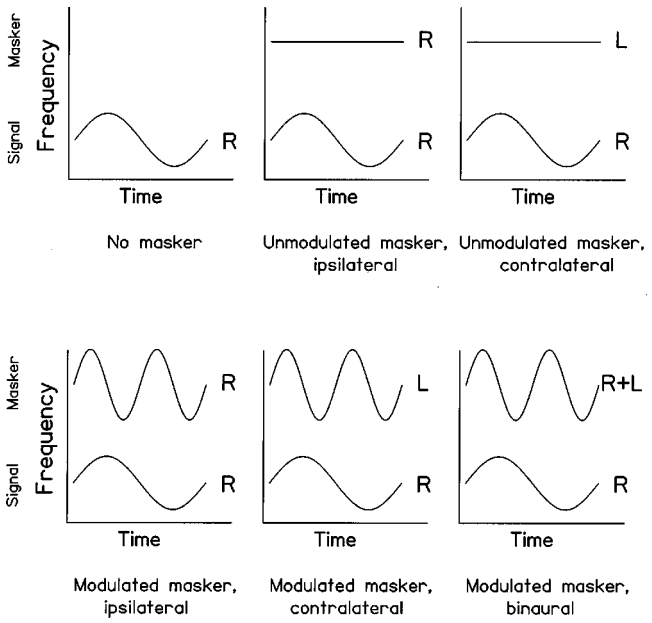


FIG. 2. Examples of the center-frequency modulations during stimulus presentation.

B. Procedure and subjects

The general measurement procedures were the same as those used, and described in detail, by Lyzenga and Carlyon

(1999). In short, we used an adaptive procedure and a three-interval, three-alternative forced-choice paradigm (3I, 3AFC) to measure the formant-frequency modulation detection thresholds for signals both in the presence and absence of maskers. When present, the maskers were presented in all three tone intervals. The silent interval between the tones was 400 ms. Immediate feedback was given. The adaptive procedure was started with the maximum modulation of 25.6% for the signals. During the procedure, the step size in the modulation depth was a factor of $\sqrt{2}$. The decision rules are described in detail in Lyzenga and Horst (1995). The procedure converges around 63% correct responses, which corresponds to a d' of 1 for the 3I, 3AFC paradigm. On average, one FMT measurement required about 80 trials.

Estimates of the FMTs were made after each set of conditions was measured four times (making 200 to 300 trials near threshold), in a pseudorandom order. After these four series, the scores for each stimulus were checked for consistency and, where the thresholds showed a spread greater than a factor of 2, one or two additional estimations were made. A small proportion (less than $\frac{1}{2}$%) of runs were more than a factor of 3 away from the others and were discarded. For each subject, the percent correct as a function of the modulation depth was estimated by averaging the scores for each target. These psychometric functions were fitted to the theoretical relation of percent correct as a function of d' by scaling the modulation depth relative to the d' [the relation between d' and modulation depth was found to be linear by Lyzenga and Carlyon (1999)]. The resulting thresholds represented the modulation depths that correspond to a d' of 1. The standard deviation of the individual FMTs was found to be smaller than 30% (on average, about a factor of 1.2). In these calculations we assumed that the subjects were not biased towards one of the three signal intervals in their answers. For the incorrect answers we found that 34% occurred for the first interval, 33% for the second interval, and 33% for the third interval, so this assumption was well justified.

Five normal-hearing subjects participated for all conditions. Three subjects were highly experienced in the task; the other two were trained before starting data collection.

II. RESULTS

For the two reference conditions with the 80- and the 240-Hz fundamentals, the average FMTs were 2.0% (with a standard deviation of 0.6) and 1.9% (standard deviation of 0.4), respectively. These values are very close to those reported for the same conditions in our previous study (Lyzenga and Carlyon, 1999).

The FMDI was evaluated for each subject from the ratio between the FMTs with and without a masker. These individual FMDI values were then averaged to obtain the average FMDI. The behavior of the individual FMDI data was very similar for all five subjects, so, only the averages are shown. A repeated measures ANOVA on the individual FMDI data (2 fundamentals×17 masker conditions) showed no effects of fundamental [$F(1,4)=2.62$, $p=0.18$]. Therefore, in the results presented in this section, we have combined the data for the two fundamentals into one set of averages.

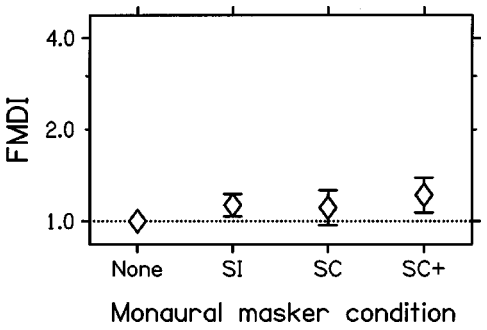


FIG. 3. Average FMDI for unmodulated maskers, calculated as the average ratio between the FMTs with and without maskers for five subjects.

A. Results for monaural maskers

The results for the unmodulated monaural maskers are shown in Fig. 3. The error bars are the standard deviations over two fundamentals and five subjects. The leftmost condition reflects the no-masker reference condition. The second entry from the left is for ipsilateral, and the rightmost ones are for contralateral unmodulated maskers. Figure 3 shows that FMDI is very small for the unmodulated maskers, though for the ipsilateral masker and the contralateral masker at +20 dB it is significantly larger than unity (one sample t-test, 95% confidence interval, in both cases: $p<0.005$).

Figure 4 shows the results for modulated monaural maskers. The left panel shows FMDI for ipsilateral, and the right panel for contralateral modulated maskers. For the modulated maskers we find substantial FMDI (significant after Bonferoni correction in one sample t-tests at $p<0.005$ for the C-condition and at $p<0.001$ for the remaining conditions). A two-way repeated measures ANOVA (sides×level) revealed, after Bonferoni correction, that FMDI is significantly larger for the ipsilateral than for the contralateral masker [$F(1,9)=16.13$, $p<0.005$], and that it grows with the level of the masker [$F(2,18)=43.42$, $p<0.001$]. The notion that FMDI is larger for ipsilateral than for contralateral modulated maskers is in agreement with the findings of Sheft and Yost (1997), who found larger MDI when the presentation schemes of signal and masker showed greater similarity (no, one, or two common ears of presentation).

For the ipsilateral and the contralateral ears, the growth of FMDI with level is more or less linear over the range of levels used. (At the relative level of −30 dB the maskers

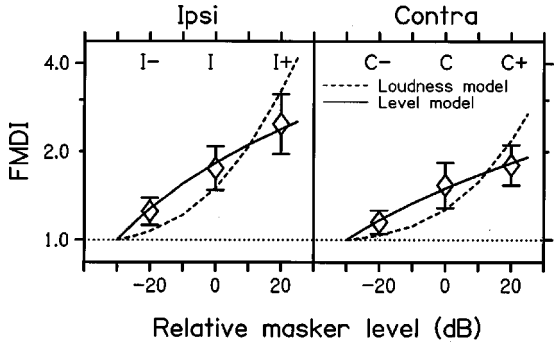


FIG. 4. Average FMDI for modulated monaural maskers, calculated as the average ratio between the FMTs for modulated (10-Hz rate) and unmodulated maskers.

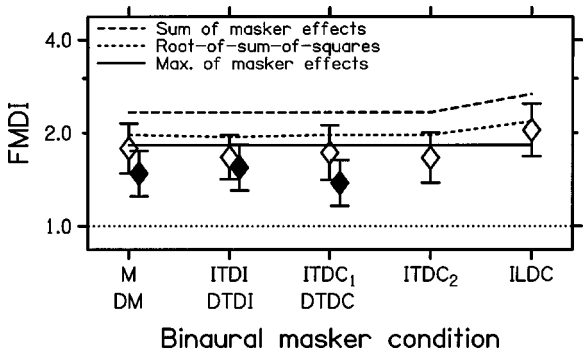


FIG. 5. Average FMDI for modulated binaural maskers. Data for masker/signal combinations with equal and different fundamentals are shown in the open and filled symbols, respectively.

disappear in the noise background, so their contributions to FMDI should disappear as well.) In order to examine the way ipsilateral and contralateral maskers combine when presented simultaneously, we have first modeled the FMDI for monaural maskers. The concept of the linear growth of FMDI with level has been implemented in a simple descriptive model:

$$\mathrm{FMDI} = 1 + W_1 \cdot L_{\mathrm{masker}}, \tag{1}$$

where L_{masker} is the level of the masker above the pink noise, and W_1 equals 0.0278 and 0.0167 for the ipsilateral and contralateral maskers, respectively. These values of W_1 have been chosen for minimal squared errors between predictions and data. The solid lines in Fig. 4 show the fit of the level model for monaural maskers (note that they are slightly bent due to the logarithmic scale along the y axis). The dashed lines describe the fit of a model based on the loudness of the tones rather than on their levels:

$$\mathrm{FMDI} = 0.95 + W_2 \cdot (I_{\mathrm{masker}})^{0.3}, \tag{2}$$

where I_{masker} is the total intensity of the masker, and W_2 equals 0.0256 and 0.0139 for the ipsilateral and contralateral maskers, respectively. The predictions of this model do not follow the general trend of the data as well as those of the level model (for the level model the sum of the six squared relative errors is 0.005, for the loudness model it is 0.20). Therefore, the level model is the better candidate for an accurate assessment of the way in which the influence of the ipsilateral and contralateral maskers combine under the binaural masking conditions.

B. Results for binaural maskers

Figure 5 shows significant FMDI for eight binaural masker conditions (one sample t-test: after Bonferoni correction $p < 0.002$ for all conditions). A one-way ANOVA on the five conditions with equal masker/signal fundamentals (open symbols) showed a significant difference between these conditions after Bonferoni correction [$F(4,36) = 6.74$, $p < 0.001$]. However, when the ILDC condition was left out from this analysis, the effect of condition disappeared completely [$F(3,27) = 1.01$, $p = 0.40$]. It was also absent for signals and maskers with different fundamentals (filled symbols) [one-way ANOVA, $F(2,18) = 2.42$, $p = 0.12$, after Bonferoni correction]. So, the different ITDs had no significant effect on FMDI. The FMDI for the ITD conditions with equal signal/masker fundamentals appears to be somewhat larger than for those with different fundamentals, but this effect just failed to reach significance [$F(1,9) = 4.39$, $p = 0.066$] after Bonferoni correction [significance was found for just the signals with $F_0 = 240$ Hz, $F(1,4) = 27.13$, $p < 0.01$]. If the perceived location of the masker had affected FMDI in the ILDC condition, the perceptual separation between signal and masker should have produced a decrease in FMDI as compared to the diotic M condition. This is clearly not the case, rather the opposite is true. So, the increase in the level of the contralateral masker, rather than its associated change in perceived location, affected FMDI. All these results are in general agreement with the findings for AM pure tones of Sheft and Yost (1997).

The predictions of three models are shown in Fig. 5 in the three lines. How well the predictions approach the data depends strongly on the rule that is used to combine the effects of the monaural maskers. The three assumptions used for the current predictions are that the amount of binaural FMDI is determined (a) by only the ear with the largest level, (b) by the root-mean-square (rms) sum of the levels on the left and on the right, and (c) by the linear sum of the left and right levels. These three rules are defined in the relations (3), (4), and (5), respectively:

$$\mathrm{FMDI}_{\mathrm{total1}} = \mathrm{Max}(\mathrm{FMDI}_{\mathrm{ipsi}},\ \mathrm{FMDI}_{\mathrm{contra}}), \tag{3}$$

$$\mathrm{FMDI}_{\mathrm{total2}} = \sqrt{(\mathrm{FMDI}_{\mathrm{ipsi}}^2 + \mathrm{FMDI}_{\mathrm{contra}}^2)}, \tag{4}$$

$$\mathrm{FMDI}_{\mathrm{total3}} = \mathrm{FMDI}_{\mathrm{ipsi}} + \mathrm{FMDI}_{\mathrm{contra}}, \tag{5}$$

where $\mathrm{FMDI}_{\mathrm{ipsi}}$ and $\mathrm{FMDI}_{\mathrm{contra}}$ were evaluated using relation (1). The relations (3) and (4) produce a good fit to the data (shown in the solid and the short dashed lines, respectively). In contrast, the direct addition of the effects of both maskers in relation (5) gives predictions that are much larger than the measurements (indicated with the dashed line). This is reflected in the sums of the squared relative errors between the five data points for equal masker/signal fundamentals and the model predictions; for relations (3)–(5) these sums are 0.034, 0.076, and 0.34, respectively. So, we find that the predictions using relations (3) and (4) are much closer to the data than those of relation (5). However, the predictions of relation (3), shown in the solid line, do not show an increase in FMDI for the ILDC condition, which does not agree with the data. On the other hand, the predictions from relation (4) are consistently larger that the data values. So, neither of the models predicts the data perfectly. It seems, therefore, that the amount of interference for binaural maskers is dominated by the monaural masker with the largest effect, but, nonetheless, some sort of addition of the effects of ipsilateral and contralateral maskers, as in relation (4), seems to take place. Since FMDI is consistently smaller for maskers with different rather than equal fundamentals as the signals, and since there is a near miss between predictions and data, the weights W_1 for the maskers [in relation (1)] appear to depend to some degree on masker/signal fundamentals and on monaural/binaural presentation scheme.

III. DISCUSSION

The FMDI for the ILDC condition (where the contralateral masker level was increased) was slightly larger than for the ITD conditions. Since the perceived location of this masker was shifted away from rather than towards the signal, the increase in FMDI has to be a level effect rather than an effect of perceived location.

For modulated maskers, the smallest amount of FMDI was found for the masker that was presented to the contralateral ear. In general, FMDI was larger when the total level of the masker tones was larger, where the maskers in the ipsilateral ear produced larger FMDI than those in the contralateral ear. Because signals and maskers were well separated by the noise band that was inserted between them, it is unlikely that this was due to peripheral interactions at the level of the basilar membrane. So, we found a clear effect of ear of entry. Since FMDI was still observed for contralateral maskers, it was not a strictly single-sided effect, as it would have been if it was purely peripheral. Therefore, the place where the FMDI occurs has to lie in a central process. We did not, however, find any evidence for an effect of perceived location (especially for the conditions with different masker/signal fundamentals). So, the outputs of localization mechanisms do not appear to affect FMDI. Therefore, it seems most likely that a central effect was involved that has a different sensitivity for each ear of entry, rather than receiving input from localization processes.

The conclusion that ear of entry does, and localization processes do not, affect processing of simultaneously presented harmonic tones is in agreement with recent findings of Gockel and Carlyon (1998). They measured detection of a mistuning in one component of a harmonic complex, in the presence of a nonsimultaneous pure-tone fringe with the same frequency as the unmodulated signal component. They varied ear of entry and perceived location of the fringes (using ILDs). In another experiment they did not use any fringes, but varied the ears of presentation and perceived locations (set using ILDs) of the mistuned harmonic and of the remainder of the complex. In both experiments, they found no effects of the perceived location of these sounds, and strong effects of ear of entry. As in the present study, they found some effects for contralateral presentation of the interfering sounds, but these effects were much reduced compared to ipsilateral presentation.

IV. CONCLUSIONS

The FMDI is larger for ipsilateral than for contralateral maskers. For the masker levels used, the amount of FMDI increases roughly linearly with the level of a monaural masker. For binaural maskers, the amount of FMDI seems to be determined by a weighted rms sum of the monaural masker effects. This sum is dominated by the masker with the largest level, with a possible, but not conclusively established, role of the smaller masker. The ITDs that we applied in the binaural maskers (which affect their perceived locations) have no direct effect on FMDI. The interference effect appears to reside in a central process that has different sensitivities for each ear of entry.

ACKNOWLEDGMENTS

This research was funded by the Wellcome Trust. The authors would like to thank John Grose and an anonymous reviewer for their constructive comments on an earlier version of this paper.

Bernstein, L. R., and Trahiotis, C. (**1985**). ''Lateralization of low-frequency, complex waveforms: The use of envelope-based temporal disparities,'' J. Acoust. Soc. Am. **77**, 1868–1880.

Broadbent, D. E., and Ladefoged, P. (**1957**). ''On the fusion of sounds reaching different sense organs,'' J. Acoust. Soc. Am. **29**, 708–710.

Buell, T. N., and Hafter, E. R. (**1991**). ''Combination of binaural information across frequency bands,'' J. Acoust. Soc. Am. **90**, 1894–1900.

Culling, J. F., and 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.

Dye, R. H. (**1990**). ''The combination of interaural information across frequencies: Lateralization on the basis of interaural delay,'' J. Acoust. Soc. Am. **88**, 2159–2170.

Gockel, H., and Carlyon, R. P. (**1998**). ''Effects of ear of entry and perceived location of synchronous and asynchronous components on mistuning detection,'' J. Acoust. Soc. Am. **104**, 3534–3545.

Hall, W. H., and Grose, J. H. (**1991**). ''Some effects of auditory grouping factors on modulation detection interference (MDI),'' J. Acoust. Soc. Am. **90**, 3028–3035.

Lyzenga, J., and Carlyon, R. P. (**1999**). ''Center frequency modulation detection for harmonic complexes resembling vowel formants and its interference by off-frequency maskers,'' J. Acoust. Soc. Am. **105**, 2792–2806.

Lyzenga, J., and Horst, J. W. (**1995**). ''Frequency discrimination of band-limited harmonic complexes related to vowel formants,'' J. Acoust. Soc. Am. **98**, 1943–1955.

Lyzenga, J., and Horst, J. W. (**1997**). ''Frequency discrimination of stylized synthetic vowels with a single formant,'' J. Acoust. Soc. Am. **102**, 1755–1767.

Mossop, J. E., and Culling, J. F. (**1998**). ''Lateralization of large interaural delays,'' J. Acoust. Soc. Am. **104**, 1574–1579.

Scharf, B. (**1970**). ''Critical bands,'' in *Foundation of Modern Auditory Theory Vol. 1*, edited by J. V. Tobias (Academic, New York), pp. 159–202.

Sheft, S., and Yost, W. A. (**1997**). ''Binaural modulation detection interference,'' J. Acoust. Soc. Am. **102**, 1791–1798.

Intensity discrimination and detection of amplitude modulation in electric hearing

Gail S. Donaldson[a] and Neal F. Viemeister
Departments of Otolaryngology and Psychology, University of Minnesota, Minneapolis, Minnesota 55455

(Received 24 February 2000; accepted for publication 5 May 2000)

Wojtczak and Viemeister [J. Acoust. Soc. Am. **106**, 1917–1924 (1999)] demonstrated a close relationship between intensity difference limens (DLs) and 4-Hz amplitude modulation (AM) detection thresholds in normal-hearing acoustic listeners. The present study demonstrates a similar relationship between intensity DLs and AM detection thresholds in cochlear-implant listeners, for gated stimuli. This suggests that acoustic and cochlear-implant listeners make use of a similar decision variable to perform intensity discrimination and modulation detection tasks. It can be shown that the absence of compression in electric hearing does not preclude this possibility. © *2000 Acoustical Society of America.* [S0001-4966(00)04508-2]

PACS numbers: 43.66.Ts, 43.66.Cb, 43.66.Mk, 43.64.Me [RVS]

INTRODUCTION

Wojtczak and Viemeister (1999) investigated the relationship between intensity discrimination thresholds and 4-Hz amplitude modulation (AM) detection thresholds in normal-hearing acoustic listeners. Using continuous pure-tone pedestals and carriers spanning a wide range of levels, they found that intensity DLs were related to 4-Hz AM detection thresholds by the equation

$$10\log(\Delta I/I) = 0.44^{*}(20\log m) + 1.7, \quad (1)$$

where $\Delta I/I$ is the Weber fraction for intensity discrimination and m is the modulation index at threshold. They demonstrated that Eq. (1) generalizes to other low modulation frequencies by adjusting the additive constant term, and showed that it holds over a range of performance (d') levels. They also argued that Eq. (1) holds for a range of pedestal/carrier frequencies, for noise carriers, and for gated pedestals/carriers of varying durations. However, they noted that Eq. (1) does not describe the relationship between psychometric functions for intensity discrimination and AM detection when gated pedestals/carriers are used.

Equation (1) suggests that the same decision variable is used for intensity discrimination and AM detection tasks in acoustic hearing. Wojtczak and Viemeister (1999) examined several theoretical models of AM processing in an attempt to identify this decision strategy, but concluded that no existing model accounts convincingly for the empirical relationship that Eq. (1) describes. In this report, we examine the relationship between intensity discrimination and AM detection in cochlear-implant listeners, using gated stimuli. The relationship that we observe is similar to Eq. (1), suggesting that whatever decision strategy is used to perform intensity discrimination and modulation detection tasks in acoustic hearing also applies to electric hearing.

[a]Author to whom correspondence should be addressed. Electronic mail: donal005@tc.umn.edu

I. PROCEDURE

Intensity DLs and 4-Hz AM detection thresholds were measured in three subjects with the Nucleus 22 cochlear implant. Three electrodes from different regions of the implanted array were tested in each subject. Electrode configuration was bipolar, with an electrode separation of 1.5 mm (BP+1) in subjects S1 and S3, and 2.25 mm (BP+2) in subject S2. Stimuli were 500-ms gated trains of 800-pps, 80-μs/ph biphasic pulses, spanning levels between 15% and 85% of the test electrode's dynamic range. Five identical levels of the standard stimulus were evaluated in both the intensity discrimination and AM detection tasks. Level was incremented or sinusoidally modulated by varying the current amplitude of pulses within the stimulus trains. Thresholds were obtained using a 3-down, 1-up 3-interval forced choice adaptive procedure that estimated 79.4% correct (Levitt, 1971), corresponding to a detection sensitivity of $d' = 1.63$.

Table I describes the subjects and shows the absolute thresholds and dynamic ranges (DRs) that were measured for each of the nine test electrodes in response to the standard stimulus (800 pps, 80-μs/ph, 500-ms pulse train).[1] Note that DRs are similar for electrodes within a given subject, but vary considerably across subjects. This suggests that the three subjects possess different numbers or patterns of surviving auditory neurons.

II. RESULTS

Figure 1 shows 4-Hz AM detection thresholds plotted as a function of current amplitude (dB μA) for each of the nine test electrodes. In each panel, individual threshold estimates are shown as shaded triangles and mean threshold values are shown by a thin solid line. For comparison, mean acoustic data obtained by Wojtczak and Viemeister (1999) for a 1-kHz continuous carrier are plotted in the upper-left panel.

The range of AM detection thresholds exhibited by the present subjects is generally similar to those reported in earlier cochlear-implant studies (Shannon, 1992; Busby *et al.*, 1993; Cazals *et al.*, 1994). Thresholds are also roughly simi-

TABLE I. Subjects.

Subj	M/F	Age	Etiology of hearing loss	Years deaf	Years implanted	Electrodes tested	Threshold (dB *re*: 1 μA)	Dynamic range (dB μA)
S1	M	32	maternal rubella	27	2	e5	41.2	20.6
						e12	41.8	18.7
						e19	37.2	16.1
S2	F	60	otosclerosis	12	2	e4	48.0	11.0
						e17	45.6	10.1
						e21	49.3	10.1
S3	M	59	meningitis	3	3	e5	48.7	6.4
						e12	46.6	5.8
						e20	50.5	4.9

lar to the acoustic data for pure-tone carriers reported by Wojtczak and Viemeister (1999) and most others (see Kohlrausch, 1993, for a summary). Differences are apparent across subjects, however: Subjects S1 and S2 show steep functions with quite sensitive thresholds (−30 to −40 dB) at the highest stimulus levels, whereas subject S3 shows flatter functions with relatively insensitive AM thresholds (−10 to −20 dB) at all levels. None of the measured thresholds was limited by the resolution of current amplitude delivery in the subjects' devices; however, thresholds for three electrodes (CJP rEL5, CJP rEL12, and DAW rEL17) approached this limit at the highest carrier level.

Figure 2 shows corresponding intensity DLs [$10\log(\Delta I/I)$] for each of the nine test electrodes, plotted as a function of current amplitude (dB μA). As in Fig. 1, individual threshold estimates are shown as shaded triangles and mean thresholds are indicated by a thin solid line. All intensity discrimination thresholds were well below the resolution limits of the subjects' devices. Mean acoustic intensity DLs from Wojtczak and Viemeister (1999) for a 1-kHz continuous pedestal are shown in the upper-left panel.

Note that, for individual electrodes, the shapes of the mean intensity discrimination functions shown in Fig. 2 are similar to the shapes of AM detection functions shown in Fig. 1. That is, subjects S1 and S2 show steep functions with quite sensitive intensity DLs (−10 to −16 dB) at the highest pedestal levels, and subject S3 shows shallower functions with relatively large intensity DLs (−4 to −8 dB) at all pedestal levels. We have previously observed considerable variability across subjects and electrodes in the size and level dependence of intensity DLs in cochlear-implant subjects with the Nucleus 22 device (Nelson *et al.*, 1996; Donaldson and Nelson, 1997a,b). The data shown Fig. 2 are typical of those we have reported in earlier studies.

The unfilled square symbols in each panel of Fig. 2 represent intensity DLs predicted from the mean 4-Hz AM detection thresholds in Fig. 1, using Eq. (1). Note that predicted values are similar to the mean, measured DLs. This suggests that Eq. (1), derived from data for acoustic listeners and continuous pedestals/carriers, describes quite accurately the relationship between intensity DLs and modulation thresholds for cochlear-implant listeners and gated pedestals/carriers.

In Fig. 3, the intensity DLs from Fig. 1 are replotted as a function of the corresponding AM detection thresholds from Fig. 2. The upper-left, upper-right, and lower-left panels show data for subjects S1, S2, and S3, respectively. In each of these panels, data for individual test electrodes are plotted with different symbols, and a single linear regression

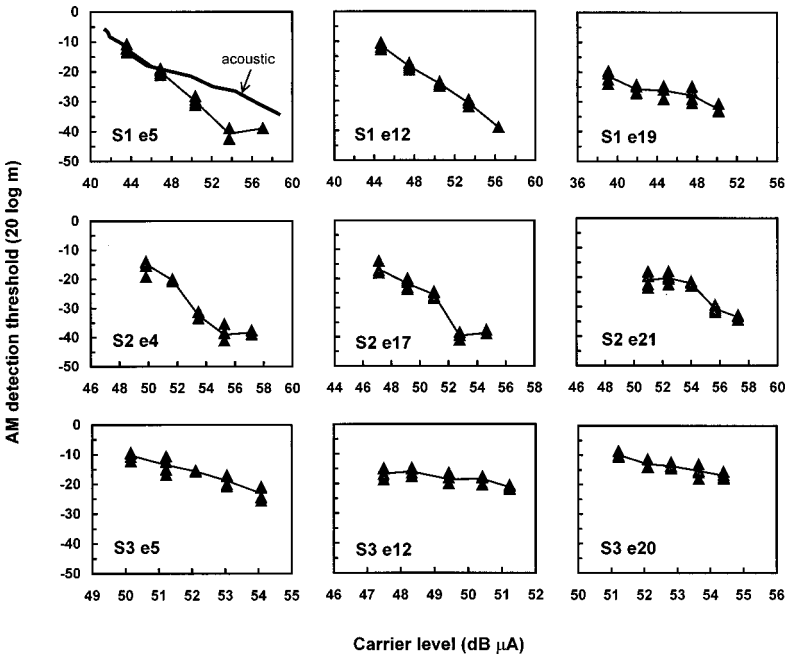


FIG. 1. 4-Hz AM detection thresholds as a function of carrier level. Individual threshold estimates are shown as shaded triangles; mean data are indicated by the solid line (without symbols). Each panel represents data for one electrode; each row of panels represents data from a different subject. Acoustic data for a 1-kHz carrier from Wojtczak and Viemeister (1999) are shown in the upper-left panel. These data are scaled along the *x* axis so as to match the cochlear-implant data in terms of percent dynamic range in dB, by assuming an acoustic threshold of 5 dB SPL and an acoustic maximum acceptable loudness of 95 dB SPL.

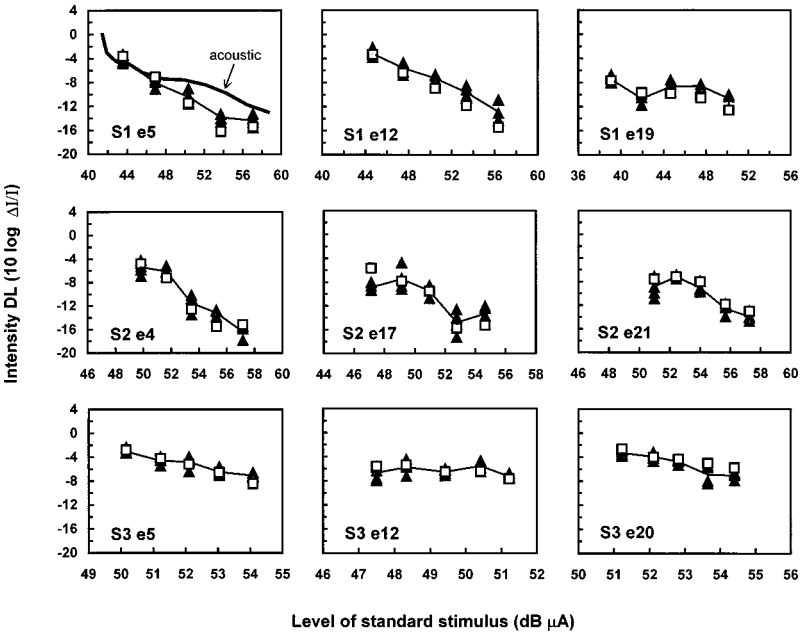


FIG. 2. Intensity DLs as a function of the level of the standard stimulus. Panels and symbols as in Fig. 1. Open squares represent intensity DLs predicted from the mean AM detection thresholds in Fig. 1, using Eq. (1).

line is fit to the combined data from all three electrodes. The data for each subject show a strong, linear relationship between intensity DLs and 4-Hz AM detection thresholds. Slopes of all three regression functions are similar (0.33–0.36) and the intercepts are all close to zero. In each case, AM detection thresholds account for a large (77%–93%) and highly significant proportion ($p<0.0001$) of the variance observed in the intensity DLs.

The data for subjects S1, S2, and S3 are combined in the lower-right panel of Fig. 3. Here, different symbols represent the data for different subjects. The dark line in this panel is the best-fitting linear regression function for the combined subject data

$$10\log(\Delta I/I) = 0.36^{*}(20\log m) + 0.9. \quad (2)$$

The lighter gray line in this panel represents Eq. (1). Both the slope and y intercept of Eq. (2) are significantly smaller than their counterparts in Eq. (1) ($p<0.005$); however, it is apparent from Fig. 3 that these differences are small.

III. DISCUSSION

The present data suggest that the empirical relationship between intensity DLs and AM detection thresholds reported by Wojtczak and Viemeister [Eq. (1)] is approximately correct in both acoustic and electric hearing. This result is at first surprising, given that stimulus coding properties are known to differ substantially in the two modes. For example, at the level of the auditory nerve it is known that electric stimulation produces steeper rate-level functions, smaller dynamic ranges, and less variable spike counts than acoustic stimulation (Moxon, 1971; Hartmann *et al.*, 1984; Javel, 1990; Javel and Viemeister, 1999). In addition, factors such as loss of cochlear compression and altered patterns of neural

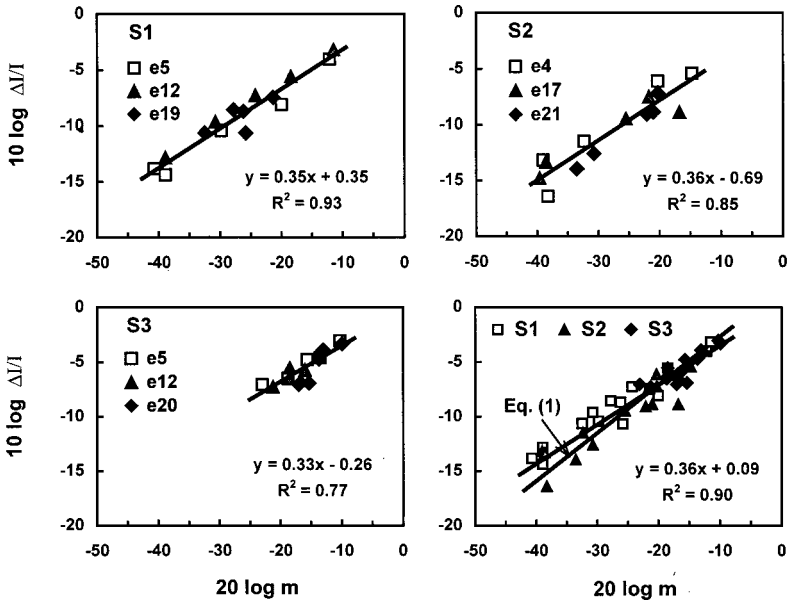


FIG. 3. Intensity DLs as a function of AM detection thresholds. Data for subjects S1, S2, and S3 are shown individually in the upper-left, upper-right, and lower-left panels, respectively, and are plotted together in the lower-right panel.

recruitment almost certainly affect psychophysical measures of intensity coding in the electric case (e.g., Zeng and Shannon, 1994; Zeng *et al.*, 1998).

A possible explanation for the present finding is that both acoustic and cochlear-implant listeners use a similar decision statistic for performing intensity discrimination and modulation detection tasks. Specifically, if the decision variables applied in acoustic and electric hearing are related by a monotonic transform such as compression, and the noise that limits performance occurs before the decision statistic is ''computed,'' then they will produce identical performance (Eagan, 1975). This would not only explain the similarity between Eqs. (1) and (2) but would also explain the similarity between the Weber fractions and between the modulation thresholds. The requirement that the internal noise precede the monotonic transformation presents a potential problem, since compression is observed prior to hair-cell transduction and other possible sources of internal noise in acoustic hearing. However, this requirement is too restrictive, since some decision variables can produce identical performance even when the noise *follows* the transform.

For example, assume that the decision variable is the effective peak value of the transformed, processed stimulus and that an internal noise is added *after* the peak is computed. Then

$$d'_{\Delta I}=\frac{f(I+\Delta I)-f(I)}{\sigma}, \quad d'_m=\frac{f(I(1+m)^2)-f(I)}{\sigma},$$

where $d'_{\Delta I}$ and d'_m are performance measures for intensity discrimination and AM detection, respectively, f is a monotonically increasing function, and σ is the standard deviation of the decision variable.

For equal ''threshold'' performance,

$$f(I+\Delta I)-f(I)=f(I(1+m)^2)-f(I).$$

This equation is satisfied if

$$I+\Delta I=I(1+m)^2,$$

or, equivalently,

$$\frac{\Delta I}{I}=2m+m^2.$$

The last equation, when expressed in logarithmic form, is similar to Eqs. (1) and (2). The important point, however, is that this equation does not depend on $f(\ )$. This demonstrates that the relationship between intensity discrimination and AM detection need not be affected by compression. It does not imply, however, that compression *cannot* affect the relationship. Indeed, it is possible that the small differences between the slopes and between the intercepts for acoustic and electric hearing [Eq. (1) vs Eq. (2)] may be the result of compression. Another possibility is that these differences reflect the use of gated stimuli in the present experiment as compared to the continuous carriers and pedestals used by Wojtczak and Viemeister. Finally, it should be noted that we are not suggesting that intensity discrimination and modulation detection, considered separately, are not affected by $f(\ )$. Clearly, in this demonstration $d'_{\Delta I}$ and d'_m depend on $f(\ )$.

ACKNOWLEDGMENTS

This research was supported by NIDCD Grant No. DC00110 and the Lions 5M International Hearing Foundation. Preliminary findings were presented at the 22nd Midwinter Research Meeting of the Association for Research in Otolaryngology (February, 1999). The authors thank John Van Essen for writing the computer programs used to obtain the psychophysical data; Tanya Grann for assisting with data collection and analysis; and Monita Chatterjee and an anonymous reviewer for their helpful comments on an earlier version of the manuscript. In addition, we extend special thanks to the three cochlear-implant subjects who participated in this work.

[1]Absolute thresholds were measured with an adaptive task, and maximum acceptable loudness levels (MALs) were measured using an ascending method of limits procedure, as described in Nelson *et al.*, 1996. Dynamic range was computed as the decibel difference between the average MAL and threshold values.

Busby, P. A., Tong, Y. C., and Clark, G. M. (**1993**). ''The perception of temporal modulations by cochlear implant patients,'' J. Acoust. Soc. Am. **94**, 124–131.

Cazals, Y., Pelizzone, M., Saudan, O., and Boex, C. (**1994**). ''Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants,'' J. Acoust. Soc. Am. **96**, 2048–2054.

Donaldson, G. S., and Nelson, D. A. (**1997a**). ''Effects of pulse duration on discrimination of current amplitude by Nucleus cochlear-implant users,'' Abstracts of the 20th Midwinter Research Meeting, Association for Research in Otolaryngology, p. 78 (unpublished).

Donaldson, G. S., and Nelson, D. A. (**1997b**). ''Intensity resolution for pulse amplitude and pulse duration: Individual differences and stimulus-parameter effects,'' Abstracts of 1997 Conference on Implantable Auditory Prostheses (Asilomar, CA), p. 24 (unpublished).

Eagan, J. P. (**1975**). *Signal Detection Theory and ROC Analysis* (Academic, New York).

Hartmann, R., Topp, G., and Klinke, R. (**1984**). ''Electrical stimulation of the cat cochlea: Discharge pattern of single auditory fibers,'' Adv. Audiol. **1**, 18–29.

Javel, E. (**1990**). ''Acoustic and electrical encoding of temporal information,'' *Models of the Electrically Stimulated Cochlea*, edited by J. M. Miller and F. A. Spelman (Springer, New York), pp. 247–292.

Javel, E., and Viemeister, N. F. (**2000**). ''Stochastic properties of cat auditory nerve rsponses to electric and acoustic stimuli and application to intensity discrimination,'' J. Acoust. Soc. Am. **107**, 908–921.

Kohlrausch, A. (**1993**). ''Comment on 'Temporal modulation transfer functions in patients with cochlear implants [J. Acoust. Soc. Am. **91**, 2086–2104 (1992)],' '' J. Acoust. Soc. Am. **93**, 1649–1650.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Moxon, E. C. (**1971**). ''Neural and Mechanical Responses to Electric Stimulation of the Cat's Inner Ear,'' Doctoral thesis, MIT.

Nelson, D. A., Schmitz, J. L., Donaldson, G. S., Viemeister, N. F., and Javel, E. (**1996**). ''Intensity discrimination as a function of stimulus level with electrical stimulation,'' J. Acoust. Soc. Am. **100**, 2393–2414.

Shannon, R. V. (**1992**). ''Temporal modulation transfer functions in patients with cochlear implants,'' J. Acoust. Soc. Am. **91**, 2156–2164.

Wojtczak, M., and Viemeister, N. F. (**1999**). ''Intensity discrimination and detection of amplitude modulation,'' J. Acoust. Soc. Am. **106**, 1917–1924.

Zeng, F.-G., and Shannon, R. V. (**1994**). ''Loudness-coding mechanisms inferred from electric stimulation of the human auditory system,'' Science **264**, 564–566.

Zeng, F.-G., Shannon, R. V., and Hellman, W. S. (**1998**). ''Physiological processes underlying psychophysical laws,'' *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 473–481.

A cross-language study of the identification of non-native nasal consonants varying in place of articulation

James D. Harnsberger[a)]
Program in Linguistics, University of Michigan, 1076 Frieze Building, Ann Arbor, Michigan 48109

(Received 27 January 1999; revised 17 April 2000; accepted 19 April 2000)

Seven listener groups, varying in terms of the nasal consonant inventory of their native language, orthographically labeled and rated a set of naturally produced non-native nasal consonants varying in place of articulation. The seven listener groups included speakers of Malayalam, Marathi, Punjabi, Tamil, Oriya, Bengali, and American English. The stimulus set included bilabial, dental, alveolar, and retroflex nasals from Malayalam, Marathi, and Oriya. The stimulus set and nasal consonant inventories of the seven listener groups were described by both phonemic and allophonic representations. The study was designed to determine the extent to which phonemic and allophonic representations of perceptual categories can be used to predict a listener group's identification of non-native sounds. The results of the experiment showed that allophonic representations were more successful in predicting the native category that listeners used to label a non-native sound in a majority of trials. However, both representations frequently failed to accurately predict the goodness of fit between a non-native sound and a perceptual category. The results demonstrate that the labeling and rating of non-native stimuli were conditioned by a degree of language-specific phonetic detail that corresponds to perceptually relevant cues to native language contrasts. © *2000 Acoustical Society of America.* [S0001-4966(00)00508-7]

PACS numbers: 43.71.An, 43.71.Hw, 43.71.Es [JMH]

INTRODUCTION

A. Background

Cross-language speech perception research has demonstrated that specific linguistic experience can limit listener sensitivity to some non-native phonemic distinctions, particularly consonantal ones (Abramson and Lisker, 1970; Miyawaki *et al.*, 1975; Werker *et al.*, 1981). Pairs of non-native consonant stimuli that are phonemically contrastive in the non-native language but are similar to a single phoneme in a listener's native language have often proven to be difficult for listeners to accurately discriminate and identify, such as /l–ɹ/ for Japanese listeners (Goto, 1971; Miyawaki *et al.*, 1975; MacKain *et al.*, 1981), Hindi dental versus retroflex stops for English listeners (Werker *et al.*, 1981; Tees and Werker, 1984; Werker and Tees, 1984b), and Salish /k′–q′/ for English (Werker and Tees, 1984a, b) and Farsi listeners (Polka, 1992). However, non-native consonant contrasts have been shown to vary in their discriminability, from chance-level to near native-level performance. For instance, Polka (1991) and Pruitt (1995) have demonstrated that the discriminability of Hindi dental versus retroflex stop contrasts for English listeners can vary significantly as a function of voicing/manner class, despite the fact that English listeners have only a single category (/t/ or /d/) that the contrasts correspond to. English listeners were quite successful in discriminating the Farsi velar-uvular stop contrast, though they did not perform as well as native listeners (Polka, 1992).

Several factors have been proposed to account for this cross-language variation in the discriminability of non-native contrasts, including the psychophysical salience of the contrast (Sheldon and Strange, 1982; Burnham, 1986; Polka, 1991); a listeners general experience with features employed in the contrast (Werker *et al.*, 1981; Polka, 1992); the particular allophonic variants, or phonetic realizations, of the native category (Henly and Sheldon, 1986; Ingram and Park, 1998); and the degree of similarity between non-native sounds and native perceptual categories (Best *et al.*, 1988; Best, 1995; Kuhl, 1991; Iverson and Kuhl, 1995, 1996), with stimuli and categories usually defined using single, abstract labels (e.g., ''dental,'' ''retroflex''). Of these factors, the last, degree of perceptual similarity, is very commonly cited in *posthoc* explanations of experimental results. For example, Best *et al.* (1988) discovered that nine Zulu click contrasts, differentiated only by place of articulation, were nevertheless highly discriminable for American English listeners. Based on their listeners' posttest questionnaires, Best *et al.* (1988) concluded that the relative discriminability of the click contrasts was due to their lack of similarity to any perceptual category in the listeners' experience, allowing listeners to process the stimuli in an auditory, rather than speech, mode. Some form of phonetic similarity is always assumed in cross-language studies, with researchers making educated guesses about the correspondence of non-native stimuli to native categories based on abstract phonological or phonetic characterizations of the stimuli and categories. For instance, Schmidt (1996), in a study of Korean listeners' identification of American English consonants in initial position, predicted that English aspirated stops, /tʃ/, and nasals, /j/, /w/, /s/, /h/, would be consistently labeled with the corresponding Korean phonemes, based on earlier acoustic studies

[a)]Now at Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, IN 47405. Electronic mail: jharnsbe@indiana.edu

of Korean consonants. However, English /s/, /j/, and /w/ were matched to their Korean counterparts in only 32%, 67%, and 67% of responses, respectively. Schmidt (1996) accounted for mixed labeling patterns such as these by making reference to both acoustic properties of the stimuli and to allophonic variants of Korean phonemes.

While virtually all cross-language phonetic perception studies, and, by extension, all cross-language speech models, assume some kind of phonetic similarity metric relating non-native stimuli to native categories, no current model currently incorporates an explicit, language-general metric of similarity. Such a metric has been difficult to devise since its primitive units are unclear. The simplest candidate primitive would be the phoneme. This is a logical choice considering that the perceptual difficulties encountered by non-native listeners are likely related in some way to the properties of their native lexicons, whose structure is shaped by the minimally contrastive units of words, namely phonemes. If we assume the phoneme as the primitive unit in cross-language speech perception, it should be quite feasible to collect similarity judgments among a large subset of phonemes, say vowels, from speakers of different languages varying in their phonemic inventory.[1] From similarity matrices constructed from a set of such studies, a language-general similarity model could be developed for use in cross-language speech perception research.

There are, however, several problems with a phoneme-based approach. While the phonemic contrasts in the language undoubtedly color our perception of most non-native sounds, listeners in a number of studies have shown a sensitivity to greater phonetic detail in non-native stimuli than can be captured in phonemic representations (Polka, 1991; Pruitt, 1995). This detail may correspond to the individual gestural or acoustic cues that play a predominant role in signaling the identity of a native sound in question. Adult listeners of different languages have been shown to weight these cues in a language-specific manner (Terbeek, 1977; Gottfried and Beddor, 1988; Rochet, 1991). In recent models of cross-language speech perception, the phoneme has been largely abandoned as a sufficient representation for native categories, replaced for the moment by another, more detailed, symbolic unit, the position-dependent allophone. This choice reflects observed differences in the acoustic cues of phonemes in different syllable positions (Strange, 1995). However, even this unit may be too abstract if languages, and dialects of languages, differ significantly in their weighting of cues to segment identity (Lisker and Abramson, 1964; Rochet, 1991). It is important for future research to know to what extent abstract representations such as phonemes or allophones are sufficiently detailed to account for the perception of non-native contrasts by a variety of listener groups, and to what extent more detailed representations, involving features or cues, are necessary.

To address the issue of the efficacy of abstract representations in cross-language research, the present study examined the identification of non-native sounds by several non-native listener groups varying in their inventory of similar perceptual categories to the stimuli, described at a phonemic and allophonic level of representation. The phonemic and allophonic levels of representation were chosen as two abstract models of perceptual categories that are commonly assumed in cross-language studies. The study used as stimuli nasal consonants varying in place of articulation (bilabial, dental, alveolar, and retroflex) from three languages, Malayalam, Marathi, and Oriya. A bilabial and coronal[2] nasal series varying in place of articulation constituted a stimulus set that had not been examined in previous cross-language speech perception studies. This set was chosen to maximize the potential for varying degrees of perceptual performance by non-native listeners with different native languages, given the potential difficulty that non-native place contrasts often present to listeners (Werker *et al.*, 1981), and given that nasals varying in terms of place have been shown to be confusable as a group relative to other contrasts (Mohr and Wang, 1968; Hura *et al.*, 1992). Of the three languages, Malayalam is the only one to possess the full complement of nasals. However, it exhibits a number of phonotactic constraints that impose limitations on the kind of stimuli that could be elicited: the bilabial, dental, alveolar, and retroflex nasals all contrast only when occurring as intervocalic geminates. To ensure that the results were generalizable beyond Malayalam intervocalic geminates, nasals from Marathi and Oriya were included, languages that allow some of these contrasts as singletons in medial and final position.

Listeners in the study were drawn from seven groups, representing a range of native bilabial and coronal nasal consonant inventories: bilabial-dental-alveolar-retroflex (Malayalam), bilabial-dental-retroflex (Marathi and Punjabi), bilabial-alveolar-retroflex (Tamil and Oriya), and bilabial-alveolar (Bengali and American English). The inventories of nasal consonants in these languages, for the purposes of this experiment, were represented at both a phonemic and position-dependent allophonic level of representation. The use of multiple listener groups and non-native contrasts increased the likelihood that the identification patterns collected in this study could be used to successfully gauge the extent to which two types of abstract representations, phonemic and allophonic, successfully capture important details in the speech perception process. In addition, the study concerns a set of sounds and a number of listener groups that have not been examined in previous cross-language speech perception studies. Thus, the results of this study make an important contribution to understanding how listeners identify non-native sounds.

Stimuli were presented to these listeners in an identification task using an orthographic classification paradigm in which the response set consisted of letters in the native orthography. Orthographic classification is a technique that has been employed in a number of identification tests. The response sets are usually either single letters in the language (Beddor and Strange, 1982; Strange *et al.*, 1998), or keywords (Polka and Bohn, 1996). In addition to labeling non-native nasal consonants using native orthography, listeners were asked to rate the goodness of the match on a five-point scale. The results of this test were used to evaluate the *phonemic* and *allophonic category* hypotheses, namely, that listeners will identify a non-native stimulus with their nearest native perceptual category, corresponding either to a pho-

TABLE I. Demographics of talkers. ''NL''=native language, either Malayalam (ML), Marathi (MR), or Oriya (OR). ''Home''=home city or district within India. ''Years''=years outside of an environment where the native language is widely spoken.

Name	NL	Sex	Age	Home	Years	Other languages spoken
Ym	ML	m	58	Malabar	29[a]	English, Hindi
Ys	ML	f	47	Malabar	26[a]	English, Hindi, Tamil
Ms	MR	m	26	Mumbai	1	English, Hindi
Mv	MR	f	35	Mumbai	6.5	English, Hindi, Gujarati
Oc	OR	f	35	Cuttack	5	English, Hindi, Marathi, Bengali
Os	OR	f	30	Bhubaneswar	9	English, Telegu, Hindi

[a]While both Malayalam talkers had spent half or more of their lifetimes outside of their home state of Kerala, where Malayalam is spoken, both reported using Malayalam regularly at home with their Malayalam-speaking spouses. Moreover, their stimuli were consistently identified correctly by three native speakers of Malayalam.

neme of the listeners' language, or the allophonic variant of a phoneme in the context in which the non-native stimulus appears. The phonemic category hypothesis represents that classic view of the role of native phonemes in cross-language speech perception as a ''sieve'' filtering out all discrepant phonetic detail (Trubetzkoy, 1969). The allophonic category hypothesis forms a part of Flege's (1995) Speech Learning Model, and is similar to the *strong phonological hypothesis* offered by Best (1994) to account for non-native speech perception by older infants. The phonemic and allophonic category hypotheses assume that non-native sounds are mapped to their respective representations, and that listeners are less sensitive to the phonetic detail that is not captured at their respective levels. If a given abstract level of representation is the appropriate level of detail for modeling cross-language speech perception, then we should be able to predict what native category a listener of a particular language might use in identifying a non-native stimulus, how closely a non-native stimulus will be identified with a native category, and whether or not listeners who differ in their native language will also differ in their identification of non-native stimuli.

This experiment is described in detail in Sec. I A. In Sec. I B the predicted identification patterns of each listener group are given based on the phonemic and allophonic representations of native categories; in Sec. I C the results of the experiment are listed, while a discussion of the results and their implications for cross-language speech perception research and modeling appear in Secs. II and III.

I. EXPERIMENT

A. Method

1. Stimulus materials

Six talkers, two each of Malayalam, Marathi, and Oriya, were recorded reading from a list of real and nonsense words from their native language, in both isolation and a sentence frame, with five repetitions each for a total of ten repetitions. Nonsense words were read in cases where the lexicon of the language did not provide a word composed of a necessary sequence of vowel(s) and nasal(s). The nasals of interest appeared in all syllable positions allowable by the individual languages, in an [a], [i], or [u] vocalic context. Table I lists some relevant characteristics of the speakers who produced the stimuli.

The recording took place in a sound-attenuated chamber in the University of Michigan Phonetics Laboratory using a Panasonic SV3500 DAT recorder. The stimuli were digitized at 44.1 kHz (filtered at 22 kHz), randomized, and played with a 3.5-s intertrial interval (ITI) to native speakers of the respective languages in an identification test in order to exclude any stimuli from use in the experiment that might be poor exemplars. Only those stimuli that were consistently identified (in 100% of all trials) by native speakers in a pilot study were used in this experiment. Of the stimuli that were recorded and evaluated, a subset was used in piloting and in the experiment: four exemplars from an isolation context, two from each talker, of 18 types of stimuli. The stimulus types are listed in Table II.[3] Particular stimuli were chosen to minimize tone, duration, and other differences between the stimuli that were deemed irrelevant to the identity of the stimuli.

All of the nasal stimuli appeared in an [a] context, with the exception of the Malayalam medial geminate series, which appeared in both the [a] and [i] contexts. Pilot testing had shown that the identification by non-native listeners of talker Ym's dental stimuli, which were actually produced as interdentals ([n̪]), differed according to vocalic context. In an [a] context, Ym's [n̪] stimuli were identified as /m/ exemplars by non-native listeners, while in an [i] context, Ym's [n̪] received the more expected label of dental or alveolar, depending on the listener group in question.[4] Both vocalic contexts were preserved in this case, given this un-

TABLE II. Stimuli and their source languages. The vocalic context was [a] for all but underlined stimuli. Underlining indicates that the stimulus appeared in [i] as well as [a] contexts. The dental nasal of Malayalam talker Ym was produced as an interdental.

		Nasal			
Language	Syllable	m	n̪	n	ɳ
Malayalam	VCV			√	√
Marathi			√		√
Oriya				√	√
Malayalam	VC:V	√	√	√	√
Marathi			√		√
Oriya					
Malayalam	VC				
Marathi			√		√
Oriya					

TABLE III. The perceptual category inventories for the seven listener groups, at the allophonic level of detail. ''Syllable''=syllabic context the nasal appears in.

Group	Syllable	Perceptual category			
		Bilabial	Dental	Alveolar	Retroflex
Malayalam	VCV	m		n	ɳ
	VC:V	m	n̪	n	ɳ
	VC	m		n	
Marathi	VCV	m	n̪		ɳ
	VC:V	m	n̪		ɳ
	VC	m	n̪		ɳ
Punjabi	VCV	m	n̪		ɳ
	VC:V	m	n̪		
	VC	m	n̪		ɳ
Tamil	VCV	m		n	ɳ
	VC:V	m		n	ɳ
	VC				
Oriya	VCV	m		n	ɳ
	VC:V				
	VC[a]	m		n	ɳ
Bengali	VCV	m		n	
	VC:V	m		n	
	VC	m		n	
English	VCV	m		n	
	VC:V				
	VC	m		n	

[a]Oriya allows/disallows final consonants, depending on the dialect spoken [see Harnsberger (1998) for a summary].

expected and interesting identification pattern.

2. Participants

Native speakers of Malayalam, Marathi, Punjabi, Tamil, Oriya, Bengali, and American English were chosen for this experiment to represent a range of nasal consonant inventories for the purpose of testing the phonemic and allophonic category hypothesis. A description of the nasal consonant inventories of these seven groups, at the allophonic level of detail, appears in Table III. Fifteen to 18 speakers of each group were recruited and tested. Some subjects were eventually excluded from the results analysis due to frequent lapses in answering within the test sequences,[5] leaving 12 to 18 subjects per listener group for a given test. Table IV lists the total number of subjects, listed by gender, that were recruited for each listener group, along with the mean age of each listener group.

TABLE IV. The number of subjects in each listener group. ''N''=number of participants, ''M''=number of male participants, and ''F''=number of female participants.

Listener group	N (M,F)	Mean age
Malayalam	18 (12,6)	24
Marathi	17 (10,7)	23
Punjabi	14 (13,1)	22
Oriya	16 (12,4)	24
Tamil	12 (9,3)	22
Bengali	15 (9,6)	24
American English	18 (3,15)	19

All but the English listeners were tested in India, in order to recruit subjects who varied little in terms of age, dialect spoken, and overall linguistic experience. This last criterion was especially important: if listeners who belonged to a particular group also had experience in another language with a richer set of nasal contrasts, their identification results might be influenced by these contrasts. To limit the effect of the bi- or multi-lingualism of the listeners as a source of significant variability in the results, only those listeners were tested who had no experience in a language which employed a greater number of coronal nasal contrasts than those in the listener's first language. The demographic background of all the listeners, including their experience in languages other than their first language, appears in the Appendix.

For the great majority of subjects from India, their experience with second languages consisted of classroom instruction in Hindi, an important national language of India, and in English, the language of instruction at higher institutions of learning. Neither Hindi nor English afforded the subjects from India experience with a richer set of nasal consonants than that found in their native language. Eleven of the 92 subjects from India also reported some degree of experience with other European languages (German, Russian, or French) or Japanese. None of the multi-lingual subjects from India reported acquiring a second language abroad, in an environment in which the second language was spoken by native speakers (e.g., instruction in French or Japanese took place in India, rather than France or Japan). Of the English subjects, all but two reported some degree of instruction in a second language, commonly Spanish or French, at the high school or college level. Moreover, two English subjects reported spending more than four months living in a country where a language other than English is commonly spoken. However, none of the English subjects had experience in a language that possessed a two- or three-way coronal nasal distinction that might aid them in identifying non-native dental or retroflex consonants.

All subjects from India were recruited by posting flyers on the campuses of local universities. The English listeners were recruited through introductory linguistics classes at the University of Michigan. The Malayalam, Oriya, Tamil, and Marathi listeners were students attending national universities in India's capital, New Delhi. The Bengali listeners were university students who were recruited and tested in Calcutta, the capital of West Bengal, where Bengali is primarily spoken. The Punjabi listeners were recruited and tested in Amritsar, the cultural center of the Punjab state in northwestern India. The Malayalam, Marathi, and Oriya listener groups were recruited to serve as controls. In addition, all listener groups served as experimental groups representing a diverse range in nasal consonant inventories to test the phonemic and allophonic category hypotheses.

3. Procedure

Stimuli were presented binaurally over Sony MDR-7506 headphones connected to a Sony TCD-D8 portable DAT recorder. Responses were made on photocopied answer sheets. The experiment consisted of an orthographic classification task in which listeners used response sets based on native

phonemic categories. Specifically, the response choices consisted of individual letters in the native orthography. Listeners were also instructed to provide category-goodness judgments on a five-point scale, a task that would allow listeners to note the degree of difference between a non-native stimulus and the nearest native category. Listeners identified and rated two exemplars each of the 18 types of stimuli, produced by two talkers each, for a total of 72 stimuli. The identification test included two repetitions of this set, presented in random order, for a total of 144 trials. The test employed a 6-s intertrial interval and a 6-s interblock interval, with ten trials per block. Listeners were instructed to ignore ''irrelevant differences'' of duration, tone, or voice quality. Before the test began, ten trials were presented to familiarize listeners with the time allotted for labeling and rating a stimulus, with no feedback from the investigator.

B. Predictions

Two sets of predicted identification patterns of non-native stimuli were generated based on the phonemic and position-dependent allophonic representations of the nasal consonants of each listener group. For the majority of listener groups, including Marathi, Tamil, Oriya, Bengali, and English, the two hypotheses did not make many different predictions concerning the top labeling choices for non-native stimulus types. Thus, these listener groups served as a general test of abstract models of the perceptual category, both in terms of their own results as well as in their comparison to another group with a similar nasal consonant inventory.

Two of the listener groups, Malayalam and Punjabi, served as good test cases to differentiate the phonetic and allophonic category hypothesis. For example, the full set of Malayalam nasal categories are not preserved in final position. Instead, only the bilabial and alveolar categories are available. Thus, non-native dental or retroflex stimuli in final position may have been labeled as alveolar, despite the fact that the Malayalam phonemic inventory includes dental and retroflex nasals. In Punjabi, the dental-retroflex nasal contrast is not preserved in medial geminates—only the dental nasal appears in that position. According to the allophonic category hypothesis, Punjabu listeners should identify medial geminate retroflex nasals with their dental category, instead of the retroflex category as the phonemic category hypothesis predicts.

The studies of listener languages used in making the predictions for the phonemic and allophonic category hypothesis can be classified into three broad categories, defined by the type of evidence used to support a particular characterization of a language's nasal consonant inventory. First, there are traditional descriptions of languages spoken in India that are based directly on the organization of the language's alphabet, which groups consonants by place of articulation (from velar to bilabial) following the ordering of Brahmi letters in Sanskrit. These traditional descriptions were often inadequate in providing sufficient phonetic detail, and were discounted in positing representations where they conflicted with more reliable kinds of evidence. Second, there are detailed phonetic descriptions by linguists who are also native speakers of the language, based on their own articulation of the sound(s) in question and/or by auditory judgments of other native speakers. Finally, there are detailed articulatory and/or acoustic studies, typically based on a very small sample of data in the case of Indo-Aryan or Dravidian languages. This final kind of study took precedence over the others in the development of representations when different studies conflicted in their characterization of a language's nasal consonants. Given these criteria, the following studies were used in characterizing the nasal consonant inventories of the listener groups' native language: Malayalam (Subramoniam, 1962; Aiyar, 1972; Kumari, 1972; Moag, 1973; Kalackal, 1985; Dart, 1991; Ladefoged and Maddieson, 1996; Asher and Kumari, 1997), Marathi (Firth, 1948; Kelkar, 1958; Jha, 1977; Pandharipande, 1997), Punjabi (Jain, 1934; Arun, 1961; Gill and Gleason, 1986; Sandhu, 1986), Oriya (Dhall, 1966; Pattanayak and Das, 1972; Bhattacharya, 1993), Tamil (Švarný and Zvelebil, 1955; Balasubramanian, 1982; Ladefoged and Bhaskararao, 1983), Bengali (Hai, 1960; Kostic and Das, 1972; Haldar, 1986; Bhattacharya, 1988, 1993), American English (Dart, 1991).

TABLE V. Predictions of identification patterns based on the phonemic category hypothesis. ''Place''=place of articulation of a stimulus type, ''ML''=Malayalam, ''MR''=Marathi, ''PN''=Punjabi, ''OR''=Oriya, ''TM''=Tamil, ''BN''=Bengali, and ''EN''=American English. Perceptual categories that ''match'' or share the same symbol with a stimulus type are underlined. Perceptual categories that are only tentatively predicted for a stimulus type appear in plain text without underlining.

Stimulus type place	Listener group						
	ML	MR	PN	OR	TM	BN	EN
m	m	m	m	m	m	m	m
n̪	n̪	n̪	n̪	n/ɳ	n/ɳ	n	n
n	n	n̪/ɳ	n̪/ɳ	n	n	n	n
ɳ	ɳ	ɳ	ɳ	ɳ	ɳ	n	n

The predicted identification patterns of each group, based on their phonemic- and allophonically based nasal consonant inventories, are presented in Tables V and VI. Many of these predictions were simple to make in cases in which the stimulus in question shared the same abstract representation as the category. For instance, in the case of allophonic category hypothesis predictions, all medial geminate alveolar nasal stimuli were predicted to be consistently labeled with medial geminate alveolar categories, and receive high category-goodness ratings. Such pairs of stimulus types and categories that shared the same abstract symbol are referred to as *matched* pairs.

For listener groups that lacked a matching native category for a given stimulus type (*mismatched* pairs), it was more problematic to make predictions because no prior acoustic or perceptual studies exist concerning similarity among dental, alveolar, and retroflex nasals, or the identification of nasal contrasts that correspond to a native contrast that appears in a different syllabic context. For example, in the case of the Bengali and English listener groups, one might predict that non-native dental and retroflex nasals would be identified with the native /n/ category rather than

TABLE VI. Predictions of identification patterns based on the allophonic category hypothesis. ''Place''=place of articulation of a stimulus type, ''NL''=the native language of the talker who produced the stimulus type, and ''Syllable''=syllabic context of the stimulus type. ''ML''=Malayalam, ''MR''=Marathi, ''PN''=Punjabi, ''OR''=Oriya, ''TM''=Tamil, ''BN''=Bengali, and ''EN''=American English. Perceptual categories that ''match'' or share the same allophonic-based symbol with a stimulus type are underlined. Perceptual categories that are only tentatively predicted for a stimulus type appear in plain text without underlining.

Stimulus type			Listener group						
Place	NL	Syllable	ML	MR	PN	OR	TM	BN	EN
m	ML	VC:V	m	m	m	m	m	m	m
n̪	ML	VC:V	n̪	n̪	n̪	n/ɳ	n/ɳ	n	n
	MR	VC	n	n̪	n̪	n/ɳ	n/ɳ	n	n
	MR	VC:V	n̪	n̪	n̪	n/ɳ	n/ɳ	n	n
	MR	VCV	n/ɳ	n̪	n̪	n/ɳ	n/ɳ	n	n
n	ML	VC:V	n	n̪/ɳ	n̪/ɳ	n	n	n	n
	ML	VCV	n	n̪/ɳ	n̪/ɳ	n	n	n	n
	OR	VCV	n	n̪/ɳ	n̪/ɳ	n	n	n	n
	ML	VC:V	n̪	n̪	n̪	ɳ	n̪	n	n
ɳ	ML	VCV	ɳ	ɳ	ɳ	ɳ	ɳ	n	n
	MR	VC	n	ɳ	ɳ	ɳ[a]	ɳ	n	n
	MR	VC:V	n̪	n̪	n̪	ɳ	n̪	n	n
	MR	VCV	n̪	n̪	n̪	n̪	n̪	n	n
	OR	VCV	ɳ	ɳ	ɳ	ɳ	ɳ	n	n

[a]The retroflex category is available if these Oriya listeners maintain the alveolar-retroflex distinction in final position, otherwise the listener group's native category for this stimulus type should be alveolar. See Harnsberger (1998) for a discussion of dialect differences in Oriya and the alveolar-retroflex distinction in final position.

their /m/, based on their gestural similarity. However, for the Marathi and Punjabi listeners, all of whom have dental and retroflex categories, it was not possible to predict whether alveolar nasals would be identified with either the /n̪/ or the /ɳ/ category. Likewise, for Tamil and Oriya listeners who possess alveolar and retroflex nasal categories, it was not possible to predict how non-native dental nasals would be identified.

There were also cases in the set of allophonic category hypothesis predictions in which a listener group lacked any nasal contrast in the syllable context that a non-native stimulus appeared in. For instance, Oriya listeners have an alveolar-retroflex distinction in medial position, but their native language lacks geminate nasals. How would Oriya listeners label alveolar and retroflex geminate nasals? One might suppose that the most similar nasal set in Oriya to medial geminate nasals would be the medial singleton series, such that non-native [Vn:V] and [Vɳ:V] would be labeled as the Oriya alveolar and retroflex nasals, respectively (V =vowel). However, there was no way to predict with confidence that such a similarity relationship between medial singletons and geminates would actually determine Oriya labeling of geminate nasals. For such cases in which a labeling prediction was only tentative, the potential labeling choice, or choices, appears in Tables V and VI in plain text. For these cases, the results of this study will serve a useful role in phonetic description, in addition to testing the phonemic and allophonic category hypothesis. Finally, for both matched and mismatched pairs, none of the listener groups were predicted to label or rate any of the stimulus types differently on the basis of talker or vocalic context.

If phoneme- or allophone-based representations are sufficient descriptions of listener categories in cross-language research, then stimulus types that matched a category's symbol (matched pairs) were predicted to be ''consistently'' labeled with their corresponding native category and receive ''high'' category-goodness ratings because the listener is presumed to map the stimulus to its corresponding category. Inconsistent labeling or low category-goodness ratings were taken to mean that the listener was sensitive to the differences between the non-native stimulus and his/her most similar native category.

The ''consistent'' labeling predicted by the phonemic and allophonic category hypotheses was not defined by a single threshold, such as 75% or 90% of the labeling responses. Instead, for each stimulus type, the most popular native label that was chosen by a listener group was determined. This label will be referred to as the *top labeling choice*, or *top label*. A listener group's top label for a given stimulus type represents a percentage of the group's labeling responses, hereafter referred to as the *percent labeling rate*. The stimulus type(s) that elicited the highest percent labeling rates from a listener group were said to be ''consistently'' labeled. In addition, stimulus types were classified as ''consistently'' labeled by a listener group if those stimulus types' percent labeling rates were statistically identical to the highest labeling rates, in *posthoc* least-square means *t*-tests in a mixed model ANOVA with an alpha level of $p>0.05$.[6] Thus, the range of predicted labeling rates could vary somewhat by listener group.

For the definition of ''high'' category-goodness ratings, a statistical comparison method was also used instead of an absolute threshold for all listener groups. ''High'' category-goodness ratings were defined in this study as significantly

greater than, or statistically identical to, those of all of the stimulus types that received a mean category-goodness rating identical to the mean rating across all stimulus types for a given listener group (rounded to one decimal point), in *posthoc* least-square means *t*-tests in a mixed model ANOVA with an alpha level of $p>0.05$. For example, if a group's mean rating across all stimulus types was 3.5, then every stimulus type with a rating of 3.5 was compared to all other stimulus types in *posthoc t*-tests. The stimulus types with ratings that were lower than and significantly different from stimulus types with a 3.5 rating were not judged as having received ''high'' ratings. Just as with the percent labeling rate data, the use of a statistical comparison method meant that the range of predicted ratings varied slightly by listener group.

The ''mean-based'' criterion in the definition of ''high'' category-goodness ratings is a more liberal measure than that used to define ''consistent'' in the labeling data, which was based on a group's highest percent labeling rate. The ''mean-based'' criterion was chosen to take into account the possibility that when subjects are presented with stimuli to rate, many may attempt to employ a large portion (or the entire range) of the rating scale. In doing so, they may give a lower rating to a stimulus that differs from their category representation along acoustic dimensions that are not exploited phonemically in the language.

Finally, the identification test results were also predicted not to vary for groups with a similar nasal consonant inventory, such as Marathi and Punjabi, Tamil and Oriya, and Bengali and American English. That is, if two listener groups had the same, phonemically or allophonically defined, perceptual category for a given stimulus type, they were not predicted to significantly differ in their top labeling choice, their top label's percent labeling rate, and their top label's mean category-goodness rating. For the cross-language comparisons of percent labeling rates and category-goodness ratings, only those stimulus types were used which received the same label from both groups. In addition, those stimulus types had to be either matched pairs for both groups, or mismatched pairs for both. The number of significant cross-language differences observed in the experiment was another gauge of the efficacy of both phonemic and allophonic representations of perceptual categories.

C. Results

1. Overall

The results for each listener group appear in Figs. 1–7 for Malayalam, Marathi, Punjabi, Tamil, Oriya, Bengali, and English, respectively. In Figs. 1–7, the results are presented by each stimulus type (i.e., ''Ym [am:a]''), with sets of stimulus types defined using large brackets with phonetic symbols denoting either the bilabial, dental, alveolar, or retroflex nasal. Next to the stimulus type, the top labeling choice for that stimulus type is listed as an IPA symbol. Following the top labeling choices are two sets of bar graphs representing each stimulus type's mean percent labeling rate and mean category-goodness rating. In both columns, shaded bars denote those values that were significantly lower than

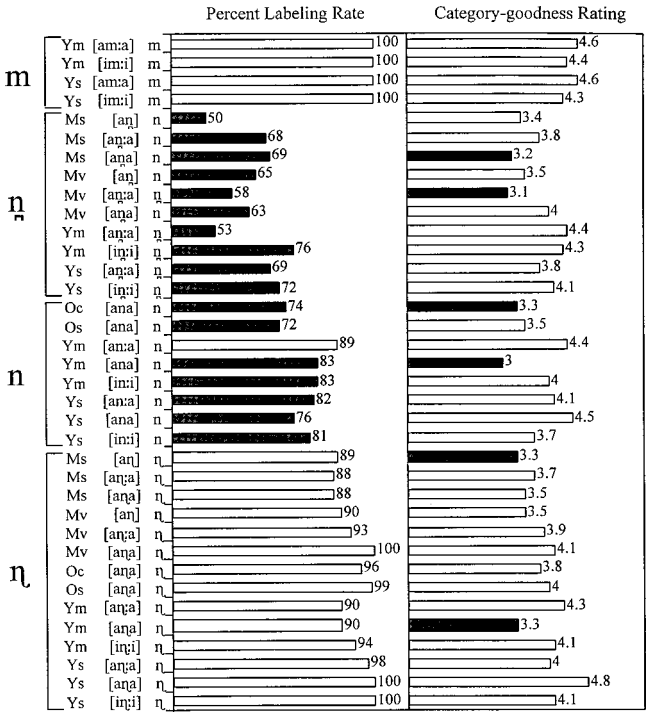


FIG. 1. The identification test results for Malayalam listeners. Shaded bars denote percent labeling rates or category-goodness ratings that did not match their predicted value.

the predicted range of values. As stated in Sec. I B, the predicted range of values for the percent labeling rates and category-goodness ratings involved a statistical comparison method based on either the highest or mean values elicited in the experiment. Such a method could result in small cross-language differences in the range of predicted values, and

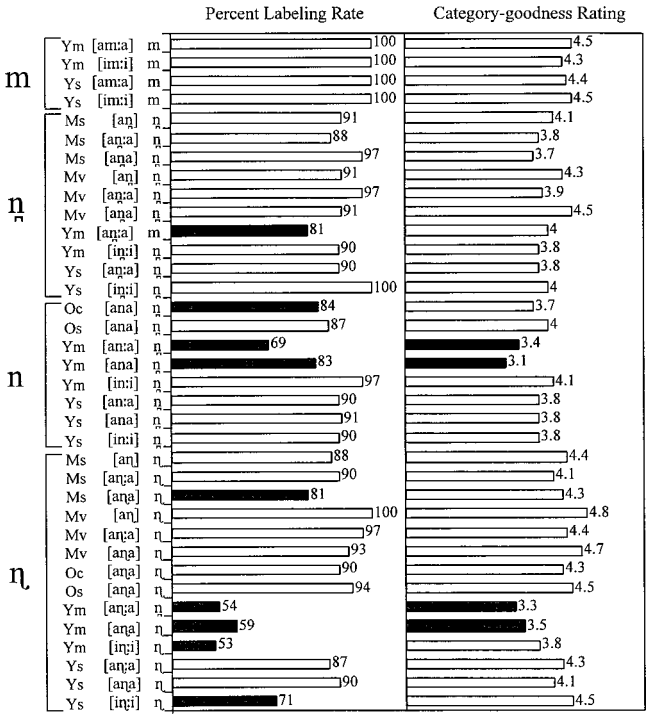


FIG. 2. The identification test results for Marathi listeners. Shaded bars denote percent labeling rates or category-goodness ratings that did not match their predicted value.

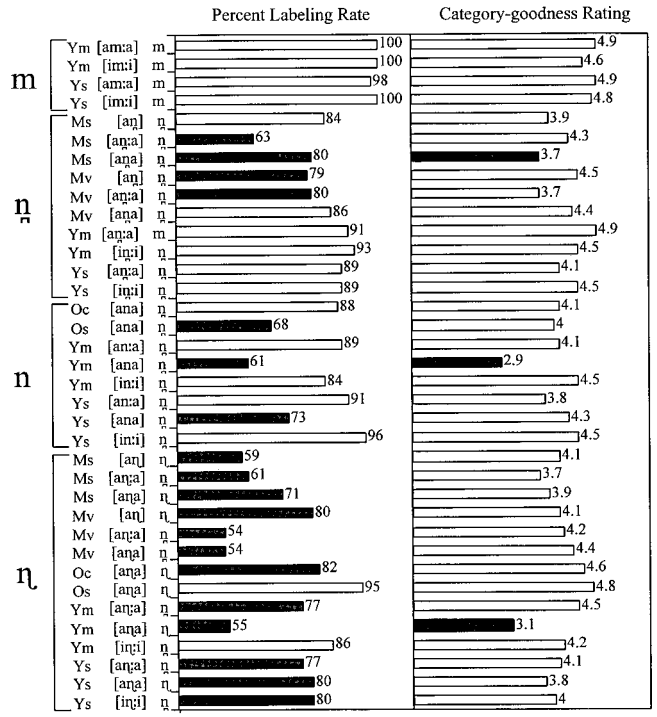


FIG. 3. The identification test results for Punjabi listeners. Shaded bars denote percent labeling rates or category-goodness ratings that did not match their predicted value.

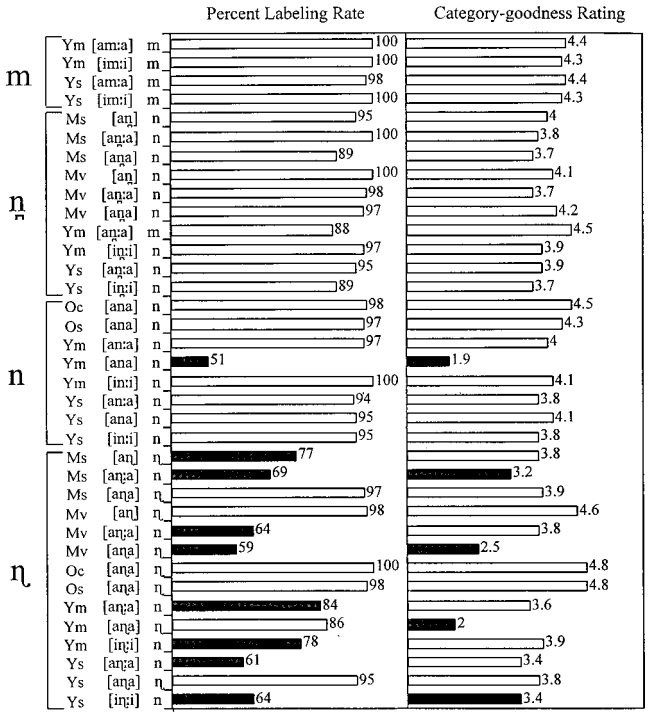


FIG. 5. The identification test results for Oriya listeners. Shaded bars denote percent labeling rates or category-goodness ratings that did not match their predicted value.

did, in fact, for this experiment. For the percent labeling rates, the ranges of predicted values were, by listener group: Malayalam: 88%–100%; Marathi: 87%–100%; Punjabi: 84%–100%; Oriya: 86%–100%; Tamil 85%–100%; Bengali 85%–100%; and English 89%–100%. For the category-goodness ratings, the ranges of predicted values were Malayalam: 3.4–5.0; Marathi: 3.7–5.0; Punjabi: 3.7–5.0; Oriya: 3.4–5.0; Tamil 3.1–5.0; Bengali 4.1–5.0; and English 3.1–5.0.

The cross-language differences are presented in Table VII for each pair of listener groups with similar coronal nasal

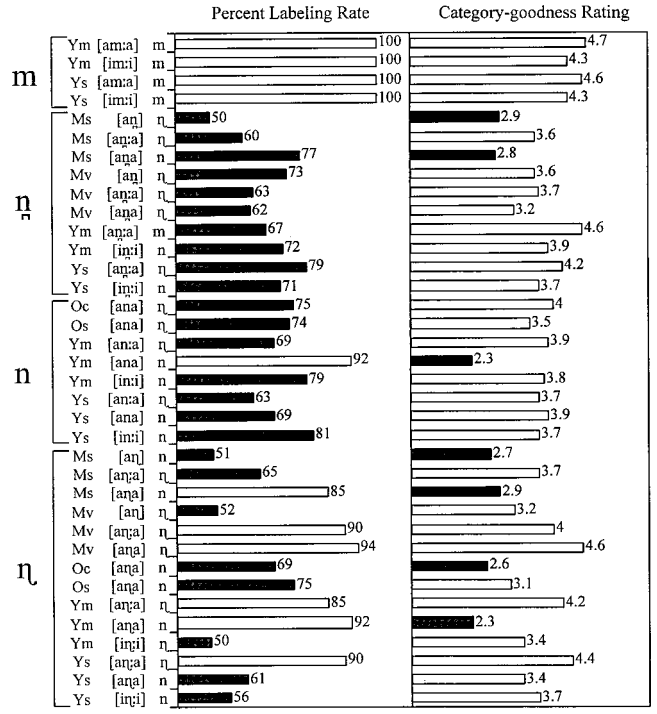


FIG. 4. The identification test results for Tamil listeners. Shaded bars denote percent labeling rates or category-goodness ratings that did not match their predicted value.

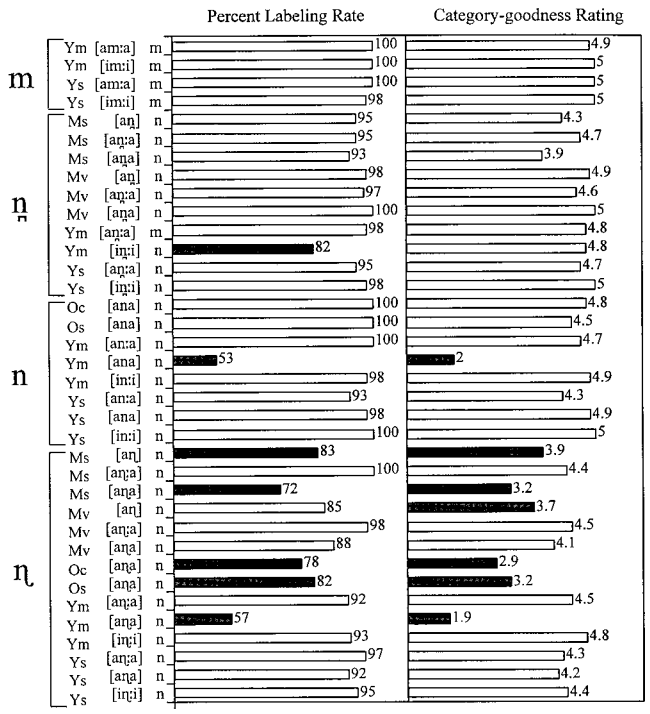


FIG. 6. The identification test results for Bengali listeners. Shaded bars denote percent labeling rates or category-goodness ratings that did not match their predicted value.

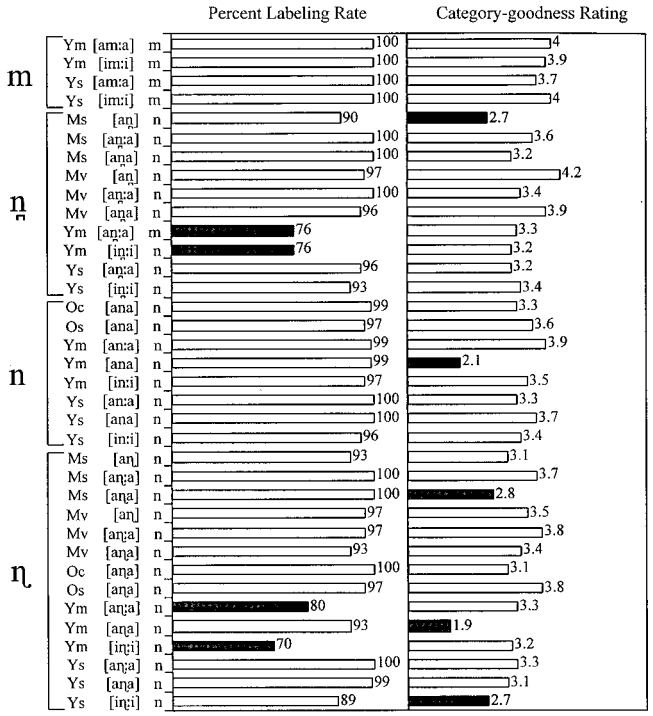


FIG. 7. The identification test results for English listeners. Shaded bars denote percent labeling rates or category-goodness ratings that did not match their predicted value.

consonant inventories (hereafter *place groups*): dental-retroflex (Marathi, Punjabi), alveolar-retroflex (Tamil, Oriya), and alveolar (Bengali, English). In Table VII, the stimulus types are grossly classified by place of articulation (m, n̪, n, ɳ). Checks (''√'') denote the percent labeling rates or category-goodness ratings that differed significantly from those predicted in *posthoc t*-tests for the three place groups.

Unlike Tables IV–VI, the stimulus types in Figs. 1–7 and Table VII are defined in terms of talker and vocalic context, in addition to the native language of the talker who produced the stimuli and the syllabic context the nasals appeared in, given the effect these variables had on the results. An examination of the percent labeling rate and category-goodness ratings results show large effects of two variables that were not predicted to affect the identification patterns of listeners, namely talker and vocalic context. The percent labeling rate and the category-goodness rating results were analyzed separately in two mixed model repeated measures ANOVAs. The factors in both ANOVAs were stimulus type (36 levels, one for each stimulus type as defined by place of articulation, syllabic context, vocalic context, and talker) and listener group (seven levels). The main effects were significant as well as their interaction in the percent labeling rate ANOVA [stimulus type: $F(6,103)=9.17$, $p\leqslant 0.0001$; listener group: $F(35,3605)=9.65$, $p\leqslant 0.0001$; interaction: $F(210,3605)=4.58$, $p\leqslant 0.0001$] as well as category-goodness rating ANOVA [stimulus type: $F(6,103)=6.52$, $p\leqslant 0.0001$; listener group: $F(35,3302)=33.51$, $p\leqslant 0.0001$; interaction: $F(210,3302)=4.61$, $p\leqslant 0.0001$].[7] Nineteen out of 26 pairs of stimulus types, which varied only in either vocalic context or talker, differed in their top labeling choice or differed significantly ($p\leqslant 0.05$) in their percent labeling rates, the latter determined by *posthoc* Tukey *t*-tests of stimulus type by listener group pairs. In addition, 17 out of 26 pairs differed significantly in the category-goodness ratings results. Only 5 out of 26 minimal pairs showed no significant differences in either percent labeling rates or category-goodness ratings: the four bilabial pairs, and Malayalam talkers' Ym and Ys's [in:i] stimuli. Given the strong effect of talker and vocalic context on the labeling and ratings data, the results will discussed at the level of these 36 stimulus types, rather than the 14 types posited in Table IV.

For each listener group and for the cross-language comparisons, the results are reported in terms of the percentage of successfully predicted top labeling choices, percent labeling rates, and category-goodness ratings by the phonemic and allophonic category hypotheses (the latter two measures are also referred to as ''goodness measures'' in the results subsections below). Specifically, these hypotheses were evaluated using the following measures:

TABLE VII. Cross-language differences by place group. ''NL''=native language, ''Tk''=talker, and ''S-V''=the syllabic and vocalic context that the nasal appears in.

Stimulus type				Dental-retroflex		Alveolar-retroflex		Alveolar	
Place	NL	Tk	S-V	%LR	Rating	%LR	Rating	%LR	Rating
m	ML	Ym	am:a						√
	ML	Ym	im:i						√
	ML	Ys	am:a						√
	ML	Ys	im:i						√
n̪	MR	Ms	an̪						√
	MR	Ms	an̪:a	√					√
	MR	Ms	an̪a	√			√		√
	MR	Mv	an̪						√
	MR	Mv	an̪:a	√					√
	MR	Mv	an̪a						√
	ML	Ym	an̪:a		√	√		√	√
	ML	Ym	in̪:i		√	√			√
	ML	Ys	an̪:a						√
	ML	Ys	in̪:i			√			√
n	OR	Oc	ana						√
	OR	Os	ana	√					√
	ML	Ym	an:a	√	√				√
	ML	Ym	ana	√		√		√	
	ML	Ym	in:i			√			√
	ML	Ys	an:a						√
	ML	Ys	ana	√		√			√
	ML	Ys	in:i		√				√
ɳ	MR	Ms	anɳ	√					√
	MR	Ms	anɳ:a						√
	MR	Ms	anɳa					√	
	MR	Mv	anɳ	√	√	√	√		
	MR	Mv	anɳ:a						√
	MR	Mv	anɳa			√	√		√
	OR	Oc	anɳa					√	
	OR	Os	anɳa						√
	ML	Ym	anɳ:a	√	√				√
	ML	Ym	anɳa					√	
	ML	Ym	inɳ:i					√	√
	ML	Ys	anɳ:a						√
	ML	Ys	anɳa						√
	ML	Ys	inɳ:i						√

(i) the percent success at predicting the top label for all matched pairs and for all mismatched pairs, *excluding* those pairs in which there were two possible predicted labels;
(ii) the percent success at predicting the goodness measures (percent labeling rates and/or category-goodness ratings) for all matched pairs, and for all mismatched pairs, *including* those pairs in which there were two possible predicted labels;[8]
(iii) the percent success at predicting any cross-language differences in the top label for stimulus types that were either matched pairs for both listener groups, or mismatched pairs for both; and
(iv) the percent success at predicting any cross-language differences in the goodness measures (percent labeling rates and/or category-goodness ratings) for stimulus types that were either matched pairs for both listener groups, or mismatched pairs for both, *and* sharing the same top label.

a. Perception of native stimuli. The Malayalam, Marathi, and Oriya listener groups, in addition to serving as representative groups having different nasal consonant inventories, acted as control groups for the stimuli used in this study. The stimuli had been screened prior to testing (see Sec. II A 1), and if Malayalam, Marathi, and Oriya listeners were each homogenous groups with respect to the perception of their own native stimuli, then category-goodness ratings and, particularly, percent labeling rates were expected to be high. Overall, this expectation was born out: Oriya, Marathi, and Malayalam listeners averaged 98%, 92%, and 87%, respectively, in their percent labeling rates with native stimulus types. None of the Oriya listeners gave any of the Oriya stimulus types significantly below-average category-goodness ratings or percent labeling rates that were significantly below the highest labeling rates. Only one Marathi stimulus type, talker Ms's [aɳa], elicited a significantly lower percent labeling rate (81%) from native speakers.

In contrast, Malayalam listeners gave 10 out of 20 of their native stimulus types either a significantly lower percent labeling rate or a significantly below-average rating. Of these stimulus types, all but one were dental or alveolar. An examination of the individual Malayalam listeners' identification performance showed that four subjects (Mal 8, 10, 12, 15; see the Appendix) performed particularly poorly with Malayalam dental and alveolar stimuli, averaging 75%, 68%, 77%, and 71% in their percent labeling rate. This inconsistent labeling by a subset of the Malayalam listeners may indicate that there exists some hitherto undescribed variation in the use of the dental-alveolar contrast in Malayalam. Though these four listeners do not share any outstanding commonalties in their background, as measured by the subject questionnaire, some sociolinguistic communities within Malayalam may have lost the dental-alveolar contrast. The phonetic realization of the dental nasal has been reported in previous work to vary by talker, with interdental, dental, and, most interestingly, alveolar instantiations reported (Dart, 1991; Ladefoged and Maddieson, 1996). If this talker variation in production mirrors listener variation in perception, then we might expect to find differences among native Malayalam speakers in their identification of dental and alveolar nasals. Excluding this subset of Malayalam subjects, the percent labeling rates for the Malayalam dental and alveolar stimuli were within the ''consistent'' range defined for this study.[9]

2. Coronal nasal inventory: Dental-alveolar-retroflex (Malayalam)

a. Top labeling choices. The top label results of the Malayalam listeners for the non-native stimulus types supported neither the phoneme nor the allophone hypothesis completely. The phonemic and allophonic category hypotheses made accurate predictions in 69% and 90% of the matched pairs, respectively. Of the four mismatched pairs for the allophonic category hypothesis (Marathi talker Ms's and Mv's [an̪] and [aɳ] stimulus types), two received the predicted top label. Each hypothesis was successful with different subsets of the data. The phonemic category hypothesis correctly predicted the labeling of all retroflex stimuli as exemplars of the Malayalam retroflex nasal, even in syllabic contexts in which it does not appear, such as final position. However, the top labels for dental stimuli were more accurately predicted by the allophonic category hypothesis. In medial and final position in Malayalam, an alveolar rather than dental nasal appears. With all non-native stimulus types in these positions, alveolar was the top label for Malayalam listeners. Out of all top labels for non-native stimulus types, only one was not predicted by either hypothesis, the Malayalam alveolar nasal for Marathi talker Ms's [an̪:a].

b. Goodness measures. The goodness measures revealed greater variation in the identification of non-native sounds by Malayalam listeners than the top labeling choices. Malayalam listeners showed particular variability in their labeling of non-native dental and alveolar stimulus types, while the bilabial and retroflex stimulus types typically elicited consistent percent labeling rates and high category-goodness ratings. The phonemic category hypothesis was only successful in predicting the percent labeling rates and/or category-goodness ratings of 44% of the non-native stimulus types, which were all matched pairs for this hypothesis. The allophonic category hypothesis fared somewhat better than the phonemic category hypothesis with its matched pairs, successfully predicting the goodness measures of 50% of these. The goodness measures for the small number of mismatched pairs were not successfully predicted but in a minority (17%) of cases. Across all non-native stimulus types, Malayalam listeners results indicate that non-native dental and alveolar stimulus types were not judged as highly similar to native dental or alveolar nasals. This performance may be related to intragroup differences in the use of the native dental-alveolar contrast (see Sec. I C 1 *a* for a discussion of this).

3. Dental-retroflex place group

a. Marathi top labeling choices. Both the phonemic and allophonic category hypotheses made the same predictions

concerning the top labeling choices of Marathi listeners. Thus, both hypotheses were strongly supported by the Marathi results: 88% of the non-native stimulus types received the predicted top label. There were just two exceptions to this trend: Malayalam talker Ym's dental (interdental) stimuli in the [a] context were identified as the Marathi bilabial nasal,[10] and Malayalam talker Ym's [aŋ̪:a] stimuli were identified with the Marathi dental nasal. In the case of the alveolar nasals, for which no predictions could be made concerning which top label they would receive, all eight stimulus types received the dental nasal as their top label. Thus, Marathi listeners, like the Malayalam listeners with Marathi [aɳa], found dental and alveolar nasals to be more similar to one another than either were to retroflex nasals.

b. Marathi goodness measures. As with the top labeling choices, both hypotheses made the same predictions concerning the goodness measures for the non-native stimulus types. For the matched pairs, only 69% of the non-native stimulus types received the predicted goodness measures. The stimulus types receiving lower percent labeling rates and/or category-goodness ratings included several Malayalam retroflex stimulus types, as well as the interdental stimuli in an [i] context. Stimulus types that received unpredicted top labels also had lower percent labeling rates and/or category-goodness ratings.

The Marathi listeners' mismatched pairs, for both hypotheses, consisted of eight stimulus types, all involving alveolar nasals: Malayalam talker Ym's and Ys's [an:a], [ana], and [in:i], and Oriya talker Oc's and Os's [ana]. For these mismatched pairs, three received percent labeling rates and/or category-goodness ratings significantly below those predicted by either hypothesis: Malayalam talker Ym's [an:a] and [ana], and Oriya talker Oc's [ana] stimuli. The non-native stimulus types that were not perceived as a good match to a Marathi category spanned a range of talkers and vocalic and syllabic contexts. Nothing in the descriptions of the stimulus types and the Marathi nasals at the phonemic or allophonic level could account for these identification patterns.

c. Punjabi top labeling choices. The Punjabi listeners' top labeling choices provided the strongest support for the allophonic over the phonemic category hypothesis. Punjabi, while maintaining a dental-retroflex distinction in medial and final singletons, only employs a dental nasal in medial geminates. The phonemic category hypothesis predicted that Punjabi listeners would be able to map geminate retroflex stimuli onto a general retroflex category, regardless of syllabic context, while the allophonic category hypothesis predicted dental top labels for such stimuli. Thus, geminate retroflexes were classified as matched pairs for the phonemic category hypothesis, and mismatched pairs for the allophonic category hypothesis. In fact, the Punjabi listeners showed the pattern predicted by the latter hypothesis: in all six [Vŋ̪:V] cases, the Punjabi dental nasal served as the top label. This labeling pattern resulted in a lower success rate for the phoneme-based versus allophone-based predictions for matched pairs, 71% vs 91%, respectively.

Of the matched pairs, there only two exceptions to the allophone-based predictions, Malayalam talker Ym's (inter)dental stimuli in the [a] context and Marathi talker Mv's [aɳa]. As with the Marathi listeners, Punjabi listeners identified the interdental stimuli with their bilabial category, which was a somewhat unusual choice given the gestural similarity between interdental and dental nasals. However, an acoustic analysis of the stimuli revealed that the interdental and bilabial stimuli shared several acoustic properties (Harnsberger, 1998).

Of the six mismatched pairs for the allophonic category hypothesis, all received the predicted top labels. Among the mismatched pairs for which there were two possible top labels, Punjabi listeners also resembled Marathi listeners in labeling the eight non-native alveolar stimulus types with their dental rather than their retroflex category, in a majority of responses. As with the Malayalam and Marathi listeners, dental and alveolar places were judged as relatively similar to one another in relation to the retroflex place.

d. Punjabi Goodness measures. The goodness measures of the Punjabi listeners, particularly the percent labeling rates, provided less support for either hypothesis than the top label choices. For matched and mismatched pairs, the phonemic category hypothesis was only successful in 43% and 63% of its predictions, respectively. The allophonic category hypothesis actually fared worse with the Punjabi results, successfully predicting the goodness measures of only 50% of the matched pairs and 43% of the mismatched pairs. The stimulus types that elicited low percent labeling rates and/or category-goodness ratings spanned a range of places, talkers, and vocalic and syllabic contexts. As a class, only the Malayalam bilabial and dental nasals (excluding talker Ym's interdental nasals) were closely identified with a Punjabi category.

e. Dental-retroflex place group differences. A comparison of the top labeling choices of the dental-retroflex groups revealed only a small number of stimulus types that elicited cross-language differences, and most of these were accounted for by the allophonic category hypothesis. Of the stimulus types that were either matched or mismatched pairs for both groups, 100% received the top labels predicted by the allophonic category hypothesis versus 88% for the phonemic category hypothesis. The lower success rate for phoneme-based predictions reflects the labeling trend of Punjabi, but not Marathi listeners, in mapping the geminate retroflexes to their dental category in a majority of responses.

In the goodness measures, additional cross-language differences emerged that lowered the success rate of both hypotheses considerably. Of the stimulus types that shared the same label and were either matched or mismatched pairs for both groups according to the allophonic category hypothesis, only 60% of the stimulus types had statistically identical percent labeling rates or category-goodness ratings. The equivalent success rate for the phoneme-based predictions was 62%. The stimulus types that elicited cross-language differences were, in many cases, dental or alveolar, indicating important cross-language differences between the dental categories of Marathi and Punjabi listeners that are not captured in either phonemic or allophonic representations of these groups' perceptual categories.

4. Alveolar-retroflex place group

a. Tamil top labeling choices. Of all listener groups, Tamil had the highest number of top labels that differed from the predictions of both the phonemic and allophonic category hypotheses. In the case of Tamil, the two hypotheses differed very little in their top label predictions for matched pairs, which constituted almost all of the stimulus types for which a single label prediction was possible. The predictions for both hypotheses were correct in only 58% of the matched pairs, with alveolar and retroflex stimulus types being frequently ''mislabeled.'' Only the bilabial stimulus types were, as a whole, consistently labeled.

Among the mismatched pairs for which there were two possible top labels (all dental stimulus types), Tamil listeners were highly variable in their top labeling choices. Six of the dental stimulus types received retroflex top labels, while three received dental top labels. As with the Marathi and Punjabi listeners, Tamil listeners identified the talker Ym's (inter)dental stimuli in an [a] context with their bilabial category in a majority of responses.

b. Tamil goodness measures. The success rates of both hypotheses were even lower in the goodness measures than in the top label choices. The Tamil listeners showed a great deal of variability in their labeling and rating of alveolar and retroflex stimulus types, whether they involved matched or mismatched pairs. For the matched pairs, the phonemic and allophonic category hypotheses made correct predictions only 31% and 33% of the time, respectively. For the mismatched pairs, the phonemic and allophonic category hypotheses failed to make any correct predictions.

These high failure rates of these two hypotheses stem from a puzzling trend that emerged across dental, alveolar, and retroflex stimulus types: those in an [a] context were labeled as retroflex in a majority of Tamil listener responses, while those in an [i] context were labeled as dental. Particularly, 10 out of 14 dental and alveolar stimulus types were labeled retroflex in the majority of responses, while three out of four alveolar and retroflex stimulus types were labeled dental in the majority of responses. There were exceptions to this pattern, including Malayalam [ana] and [iɳːi] stimuli, and all [aɳa] stimuli except those produced by talker Mv, which were almost split between the Tamil dental and retroflex labels. These unexpected results for a listener group whose native language possesses a dental-retroflex nasal contrast indicate that many of the non-native dental and retroflex stimuli were poor matches to Tamil perceptual categories. An examination of individual listener scores reveals that these labeling trends were not the result of two or more well-defined subsets of speakers with different labeling patterns, but reflected most of the subjects' choices.

c. Oriya top labeling choices. The Oriya listener group provided an interesting test case for the phonemic and allophonic category hypotheses. The phonemic category hypothesis made a set of simple top label predictions for the non-native bilabial, alveolar, and retroflex stimulus pairs, regardless of syllable type. The allophonic category hypothesis made the same type of predictions, only more tentatively in the case of medial geminate retroflexes, for which there are no directly corresponding Oriya allophonic categories. The Oriya top label results with the medial geminate retroflexes support the tentativeness of the allophonic predictions: Oriya listeners opted for an alveolar over a retroflex category in a majority of their responses to geminate, but not singleton, retroflex stimuli. This result weighed heavily in the success rates of the two category hypotheses, since the medial geminate retroflexes counted as matched pairs for the phonemic, but not allophonic, category hypothesis. With matched pairs, the phonemic and allophonic category hypotheses were successful in 73% and 100% of their predictions, respectively. However, the allophonic category hypothesis' success rate with mismatched pairs was relatively low, 57%, due to its failure to predict the Oriya results with the medial geminate retroflexes.

Beyond the retroflex stimulus types, both hypotheses accurately predicted the top labels for the bilabial and alveolar nasal stimuli. The dental stimuli, which could have been identified as either alveolar or retroflex, consistently received alveolar top labels from this group, with the exception of talker Ym's (inter)dental stimuli in an [a] context. This result mirrors that of the Malayalam, Marathi, and Punjabi listeners in grouping together the dental and alveolar places in mapping stimuli to perceptual categories.

d. Oriya goodness measures. The success rates of the phonemic and allophonic category hypotheses for the matched pairs were 55% and 50%, respectively, and 100% and 75% for the mismatched pairs, respectively. All but one of the stimulus types with a significantly lower goodness measure(s) were retroflex stimulus types, particularly geminate retroflexes. The bilabial, dental, and alveolar nasals were in almost all cases consistently labeled and received average to high ratings. The difference between the two hypotheses' success rates for mismatched pairs was largely a function of the range of stimulus types classified as mismatched pairs by them. In the case of phonemic predictions, only the dental stimulus types were classified as mismatched pairs, while for the allophonic predictions, the mismatched pairs set also included the geminate retroflexes, which were variably identified by Oriya listeners.

e. Alveolar-retroflex place group differences. Both the phonemic and allophonic category hypotheses predicted that Tamil and Oriya listeners would show no significant differences in their identification of many of the non-native stimulus types by any measure. In fact, Tamil and Oriya listeners differed in 43% or 46% of the predicted top labels according the phonemic or allophonic category hypothesis, respectively. These high failure rates for top label predictions stem largely from the variability observed in the Tamil results. With so many mismatches between the Oriya and Tamil top labels, a limited number of stimulus types with the same top labels remained for the purpose of comparing Oriya and Tamil goodness measures: 15 stimulus types for the phonemic category hypothesis and 7 for the allophonic category hypothesis. For these stimulus types, the success rates of the phonemic and allophonic category hypotheses were 40% and 0%, respectively.

5. Alveolar place group

a. Bengali top labeling choices. Of the seven listener groups tested, the Bengali and American English groups possessed the fewest number of possible native coronal categories (one) relevant to the stimuli presented in this experiment. Thus, it was not surprising to find that the Bengali listeners' top labeling choices closely matched those predicted by the phonemic and allophonic category hypotheses, which did not differ in their predictions for this group. One-third of the non-native stimuli were classified as matched pairs for this group, namely the bilabial and alveolar stimuli. The dental and retroflex stimuli were classified as mismatched pairs, but were expected to be identified with the Bengali alveolar nasal in a majority of trials. In fact, all of the matched and mismatched pairs predictions were upheld by the results, with the one exception of talker Ym's (inter)dental stimuli in an [a] context. As with the other non-native listener groups, Bengali listeners identified this stimulus type with their bilabial category. Overall, the two hypotheses were very successful in predicting the top labeling choices of this listener group, which differed from the dental-retroflex and alveolar-retroflex in having a smaller set of relevant perceptual categories to use in labeling.

b. Bengali goodness measures. Both hypotheses were also successful in predicting the high percent labeling rates and goodness ratings from a listener group with a relatively sparse nasal consonant inventory. Of the 12 matched pairs (which were the same for both hypotheses), only one stimulus type received low goodness measures, talker Ym's [ana] stimuli. Seven of the 24 mismatched pairs were also poorer matches to the most similar Bengali category, five of which were retroflex stimulus types. Overall, the phonemic and allophonic category hypotheses were relatively successful in predicting Bengali listeners' identification of these non-native stimulus types.

c. English top labeling choices. As with the Bengali listeners, the phonemic and allophonic category hypotheses did not differ in their predictions for the English listeners. Both were highly successful in predicting the top labeling choices of English listeners for 35 of the 36 stimulus types, with the one exception, again, being talker Ym's (inter)dental stimuli in an [a] context. English listeners in mismatched pairs identified non-native coronal stimuli with the English alveolar nasal, rather than the bilabial nasal, as predicted.

d. English goodness measures. The two hypotheses were also successful in predicting the percent labeling rates and category-goodness ratings for most stimulus-category pairs. In the case of English listeners, the phonemic and allophonic category hypotheses did differ in their classification of the stimulus types as matched or mismatched. For the allophone-based predictions, all of the geminate nasals were classified as mismatched pairs due to the absence of contrastive gemination in English. However, for the phoneme-based predictions, the bilabial and alveolar geminates were classified as matched pairs, since English possesses a bilabial and an alveolar phoneme.

The success rates of the two hypotheses for matched and mismatched pairs were, nevertheless, similar: 92% and 67% of the phoneme-based predictions were upheld for matched and mismatched pairs, respectively. For the allophone-based predictions, the equivalent success rates were both 75% and 75%. While a 75% success rate for the matched pairs predictions of the allophonic category hypothesis was much lower than that of the phonemic category hypothesis, the former represents only four stimulus types, namely the bilabial and alveolar medial singleton nasals. As with the Bengali listeners, English listeners provided low goodness measures for a small number of stimulus types. Overall, the results of the English listener group support the phonemic and allophonic category hypotheses.

e. Alveolar place group differences. While the top labeling choices and goodness measures of the alveolar groups strongly support the phonemic and allophonic category hypotheses, the cross-language differences in the goodness measures provide a very different picture. Thirty-five out of 36 stimulus types were labeled or rated in a significantly different manner by these two listener groups, in contrast with both hypotheses. A very large proportion of these differences, 30 out of 35, were a product of significantly lower ratings given by the English listeners. This pattern could not have been predicted from abstract representations (phonemic or allophonic) of these listener groups' nasal categories. The lower ratings suggest profound differences in the perceptual weighting of the phonetic properties of the stimuli. The nearly consistent dichotomy between the listener group's ratings raises the possibility that some properties common to the entire stimulus set were relatively more foreign to the English listeners. One possible foreign property could be

TABLE VIII. Summary of the percentage of stimulus types that received the top labels predicted by the phonemic and allophonic category hypotheses. ''Phoneme matched'' and ''Allophone matched''=stimulus-category pairs that shared a common phonemic or allophonic symbol, ''Phoneme mismatched'' and ''Allophone mismatched''=pairs that did not.

Listener group	Phoneme matched	Phoneme mismatched	Allophone matched	Allophone mismatched
Malayalam	69% ($N=16$)	···	90% ($N=10$)	50% ($N=4$)
Marathi	88% ($N=16$)	···	88% ($N=16$)	···
Punjabi	71% ($N=28$)	···	91% ($N=22$)	100% ($N=6$)
Oriya	73% ($N=22$)	···	100% ($N=8$)	57% ($N=14$)
Tamil	58% ($N=26$)	···	58% ($N=24$)	50% ($N=2$)
Bengali	100% ($N=12$)	96% ($N=24$)	100% ($N=12$)	96% ($N=24$)
English	100% ($N=12$)	96% ($N=24$)	100% ($N=4$)	97% ($N=32$)

TABLE IX. Summary of the percentage of stimulus types that received the percent labeling rates and category-goodness ratings predicted by the phonemic and allophonic category hypotheses. ''Phoneme matched'' and ''Allophone matched''=stimulus-category pairs that shared a common phonemic or allophonic symbol. ''Phoneme Mismatched'' and ''Allophone Mismatched''=pairs that did not.

Listener group	Phoneme matched	Phoneme mismatched	Allophone matched	Allophone mismatched
Malayalam	44% ($N=16$)	···	50% ($N=10$)	17% ($N=6$)
Marathi	69% ($N=16$)	63% ($N=8$)	69% ($N=16$)	63% ($N=8$)
Punjabi	43% ($N=28$)	63% ($N=8$)	50% ($N=22$)	43% ($N=14$)
Oriya	55% ($N=22$)	100% ($N=10$)	50% ($N=8$)	75% ($N=24$)
Tamil	31% ($N=26$)	0% ($N=10$)	33% ($N=24$)	0% ($N=12$)
Bengali	92% ($N=12$)	71% ($N=24$)	92% ($N=12$)	71% ($N=24$)
English	92% ($N=12$)	67% ($N=24$)	75% ($N=4$)	75% ($N=32$)

gemination: 20 of 36 of the stimulus types involved geminate nasals, which appear in Bengali but not English (see Table IV). However, mean English ratings for medial geminate stimuli were actually higher than those for medial singleton (3.5 versus 3.2), casting doubt on that hypothesis. Alternate possible sources of the foreignness of the stimuli for English listeners include the quality of the adjacent vowel(s), or possibly greater Bengali experience with retroflex and dental variants of their /n/ phoneme before native dental and retroflex oral stops.

II. DISCUSSION

The purpose of this experiment was twofold: to examine the extent to which perceptual categories, described abstractly in terms of phonemes and allophones, could account for the identification of non-native nasal consonants, and to describe the identification of a previously untested set of non-native sounds, nasal consonants varying in place of articulation, by listener groups varying in nasal consonant inventory. While prior work had indicated that phonemes, and perhaps even allophones, were too abstract to account for non-native listener's perceptual sensitivities in traditional identification and discrimination tests, these conclusions were based on studies using a limited range of listener groups, such as English or Japanese speakers.

A series of measures was taken in this experiment that constituted tests of the phonemic and allophonic category hypotheses, namely in the predictions of the top labeling choices, percent labeling rates, and mean category-goodness ratings for seven listener groups with up to 36 non-native stimulus types. The results of all of these tests are summarized in Tables VIII (phonemic category hypothesis) and IX (allophonic category hypothesis) for single listener groups, and Table X for cross-language comparisons.

For single listener groups, the top labeling choice was the most liberal measure supporting the two hypotheses, given that a listener group's actual top label only had to match its predicted top label in a majority of responses. A weighted average across all listener groups with matched and mismatched pairs showed that the phonemic and allophonic category hypotheses were successful in predicting top labeling choices in 81% and 85% of the non-native stimulus types, respectively. The closeness in these weighted averages may, of course, be partly due to the listener groups chosen for the study. Five of these listener groups, Marathi, Tamil, Oriya, Bengali, and English, showed few differences in the top labeling choices predicted by the two models. If we compare the weighted means of the two hypotheses using simply the Malayalam and Punjabi results, the discrepancy between them widens to 70% (phonemic) and 91% correct (allophonic). Thus, it appears that a category representation with greater phonetic detail offers an advantage in predicting the top labeling choices of non-native listeners. However, for groups such as Oriya and especially Tamil, top labeling choices deviated from those predicted very frequently, indicating that for a significant minority of listener groups, it may be difficult to predict *a priori* their top label choices given a phonemic or even an allophonic description of their language's sound categories.

More rigorous tests of both hypotheses were the percent labeling rate and mean category-goodness ratings measures, both of which had the potential for providing more graded information about perceptual categories than the top labeling choice. The results in Table IX are presented in terms of the two hypotheses and ''matched'' and ''mismatched'' pairs. Both hypotheses predicted that for stimulus-category pairs

TABLE X. Summary of the success of the phonemic and allophonic category hypotheses in predicting cross-language differences in the identification test results. ''%LR or rating'' refers to the percentage of stimulus-category pairs that did not elicit significant cross-language differences in their percent labeling rate and/or mean ratings.

	Phoneme		Allophone	
Place group	Top label	%LR or rating	Top label	%LR or rating
Dental-retroflex	88% ($N=24$)	62% ($N=21$)	100% ($N=20$)	60% ($N=20$)
Alveolar-retroflex	47% ($N=32$)	40% ($N=15$)	44% ($N=16$)	0% ($N=7$)
Alveolar	100% ($N=36$)	3% ($N=36$)	100% ($N=28$)	4% ($N=28$)

that ''match'' or share a common phonemic or allophonic symbol, the percent labeling rates and mean category-goodness ratings should be ''consistent'' and ''high,'' respectively, given the definition of these terms used in this study. In fact, the percentage of matched pairs whose percent labeling rates and/or category-goodness ratings were successfully predicted ranged widely, between just 31% for Tamil listeners to 92% for Bengali listeners, averaging 55% across all stimulus-category pairs from all listener groups, for both hypotheses. For ''mismatched'' stimulus-category pairs, the percentage of successful predictions ranged from 0% to 100% (mean=69% with phoneme-based predictions, 63% with allophone-based predictions).

The most rigorous test of the two hypotheses turned out to be the cross-language comparisons. Both predicted that no cross-language differences in the top labeling choice, percent labeling rate, or category-goodness ratings would be observed between groups that share a common nasal consonant inventory, for stimulus types for which both groups have the same nearest category. A summary of the success of the hypothesis in predicting cross-language differences is shown in Table X. In terms of the top labeling choices, small or no cross-language differences emerged for the groups whose coronal nasal inventory contained either dental and retroflex nasals (Marathi and Punjabi) or just alveolar nasals (Bengali and English). The small difference between the two hypotheses for the dental-retroflex group was due almost entirely to the Punjabi labeling pattern with geminate retroflexes discussed earlier. The low percentages of successful predictions for the alveolar-retroflex group were due almost exclusively to the unexpected labeling patterns of the Tamil listeners.

In terms of the percent labeling rates and category-goodness ratings, the hypotheses fared much worse. They failed to predict the large differences between the Bengali and English listeners in their rating of non-native stimuli, and failed to predict the frequency with which the other listener groups used multiple labels for a given stimulus type, reflected in lower percent labeling rates. These failures indicate that, with a number of listener groups, it may be difficult to predict how well non-native stimuli will match a listener group's phonemically or allophonically defined native category. These results also have implications for predictions of the *discriminability* of some non-native contrasts, since in some cross-language speech perception models, category-goodness is a strong predictor of the discriminability of two stimuli that are identified with the same native category (Best, 1995; Kuhl and Iverson, 1995).

The frequent failure of abstract representations of perceptual categories in predicting the identification patterns of different listener groups also made it difficult to generalize about the effect of particular native nasal consonant inventories on the perception of nasal consonants. A few broad trends did emerge in the similarity relations between nasals of different places that constitute important contributions to our understanding of cross-language speech perception. First, non-native coronal nasal stimuli were identified in a majority of responses with coronal, as opposed to bilabial, categories, in all but one case, the interdental stimuli of Malayalam talker Ym in the [a] context. Second, dental and alveolar stimuli were identified with dental or alveolar categories, rather than retroflex categories. This trend was observed in four of the five groups with a two- or three-way coronal place distinction. Thus, Marathi and Punjabi listeners, whose native language possesses a dental-retroflex distinction, identified non-native alveolars as dental, rather than retroflex. Malayalam listeners, whose native language possesses a dental-alveolar-retroflex distinction, identified non-native dentals and alveolars with their dental and alveolar categories. In a similar manner, Oriya listeners identified non-native dental nasals with their alveolar, rather than retroflex, category. Only Tamil listeners failed to follow this trend, showing a vowel context effect in their identification of non-native coronal nasal stimuli that could not be explained from existing descriptions of the language.

Beyond such broad statements, the complexity of the percent labeling rate and category-goodness ratings results undermine any attempt to succinctly explain how a particular listener group's nasal consonant inventory, described in terms of phonemes or allophones, affects their perception of non-native nasal stimuli. Given that the phonemic and allophonic levels of representation are too abstract to account for many of the results described here, more detailed descriptions of the nasal consonants of these listener groups are needed, corresponding perhaps to the language-specific acoustic cues or gestures of particular nasal consonants. Currently, such detailed descriptions are lacking for the languages of most of the listener groups used in this study. However, the existing phonetic literature of some of these languages provides some interesting clues that could be used to explain *posthoc* some of the test results in terms of a cue- or gesture-based description. There are at least three cases in the results that can be discussed in these terms. In these cases, I will refer to some of the acoustic properties of the stimuli measured by Harnsberger (1998). Space does not permit a discussion of all the acoustic properties measured for all stimulus types.

The first case involves the labeling of Malayalam talker Ym's [aɳːa] stimuli as dental by Marathi listeners in 54% of their responses. This result was surprising given the fact that Marathi makes a dental-retroflex distinction in medial geminates, and that other medial geminate retroflex stimulus types were labeled as retroflex in a majority of responses. Why would this geminate retroflex stimulus type, produced by this talker, be more closely identified with the Marathi dental nasal? The answer may lie in cross-language differences in the acoustic properties of retroflexes in Malayalam versus Marathi. In an acoustic analysis of the Marathi retroflex stimuli used in this study, one property common to them all was a convergence of the second and third formants at the vowel-nasal boundary, usually due to a lowering of $F3$, a property that has been observed in retroflex oral stops in prior work (Stevens and Blumstein, 1975). In contrast, Ym's [aɳːa] stimuli lacked this $F2-F3$ convergence, with its $F2$ and $F3$ transitions more closely resembling the talker's dental, or more precisely, interdental, stimuli in the [a] context. Thus, if the Marathi retroflexes used in this study are good examples of the retroflexes typically used by Marathi speakers, then Marathi listeners may rely heavily on $F2$ and $F3$

transitions in identifying the Marathi retroflex nasal. Coronal nasals lacking such a characteristic transition would then be identified by the (presumably) next most similar category, the Marathi dental nasal.

The most obvious explanation for this difference between Marathi retroflexes and those produced by Ym in the [a] context would be that Ym failed to correctly produce a retroflex nasal. However, this explanation is highly unlikely considering that these stimuli were consistently judged both by native speakers in the pilot study and in the main experiment as excellent exemplars of the Malayalam retroflex nasal. What Ym likely produced was a retroflex nasal without the same degree of retroflexion that results in an $F2-F3$ convergence, which produces impressionistically a strong rhotic quality in the preceding vowel. Retroflexes have been observed to vary in degree of retroflexion, though no study has examined the effect of this variable on the identification of non-native retroflex sounds (Ladefoged and Bhaskararao, 1983). In summary, if we analyze the acoustic properties of retroflex nasals in Malayalam and Marathi, we may find systematic differences that account for identification patterns that would otherwise remain unexplained. The same symbol for both languages' retroflex nasals may be hiding perceptually pertinent information.

The second case concerns Punjabi listeners' labeling of geminate retroflex stimuli from Malayalam and Marathi as the Punjabi dental nasal in a majority of responses. The Punjabi language, while possessing a dental-retroflex phonemic contrast, does not make this contrast in medial geminates. The only medial geminate coronal nasal in Punjabi is dental. Thus, the Punjabi dental nasal is the most similar perceptual category to all coronal nasals that are geminates in medial position. However, a valid alternative explanation exists if we consider the possible acoustic cues used to signal the Punjabi retroflex nasal. In this study, actual Punjabi nasals, which would have allowed for acoustic analysis, were not collected. However, in the existing phonetic literature on Punjabi, a number of scholars claim that the Punjabi retroflex nasal freely varies with a flapped retroflex nasal stop in medial position, resulting in a relatively short nasal murmur (Dulai and Koul, 1980; Masica, 1991). If flapping is common in medial position, then the retroflex nasal categories of Punjabi listeners might reflect a preference for a short murmur duration, with relatively long murmurs treated as a characteristic property of the Punjabi dental nasal. This durational property of retroflexes in Punjabi could influence Punjabi listeners' identification of non-native retroflex nasals that have a relatively long murmur, such as geminates. In this scenario, Punjabi listeners would treat the duration of the murmur as somewhat more important than other cues, such as formant transitions, in identifying non-native coronal nasals. This small difference in the weighting of temporal and spectral cues to the Punjabi retroflex nasal would be mirrored in the mixed labeling patterns of geminate retroflex by Punjabi listeners.

The third and final case also concerns the perception of geminate retroflexes, this time by Oriya speakers. Like the Punjabi listeners, Oriya listeners labeled geminate retroflex stimuli as the Oriya alveolar nasal in a majority of responses. Unlike Punjabi, the Oriya language lacks geminate nasals of any kind (Pattanayak and Das, 1972; Bhattacharya, 1993). Thus, Oriya listeners cannot be said to be filtering medial geminate retroflex stimuli through a medial geminate non-retroflex perceptual category. The most similar set of potential native categories would be the singleton alveolar and retroflex nasals that appear in medial position. In this study, because Oriya medial singleton retroflex nasals were used as stimuli, they could be acoustically analyzed (Harnsberger, 1998). Their formant transition patterns from the vowel into the nasal murmur closely resembled those of most of the retroflex stimuli used in this study, showing an $F2-F3$ convergence that corresponded impressionistically to a rhotic quality to the preceding vowel. If these stimuli were representative of retroflex nasals used by Oriya speakers, then the Oriya retroflex nasal closely resembles the geminate retroflex stimuli from Marathi and Malayalam talker Ys in terms of their spectral qualities. So why were these non-native geminate retroflexes identified with the Oriya alveolar nasal in a majority of responses? If we consider the temporal properties of the Oriya alveolar and retroflex nasals, then this identification pattern can be explained. In the acoustic analysis of the stimuli, the nasal murmurs of Oriya medial alveolars were found to be approximately three times longer in duration than those of Oriya medial retroflexes (Harnsberger, 1998). If these stimuli are representative of Oriya alveolar and retroflex nasals, then geminate retroflexes, by virtue of their lengthy nasal murmurs, may have been more frequently labeled as alveolar, despite the spectral similarities between the geminate retroflexes and the Oriya retroflex nasal, if Oriya listeners were treating murmur duration as a critical cue in identifying nasals as either alveolar or retroflex. This example, along with the Marathi and Punjabi cases, illustrate the type of explanations, focusing on language-specific phonetic detail, that may account for the kinds of complex cross-language differences that do not lend themselves to analysis at the phonemic or allophonic level of detail.

III. CONCLUSIONS

The results of this identification test demonstrated that, for several different types of measures, perceptual categories described in terms of abstract units such as phonemes or allophones lack a degree of acoustic-phonetic detail that is available to listener groups. This degree of phonetic detail is largely undescribed for the perceptual categories of most linguistic communities. If the results of this cross-language study generalize to other listener groups, then the detailed description of perceptual categories constitutes an important line of future cross-language phonetic research. The candidate primitive for such a level of description might be the phonetic feature, a property of an allophone that plays a special role in the identity of the allophone. Such a constituent could be defined as an acoustic cue or a gesture, depending upon one's assumptions concerning the ''object'' of speech perception (Fowler, 1989; Diehl and Kluender, 1989). In three cases discussed earlier, an attempt was made to explain the identification of non-native stimuli with native categories by examining such cues of non-native stimuli and native categories. An acoustic analysis was referenced to provide

some potentially important stimulus properties. For the properties of the perceptual categories, existing phonetic studies and/or acoustic analysis were used to hypothesize about the specific characteristics of perceptual categories. These explanations are, of course, speculative, given the small number of actual stimuli examined, and the small number of phonetic studies of some of the languages in question. However, the cases serve as examples of a comparison between non-native stimuli and native categories at a level of phonetic detail (the cue or feature) that is warranted by the complex results found in this study. An attention to language-specific differences at this level, rather than at the level of the phoneme or allophone, may prove to be essential in predicting and explaining the identification of non-native stimuli.

ACKNOWLEDGMENTS

I would like to thank Patrice Beddor, who advised me throughout the design and execution of this study, and Catherine Best for her many helpful comments at early and late stages of this work. I would also like to thank Rebecca Herman and David Pisoni for their comments on earlier drafts of this paper.

APPENDIX

Note: Each table in this Appendix lists the individual subjects by a number (''S#''), his/her sex and age, the city and state in which the subject lived prior to attending a national university, and his/her self-reported experience with second languages and any experience living outside of his/her home country (''Linguistic background''). Under ''Linguistic background,'' ''s''=speak, ''r''=read, ''w''=write, referring to different types of mastery of a second language. Subjective evaluations of ability in a second language (''average,'' ''fluent'') are purely those provided by the subject, and do not reflect any kind of formal evaluation.

TABLE AI. Malayalam subjects.

S#	Sex	Age	City, state	Linguistic background
1	m	27	Calicut, Kerala	English average
2	m	25	Thrissur, Kerala	English
3	m	22	Thiruvalla, Kerala	English good; Hindi fair
4	m	21	Trivandrum, Kerala	English fluent, Hindi fair
5	m	22	Trichur, Kerala	English fluent, Hindi good
6	m	24	Calicut, Kerala	English
7	f	21	Trivandrum, Kerala	English and Russian fluent; Hindi fair
8	m	21	Kottayam, Kerala	English and Hindi—fluent
9	f	21	Kannur, Kerala	English, Hindi, Tamil s,r,w
10	m	25	Trivandrum, Kerala	English s,r,w; Hindi r,w
11	f	24	Kottayam, Kerala	English good, Hindi fair
12	m	24	Kochi, Kerala	English r,w
13	m	27	Ernakulam, Kerala	English
14	f	23	Trivandrum, Kerala	English, Hindi; a little French
15	m	27	Kochi, Kerala	English and Hindi—fair
16	f	26	Trivandrum, Kerala	English, Hindi; German (diploma)
17	f	23	Ernakulam, Kerala	English s,r,w; Hindi s,r,w
18	m	23	Thiruvalla, Kerala	English and Tamil fluent; Hindi s,r,w

TABLE AII. Marathi subjects.

S#	Sex	Age	City, state	Linguistic background (self-reported)
1	f	20	Pune, Maharashtra	English, Hindi, German
2	f	24	Pune, Maharashtra	English, Hindi s,r,w; Tamil s
3	f	24	Bombay, Maharashtra	Hindi, English, German, fluent;
4	m	22	Pune, Maharashtra	English, Hindi s,r,w; Gujrati s
5	m	25	Aurangabad, Maharashtra	English, Hindi
6	m	23	Pune, Maharashtra	English, Hindi s,r,w
7	m	23	Karad, Maharashtra	English, Hindi s,r,w; Kannada s
8	m	23	Nanded, Maharashtra	English, Hindi
9	m	22	Bombay, Maharashtra	English and Hindi very good
10	f	21	Pune, Maharashtra	English, Hindi, German
11	m	23	Thane, Maharashtra	English, Hindi s,r,w
12	f	23	Pune, Maharashtra	Hindi, English, German, Japanese fluent
13	m	22	Bombay, Maharashtra	English, Hindi proficient
14	m	22	Kohlapur, Maharashtra	English, Hindi proficient
15	f	21	Bombay, Maharashtra	English, German good
16	m	24	Pune, Maharashtra	English, Hindi s,r,w
17	f	23	Pune, Maharashtra	English, Hindi

TABLE AIII. Punjabi subjects.

S#	Sex	Age	City, state	Linguistic background (self-reported)
1	m	20	Amritsar, Punjab	English, Hindi s,r,w
2	m	21	Amritsar, Punjab	English, Hindi s,r,w
3	m	18	Amritsar, Punjab	Hindi
4	m	20	Amritsar, Punjab	English, Hindi s,r,w
5	m	22	Amritsar, Punjab	English, Hindi
6	m	18	Amritsar, Punjab	English, Hindi s,r,w
7	m	26	Amritsar, Punjab	English, Hindi s,r,w
8	m	27	Amritsar, Punjab	English, Hindi s,r,w
9	m	18	Amritsar, Punjab	English, Hindi s,r,w
10	m	27	Amritsar, Punjab	English, Hindi s,r,w
11	m	21	Amritsar, Punjab	English, Hindi s,r,w
12	m	23	Ludhiana, Punjab	English, Hindi s,r,w
13	f	21	Patiala, Punjab	English, Hindi, German s,r,w
14	m	22	New Delhi	English fluent; Hindi average

TABLE AIV. Oriya subjects.

S#	Sex	Age	City, state	Linguistic background (self-reported)
1	m	27	Sambalpur, Orissa	English good, Hindi fair
2	m	23	Sambalpur, Orissa	English, Hindi s,r,w; Sanskrit r
3	m	24	Bhubaneswar, Orissa	English fluent; Hindi s,r
4	f	22	Bhubaneswar, Orissa	English, Hindi s,r,w
5	m	26	Baripada, Orissa	English, Hindi r,w
6	m	23	Konark, Orissa	English, Hindi s,r,w
7	m	23	Balasore, Orissa	English s,r,w, Hindi s
8	f	22	Sambalpur, Orissa	English, Hindi s,r,w
9	f	22	Bhubaneswar, Orissa	English, Hindi s,r,w; Bengali s
10	m	25	Bhubaneswar, Orissa	English, Hindi
11	f	24	Cuttack, Orissa	English, Hindi s,r,w; Bengali s
12	m	33	Cuttack, Orissa	English, Hindi fluent
13	m	25	Cuttack, Orissa	English, Hindi s,r,w; Bengali s
14	m	21	Cuttack, Orissa	English, Hindi s,r,w
15	m	25	Cuttack, Orissa	English, Hindi s,r,w; Bengali s
16	m	20	Balesore, Orissa	English, Hindi

TABLE AV. Tamil subjects.

S#	Sex	Age	City, state	Linguistic background (self-reported)
1	f	21	Madras, TamilNadu	English s,r,w; Hindi s,r,w
2	f	18	Madras, TamilNadu	English fluent
3	m	21	Madras, TamilNadu	English
4	m	26	Madras, TamilNadu	Telugu, English French fluent; Hindi r,w
5	m	23	Madras, TamilNadu	English, Telegu
6	m	23	Salem, TamilNadu	English good
7	m	21	Madras, TamilNadu	English, Hindi, and German s,r,w
8	m	24	Madras, TamilNadu	English, Hindi s,r,w
9	f	21	Madras, TamilNadu	English and Hindi fluent; German s,r,w
10	m	25	Madras, TamilNadu	English fluent
11	m	24	Madras, TamilNadu	English; French, Russian average
12	m	20	Madras, TamilNadu	English s,r,w

TABLE AVI. Bengali subjects.

S#	Sex	Age	City, state	Linguistic background (self-reported)
1	m	24	Calcutta, West Bengal	English, Hindi
2	m	24	Calcutta, West Bengal	English, Hindi
3	m	23	Calcutta, West Bengal	English
4	f	22	Calcutta, West Bengal	English, Hindi
5	m	25	Calcutta, West Bengal	English, Hindi
6	f	22	Calcutta, West Bengal	English
7	m	21	Calcutta, West Bengal	English, a little Hindi
8	m	33	Calcutta, West Bengal	English, Hindi
9	f	23	Calcutta, West Bengal	English, a little Hindi
10	m	25	Calcutta, West Bengal	English
11	f	22	Calcutta, West Bengal	English, Hindi
12	m	22	Calcutta, West Bengal	English
13	f	22	Calcutta, West Bengal	English, Hindi
14	f	29	Calcutta, West Bengal	English, Hindi
15	m	20	Calcutta, West Bengal	English, Hindi

TABLE AVII. English subjects.

S#	Sex	Age	City, state	Linguistic background (self-reported)
1	f	20	Detroit, MI	2 years Spanish
2	f	19	Warren, MI	4 years French; 1 year Italian
3	f	18	Detroit, MI	1 year French
4	f	20	Owosso, MI	none
5	f	21	Farmington, MI	2 years Spanish
6	m	18	Flint, MI	1 year Spanish
7	f	19	Owosso, MI	none
8	f	18	Southfield MI	introductory Spanish
9	f	18	Detroit, MI	French proficient
10	f	21	St. Clair Shores, MI	German, lived in Germany from age 6–10
11	m	20	Detroit, MI	lived in Netherlands for 1 year
12	f	18	Detroit, MI	4 years high school Spanish
13	f	19	Southfield, MI	none
14	f	19	Freeland, MI	some German, Spanish
15	f	20	Greenville, MI	2 years of Spanish
16	f	19	Battle Creek, MI	some French
17	f	20	Traverse City, MI	introductory French
18	m	18	Detroit, MI	introductory French

[1]Terbeek (1977) conducted an experiment such as this, presenting a set of naturally produced vowels from different portions of an $F1/F2$ space to native speakers of English, German, Turkish, Thai, and Swedish. Terbeek's object was not to construct a language-general similarity metric, but rather to discover what distinctive features are used by different listener groups to perceive vowels.

[2]The term *coronal* refers to any sound produced by the tongue tip or blade within the dental, alveolar, and postalveolar regions (Ladefoged and Maddieson, 1996).

[3]In order to compose the stimulus set in Table II, both word and nonword stimuli from Malayalam, Marathi, and Oriya were employed. In reading nonwords in isolation, none of the talkers produced any disfluencies in the author's judgment. Moreover, both word and nonword stimuli were consistently identified by native speaker participants in piloting. No systematic differences were observed in the identification of words and nonwords by the control groups in the experiment. All of the stimuli that were non-native for a given listener group were assumed to be treated as nonwords. The extent to which *non-native* stimuli were judged as words in a listener's native language, and the possible effect of such judgments on the identification patterns elicited, could not be predicted by a simple comparison of the stimuli and the listener's native nasal consonant inventory, and was not measured by a specific test in this experiment (that is, non-native listeners were not asked if any of the non-native stimuli sounded similar to a word in their native language).

[4]The vowel context effect on the identification of the interdentals is reminiscent of studies of nasal identification in noise by English speakers, in which an [i] context elicited a greater number of alveolar responses relative to bilabial responses (Sharf and Ostreicher, 1973; Zee, 1981).

[5]Frequent lapses in answering within the test sequence were observed in test sheets in which the subject failed to circle the last trial on a test page or noted in writing that the interblock interval followed the penultimate trial that they had answered, rather than the final trial, indicating that they had failed to mark a response to one of the trials in the block.

[6]''Listener group,'' ''stimulus type,'' and their interaction were specified in the ANOVA.

[7]Both ANOVAs were calculated using SAS version 6.11. In a repeated measures mixed model ANOVA, SAS computes denominator degrees of freedom for a given effect or interaction by dividing the residual degrees of freedom into between- and within-subject portions, and then either assigning the within-subjects portion to the effect if it changes within any subject, or the between-subjects portion if it does not (SAS Institute, Inc., 1996).

[8]The pairs for which there were top possible predicted values were included in this measure because, though it was not possible *a priori* to determine how such stimulus types would be identified, they should be consistently labeled with whatever category they are mapped to, and receive high ratings, according to either hypothesis.

[9]There was one set of stimuli whose percent labeling rate by Malayalam listeners remained below the threshold for ''consistent,'' namely talker Ym's interdental stimuli, which were produced in an [a] context. Four additional Malayalam listeners were inconsistent in their labeling of the interdental stimuli. Thus, out of the 18 Malayalam listeners, only 10 consistently identified the interdental stimuli as examples of the Malayalam dental nasal. This is likely due to the variability in the realization of the dental nasal in Malayalam. These other four Malayalam listeners may not have encountered interdental realizations in their respective linguistic communities within Kerala. Given that interdental nasals are attested in Malayalam (Ladefoged and Maddieson, 1996), and given that the three Malayalam listeners in piloting and ten of the Malayalam listeners in the experiment consistently identified the interdental nasals with the Malayalam alveolar nasal, the interdental nasal results were used to evaluate the allophonic category hypothesis.

[10]This pattern was not entirely unexpected given the results of pilot tests in which non-native listeners also associated this talker's (inter)dental nasal in an [a] context with their native /m/. For the acoustic properties of the interdental stimuli in an [a] context, showing their similarity to bilabial nasals, see Harnsberger (1998).

Abramson, A. S., and Lisker, L. (**1970**). ''Discriminability along the voicing continuum: Cross-language tests,'' in *Proceedings of the Sixth International Congress of Phonetic Sciences* (Academia, Prague), pp. 569–573.

Aiyar, L. V. R. (**1972**). ''Two Malayalam phonemes,'' Int. J. Dravidian Linguistics **1**, 10–18.

Arun, V. B. (**1961**). *A Comparative Phonology of Hindi and Panjabi* (Panjabi Sahitya Akademi, Ludhiana).

Asher, R. E., and Kumari, T. C. (**1997**). *Malayalam* (Routledge, London).
Balasubramanian, T. (**1982**). ''The two r's and the two n's in Tamil,'' J. Phonetics **10**, 89–97.
Beddor, P. S., and Strange, W. (**1982**). ''Cross-language study of the oral-nasal distinction,'' J. Acoust. Soc. Am. **71**, 1551–1561.
Best, C. T. (**1994**). ''The emergence of native language phonological influences in infants: A perceptual assimilation model,'' in *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, edited by J. Goodman and H. C. Nusbaum (MIT, Cambridge), pp. 167–224.
Best, C. T. (**1995**). ''A direct realist view of cross-language speech perception,'' in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Baltimore), pp. 171–203.
Best, C. T., McRoberts, G. W., and Sithole, N. M. (**1988**). ''Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants,'' J. Exp. Psychol. **14**, 345–360.
Bhattacharya, K. (**1988**). *Bengali Phonetic Reader* (Central Institute of Indian Languages, Mysore).
Bhattacharya, K. (**1993**). *Bengali-Oriya Contrastive Morphology* (Gupta, Calcutta).
Burnham, D. K. (**1986**). ''Developmental loss of speech perception: Exposure to and experience with a first language,'' Appl. Psycholinguist. **7**, 207–239.
Dart, S. N. (**1991**). ''Articulatory and acoustic properties of apical and laminal articulations,'' doctoral dissertation, University of California, Los Angeles.
Dhall, G. B. (**1966**). *Aspiration in Oriya* (Utkal Univ., Bhubaneswar).
Diehl, R. L., and Kluender, K. R. (**1989**). ''On the objects of speech perception,'' Ecological Psychol. **1**, 121–144.
Dulai, N. K., and Koul, O. N. (**1980**). *Punjabi Phonetic Reader* (Central Institute of Indian Languages, Mysore).
Firth, J. R. (**1948**). ''Word-palatograms and articulation,'' Bulletin of the School of Oriental African Stud. **12**, 857–864.
Flege, J. E. (**1995**). ''Second language speech learning: Theory, findings, and problems,'' in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Baltimore), pp. 233–277.
Fowler, C. (**1989**). ''Real objects of speech perception: A commentary on Diehl and Kluender,'' Ecological Psychol. **1**, 145–169.
Gill, H. S., and Gleason, H. A. (**1986**). *A Reference Grammar of Punjabi* (Publication Bureau, Punjabi Univ., Patiala).
Goto, H. (**1971**). ''Auditory perception by normal Japanese adults of the sounds 'l' and 'r','' Neuropsychologia **9**, 317–323.
Gottfried, T. L., and Beddor, P. S. (**1988**). ''Perception of temporal and spectral information in French vowels,'' Lang Speech **31**, 57–75.
Hai, M. A. (**1960**). *Nasals and Nasalization in Bengali* (Univ. of Dacca, Dacca).
Haldar, G. (**1986**). *A Comparative Grammar of East Bengali Dialects* (Puthipatra, Calcutta).
Harnsberger, J. D. (**1998**). ''The perception of non-native nasal contrasts: a cross-linguistic perspective,'' doctoral dissertation, The University of Michigan.
Henly, E., and Sheldon, A. (**1986**). ''Duration and context effects on the perception of English /r/ and /l/: A comparison of Cantonese and Japanese speakers,'' Lang. Learning **36**, 505–521.
Hura, S. L., Lindblom, B., and Diehl, R. L. (**1992**). ''On the role of perception in shaping phonological assimilation rules,'' Lang Speech **35**, 59–72.
Ingram, J., and Park, S-G. (**1998**). ''Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners,'' J. Acoust. Soc. Am. **103**, 1161–1174.
Iverson, P., and Kuhl, P. K. (**1995**). ''Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling,'' J. Acoust. Soc. Am. **97**, 553–562.
Iverson, P., and Kuhl, P. K. (**1996**). ''Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/,'' J. Acoust. Soc. Am. **99**, 1130–1140.
Jain, B. D. (**1934**). *A Phonology of Punjabi as Spoken about Ludhiana* (Punjab Univ., Lahore).
Jha, A. (**1977**). *An Outline of Marathi Phonetics* (Deccan College, Poona).
Kalackal, T. (**1985**). ''A contrastive analysis of the phonological systems of English and Malayalam,'' doctoral dissertation, Northern Illinois University.
Kelkar, A. (**1958**). ''Marathi phonology and morphophonemics,'' doctoral dissertation, Cornell University.
Kostic, D., and Das, R. (**1972**). *A Short Outline of Bengali Phonetics* (Statistical Publishing Society, Calcutta).
Kuhl, P. K. (**1991**). ''Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not,'' Percept. Psychophys. **50**, 93–107.
Kuhl, P. K., and Iverson, P. (**1995**). ''Linguistic experience and the 'perceptual magnet effect,''' in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Baltimore), pp. 121–154.
Kumari, B. S. (**1972**). *Malayalam Phonetic Reader* (Central Institute of Indian Languages, Mysore).
Ladefoged, P., and Bhaskararao, P. (**1983**). ''Non-quantal aspects of consonant production: A study of retroflex consonants,'' J. Phonetics **11**, 291–302.
Ladefoged, P., and Maddieson, I. (**1996**). *The Sounds of the World's Languages* (Blackwell, Cambridge, MA).
Lisker, L., and Abramson, A. S. (**1964**). ''A cross-language study of voicing in initial stops: Acoustical measurements,'' Word **20**, 384–422.
MacKain, K. S., Best, C. T., and Strange, W. (**1981**). ''Categorical perception of English /r/ and /l/ by Japanese bilinguals,'' Appl. Psycholinguist. **2**, 369–390.
Masica, C. (**1991**). *The Indo-Aryan Languages* (Cambridge U.P., Cambridge).
Miyawaki, K., Strange, W., Verbrugge, R. R., Liberman, A. M., Jenkins, J. J., and Fujimura, O. (**1975**). ''An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English,'' Percept. Psychophys. **18**, 331–365.
Moag, R. F. (**1973**). ''A phonological grammar of style variation in Malayalam,'' doctoral dissertation, University of Wisconsin.
Mohr, B., and Wang, W. S.-Y. (**1968**). ''Perceptual distance and the specification of phonological features,'' Phonetica **18**, 31–45.
Pandharipande, R. V. (**1997**). *Marathi* (Routledge, London).
Pattanayak, D. P., and Das, G. N. (**1972**). *Conversational Oriya* (Kapila Power, Mysore).
Polka, L. (**1991**). ''Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions,'' J. Acoust. Soc. Am. **89**, 2961–2977.
Polka, L. (**1992**). ''Characterizing the influence of native language experience on adult speech perception,'' Percept. Psychophys. **52**, 37–52.
Polka, L., and Bohn, O-S. (**1996**). ''A cross-language comparison of vowel perception in English-learning and German-learning infants,'' J. Acoust. Soc. Am. **100**, 577–592.
Pruitt, J. S. (**1995**). ''The perception of Hindi dental and retroflex stop consonants by native speakers of Japanese and American English,'' doctoral dissertation, University of South Florida, Tampa.
Rochet, B. L. (**1991**). ''Perception of the high vowel continuum: A cross-language study,'' in *Proceedings of the Twelfth International Congress of Phonetic Sciences* (Univ. of Provence, Aix-en-Provence, France), Vol. 4, pp. 94–97.
Sandhu, B. S. (**1986**). *The Articulatory and Acoustic Structure of the Panjabi Consonants* (Punjabi University Publication Bureau, Patiala).
SAS Institute Inc. (**1996**). *SAS/STAT Software: Changes and Enhancements through Release 6.11* (SAS Institute, Inc., Cary, NC).
Schmidt, A. M. (**1996**). ''Cross-language identification of consonants. Part I. Korean perception of English,'' J. Acoust. Soc. Am. **99**, 3201–3211.
Sharf, D. J., and Ostreicher, H. (**1973**). ''Effects of forward and backward coarticulation on the identification of speech sounds,'' Lang. Speech **16**, 196–206.
Sheldon, A., and Strange, W. (**1982**). ''The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception,'' Appl. Psycholinguist. **3**, 243–261.
Stevens, K. N., and Blumstein, S. E. (**1975**). ''Quantal aspects of consonant production and perception: A study of retroflex consonants,'' J. Phonetics **3**, 215–233.
Strange, W. (**1995**). ''Cross-language studies of speech perception: A historical review,'' in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Baltimore), pp. 3–45.
Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S., Nishi, K., and

Jenkins, J. (**1998**). ''Perceptual assimilation of American English vowels by Japanese listeners,'' J. Phonetics **26**, 311–344.
Subramoniam, V. I. (**1962**). ''Phonemic outline of a dialect of Malayalam,'' Indian Linguistics **23**, 99–116.
Švarný, O., and Zvelebil, K. (**1955**). ''Some remarks on the articulation of the 'cerebral' consonants in Indian languages, especially in Tamil,'' Archiv. Orientalni **23**, 374–434.
Tees, R. C., and Werker, J. F. (**1984**). ''Perceptual flexibility: maintenance or recovery of the ability to discriminate non-native speech sounds,'' Can. J. Psychol. **38**, 579–590.
Terbeek, D. (**1977**). ''A cross-language multidimensional scaling study of vowel perception,'' doctoral dissertation, University of California, Los Angeles.
Trubetzkoy, N. S. (**1969**). *Principles of Phonology* (Univ. of California, Berkeley).
Werker, J. F., and Tees, R. C. (**1984a**). ''Cross-language speech perception: evidence for perceptual reorganization during the first year of life,'' Infant Behav. Dev. **7**, 49–63.
Werker, J. F., and Tees, R. C. (**1984b**). ''Phonemic and phonetic factors in adult cross-language speech perception,'' J. Acoust. Soc. Am. **75**, 1866–1878.
Werker, J. F., Gilbert, J. H. V., Humphrey, K., and Tees, R. C. (**1981**). ''Developmental aspects of cross-language speech perception,'' Child Dev. **52**, 349–353.
Zee, E. (**1981**). ''Effect of vowel quality on perception of post-vocalic nasal consonants in noise,'' J. Phonetics **9**, 35–48.

Tests of auditory–visual integration efficiency within the framework of the fuzzy logical model of perception

Dominic W. Massaro[a] and Michael M. Cohen
University of California, Santa Cruz, California 95060

(Received 12 October 1999; revised 22 February 2000; accepted 21 April 2000)

The fuzzy logical model of perception [FLMP, Massaro, *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle* (MIT Press, Cambridge, MA, 1998)] has been extremely successful at describing performance across a wide range of ecological domains as well as for a broad spectrum of individuals. Because the model predicts optimal or maximally efficient integration, an important issue is whether this is the case for most individuals. Three databases are evaluated to determine to what extent a significant quantitative improvement in predictive ability can be obtained if integration is assumed to be somewhat inefficient. For the most part, there were no significant signs of inefficient integration. The previous differences found by Grant and Seitz [J. Acoust. Soc. Am. **104**, 2438–2450 (1998)] must be due to their measures of efficiency, which appear to be invalid and/or conflate information with integration efficiency. Finally, the descriptive ability of the FLMP is shown to be theoretically informative and not simply the model's ability to describe any possible outcome. © *2000 Acoustical Society of America.* [S0001-4966(00)01008-0]

PACS numbers: 43.71.An, 43.71.Ma [CWT]

I. TESTS OF AUDITORY–VISUAL INTEGRATION WITHIN THE FRAMEWORK OF THE FLMP

Grant and Seitz (1998) should be applauded for initiating an important study of the individual's ability to integrate auditory and visual (AV) information. As pointed out in their discussion, this assessment has important implications for treatment of hearing-impaired and even visual-impaired subjects (see Grant and Walden, 1995; Grant, Walden, and Seitz, 1998). However, before we can reach any conclusion about a person's ability to integrate, a more precise treatment within a validated model of integration must be carried out.

Grant and Seitz propose four measures of AV integration efficiency. The first is an integration efficiency (IE) measure from Braida's (1991) Pre-Labeling Model (PRE). From what we can tell, however, this measure uses only correct responses rather than all of the cells of the bimodal confusion matrix. We believe that all of the data should be used in measuring IE. The McGurk susceptibility measure is confounded with the quality of the auditory and visual information. A person will be more susceptible to McGurk effects to the extent she has poor auditory information and good visual information, independently of her IE (Massaro, 1998, pp. 12–14). The two measures of auditory delay are also debatable in that we might expect better integrators to be less susceptible to asynchrony of the two sources rather than more susceptible.

Their two measures of AV benefit certainly reflect the quality of the auditory and visual information. The measures $AV-A$ and $(AV-A)/(1-A)$ are measures of the benefit gained from the addition of visible speech. The measure $(AV-A)/(1-A)$ adjusts for overall performance level and provides a coarse measure of the visual contribution. However, this measure is influenced by both the amount of visual information and the efficiency of integrating it with the auditory information. Thus, this measure is necessarily confounded with the amount of visual information the subject has available. Given these limitations, we propose that the Grant and Seitz (1998) analyses do not provide valid measures of integration efficiency.

II. A FORMAL MODEL OF INFORMATION INTEGRATION

We argue that any measure of integration efficiency requires a formal model of performance that specifies exactly how integration takes place. Performance of a given individual can then be assessed within the framework of the model to address the question of integration efficiency. We propose the fuzzy logical model of perception (FLMP) as an ideal model for this type of analysis. The assumptions central to the model are: (1) each source of information is evaluated to determine the degree to which that source specifies various alternatives, (2) the sources of information are evaluated independently of one another, (3) the sources are integrated to provide an overall degree of support for each alternative, and (4) perceptual identification and interpretation follows the relative degree of support among the alternatives. In a two-alternative task with /ba/ and /da/ alternatives, the degree of auditory support for /da/ can be represented by a_i, and the support for /ba/ by $(1-a_i)$. Similarly, the degree of visual support for /da/ can be represented by v_j, and the support for /ba/ by $(1-v_j)$. The probability of a response to the unimodal stimulus is simply equal to the feature value. For bimodal trials, the predicted probability of a response, $P(/\text{da}/)$ is equal to

$$P(/\text{da}/) = a_i v_j / [a_i v_j + (1-a_i)(1-v_j)]. \quad (1)$$

[a]Electronic mail: massaro@fuzzy.ucsc.edu

TABLE I. The amount of support for each of the three alternatives as a function of the auditory and visual sources of information. These degrees of support illustrate that overall performance on the bimodal conditions can be significantly worse than overall performance on the auditory conditions.

Support for alternative	Auditory /ba/	Auditory /va/	Auditory /da/	Visual /ba/	Visual /va/	Visual /da/
/ba/	0.7	0.1	0.2	0.3	0.2	0.5
/va/	0.3	0.6	0.1	0.3	0.3	0.4
/da/	0.1	0.2	0.7	0.5	0.3	0.2

$$P(/\text{ba}/|A_bV_b)=\frac{0.7\times0.3}{(0.7\times0.3)+(0.3\times0.3)+(0.1\times0.5)}$$

$$=0.21/0.35=0.60,$$

$$P(/\text{va}/|A_vV_v)=\frac{0.6\times0.3}{(0.6\times0.3)+(0.1\times0.2)+(0.2\times0.3)}$$

$$=0.18/0.26=0.69,$$

$$P(/\text{da}/|A_dV_d)=\frac{0.7\times0.2}{(0.7\times0.2)+(0.2\times0.5)+(0.1\times0.4)}$$

$$=0.14/0.28=0.50,$$

Average auditory unimodal$=(0.7+0.6+0.7)/3=2.0/3=0.67$,
Average bimodal$=0.6+0.69+0.50=0.60$.

In the course of our research, we have found that the FLMP accurately describes human pattern recognition. We have learned that people use many sources of information in perceiving and understanding speech, emotion, and other aspects of the environment. The experimental paradigm that we have developed also allows us to determine which of the many potentially functional cues are actually used by human observers (Massaro, 1998, Chap. 1). This paradigm has already proven to be effective in the study of audible, visible, and bimodal speech perception (Massaro, 1987, 1998; Massaro and Cohen, 1976; Oden and Massaro, 1978).

How can the FLMP be used to assess integration efficiency? As can be seen in Eq. (1), the auditory and visual sources of support are multiplied to give an overall degree of support for each response alternative. The value a_i representing the degree of auditory support is assumed to be the same on both unimodal auditory and bimodal trials. This same property holds for the visual support. This property and the multiplicative integration rule, followed by the relative goodness rule (RGR), entail the process to be optimal and thus maximally efficient (see Massaro, 1998, pp. 115–117; Massaro and Stork, 1998; Massaro and Friedman, 1990).

Grant and Seitz (1998) express an ''apparent belief that a second source of information can only improve overall performance.'' Although it seems unintuitive with an optimal integration rule, a second source can indeed lower performance, even though (in fact because) integration is operating efficiently. In order to demonstrate this possibility within the context of the FLMP, assume three alternatives, /ba/, /va/, and /da/, which are differentially supported by the auditory and visual sources by the parameter values given in Table I. Assume that a bimodal /da/ is presented and auditory /da/ supports /da/ 0.7, /va/ 0.1, and /ba/ 0.2. Visual /da/ supports /da/ 0.5, /va/ 0.4, and /ba/ 0.2. Correcting identifying /da/ on unimodal auditory and on unimodal visual trials would be 0.7 and 0.5, respectively. As shown in Table I, performance on bimodal trials would be equal to 0.50, which is well below performance of 0.7 on unimodal auditory trials. This somewhat surprising outcome can be true for overall performance in the task (see Table I). This theoretical demonstration, which is also found in empirical results, challenges Grant and Seitz's apparent belief that the FLMP predicts that a second source of information can only improve overall performance.

In our previous work, we have contrasted a fuzzy logical model of perception with a single-channel model (SCM) of perception. These represent integration and nonintegration models, respectively, and therefore a test between these models at the individual subject level indicates whether a person integrates the auditory and visual speech. The SCM is mathematically equivalent to a weighted averaging model (WTAV), which is an inefficient algorithm for combining the auditory and visual sources. This model predicts that two sources can never be more informative than one. Thus, previous contrasts of the two models have addressed the issue of integration efficiency. Given that the FLMP has consistently provided a significantly better description of a variety of results from several different types of experiments, people generally must be fairly efficient information integrators. In further simulations of the two-alternative task (Massaro, 1998, Chap. 10), however, we found that the fit of the FLMP fell slightly shy of a benchmark criterion indicating a perfectly accurate fit. The addition of decision noise (noise added at the response selection stage) with a standard deviation of 0.1 was necessary to bring the FLMP in line with what would be expected from the data being generated by this model. Thus, there is some hint that the perceivers might not be perfectly efficient integrators. In this letter, we pursue this issue more directly.

A direct way to measure integration efficiency in the FLMP is to determine whether there is any loss of information on bimodal trials relative to unimodal trials. A potentially inefficient integration model can be formalized within the FLMP framework. One simply assumes that reduced information from the auditory and visual sources can be present on bimodal trials relative to unimodal trials. In this case, the degree of auditory support on bimodal trials, a_{iB}, is compromised by the function

$$a_{iB}=w_a a_{iU}+(1-w_a)0.5. \tag{2}$$

An analogous function describes the visual information

$$v_{jB}=w_v v_{jU}+(1-w_v)0.5, \tag{3}$$

where w_a and w_v correspond to the weights given the auditory and visual feature values, respectively.

For tasks with two response alternatives or for models with features that lie between 0 and 1, the feature values represent more support for an alternative to the extent the value is greater than 0.5. The value 0.5 is completely ambiguous. Because the weights can lie between 0 and 1, the smaller the weight value the less the support is controlled by the unimodal feature value and the more it is controlled by 0.5 (complete ambiguity). A weight value of 1 makes the same prediction as the original FLMP.

Tests of efficiency, therefore, simply involve testing this new model and determining the weight values and also to what extent this model gives a better description of perfor-

mance compared to the standard FLMP. We apply this model and two related models to two data sets given in Massaro (1998, Chaps. 2, 6, 10), and the data set described by Grant and Seitz (1998).

III. EXPANDED FACTORIAL DESIGN WITH TWO RESPONSE ALTERNATIVES

A typical manipulation is to systematically vary the ambiguity of each of the sources of information. We used synthetic speech to cross five levels of audible speech varying between /ba/ and /da/ with five levels of visible speech varying between the same alternatives in an expanded factorial design (Massaro, 1998). There were 24 observations at each of the 35 unique experimental conditions. Eighty-two subjects were instructed to listen and to watch the speaker, and to identify the syllable as /ba/ or /da/.

The mean observed proportion of /da/ identifications was computed for each subject for the 35 unimodal and bimodal conditions. Both the auditory and the visual sources of information had a strong impact on the identification judgments in both the unimodal and bimodal conditions. Most importantly, the auditory and visual effects were not additive in the bimodal condition, as demonstrated by a significant auditory–visual interaction. This result is consistently obtained in this type of experiment. It means that the influence of one source of information is greatest when the other source is neutral or ambiguous.

The FLMP gave a better description than the WTAV model for 94% of these 82 participants, with average root-mean-square deviations (RMSDs) of the individual model fits of 0.051 and 0.097, respectively. To further address the issue of efficiency, we tested the weighted bimodal FLMP given by Eqs. (2) and (3) against these same results. The average RMSD was 0.0449, smaller than the standard fit of 0.051. An analysis of variance on the RMSDs from the two models showed that this difference was significant, $p < 0.001$. This statistical result should not be surprising because the weighted bimodal FLMP is identical to the standard FLMP, but with two additional free parameters. Rather than conclude that some of the observers are inefficient, the better fit may simply reflect the addition of two free parameters rather than a real loss of information in the bimodal condition. We are confident that a goodness-of-fit measure which takes into account model flexibility (Myung and Pitt, 1997; Massaro *et al.*, in press) would not find a significant advantage of the inefficient integration model relative to the FLMP.

Given this possibility, we felt that a more appropriate model would be one that simply compromises the outcome of integration rather than the separate information values. In this case, the predicted compromised probability of a response, $P(/\text{da}/)$ on bimodal trials is equal to

$$P(/\text{da}/) = wP(/\text{da}/) + (1-w)0.5, \tag{4}$$

where $P(/\text{da}/)$ is equal to Eq. (1), and the a_i and v_j values are identical to those on the unimodal trials.

This more direct model of efficiency gave an average RMSD of 0.0491, which was very little improvement over the standard FLMP. Most of the estimated weight values were close to 1, indicating that there was very little loss of information in the bimodal condition relative to the unimodal conditions. It appears that only 6 or 8 of the 82 subjects appear to be inefficient integrators.

IV. EXPANDED FACTORIAL DESIGN WITH EIGHT RESPONSE ALTERNATIVES

For the second database, we replicated this same experiment with 36 subjects, given eight rather than just two response alternatives. Subjects were instructed to listen to and watch the talker, and to identify the syllable as /ba/, /da/, /bda/, /dba/, /tha/, /va/, /ga/, or ''other.'' The category other was to be used by the subject whenever none of the other seven responses seemed suitable. Although the test continua were between /ba/ and /da/, we obtained several other response alternatives (see Massaro, 1998, pp. 184–188). These judgments reflect the contribution of both auditory and visual speech, even when observers are permitted a larger permissible set of response alternatives.

The FLMP is tested against results with multiple response alternatives in the same manner as with just two response alternatives. With more than two alternatives, it is necessary to estimate a unique parameter to represent the degree to which each source of information supports each alternative. The fit of this model requires $5a_i$ and $5v_j$ parameters for each of the 8 response alternatives, for a total of 80 free parameters. This might seem like a large amount but the number of data points to be predicted has increased by the same factor. We are now predicting $35\times8=280$ data points. The fit of this model to each of the 36 subjects produced an average RMSD of 0.0507. To assess whether the FLMP maintains its advantage with multiple response alternatives, we compared this fit with that of a single-channel model (or equivalently, a weighted averaging model). The fit of this competing model was about two times poorer, giving an RMSD of 0.1049. The FLMP gave a better description than the WTAV model for all but one of the 36 subjects.

To address the issue of efficiency, we tested the two weighted bimodal FLMPs against these same results. The predictions are made in the same way except now the neutral response probability is 1/8. For the two-weight model, the average RMSD was 0.0501, very close to the original average RMSD of 0.0507. The weight values were equal to 1 for most of the subjects and close to 1 for the others. For the one-weight model, the average RMSD was 0.0500. These weight values were equal to 1 for most of the subjects and close to 1 for the others.

V. IDENTIFICATION CONFUSION MATRIX WITH 18 ALTERNATIVES

Finally, we analyzed Grant and Seitz's (1998) 40 subjects in the vCv consonant identification task. We have carried out two different types of model descriptions of this type of experiment with A, V, and AV confusion matrices: modality analysis and feature analysis (Massaro and Cohen, 1999).

A. Modality-analysis implementation

As in the eight-alternative task, it is necessary to estimate a unique parameter to represent the degree to which each source of information supports each alternative. We use a_{Bi} to represent the degree to which the audible speech supports the alternative /ba/. The term v_{Dj} would represent the degree to which the visible speech supports the alternative /da/, and so on for the other response alternatives. Given both audible and visible speech, the total support for the alternative /ba/, $s(/\text{ba}/)$, would be

$$s(/\text{ba}/) = a_{Bi} v_{Bj}, \tag{5}$$

and so on for the other test conditions and the other alternatives.

As in the case of just two alternatives, the probability of a particular categorization is assumed to be equal to the relative goodness of match of that alternative relative to the sum of the goodness-of-match values of all possible response alternatives. With 18 stimulus–response alternatives, each test stimulus provides different degrees of support for each alternative. It is necessary to estimate 18 free parameters for each of the 18 test stimuli in each modality. Thus, 324 free parameters are required for the auditory modality and 324 for the visual modality. We are able to test the model by predicting 3×324 data points with 2×324 free parameters.

As in the previous tests, the standard FLMP and the two-weighted bimodal FLMPs were fit to the results. The predictions are made in the same way except that the neutral response probability is 1/18. The average RMSD for the two-weight model was equal to the average RMSD of the standard FLMP (0.0111). The weight values were equal to 1 for all of the subjects. For the one-weight model, the average RMSD was 0.0110. The weight values were equal to 1 for most of the subjects and close to 1 for the others.

B. Feature analysis implementation

The model test we have just presented makes no assumptions about the psychophysical relationship among the different test items. A unique parameter is estimated for each possible stimulus–response pairing. For example, a unique parameter is estimated to represent the amount of support a visual /b/ provides for the response alternative /d/. To test the psychological reality of various linguistic features and to reduce the number of free parameters, we have articulated the FLMP in terms of audible and visible support for these features (Massaro and Cohen, 1999). This formulation has the potential to save a large number of features, because it is assumed that a given feature in a given modality has the same impact regardless of what segment it is in. Following the tradition begun with Miller and Nicely (1955), we can define the 18 consonants by five features: voicing, nasality, place, duration, and frication (see Massaro and Cohen, 1999).

We assume that features for the 18 consonants are simply sensory primitives that distinguish speech categories. Although the features used in the following tests are chosen to be equivalent to the linguistic features, they should be thought of simply as handy labels for the underlying sensory features. Thus, for example, the auditory feature for place would not necessarily be equivalent to the parameter value for the visible feature for place. In fact, the feature values for one modality should be independent of the feature values for another modality. For example, we would expect that voicing and nasality would have informative feature values for auditory speech and relatively neutral feature values for visible speech. The place feature, on the other hand, would give relatively informative values for visible speech. Thus, the features at the evaluation stage are not linguistic, but perceptual.

Thus, each of the 18 syllables would be described by the conjunction of five features for unimodal speech and the conjunction of ten features for bimodal speech. Even though each feature is defined as a specific value or its complement (e.g., voiced or voiceless), its influence in the perception of visible speech is represented by a value between 0 and 1. The parameter value for the feature indicates the amount of influence that feature has. Therefore, if the /ma/ and /na/ prototypes are each expected to have a nasal feature and the calculated parameter value for this feature is 0.90, then the nasal feature is highly functional in the expected direction. Alternatively, if the calculated parameter value for the nasal feature is 0.50, then the nasal feature is not functional at all. Because of the definition of negation as 1 minus the feature value, a feature value of 0.5 would give the same degree of support for an alternative that has the feature, as it should for an alternative that doesn't have the feature. Finally, if the calculated parameter value is 0.20 then the nasal feature is functional but the opposite of the expected direction. Finally, it should be noted that the features are not marked in this formulation: absence of nasality is as informative as presence of nasality. Thus, if a nasal stimulus supports a nasal response alternatives to the degree 0.9, then a non-nasal stimulus also supports a non-nasal alternative to the degree 0.9.

The overall match of a test stimulus to each syllable prototype was calculated by combining the feature matches according to the assumptions of the FLMP. These constraints dictate that (1) the features are the sources of information that are evaluated independently of one another, and (2) the features are integrated multiplicatively to give the overall degree of support for a syllable alternative. Thus, the overall degree of support for /ba/, $s(/\text{ba}/)$, given the presentation of a /ba/ syllable, is

$$s(/\text{ba}/|/\text{ba}/) = a_v a_n a_p a_d a_f v_v v_n v_v v_d v_f, \tag{6}$$

where each feature value indexes a match between the feature in the stimulus and the corresponding feature in the /ba/ prototype. The feature a_v correspond to auditory voicing, v_n to visual nasality, and so on. A mismatch between the feature in the stimulus and the corresponding feature in the prototype would be indexed by $(1-f_i)$, where f_i corresponds to the modality's feature value. Thus, the support for the /ka/ prototype given presentation of a /ba/ syllable, is

$$s(/\text{ka}/) = (1-a_v)a_n(1-a_p)a_d a_f(1-v_v) \times v_n(1-v_p)v_d v_f, \tag{7}$$

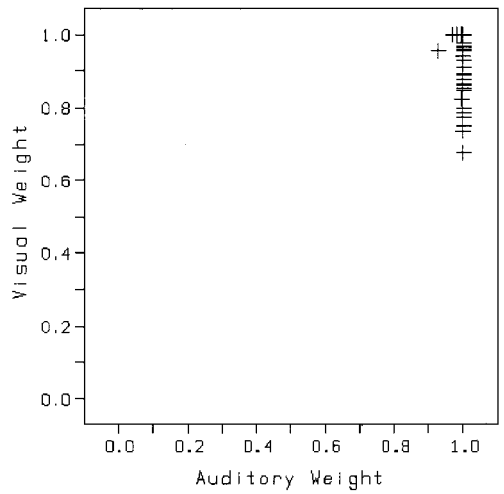


FIG. 1. The estimated auditory and visual weight values for the two-weight model for each of the 40 subjects in the Grant and Seitz (1998) 18-alternative task, for the feature-analysis implementation.

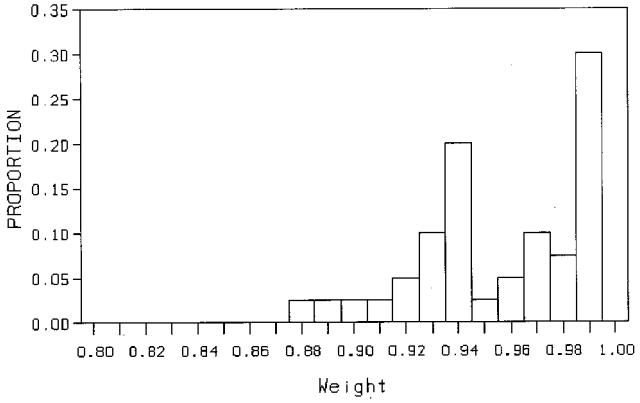


FIG. 2. The distribution of weight values for the one-weight model for the 40 subjects in the Grant and Seitz (1998) 18-alternative task, for the feature-analysis implementation.

where $(1-f_i)$ indexes a mismatch between the feature in the /ba/ stimulus and the corresponding feature in the /ka/ prototype. This same formulization is used for the place feature which has six levels rather than just two.

After the overall degree of support for each syllable is calculated, the stimulus is categorized according to the RGR, which states that the relative probably of choosing an alternative is the goodness of match of that alternative divided by the sum of the goodness of match of all alternatives. Thus, this model implementation parallels the previous one in all aspects except in terms of the featural description of the stimulus and response alternatives. The FLMP can thus be tested against the confusion matrix by estimating the amount of information in each feature and the featural correspondence between the stimulus and response prototypes. Thus, five parameters are necessary to describe the auditory information and the same number to describe the visual.

The standard FLMP and the two weighted bimodal FLMPs were fitted to the results in the same manner, with the neutral feature value of 0.5. The average RMSD of the standard FLMP fit was 0.1001. For the two-weight and one-weight models, the average RMSDs were equal to 0.1000 and 0.1001, respectively, essentially equal to the fit of the standard FLMP. The auditory and visual weight values for each subject for the two-weight model are given in Fig. 1. These weights were equal to one for most of the subjects and close to one for the others. The distribution of the weight values for the one-weight model is shown in Fig. 2. These values were also very close to one.

In summary, our test of integration efficiency revealed very little support for the thesis that some individuals might be less-efficient integrators than others. This result clearly held for seven of eight model tests across three different data sets. The only hint of inefficient integration was for the data set of 82 subjects from our expanded factorial design with two response alternatives. Possible explanations for this outcome are given in Massaro (1998, pp. 313–318).

VI. SENSITIVITY OF THE FLMP

A second important issue is raised by Grant and Seitz's (1998) claim that ''the consistently excellent fits achieved by the FLMP may also suggest that the model is less sensitive in recognizing subtle changes in integration efficiency'' (p. 2439). To support their claim, Grant and Seitz fit the FLMP to two data sets, one original set of data and a second that involved some modification of the first data set to give 16% better AV performance. The RMSDs of the fit to the two data sets were 0.013 and 0.020, respectively. Although the authors interpreted this small difference as a nonsignificant one, there was no justification for this conclusion and we believe that the observed difference is significant. To pursue this possibility, we made a similar modification of each of their 40 data sets and tested the FLMP against these two sets of data. To achieve the new hypothetical data sets, we generated a new set of bimodal results for each subject by differentially modifying the correct and incorrect proportions in the bimodal confusion matrix. Every correct cell along the negative diagonal was multiplied by 1.8. For the other cells, we multiplied their proportions by 0.2. Each entry in the final bimodal matrix was determined by normalizing each cell value by the total of all 18 cells in that row. The overall accuracy of this new set of bimodal results for each subject averaged about 14% more accurate than the original data set. Thus, the new data set has the original unimodal results and hypothetical bimodal results that are more accurate than would be expected from the FLMP.

If Grant and Seitz are correct, then the FLMP should fit these hypothetical results about as well as the original results. We have already reported the RMSD of the original data set, which was 0.0112. The average RMSD for the simulated data set was 0.0130, significantly larger that the fit to the original data, $F(1,28)=8.28$, $p=0.008$. As an additional test of sensitivity, we carried out the same contrast but with a parameter-free test in which the unimodal response probabilities were used to predict the bimodal judgments. The RMSD values for the original data set and the enhanced data set were 0.0522 and 0.0658, a significant difference, $p<0.001$. Thus, we can conclude that the FLMP is indeed sensitive to small differences in categorization behavior, and

as demonstrated, provides a powerful framework for evaluating the efficiency of integration.

ACKNOWLEDGMENTS

The research was supported by grants from the National Institute of Deafness and Other Communicative Disorders (Grant No. PHS R01 DC00236), the National Science Foundation (Challenge Grant No. CDA-9726363), Intel Corporation, and the University of California Digital Media Innovation Program.

Braida, L. D. (**1991**). ''Crossmodal integration in the identification of consonant segments,'' Quarterly J. Exp. Psychol. **43**, 647–677.

Grant, K. W., and Seitz, P. F. (**1998**). ''Measures of auditory–visual integration in nonsense syllables and sentences,'' J. Acoust. Soc. Am. **104**, 2438–2450.

Grant, K. W., and Walden, B. E. (**1995**). ''Predicting auditory–visual speech recognition in hearing-impaired listeners,'' in Proceedings of the XIIIth International Congress of Phonetic Sciences, Vol. 3, pp. 122–129 (unpublished).

Grant, K. W., Walden, B. E., and Seitz, P. F. (**1998**). ''Auditory–visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory–visual integration,'' J. Acoust. Soc. Am. **103**, 2677–2690.

Massaro, D. W. (**1987**). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry* (Erlbaum, Hillsdale, NJ).

Massaro, D. W. (**1998**). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle* (MIT Press, Cambridge, MA).

Massaro, D. W., and Cohen, M. M. (**1976**). ''The contribution of fundamental frequency and voice onset times to the /zi/–/si/ distinction,'' J. Acoust. Soc. Am. **60**, 704–717.

Massaro, D. W., and Cohen, M. M. (**1999**). ''Speech perception in hearing-impaired perceivers: Synergy of multiple modalities,'' J. Speech Lang. Hear. Sci. **42**, 21–41.

Massaro, D. W., Cohen, M. M., Campbell, C. S., and Rodrigues, T. (**in press**). ''Bayes factor of model selection balidates FLMP,'' Psychonom. Bull. Rev.

Massaro, D.W., and Freidman, D. (**1990**). ''Models of integration given multiple sources of information,'' Psychol. Rev. **97**, 225–252.

Massaro, D. W., and Stork, D. G. (**1998**). ''Speech recognition and sensory integration,'' Am. Sci. **86**, 236–244.

Massaro, D. W. (submitted).

Miller, G. A., and Nicely, P. E. (**1955**). ''An analysis of perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–352.

Myung, I. J., and Pitt, M. A. (**1997**). ''Applying Occam's razor in modeling cognition: A Bayesian approach,'' Psychonom. Bull. Rev. **7**, 79–95.

Oden, G. C., and Massaro, D. W. (**1978**). ''Integration of featural information in speech perception,'' Psychol. Rev. **85**, 172–191.

The effect of parametric variations of cochlear implant processors on speech understanding

Philipos C. Loizou[a] and Oguz Poroy
Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas 75083-0688

Michael Dorman
Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287

(Received 13 January 2000; accepted for publication 1 May 2000)

This study investigated the effect of five speech processing parameters, currently employed in cochlear implant processors, on speech understanding. Experiment 1 examined speech recognition as a function of stimulation rate in six Med-El/CIS-Link cochlear implant listeners. Results showed that higher stimulation rates (2100 pulses/s) produced a significantly higher performance on word and consonant recognition than lower stimulation rates (<800 pulses/s). The effect of stimulation rate on consonant recognition was highly dependent on the vowel context. The largest benefit was noted for consonants in the /uCu/ and /iCi/ contexts, while the smallest benefit was noted for consonants in the /aCa/ context. This finding suggests that the /aCa/ consonant test, which is widely used today, is not sensitive enough to parametric variations of implant processors. Experiment 2 examined vowel and consonant recognition as a function of pulse width for low-rate (400 and 800 pps) implementations of the CIS strategy. For the 400-pps condition, wider pulse widths (208 μs/phase) produced significantly higher performance on consonant recognition than shorter pulse widths (40 μs/phase). Experiments 3–5 examined vowel and consonant recognition as a function of the filter overlap in the analysis filters, shape of the amplitude mapping function, and signal bandwidth. Results showed that the amount of filter overlap (ranging from −20 to −60 dB/oct) and the signal bandwidth (ranging from 6.7 to 9.9 kHz) had no effect on phoneme recognition. The shape of the amplitude mapping functions (ranging from strongly compressive to weakly compressive) had only a minor effect on performance, with the lowest performance obtained for nearly linear mapping functions. Of the five speech processing parameters examined in this study, the pulse rate and the pulse width had the largest (positive) effect on speech recognition. For a fixed pulse width, higher rates (2100 pps) of stimulation provided a significantly better performance on word recognition than lower rates (<800 pps) of stimulation. High performance was also achieved by jointly varying the pulse rate and pulse width. The above results indicate that audiologists can optimize the implant listener's performance either by increasing the pulse rate or by jointly varying the pulse rate and pulse width. [S0001-4966(00)02308-0]

PACS numbers: 43.71.Ky, 43.66.Sr [CWT]

INTRODUCTION

To account for the variability in performance among cochlear implant subjects, cochlear implant manufacturers started incorporating several speech-processing strategies in their implant processors. The Advanced Bionics Corporation, for example, supports among other strategies the SAS and the continuous interleaved sampling (CIS) strategies in their Clarion device, while the Cochlear Corporation supports the ACE, SPEAK, and CIS strategies in their Nucleus 24 device. Increasing the number of speech strategies available in implant processors not only increases the chances that one of those strategies might be more beneficial than others (e.g., Osberger and Fisher, 1999), but also increases the complexity in choosing the right set of parameters associated with each strategy.

Most speech processing strategies today can be configured using a multitude of parameters, which can be easily modified by the audiologist using existing fitting software. In the SPEAK strategy, for example, one can change the pulse width, the number of maxima selected, the pulse rate, the filter allocation table, etc. In the CIS strategy, currently supported by all implant devices (Nucleus 24, Med-El and Clarion) in the United States, one can change the pulse width, the pulse rate, the electrode stimulation order, and the compression function that maps the acoustical input to electrical output. A series of studies by Wilson and colleagues (e.g., Wilson *et al.*, 1991, 1993, 1995, 1999) has shown that the CIS parameters (e.g., pulse rate, pulse width, etc.) can be varied to optimize individual subjects' performance. They reported, for example, significant improvements in speech recognition for one subject over the course of three successive visits to their laboratory. The subject's mean score on consonant recognition improved from 56% correct (at the initial visit) using a 167-μs/phase, 500-pps processor with a staggered order of stimulation, to 79% correct (at the second visit) using 33-μs/phase, 833-pps processor with a staggered order of stimulation, to 85% correct (at the third visit) using 33-μs/phase, 833-pps processor with an apex-to-base order of stimulation. Dorman and Loizou (1997) also showed simi-

[a]Electronic mail: loizou@utdallas.edu

lar improvements for another Ineraid subject fitted with a CIS processor. The subject's sentence scores improved from 73% correct using a 100-μs/phase, 823-pps processor with a staggered stimulation order to 100% correct using a 40-μs/phase, 2020-pps processor with an apex-to-base stimulation order. These studies demonstrated how the electrode stimulation order and the stimulation rate could greatly affect the subjects' performance. Other parameters that were found to affect performance include the number of channels (Wilson *et al.*, 1991, 1995; Lawson *et al.*, 1996), pulse duration (Wilson *et al.*, 1993), envelope cutoff frequency (Lawson *et al.*, 1993), and signal bandwidth (Zerbi *et al.*, 1998). As it turns out, some of these parameters interact with each other. Wilson *et al.* (1993) reported that subjects obtained the highest performance at different combinations of pulse rate and pulse width. Other studies (Brill *et al.*, 1997; Kiefer *et al.*, 1997) showed a tradeoff between number of channels and pulse rate. Brill *et al.* (1997) found that trading channels for higher stimulation rates improved performance. Higher stimulation rates produced significantly higher performance than lower stimulation rates. For implant devices with a large number of electrodes, the choice of electrodes to be stimulated is yet another parameter. The principal advantage of such electrode arrays having many electrodes is not that they can support a large number of channels, but rather that they allow the selection of subsets of electrodes to optimize individual subject's performance. Lawson *et al.* (1996) and Zwolan *et al.* (1997) have demonstrated that the selection of electrodes can significantly affect performance.

Given the large number of speech processing parameters available and the effect of some of these parameters on speech understanding, it is becoming increasingly more important to identify the set of parameters which is most likely to affect speech recognition. This is an extremely important issue particularly for fitting implant patients, since in a clinical setting the audiologists cannot devote too much time to select these parameters.

The aim of this study is to identify the speech processing parameter(s) that affects speech recognition the most. The results of our study will greatly facilitate the fitting of new implant patients, as it will provide to audiologists a good starting point for fitting. We use the CIS speech processing strategy in this article, however, the parameters examined here are employed in all speech processing strategies, and most of these parameters can be changed manually by audiologists using existing fitting software. Five parameters will be varied systematically: pulse rate, pulse width, filter overlap, compression function, and signal bandwidth. The effect of these parameters on speech recognition will be examined.

I. EXPERIMENT 1: EFFECT OF PULSE RATE

Cochlear implant devices present information to the electrodes in analog or pulsatile form. In pulsatile stimulation, the information is delivered to the electrodes using a set of narrow pulses. In some devices, the amplitudes of these pulses are extracted from the envelopes of the filtered waveforms. The advantage of this type of stimulation over analog stimulation is that the pulses can be delivered in a nonoverlapping fashion, thereby minimizing channel interactions. One of the issues associated with pulsatile stimulation is how fast, in terms of number of pulses per second, we need to stimulate the electrodes. Whether higher rates of stimulation provide more benefit to speech understanding than lower rates of stimulation has been the subject of debate. Presently, the stimulation rate employed by commercial implant processors varies from a low of 200 pulses/s to a high of 2400 pulses/s per electrode (see review by Loizou, 1998). Does it make a difference in speech understanding, however, whether we stimulate the electrodes at a low rate of 200 pulses/s or at a high rate of 2400 pulses/s? Wilson *et al.* (1998) demonstrated, using intracochlear-evoked potential recordings, that higher rates of stimulation have the potential of disrupting the high synchrony observed with low-rate stimulation, allowing more normal ''stochastic'' firing patterns. Higher rates of stimulation also provide a better temporal representation of the speech envelope, but how important is that for speech understanding? These are the questions we are trying to address in this experiment. To answer these questions, we varied systematically the pulse rate from 400 to 2100 pulses/s and examined vowel, consonant, and monosyllabic word recognition.

A. Method

1. Subjects

The subjects were six postlingually deafened adults who had used a six-channel CIS processor for periods ranging from 3 to 4 years. All the patients had used a four-channel, compressed-analog signal processor (Ineraid) for at least 4 years before being switched to a CIS processor. The patients ranged in age from 40 to 68 years and they were all native speakers of American English. Biographical data for each patient are presented in Table I.

2. Speech material

The test material included monosyllabic words, consonants in /vCv/ environment, and vowels in /hVd/ environment. The word test consisted of four different lists with 50 monosyllabic (CNC) words in each list. The consonant test consisted of 20 /vCv/ consonants in three vowel environments, /i a u/, produced by a single female speaker, and was taken from the consonant database recorded at the House Ear Institute (Shannon *et al.*, 1999). The 20 consonants were /b p d t g k f v s z ʃ ð t ʃ d ʒ m n r l j w/ in i/C/i, u/C/u, and a/C/a format. We chose to use consonants produced by a female speaker to avoid possible ceiling effects, since the vowels produced by female speakers were found by implant patients to be relatively harder to identify than the vowels produced by male speakers (Loizou *et al.*, 1998). The vowel material consisted of the vowels in the words: ''heed, hid, hayed, head, had, hod, hud, hood, hoed, who'd, heard.'' Each word was produced by three men, three women, and three girls. The stimuli were drawn from a set used by Hillenbrand *et al.* (1995).

3. Experimental setup

All the experiments were performed on our laboratory cochlear implant processor, which was based on the design

TABLE I. Biographical data of the six cochlear-implant users who participated in this study.

Subject	Gender	Age (years) at detection of hearing loss	Age at which hearing aid gave no benefit	Age fit with Ineraid	Age at testing	Etiology of hearing loss	Score on H.I.N.T sentences in quiet	Score on NU-6 words in quiet
S1	F	10	46	47	55	unknown	44	20
S2	M	5	43	48	58	unknown	92	43
S3	F	7	31	33	40	unknown/ hereditary	100	80
S4	F	23	48	51	57	unknown	100	71
S5	M	20	46	63	68	unknown	88	46
S6	M	19	19	29	41	Cogan's syndrome	100	93

of the Geneva/RTI/MEEI wearable processor (Francois *et al.*, 1994). Several modifications were made to the Geneva design, the most important of which was the addition of five current sources. (The wearable Geneva/RTI/MEEI processor was originally designed with one current source and could therefore provide only nonsimultaneous stimulation.) These modifications enabled us to investigate nonsimultaneous as well as simultaneous stimulation. The block diagram of the laboratory processor is shown in Fig. 1. The input analog circuit consists of an audio multiplexer that selects the source of the input signal to the processor, several fixed-gain amplifiers, one variable-gain amplifier (adjusted externally by a sensitivity knob), an antialiasing filter, and a 16-bit A/D converter. The sampling rate of the A/D converter is controlled by the DSP chip, and for this study it was fixed at 22 kHz. The cutoff frequency of the antialiasing filter was set at 6.7 kHz. Once the signal is digitized, it is transmitted to the Motorola DSP56002 chip, where it is processed through the CIS strategy (see description in the following section). The CIS outputs are finally fed through a SSI port to the current sources built around a digital-to-analog converter. Biphasic pulses are generated, with amplitudes equal to the CIS envelope outputs, and sent to the electrodes for stimulation. The pulse width as well as the stimulation rate was controlled through software. More information about the hardware of the laboratory processor can be found in Poroy and Loizou (2000).

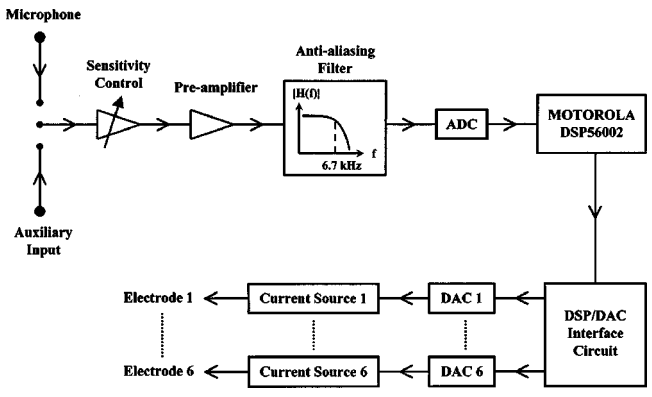


FIG. 1. Block diagram of the laboratory cochlear implant processor used in this study.

4. CIS implementation

Signals were first processed through a preemphasis filter (2000-Hz cutoff), with a 3-dB/oct roll-off, and then bandpassed into six frequency bands using sixth-order Butterworth filters. The center frequencies of the six bandpass filters were 461, 756, 1237, 2025, 3316, and 5428 Hz. The envelopes of the filtered signals were extracted by full-wave rectification and low-pass filtering (second-order Butterworth) with a 100-, 200-, or 400-Hz cutoff frequency, depending on the pulse rate. The six envelope amplitudes A_i $(i=1,2,...,6)$ were mapped to electrical amplitudes E_i using a power-law transformation:

$$E_i = cA_i^p + d, \tag{1}$$

where c and d are constants chosen so that the electrical amplitudes fall within the range of threshold and most-comfortable level, and p is the power exponent. The power exponent p was set equal to -0.0001 to obtain a compression function similar to the logarithmic function found in the Med-El/CIS link device.[1] The power-law mapping was implemented using a table-lookup procedure using a table with 512 entries (for each electrode). The mapped envelope amplitudes were finally used to modulate biphasic pulses of duration 40 μs/phase at stimulation rates ranging from 400 to 2100 pulses/s. The electrodes were stimulated in the same order as in the subject's daily processors. For most subjects, the electrodes were stimulated in ''staggered'' order.

The stimulation rates investigated were 400, 800, 1400, and 2100 pulses/s. This range was chosen because it corresponds to the range of pulse rates currently supported by the three commercial implant devices, Clarion, Nucleus 24, and Med-El. The envelope (low-pass) cutoff frequencies were set to 100 Hz for the 400-pps processor, 200 Hz for the 800-pps processor, and 400 Hz for the 1400- and 2100-pps processors. These frequencies were chosen to avoid aliasing effects. The pulse duration for all four rate conditions was fixed at 40 μs/phase. Note that the 400 or 800-pps rate conditions could alternatively be implemented by widening the pulse width (see experiment 2). Changing the pulse width, however, would not only change the pulse rate but would also affect the threshold and most-comfortable level (MCL) values, and consequently the electrical dynamic range. In this experi-

ment, we only wanted to vary a single parameter, the pulse rate.

5. Procedure

The test was divided into four sessions, one for each rate condition. The four conditions were counterbalanced among subjects to avoid any order effects. Each session consisted of a consonant, a vowel, and a monosyllabic-word test. In the vowel test, there were 12 repetitions of each vowel, and in the consonant test there were 9 repetitions of each consonant. The stimuli were presented in blocks of three repetitions each. The monosyllabic words were presented only once. A different word list was used for each condition. The vowels and the consonants were completely randomized. All test sessions were preceded by one practice session in which the identity of the vowels/consonants was indicated to the listeners.

The stimuli were presented directly to the subjects through our laboratory processor at a comfortable listening level. To collect responses, a graphical interface was used that allowed the subjects to identify the words they heard by clicking on the corresponding button in the graphical interface. For the monosyllabic-word test, the subjects wrote down the word they heard.

B. Results and discussion

The results on monosyllabic word, vowel, and consonant recognition are shown in Fig. 2.

1. Monosyllabic words

The monosyllabic words were scored in percent words correct [Fig. 2(a)]. Repeated measures analysis of variance indicated a significant main effect of rate [$F(3,15)=7.197$, $p<0.005$]. *Post hoc* analysis (according to Scheffe) showed that the score obtained at 2100 pulses/s was significantly ($p<0.05$) higher than the score obtained at 800 pulses/s. The scores at 2100 and 1400 pulses/s did not differ.

The individual subject's performance is given in Fig. 3. All subjects benefited from higher rates, some more than others. The rate at which the open-set performance reached an asymptote varied across subjects. Some subjects (S5, S6) showed a significant improvement starting at a rate of 1500 pulses/s, while other subjects (S3, S4) showed a significant improvement starting at a rate of 800 pulses/s. Other subjects (S1, S2) did not show an improvement until using the rate of 2100 pulses/s.

The above results clearly show that cochlear implant users can receive a significant benefit on open-set speech recognition from higher rates of stimulation. There does not seem to be a single ''critical'' rate, however, above which all subjects receive significant benefit. The ''critical'' rate seems to be dependent on the subject. In our study, all patients achieved maximum performance with a rate of 2100 pulses/s.

2. Multi-talker vowels

The results for vowel recognition, scored in terms of percent correct, are given in Fig. 2(b). The mean vowel scores for the rates 400, 800, 1400, and 2100 pulses/s were

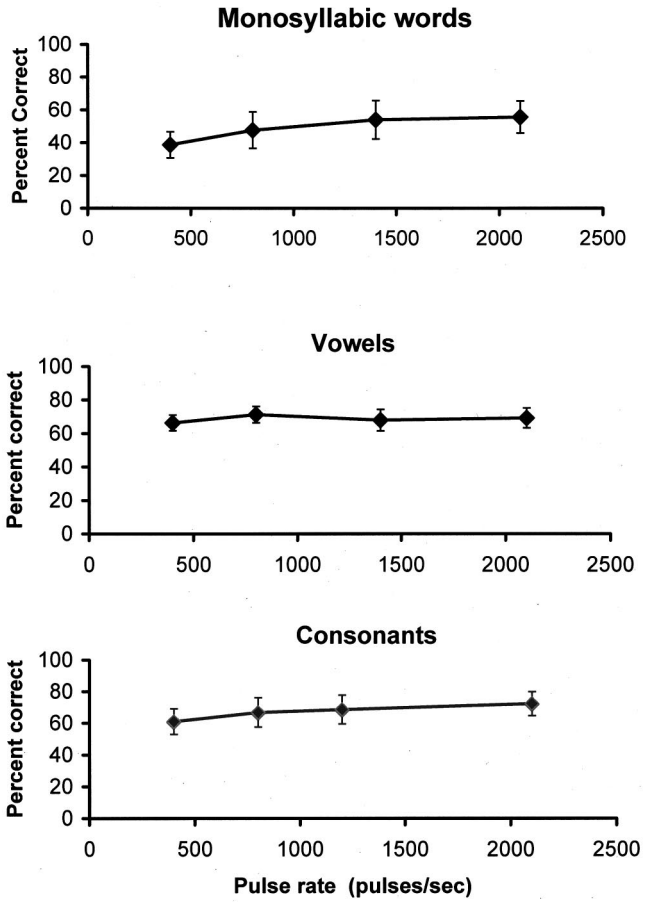


FIG. 2. Speech recognition as a function of stimulation rate. The word and consonant recognition scores obtained at 2100 pps were significantly higher than the corresponding scores obtained at 800 or 400 pps. Higher stimulation rates did not provide any benefits for vowel recognition. Error bars indicate standard errors of the mean.

66.3%, 71.3%, 68%, and 69.3%, respectively. Repeated measures analysis of variance indicated no significant main effect [$F(3,15)=1.623$, $p=0.226$] of rate on vowel recognition. The finding that rate does not seem to affect vowel recognition is not surprising, since higher rates do not improve spectral resolution, which is needed for the perception of vowels. Vowels are characterized by slowly changing formant transitions which can be adequately captured even with slow

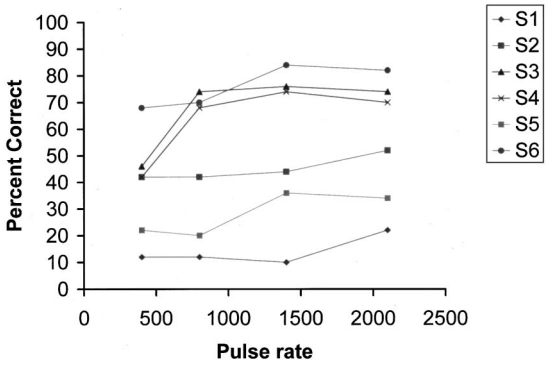


FIG. 3. Individual subject's performance on word recognition as a function of stimulation rate.

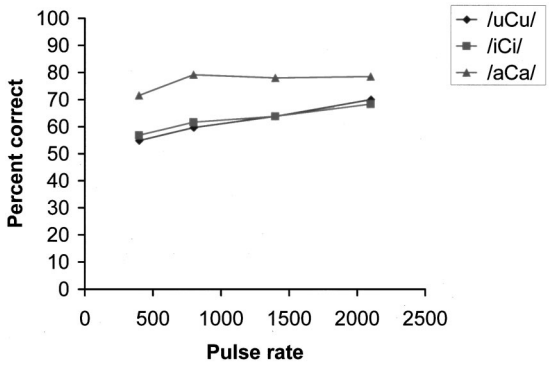


FIG. 4. Consonant recognition as a function of stimulation rate for three different vowel contexts. The standard errors of the mean for /uCu/ recognition were 9.7%, 10.7%, 12.3%, and 10.3% for the rates of 400, 800, 1400, and 2100 pps, respectively. The standard errors of the mean for /iCi/ recognition were 8.3%, 9.3%, 8.2%, and 6.0%, and for /aCa/ recognition were 7.3%, 8.5%, 7.1%, and 7.2%.

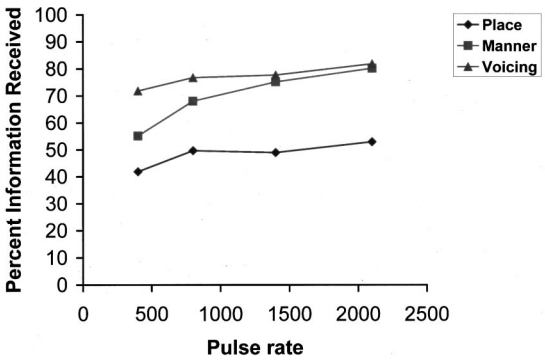


FIG. 5. Mean percent information received on the features of voicing, manner, and place of articulation as a function of stimulation rate. The standard errors of the mean for ''place'' were 8.6%, 10.1%, 5.5%, and 10.2% for the stimulation rates of 400, 800, 1400, and 2100 pps, respectively. The standard errors of the mean for ''manner'' were 9.9%, 11.7%, 5.5%, and 8.6%, and for ''voicing'' were 7.8%, 8.2%, 4.5%, and 6.9%.

rates of stimulation. Higher rates affect the temporal representation of the speech envelopes which is more important for the perception of consonants.

3. Consonants

The results on consonant recognition, scored in terms of percent correct, are given in Fig. 2(c). Repeated measures analysis of variance indicated a significant main effect $[F(3,15)=12.273,\ p<0.0001]$ of rate on consonant recognition. *Post hoc* analysis (Scheffe's test) showed a significant $(p=0.05)$ difference between the mean scores obtained at 2100 and 800 pulses/s, and a significant $(p=0.022)$ difference between the scores at 800 and 400 pulses/s. There was no significant difference between the scores obtained at 1400 and 2100 pulses/s.

To examine the benefits of high-rate stimulation on multi-vowel consonant recognition, we analyzed the results for each vowel context separately. Figure 4 shows the mean scores on consonant recognition for each of the three vowel contexts. As can be seen, the effect of rate on consonant recognition is highly dependent on the vowel context. Higher rates of stimulation seem to benefit mostly consonants in the /iCi/ and /uCu/ contexts. The effect of higher rates of stimulation on the recognition of consonants in the /aCa/ context was small. Interestingly enough, the /aCa/ consonant test is probably the most widely used test today in the speech community. Yet, this test is not sensitive to parametric variations, such as pulse rate, of CIS processors.

We also analyzed the consonant confusion matrices using information transmission analysis (Miller and Nicely, 1955). The consonant features ''manner of articulation,'' ''place of articulation,'' and ''voicing'' were computed and scored in terms of percent information transferred. The results from the feature analysis are shown in Fig. 5. Large (monotonic) improvements in ''manner'' were obtained as the pulse rate increased, while moderate improvements in ''place'' were obtained. There was a small improvement in ''voicing.'' Repeated measures analysis of variance showed a significant main effect of rate on manner $[F(3,15)=22.2,\ p<0.0005]$, a significant main effect of rate on place $[F(3,15)=5.38,\ p=0.01]$, and a nonsignificant effect of rate on voicing $[F(3,15)=2.26,\ p=0.123]$. *Post hoc* tests showed that the manner scores reached an asymptote at 1400 pulses/s. Place scores improved and reached a plateau at 800 pulses/s, however, the improvement from 400 to 800 pps was nonsignificant $(p=0.06)$.

Given the large improvements obtained in ''manner'' with higher rates of stimulation, we decided to analyze the confusion matrices further in terms of consonant class identification. We wanted to know whether higher rates of stimulation improve, say, stop, fricative or semivowel identification. The 20 /vCv/ consonants were divided into five consonant classes: stops /b p d t g k/, fricatives /f v s z ʃ ð/, affricates /tʃ dʒ/, nasals /m n/, and semivowels /r l j w/, and scored in terms of percentage of consonants identified correctly within each class. The mean scores for each vowel context are given in Fig. 6. As shown, consonant class (manner) identification improves with higher rates of stimulation and that improvement seems to be dependent on the vowel context. In the /aCa/ context, for instance, stop and affricate identification does not seem to be affected by the rate of stimulation. In contrast, in the /iCi/ context, there were large improvements in stop and affricate identification with increasing rate of stimulation. There was improvement in fricative identification in all vowel contexts with increasing rate of stimulation. For the /aCa/ and /iCi/ contexts, fricative identification reached an asymptote at 800 pulses/s, whereas for the /uCu/ context fricative identification kept increasing even up to 2100 pulses/s. There was also a large improvement in nasal discrimination in the /uCu/ context. In summary, the identification of stops, fricatives, affricates, and nasals improved with higher rates of stimulation. The improvement was more evident in the /uCu/ and /iCi/ contexts.

The above results demonstrated that higher rates of stimulation benefit consonant identification. This benefit stemmed largely from improved manner identification. We believe that this is because higher rates improve the temporal representation of the speech envelopes, thereby accentuating temporal cues that are important for consonant perception (Fig. 7). Envelope information is known to cue manner distinction. Stops, for instance, are characterized by a period of

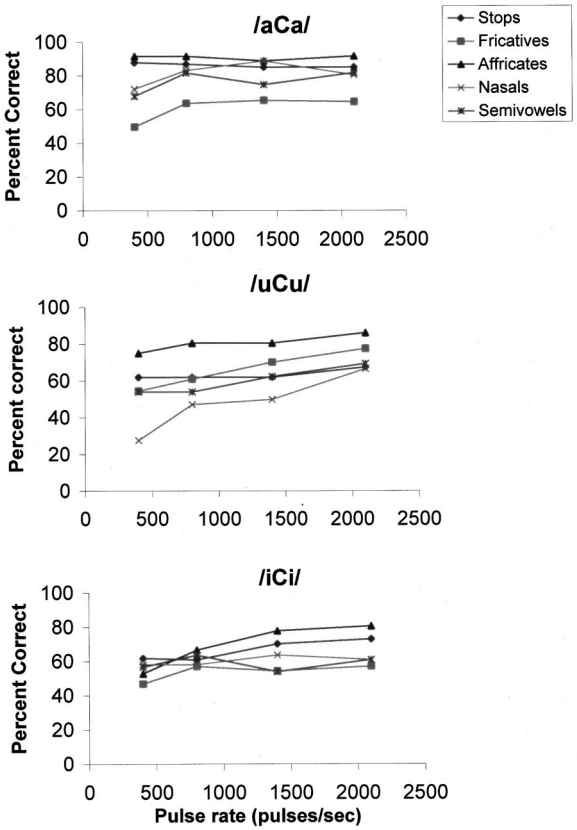


FIG. 6. Consonant class identification as a function of stimulation rate for three vowel contexts.

silence (closure) prior to signal onset. In contrast, nasals, semivowels, and fricatives lack that distinctive interval of silence before signal onset. In addition, nasals, fricatives, and semivowels are marked with relatively larger envelope amplitudes (containing more energy) during the release compared to the stop consonants. Therefore, assuming that implant devices transmit enough temporal information, the stop consonants should not be confused with fricatives, nasals, or semivowels. Yet, the low scores in manner identification obtained with low rates of stimulation suggest that some patients did confuse stops with fricatives, or stops with nasals, etc. Higher rates improved the temporal representation of the speech envelopes (see the example in Fig. 7), which in turn improved manner identification.

It is also important to note that since the speech envelopes are not sampled very often at the lower rates, it is possible that certain short-duration segments (e.g., burst) of the speech waveform may be missed altogether. This is illustrated in Fig. 7, which shows the pulsatile waveforms of the syllable /t i/ obtained at different rates. As shown in Fig. 7, the unvoiced stop consonant /t/ is marked by a period of silence (closure), followed by a burst and aspiration. As the pulse rate increases, the burst becomes more distinctive, and perhaps more salient perceptually. There seems to be no evidence of the burst at low rates, 200 or 400 pulses/s. This example clearly demonstrates that lower rates do not provide a good, if any at all, temporal representation of the burst in stop consonants. We believe that this is the reason we observed larger improvements in stop-consonant identification in the /iCi/ context than in the /aCa/ context with higher rates of stimulation. It is known from the speech perception literature that the burst is the dominant cue for the perception of stops in front-vowel environments (Smits *et al.*, 1996; Dorman *et al.*, 1977). In contrast, in the /a/ context the formant transitions are perceptually the most important for stop identification.

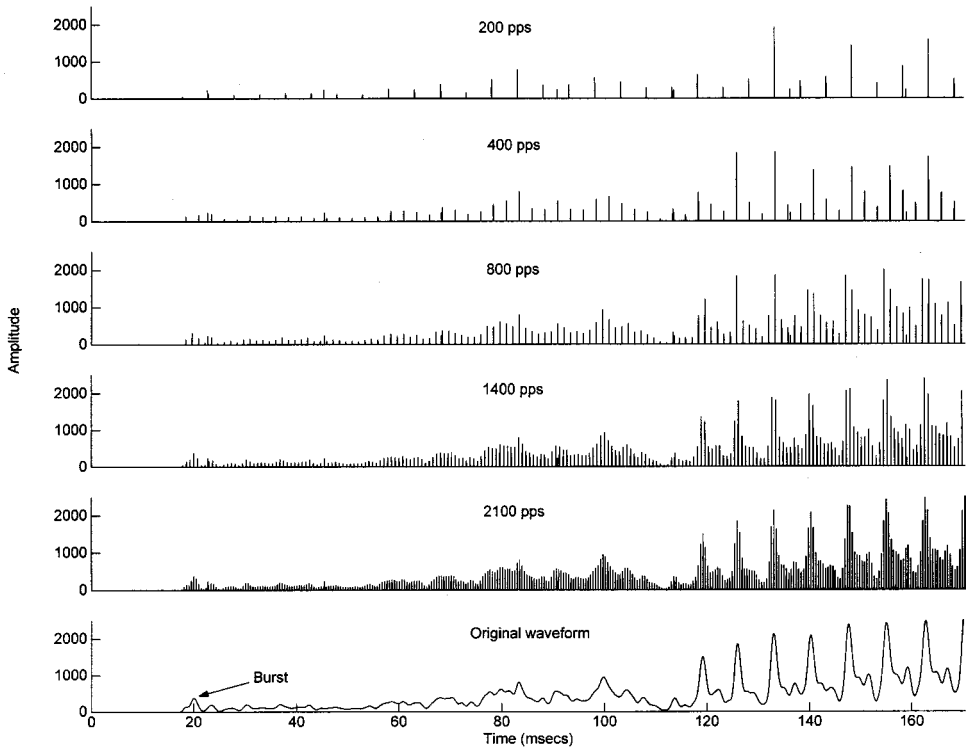


FIG. 7. The pulsatile waveforms for channel 5 of the syllable /t i/ obtained at five different stimulation rates. These waveforms were obtained by bandpass filtering the syllable /t i/ into six channels, performing envelope detection, and sampling the rectified envelopes at the rates indicated. Only the waveforms for channel 5 (with a center frequency of 3316 Hz) are shown. The bottom panel shows the original speech envelopes of channel 5. This figure shows the effect of stimulation rate in detecting short-duration segments (e.g., burst) of speech. As the pulse rate increases, the burst becomes more distinctive, and perhaps more salient perceptually.

4. Comparison with previous studies

Our findings on the effect of pulse rate on speech understanding are consistent with those by Wilson and colleagues (Wilson *et al.*, 1994; Lawson *et al.*, 1996). Lawson *et al.* (1996) reported that five Nucleus-22 patients with percutaneous access, who were fitted with a six-channel CIS processor, obtained a maximum performance on consonant recognition with a pulse rate of 833 pulses/s. There was a significant increase in performance when the rate increased from 250 to 833 pps. There were small, but not significant, differences in performance when the rate increased from 833 to 2525 pps. We suspect that the reason that Lawson *et al.* (1996) did not observe further improvements in performance for rates above 833 pps is because they used medial consonants in the /a/ context for testing. As discussed earlier, the /aCa/ consonant test is not sensitive enough to parametric variations of CIS processors.

Lawson *et al.* (1996) also compared the performance of the SPEAK strategy with a high-rate *n*-of-*m* strategy. Consonant recognition performance obtained with the high-rate (833 pps) 6-of-18 strategy was found to be significantly higher than the performance obtained with the 250-pps SPEAK strategy. The benefits in using high rates of stimulation with the CIS strategy compared to the relatively lower rates with the SPEAK strategy has been demonstrated by others in between-subjects comparisons (e.g., Kiefer *et al.*, 1996; Loizou *et al.*, 1997). Loizou *et al.* (1997) compared the performance of 11 Nucleus-22 patients using the SPEAK strategy with the performance of 7 Med-El/CIS-Link patients using the CIS strategy on consonant recognition in quiet and in +5-, +10-, and +15-dB S/N noise. Performance on consonant recognition was significantly higher with the high-rate CIS strategy than the SPEAK strategy in both quiet and noisy conditions. Feature analysis revealed that the information transmitted in manner and voicing was significantly higher with the CIS strategy. This is consistent with our findings in this study of improved manner identification with higher rates of stimulation.

Kiefer *et al.* (1997) also noted similar improvements in speech understanding with higher rates of stimulation for 13 patients using Med-El's COMBI-40 and COMBI-40+ devices. With eight channels active, Kiefer *et al.* (1997) examined vowel, consonant, and monosyllabic word recognition as a function of pulse rate (600, 1200, and 1500 pps). They found a statistically significant difference between the mean scores for monosyllabic word and consonant recognition at 600 and 1500 pps. There was no statistically significant difference between the mean vowel recognition scores at 1500 and 600 pps; an outcome consistent with our findings on vowel recognition. Brill *et al.* (1997) also examined the effect of stimulation rate for four subjects using Med-El's COMBI-40+ device, with only four channels active. Only four (out of 12) channels were activated in their study in order to examine extremely high rates of stimulation. Significant differences in word recognition were found between stimulation rates of 4545 and 800 pps.

Fu and Shannon (2000) recently investigated the effect of stimulation rate in six Nucleus-22 implant listeners fitted with a four-channel CIS processor. They varied the stimulation rate from 50 pulses/s to a maximum of 500 pulses/s and examined vowel and consonant recognition. Their results showed no effect on phoneme recognition for rates between 150 to 500 pulses/s, and a significant decrement in performance for rates lower than 150 pulses/s. Our results cannot be compared with theirs, since the maximum rate of stimulation in their study was constrained, due to hardware limitations, to rates lower than 500 pulses/s. In our study, the stimulation rate of 400 pulses/s was the lowest rate examined.

II. EXPERIMENT 2: EFFECT OF PULSE DURATION

In nonsimultaneous pulsatile stimulation, the pulse rate is directly related to the pulse duration. In general, short-duration pulses allow high rates of stimulation, and the smaller the pulse duration, the higher the rate. Therefore, in order to implement high-rate processors short-duration pulses need to be used. In contrast, low-rate processors can be implemented either by using short duration pulses and inserting an appropriate temporal gap between consecutive cycles (as in experiment 1), or by widening the pulse width. For instance, the 400-pps processor can be implemented either by using 40-μs/phase duration pulses and inserting a 2020-μs temporal gap between consecutive cycles, or by using 208-μs/phase pulses with no temporal gap between consecutive cycles. In both implementations the rate is the same; however, the pulse widths are different. This raises the question then: ''Do wider pulses offer any advantages over short pulses for a fixed rate?'' It is known from psychophysics studies (Shannon, 1993) that as the pulse duration increases, the thresholds decrease and the dynamic range increases. Given the increase in dynamic range associated with wider pulses, it is therefore of question as to whether wider pulses might be more beneficial for speech recognition for low-rate implementations of the CIS strategy. This question is addressed in this experiment by implementing 400- and 800-pps processors using wide and short duration pulses.

A. Method

1. Subjects

The subjects were the same as in experiment 1. Five of the six subjects were available for this experiment.

2. Speech material

The same /vCv/ consonants and multi-talker vowels of experiment 1 were used.

3. Procedure

The thresholds and most-comfortable levels were measured for the 104 and 208 μs/phase pulses using the ascending and descending methods of limits. We measured most-comfortable levels with an ascending method of limits, asking the subjects to tell us when the stimulus levels became most comfortable to them. Trains of biphasic pulses were presented in 50-ms bursts with a 500-ms interval between bursts.

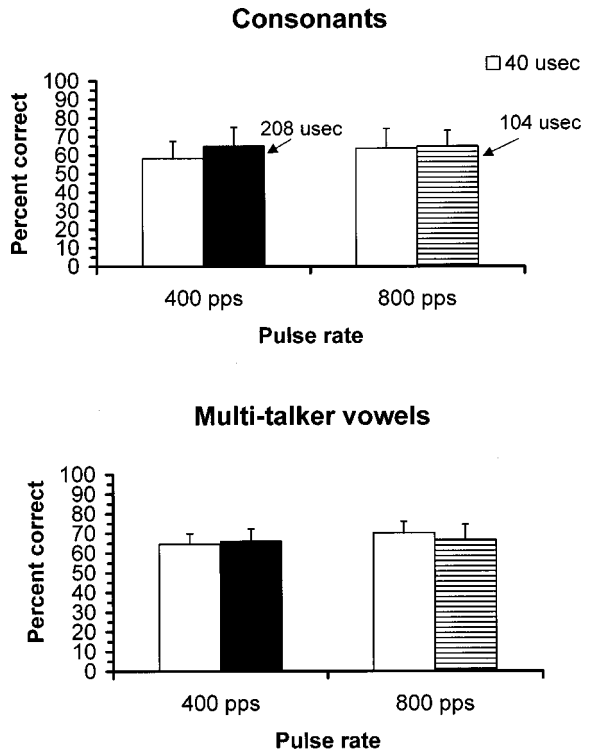


FIG. 8. Vowel and consonant recognition for 400- and 800-pps processors using short and wide pulse widths. Error bars indicate standard errors of the mean.

B. Results and discussion

The mean vowel and consonant scores for the two pulse-width conditions (208 and 104 μs/phase) are shown in Fig. 8, and are compared with the corresponding 400 and 800 rate conditions obtained using 40-μs/phase pulses in experiment 1. A two-factor (rate and pulse width) ANOVA with repeated measures showed a significant main effect of pulse width [$F(1,4)=10.4$, $p<0.05$], a nonsignificant effect of pulse rate [$F(1,4)=2.7$, $p=0.174$], and no interaction between pulse rate and pulse width [$F(1,4)=2.8$, $p=0.169$] on consonant recognition. *Post hoc* tests performed on the 400-pulses/s rate condition showed that wider pulses produced a significantly ($p=0.02$) higher score on consonant recognition. There was no significant difference between the scores obtained for the 800-pulses/s condition. A two-factor ANOVA with repeated measures showed no main effect of pulse width [$F(1,4)=10.4$, $p<0.05$], no effect of pulse rate [$F(1,4)=2.7$, $p=0.174$], and no interaction between pulse rate and pulse width [$F(1,4)=2.8$, $p=0.169$] on vowel recognition.

The above results demonstrated an advantage in using wider pulses at low rates for consonant recognition. For the same rate of stimulation (400 pps) conveying the same amount of temporal-envelope information within each channel, wider pulses (208 μs/phase) yielded a significantly higher performance on consonant recognition than short-duration (40 μs/phase) pulses. We suspect that this benefit is partly due to the increased dynamic range associated with wider pulses (Shannon, 1993). This outcome is in agreement with the findings by Loizou *et al.* (2000) about the significant effect of reduced dynamic range on speech recognition. Feature analysis of the consonant confusion matrices showed that the dynamic range affected mostly the reception of place information.

In this study, we only explored the effect of two different pulse widths per given rate. Other combinations of pulse rate and pulse duration are possible. In fact, one may construct a two-dimensional parameter space, in which the first dimension represents pulse duration and the second dimension represents pulse rate, and find the pulse rate–width combination that will yield the highest performance. Wilson and colleagues (Wilson *et al.*, 1993; Lawson *et al.*, 1993) performed such a two-dimensional parametric study of pulse rate and pulse duration. They reported an improvement in consonant recognition for 400-pps processors when wider pulses were used (Lawson *et al.*, 1993). For a four-channel processor, consonant recognition improved from 76% correct using 33-μs/phase pulses to 89% correct using 100-μs/phase pulses. For a six-channel processor consonant recognition improved from 85% correct with 33-μs/phase pulses to 90% correct with 200-μs/phase pulses. These results are consistent with our findings, in that wider pulses benefit consonant recognition for low-rate processors. The optimal combination of rate and pulse width, however, seemed to be dependent on the speaker (male versus female voice) as well as on the subject (Lawson *et al.*, 1993; Wilson *et al.*, 1993).

III. EXPERIMENT 3: EFFECT OF FILTER OVERLAP

The envelope signals in implant processors are typically extracted by processing the acoustic signal through a bank of bandpass filters spanning the signal bandwidth. The filter order (i.e., the number of filter coefficients) of the bandpass filter affects the filter roll-off, and consequently the overlap between contiguous filters. Higher-order filters have a steep roll-off, hence smaller overlap between adjacent channels. Conversely, lower-order filters are characterized by a shallow roll-off and therefore larger overlap between adjacent channels. A large filter overlap affects the spectral representation of the signal since it smears spectral information. In fact, broad (overlapping) filters have been used in the past to simulate broadened auditory filters in impaired hearing (e.g., Moore *et al.*, 1992) or channel interaction in cochlear implants (e.g., Shannon *et al.*, 1998). In this experiment we investigated the effect of spectral smearing on consonant identification by systematically varying the filter order to produce filters with varying degrees of overlap.

A. Subjects and test materials

Four implant subjects participated in this experiment. The same /vCv/ consonants in experiment 1 were used.

B. Signal processing and procedure

The CIS implementation was the same as in experiment 1. The only parameter that changed in this experiment was the filter order. Three different filter orders, fourth, eighth, and tenth, were investigated [sixth-order filters were used in experiment 1]. All the filters were Butterworth, and the roll-off of the fourth, sixth, eighth, and tenth filters was -20, -30, -45, and -60 dB/oct, respectively.

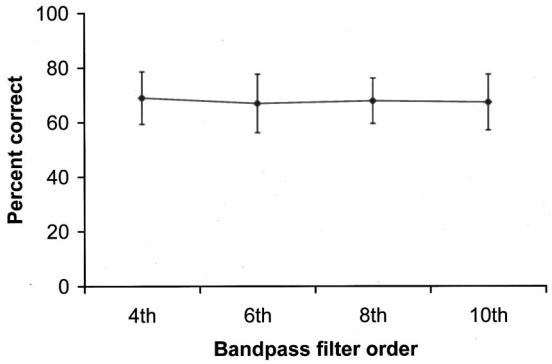


FIG. 9. Consonant recognition as a function of bandpass filter order. Error bars indicate standard errors of the mean.

The test was divided into three sessions, one for each filter condition. The three conditions were counterbalanced across subjects. All test sessions were preceded by one practice session in which the identity of the consonants was indicated to the listeners. The stimuli were presented directly to the subjects, through our laboratory processor, at a comfortable listening level. The pulse width was fixed at 40 μs/phase in all filter conditions yielding a stimulation rate of 2100 pulses/s.

C. Results and discussion

The mean consonant scores as a function of the filter order are shown in Fig. 9 and are compared with the scores obtained in Experiment 1 using sixth-order filters. As can be seen, the filter overlap seems to have no effect on consonant recognition. Repeated measures analysis of variance indicated no significant main effect of filter order on consonant recognition [$F(3,9)=0.276$, $p=0.841$].

Shannon *et al.* (1998), using noise-band simulations, did a similar experiment investigating the effect of filter overlap. Speech material was processed through four bandpass filters with varying degrees of overlap, and the envelope of each speech band was extracted through rectification and low-pass filtering. The four envelopes were modulated with band-limited noise, recombined, and presented to normal-hearing listeners for identification. Increasing the filter overlap produced a significant effect on consonant and sentence recognition; however, the absolute level of performance remained quite high even with -18-dB/oct filters. This is in agreement with our results obtained using -20-dB/oct filters (fourth order). The mean consonant recognition scores remained high even with -20-dB/oct filters.

The findings of this experiment have implications for the implementation of the CIS strategy. The filter order seemed to have no significant effect on consonant recognition, and the subjects reported no degradation in speech quality or timbre with lower-order filters. This suggests that one could implement the CIS strategy using fourth-order filters without sacrificing performance. This could reduce considerably the amount of computation required per cycle, and consequently reduce battery power.

This experiment has only investigated the effect of filter overlap on consonant recognition. It may well be that the amount of filter overlap is more important for vowel recognition. One would expect that steeper filters would be better in capturing formant information, particularly when the formants are very close to each other. Further studies are therefore needed to investigate the effect of filter overlap on vowel recognition.

IV. EXPERIMENT 4: EFFECT OF AMPLITUDE MAPPING FUNCTION

A major concern in designing cochlear implant processors is in the proper transformation of acoustic amplitude to electrical amplitude. Speech sounds in a normal conversation can range from 40 to 60 dB; however, implant listeners have only a dynamic range of 6–20 dB in electrical current. To accommodate for the smaller electrical dynamic range, the acoustic amplitudes are typically compressed using a logarithmic or power-law transformation. Logarithmic compression has been used because it has been shown to match loudness between electrical and acoustical amplitudes (Eddington *et al.*, 1978; Dorman *et al.*, 1993). Although the choice of compression function is important, only a few studies (e.g., Fu and Shannon, 1998; Wilson *et al.*, 1999) were performed to investigate the effect of the shape of the compression function on speech recognition.

The present study investigated the importance of the shape of the amplitude mapping function on consonant recognition. A range of amplitude mapping functions was created, from a strongly compressive function to a weakly—almost linear—compressive function, by varying the exponents of the power-law transformation [Eq. (1)]. Consonant recognition was examined as a function of the power exponent.

A. Subjects and test materials

Four implant subjects participated in this experiment. The same /vCv/ consonants in experiment 1 were used.

B. Signal processing and procedure

The CIS implementation was the same as in experiment 1. The pulse width was fixed at 40 μs/phase for all mapping conditions yielding a stimulation rate of 2100 pulses/s. The only parameter that changed in this experiment was the power exponent, p, used in Eq. (1). The exponent of the power function was systematically changed to -0.1, 0.2, and 0.6. Figure 10 shows the amplitude mapping functions obtained using different power exponents. The most compressive function was obtained using $p=-0.1$ and the least compressive (nearly linear) function was obtained using $p=0.6$. The function obtained using $p=-0.0001$ was logarithmic, and was the default exponent value used in our CIS implementation (experiment 1).

The test was divided into three sessions, one for each mapping condition. The three conditions were counterbalanced across subjects. All test sessions were preceded by one practice session in which the identity of the consonants was indicated to the listeners. The stimuli were presented directly to the subjects, through our laboratory processor, at a comfortable listening level.

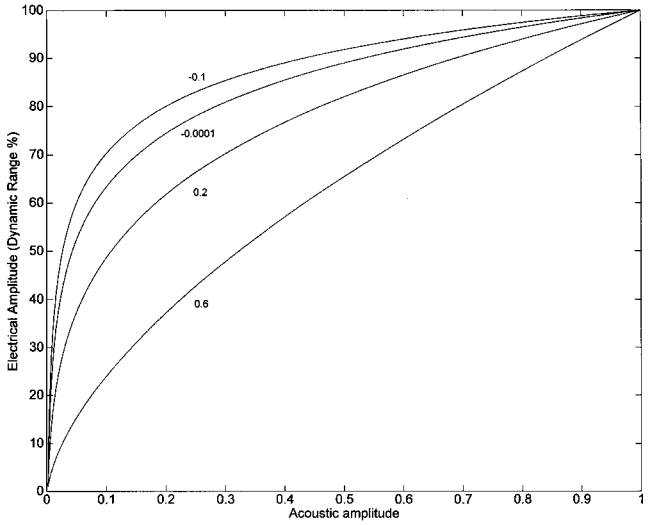


FIG. 10. The set of amplitude mapping functions used in this study.

C. Results and discussion

The results are shown in Fig. 11. Performance remained constant at about 70% correct when the value of the power exponent ranged from -0.1 to 0.2, and dropped to 40% correct when $p=0.6$. Repeated measures ANOVA showed a significant main effect $[F(3,9)=27.335,\ p<0.0005]$ of power exponent on consonant recognition. *Post hoc* tests showed no significant differences ($p>0.5$) between the scores obtained with $p=-0.0001$ (from experiment 1), $p=-0.1$, and $p=0.2$. There was, however, a significant difference ($p<0.05$) between the scores obtained with $p=0.2$ and $p=0.6$.

The present results indicate only a mild dependence of performance on the value of the power exponent, and consequently the shape of the amplitude mapping function, over a broad range of exponents. These findings are consistent with those by Boëx *et al.* (1995), Fu and Shannon (1998), Wilson *et al.* (1999), and Zeng and Galvin (1999). Boëx *et al.* (1995, 1997) investigated the use of mapping functions that preserved normal loudness growth in cochlear implants. They fitted four subjects with wearable Geneva/RTI/MEEI processors, and they reported that two subjects preferred and performed better with normal-loudness growth functions than the standard logarithmic functions. Fu and Shannon (1998) varied the exponent of the power-law mapping function from a strongly compressive ($p=0.05$) to a weakly compressive value ($p=0.75$) and investigated the effect of compression on phoneme recognition using 3 Nucleus-22 implant users fitted with a four-channel CIS processor. Their study showed that the function relating power exponent and performance was relatively flat, suggesting that phoneme recognition was only mildly affected by the shape of the compression function, at least for power exponents $p\leqslant 0.5$. There was a significant drop in performance when the mapping function was nearly linear ($p=0.75$), consistent with our findings. Best performance was obtained for power exponents that restored normal loudness growth. Wilson *et al.* (1999) recently replicated the study by Fu and Shannon (1998) and also found a relatively uniform performance on consonant recognition over a wide range of exponent values for two subjects wearing the Med-El/CIS-link device. Vowel recognition was hardly affected by the different manipulations in mapping functions. Zeng and Galvin (1999) reported similar findings when they varied the amount of compression by manipulating the Q-value in the Nucleus-22 processor with four implant listeners. The Q-values used (20, 30, 40, and 50) corresponded to power exponents 0.24, 0.4, 0.51, and 0.63, respectively. Vowel and consonant recognition were not significantly affected by the amount of compression.

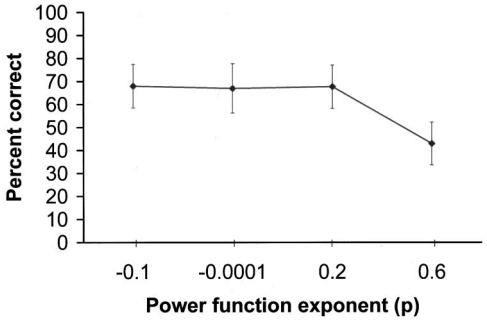


FIG. 11. Consonant recognition as a function of the power exponent of the mapping function. Error bars indicate standard errors of the mean.

The above experiments were performed in quiet. A different pattern of performance was observed for experiments conducted in noise. Results from a single subject in the Wilson *et al.* (1999) study indicated that less compressive mapping functions might be more helpful for listening to speech in noise. For consonants in +15-dB S/N speech-shaped noise, the power exponent of 0.4 produced the highest score. Fu and Shannon (1999) observed a different outcome in a similar study performed in noise. Maximum consonant performance in both quiet and noisy conditions was obtained with the exponent of 0.2 corresponding to logarithmic mapping. The overall effect of the amplitude mapping function, however, was different in noise. Performance with weakly compressive mapping functions declined mildly in noise, whereas performance with strongly compressive amplitude mapping declined dramatically in noise. In contrast to the above two studies, Zeng and Galvin (1999) found no significant effect of compression in noise, for consonant and vowel recognition.

In summary, our results and those by others confirm that in quiet logarithmic mapping functions are needed for accurate consonant and vowel recognition. More studies are needed, however, to investigate how the acoustic signal mapping should be done in noise. It may well be that different mapping functions are needed for noisy conditions.

V. EXPERIMENT 5: EFFECT OF SIGNAL BANDWIDTH

The signal bandwidth is constrained, according to Nyquist's theorem, to half the sampling frequency. Accordingly, the higher the sampling frequency, the higher the signal bandwidth. Although a high bandwidth (>10 kHz) is needed for music perception (and enjoyment), a considerably smaller bandwidth is needed for speech perception. For vowels, a bandwidth of 3 kHz is needed for the perception of the first three formants (Peterson and Barney, 1952; Hillenbrand *et al.*, 1995). A relatively higher bandwidth is needed for the

perception of consonants, such as fricatives and stops. The spectral peaks of /f/ and /s/, for instance, could be as high as 8.5 kHz (Manrique and Massone, 1981). Limiting the spectrum to a small bandwidth, such as the 3.2-kHz telephone bandwidth, for instance, can potentially produce consonant confusions between f/s, t/k, etc. It seems reasonable, then, to expect that a wide bandwidth might be beneficial for consonant recognition.

Indeed, recent experiments by Zerbi *et al.* (1998) showed that a wide bandwidth was beneficial for the perception of consonants, and, in particular, consonants produced by female talkers. Significant improvements on consonant recognition were obtained when the bandwidth increased from 5.5 to 9.5 kHz. Zerbi *et al.* (1998) found that the addition of a high-frequency band spanning from 5.6 to 9.5 kHz helped reduce principal confusions (e.g., t/k, f/s, p/t) among medial consonants uttered by a female talker. No significant improvements were found, however, for consonants produced by a male talker.

The two bandwidths (5.5 and 9.5 kHz) examined in the Zerbi *et al.* study were one octave apart, hence it was not clear whether any bandwidth in between would improve consonant recognition. We therefore investigated two additional bandwidths, in the present experiment, in the range of 5.5 to 9.9 kHz. The results on the 6.7-kHz bandwidth were reported in experiment 1, and are included here for comparative purposes. The two additional bandwidths examined were 8.4 and 9.9 kHz. At question in this experiment was whether a high ($>$ 5.5 kHz) signal bandwidth would benefit consonant and/or vowel recognition.

A. Subjects and test materials

Five implant subjects participated in this experiment. The same /vCv/ consonants and multi-talker vowels of experiment 1 were used.

B. Signal processing and procedure

The CIS implementation was the same as in experiment 1. Two different antialiasing filters were used to limit the signal bandwidth to 8.4 and 9.9 kHz, respectively. The sampling frequency was fixed at 22 kHz for both bandwidth conditions. The center frequencies of the six analysis filters were obtained by dividing the bandwidth into six logarithmically equal bands. The center frequencies for the 8.4-kHz bandwidth were 472, 803, 1365, 2320, 3943, and 6703 Hz. The center frequencies for the 9.9-kHz bandwidth were 481, 839, 1464, 2555, 4460, and 7786 Hz. Sixth-order Butterworth filters were used.

The test was divided into two sessions, one session for each bandwidth condition. The two conditions were counterbalanced across subjects. All test sessions were preceded by one practice session in which the identity of the consonants/vowels was indicated to the listeners. The stimuli were presented directly to the subjects, through our laboratory processor, at a comfortable listening level. The pulse width was fixed at 40 μs/phase in all bandwidth conditions yielding a stimulation rate of 2100 pulses/s.

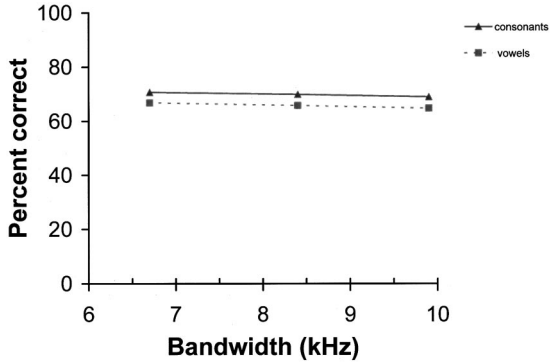


FIG. 12. Mean vowel and consonant recognition performance as a function of signal bandwidth. The standard errors of the mean for the consonants were 9.1%, 8.9%, and 7.3% for the bandwidths of 6.7, 8.4, and 9.9 kHz, respectively. The standard errors of the mean for the vowels were 6.5%, 9.1%, and 7.3% for the bandwidths of 6.7, 8.4, and 9.9 kHz, respectively.

C. Results and discussion

The results are shown in Fig. 12. Performance on consonant recognition remained constant at roughly 70% correct for all three bandwidth conditions, while performance on vowel recognition remained constant at roughly 65% correct. Repeated measures ANOVA showed no significant main effect $[F(2,8)=0.53,\ p=0.608]$ of signal bandwidth on consonant recognition, and no significant main effect $[F(2,8)=0.54,\ p=0.5]$ on vowel recognition.

Our results indicate that the recognition of consonants and vowels is not affected by the signal bandwidth, at least in quiet and for bandwidths in the range of 6.7–9.9 kHz. Zerbi *et al.* (1998) found a significant difference between the performance on consonant recognition obtained at 5.5 and 9.5 kHz. In the present study we showed that a 6.7-kHz bandwidth is sufficient for accurate consonant recognition. Using a higher ($>$6.7 kHz) bandwidth did not seem to benefit consonant recognition. We suspect that this is because the number of filters allotted for high frequencies was not increased when the bandwidth increased. In all three conditions, two filters were allotted for frequencies above 3 kHz. One could conceivably choose a different filter allocation allowing more filters in the high frequencies. Such an arrangement, however, would most likely impair vowel recognition where more filters are needed to capture low-frequency information in the F1–F2 region.

VI. CONCLUSIONS

(i) Of the five speech processing parameters examined in this study, the pulse rate and the pulse width had the largest (positive) effect on speech recognition. For a fixed pulse width, higher rates (2100 pps) of stimulation provided significantly more benefit on speech recognition compared to lower rates (400 pps) of stimulation, consistent with the findings by Lawson *et al.* (1996), Brill *et al.* (1997), and Kiefer *et al.* (1997). When we jointly varied the pulse rate and the pulse width, we found that the pulse width had a significant effect on consonant recognition, at least for low-rate (400 pps) processors. This interaction between pulse rate and pulse width suggests that it is

possible to achieve high levels of consonant recognition by a slower pulse rate (400 ps) and an appropriate choice of pulse width. Audiologists or clinicians can therefore optimize a subject's performance either by (1) using a short-duration pulse (say, 40 μs/phase) and a high pulse rate (as per experiment 1), or (2) by jointly varying the pulse width and pulse rate (as per experiment 2). Of the two methods that can be used to optimize implant listener's performance, the former is less time consuming.

(ii) The finding that the 2100-pps processors produced a significantly higher performance on open-set recognition than the 800-pps processors suggests that implant manufacturers ought to support high stimulation rates (>800 pps) to improve the benefits of cochlear implants.

(iii) There was a significant effect of stimulation rate on consonant recognition. Higher rates of stimulation produced a significant benefit to consonant recognition. This benefit stemmed primarily from improved manner identification.

(iv) The effect of stimulation rate on consonant recognition was highly dependent on the vowel context. Higher rates benefited mostly consonants in the /uCu/ and /iCi/ contexts. A small benefit was observed for consonants in the /aCa/ context. This finding strongly recommends the use of multi-vowel consonants for testing parametric variations of implant processors. The /aCa/ consonant test is not sensitive enough to parametric variations of implant processors.

(v) For low-rate (<800 pps) processors, wider pulse widths provided a significantly higher benefit on consonant recognition than narrower pulse widths. We suspect that this is due to the increased dynamic range associated with wider pulses. There was no effect of pulse width on vowel recognition. As demonstrated in this study, as well by others (Wilson *et al.*, 1993; Lawson *et al.*, 1993), it is possible to optimize a subject's performance by varying both pulse rate and pulse width. The best combination of pulse rate and pulse width, however, might be subject dependent.

(vi) The amount of filter overlap seemed to have no effect on consonant recognition. Filters with a -20-dB/oct roll-off produced the same performance as filters with a -60-dB/oct roll-off.

(vii) The shape of the amplitude mapping function had only a minor effect on performance, with the lowest performance obtained when a nearly linear mapping function ($p=0.6$) was used. Performance did not seem to be affected by the value of the power exponents over a broad range of exponents, -0.1 to 0.2, including the exponent -0.0001 corresponding to logarithmic mapping. The present logarithmic mapping function used in current implant devices appears to be a good choice for many subjects, at least in quiet.

(viii) Consonant and vowel recognition was not affected by the signal bandwidth. There was no statistically significant difference in performance between the 6.7-, 8.4-, and 9.9-kHz signal bandwidths.

ACKNOWLEDGMENTS

This research was supported by Grant No. R01 DC03421 from the National Institute of Deafness and other Communication Disorders, NIH.

[1]The power exponent p was set equal to -0.0001 to match the logarithmic compression function, of the form $y=A\log(1+cx)+B$, used in the Med-El device, where A and B are constants used for mapping the acoustic signal x between threshold and most-comfortable level, and c is a constant that controls the shape of the compression function ($c=512$, in our case). It should be noted that the value of the power exponent p corresponding to logarithmic mapping depends on the input range, and, in particular, on the minimum value of x (Xmin). If Xmin=0, then the value of $p=0.2$ will yield a logarithmic mapping similar to the one used in the Med-El device, whereas if Xmin=1, then the value of $p=-0.0001$ will yield a logarithmic mapping. In our implementation, Xmin=1.

Boëx, C., Pelizzone, M., Piloux, V., and Montandon, P. (**1995**). ''Use of loudness scaling measurements to determine compressive mapping in speech processing for cochlear implants,'' Abstracts of 1995 Conference on Implantable Auditory Prostheses, p. 57.

Boëx, C., Eddington, D., Noel, V., Rabinowitz, W., Tierney, J., and Whearty, W. (**1997**). ''Restoration of normal loudness growth for CIS sound coding strategies,'' Abstracts of 1997 Conference on Implantable Auditory Prostheses, p. 26.

Brill, S., Gstottner, W., Helms, J., Ilberg, C. v., Baumgartner, W., Muller, J., and Kiefer, J. (**1997**). ''Optimization of channel number and stimulation rate for the fast CIS strategy in the COMBI 40+,'' Am. J. Otol. **18**, S104–S106.

Dorman, M., and Loizou, P. (**1997**). ''Changes in speech intelligibility as a function of time and signal processing strategy for an Ineraid patient fitted with CIS processors,'' Ear Hear. **18**(2), 147–155.

Dorman, M., Smith, L., and Parkin, J. (**1993**). ''Loudness balance between acoustic and electric stimulation by a patient with a multichannel cochlear implant,'' Ear Hear. **14**, 290–292.

Dorman, M., Studdert-Kennedy, M., and Raphael, L. (**1977**). ''Stop-consonant recognition: Release burst and formant transitions as functionally equivalent, context-dependent cues,'' Percept. Psychophys. **22**, 109–122.

Eddington, D., Dobelle, W., Brachman, D., Mladevosky, M., and Parkin, J. (**1978**). ''Auditory prosthesis research using multiple intracochlear stimulation in man,'' Ann. Otol. Rhinol. Laryngol. **87**(S53), 1–39.

Francois, J., Tinembart, J., Bessat, C., Leone, P., Rossman, F., and Pelizzone, M. (**1994**). ''Implants cochle aires: Un processeur portable pour le developpement de l'algorithme CIS,'' *Actes de le conference DSP 94*, Paris.

Fu, Q-J., and Shannon, R. (**1998**). ''Effect of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners,'' J. Acoust. Soc. Am. **104**, 2570–2577.

Fu, Q-J., and Shannon, R. (**1999**). ''Phoneme recognition by cochlear implant users as a function of signal-to-noise ratio and nonlinear amplitude mapping,'' J. Acoust. Soc. Am. **106**, L18–L23.

Fu, Q-J., and Shannon, R. (**2000**). ''Effect of stimulation rate on phoneme recognition by Nucleus-22 cochlear implant listeners,'' J. Acoust. Soc. Am. **107**, 589–597.

Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

Kiefer, J., Muller, J., Pfennigdorff, T., Schon, F., Helms, J., von Ilberg, C., Baungartner, W., Gstottner, W., Ehrenberger, K., Arnold, W., Stephan, K., Thumfart, W., and Baur, S. (**1996**). ''Speech understanding in quiet and in noise with the CIS speech coding strategy (Med-El Combi 40) compared to the Multipeak and Spectral Peak strategies (Nucleus),'' ORL, **58**, 127–135.

Kiefer, J., Ilberg, C. v., Rupprecht, V., Hubnet-Egener, J., Baumgartner, W., Gstottner, W., Forgasi, K., and Stephan, K. (**1997**). ''Optimized speech understanding with the CIS speech coding strategy in cochlear implants:

the effect of variations in stimulus rate and number of channels,'' Abstracts of Vth International Cochlear Implant Conference, New York, NY.
Lawson, D., Wilson, B., and Zerbi, M. (**1993**). ''Speech Processors for Auditory Prostheses,'' NIH Project N01-DC-2-2401, Second Quarterly Progress Report.
Lawson, D., Wilson, B., Zerbi, M., and Finley, C. (**1996**). ''Speech Processors for Auditory Prostheses,'' NIH Project N01-DC-5-2103, Third Quarterly Progress Report.
Loizou, P. (**1998**). ''Mimicking the human ear: An overview of signal processing techniques for converting sound to electrical signals in cochlear implants,'' IEEE Signal Process. Mag. **15**(5), 101–130.
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.
Loizou, P., Dorman, M., and Powell, V. (**1998**). ''The recognition of vowels produced by men, women, boys and girls by cochlear implant patients using a six-channel CIS processor,'' J. Acoust. Soc. Am. **103**, 1141–1149.
Loizou, P., Graham, S., Dickins, J., Dorman, M., and Poroy, O. (**1997**). ''Comparing the performance of the SPEAK strategy (Spectra 22) and the CIS strategy (Med-El) in quiet and in noise,'' Abstracts of 1997 Conference on Implantable Auditory Prostheses.
Manrique, A., and Massone, M. (**1981**). ''Acoustic analysis and perception of Spanish fricative consonants,'' J. Acoust. Soc. Am. **69**, 1145–1153.
Miller, G., and Nicely, P. (**1955**). ''An analysis of perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–352.
Moore, B., Glasberg, B., and Simpson, A. (**1992**). ''Evaluation of a method of simulating reduced frequency selectivity,'' J. Acoust. Soc. Am. **91**, 3402–3423.
Osberger, M., and Fisher, L. (**1999**). ''SAS-CIS preference study in postlingually deafened adults implanted with the Clarion cochlear implant,'' Ann. Otol. Rhinol. Laryngol. **108**, 74–79.
Peterson, G., and Barney, H. (**1952**). ''Control methods used in a study of vowels,'' J. Acoust. Soc. Am. **24**, 175–184.
Poroy, O., and Loizou, P. (**2000**). ''Development of a speech processor for laboratory experiments with cochlear implant patients,'' *IEEE International Conference on Acoustics Speech and Signal Processing*, Istanbul, Turkey.
Shannon, R. (**1993**). ''Psychophysics,'' in *Cochlear Implants: Audiological Foundatations*, edited by R. Tyler (Singular Publishing Group).
Shannon, R., Zeng, F-G., and Wygonski, J. (**1998**). ''Speech recognition with altered spectral distribution of envelope cues,'' J. Acoust. Soc. Am. **104**, 2467–2476.
Shannon, R., Jensvold, A., Padilla, M., Robert, M., and Wang, X. (**1999**). ''Consonant recordings for speech testing,'' J. Acoust. Soc. Am. **106**, L71–L74.
Smits, R., Bosch, L., and Collier, R. (**1996**). ''Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. I. Perception experiment,'' J. Acoust. Soc. Am. **100**, 3852–3864.
Wilson, B., Lawson, D., and Zerbi, M. (**1993**). ''Speech Processors for Auditory Prostheses,'' NIH Project N01-DC-2-2401, Fifth Quarterly Progress Report.
Wilson, B., Lawson, D., and Zerbi, M. (**1995**). ''Advances in coding strategies for cochlear implants,'' Advances in Otolaryngology–Head and Neck Surgery **9**, 105–129.
Wilson, B., Finley, C., Lawson, D., and Zerbi, M. (**1998**). ''Temporal representations with cochlear implants,'' Am. J. Otology **18** (Suppl.), S30–S34.
Wilson, B., Lawson, D., Finley, C., and Zerbi, M. (**1991**). ''Speech Processors for Auditory Prostheses,'' NIH Project N01-DC-9-2401, Tenth Quarterly Progress Report.
Wilson, B., Lawson, D., Zerbi, M., and Wolford, R. (**1999**). ''Speech Processors for Auditory Prostheses,'' NIH Project N01-DC-8-2105, Third Quarterly Progress Report.
Zeng, F-G., and Galvin, J. (**1999**). ''Amplitude mapping and phoneme recognition in cochlear implant listeners,'' Ear Hear. **20**, 60–74.
Zerbi, M., Lawson, D., and Wilson, B. (**1998**). ''Speech Processors for Auditory Prostheses,'' NIH Project N01-DC-5-2103, Tenth Quarterly Progress Report.
Zwolan, T., Collins, L., and Wakefield, G. (**1997**). ''Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects,'' J. Acoust. Soc. Am. **102**, 3673–3685.

Vibrational mode shapes in Caribbean steelpans. I. Tenor and double second

Thomas D. Rossing, Uwe J. Hansen,[a] and D. Scott Hampton[b]
Physics Department, Northern Illinois University, DeKalb, Illinois 60115

(Received 29 November 1999; accepted for publication 10 May 2000)

Vibrational mode shapes have been studied in several tenor (soprano range) and double-second (alto range) steelpans by three well-known makers. Normal modes are determined from operating deflection shapes recorded by means of electronic TV holography. Vibrational modes of the various note areas are designated by the number of radial and circumferential nodal lines. Tuners generally tune one mode an octave above the fundamental and, if possible, a third mode is tuned either a twelfth or two octaves above the fundamental. Note dimensions follow an approximate scaling law $L = Kf^{2/3}$, and the larger note areas can be roughly modeled as rectangular plates with simply supported edges. Modal shapes in the skirt correspond to standing bending waves propagating around the ring. © *2000 Acoustical Society of America.* [S0001-4966(00)03808-X]

PACS numbers: 43.75.Kk [WJS]

INTRODUCTION

One of the most important musical instruments developed in the 20th century is the Caribbean steelpan. In addition to being the foremost musical instrument in its home country, Trinidad and Tobago, the steelpan is becoming increasingly popular in North America and Europe. The modern family of steelpans covers a five-octave range, and steel bands of today use them to perform calypso, jazz, popular, and classical music.

Most steelpans are hand-crafted from new or used 55-gallon steel drums. Although the steps in the process depend upon the maker, they nearly always include: ''sinking'' the pan with a sledge hammer to create a concave playing surface, scribing the surface to define the note areas, heating the pan over a bonfire to relieve stress, cutting the ''skirt'' to the desired length, and carefully tuning the various note areas. The number of steelpan makers and tuners is growing rapidly; the Pan Page ⟨www.smus.se/musikmuseet/pan/tuners.html⟩ lists 86 makers and tuners worldwide.

Scientific studies on the instrument have not kept pace with the growing popularity of the instrument. The acoustical behavior has been studied by Hampton *et al.* (1987), Rossing *et al.* (1996), Hansen *et al.* (1995), and Kronman (1991); the mechanical features by Achong (1996), Achong and Sinanan-Singh (1997), and Achong (1998) and the metallurgy by Murr *et al.* (1999) and Ferreyra *et al.* (1999).

The purpose of this paper is to report measurements of modal shapes in several tenor and double second steelpans by prominent makers, and to interpret these results in terms of the geometry and fabrication history of these pans. Some preliminary attempts to model the physical behavior of the instruments will be reported.

I. THE STEELPAN

Modern steelpans are known by various names, such as tenor, double second, double tenor, guitar, cello, quadrophonic, and bass. The overlapping ranges of these instruments are shown in Fig. 1. The tenor pan has from 28 to 31 different notes, but each bass pan has only 3 to 5; the note areas increase in size as the frequency decreases. Most steelpans are played with a pair of short wood or aluminum sticks wrapped with surgical rubber; the bass pans are hit with a beater consisting of a sponge rubber ball on a stick.

The pans used in these measurements were constructed and tuned by three leading pan makers, Clifford Alexis (Illinois), Ellie Mannette (West Virginia), and Felix Rohner (Switzerland). Alexis' and Mannette's pans were sunk with a sledgehammer, so that the 18-gauge (1.15-mm thick) low-carbon steel in the playing surface was thinned to about 0.6 mm at the center. Rohner's pans were constructed from steel of about the same final thickness shaped by cold drawing.

Sinking the drum by sledgehammer blows changes its metallurgical structure. Ferreyra *et al.* (1999) found that the Vickers hardness of hammered drums increased from 113 in the undeformed material to 170 after sinking. Heat treatment of the drum increases the hardness by another 5% or more due to strain aging, the amount of change depending upon the carbon content of the steel, among other things. Modal frequencies in the note areas increase 10% to 30% after firing (Rossing *et al.*, 1996).

Rohner, on the other hand, draws the drumhead to shape and then increases the surface hardness of the material by ''nitriding'' it in an atmosphere of hot nitrogen-rich gas, which increases the surface hardness while maintaining a softer interior, a property that makes tuning easier but maintains durability and tuning stability. He has measured the Vickers surface hardness to be 400 or more.

The skirts on the drums were quite different as well. Alexis and Mannette form the skirt by cutting the cylindrical portion of the oil drum to the desired height. Generally this portion of the oil drum is fabricated from 20-gauge steel, having a nominal thickness of 0.9 mm. Rohner's pans, on the

[a]Permanent address: Dept. of Physics, Indiana State University, Terre Haute, IN 47809.
[b]Permanent address: Containerless Research, Evanston, IL 60201.

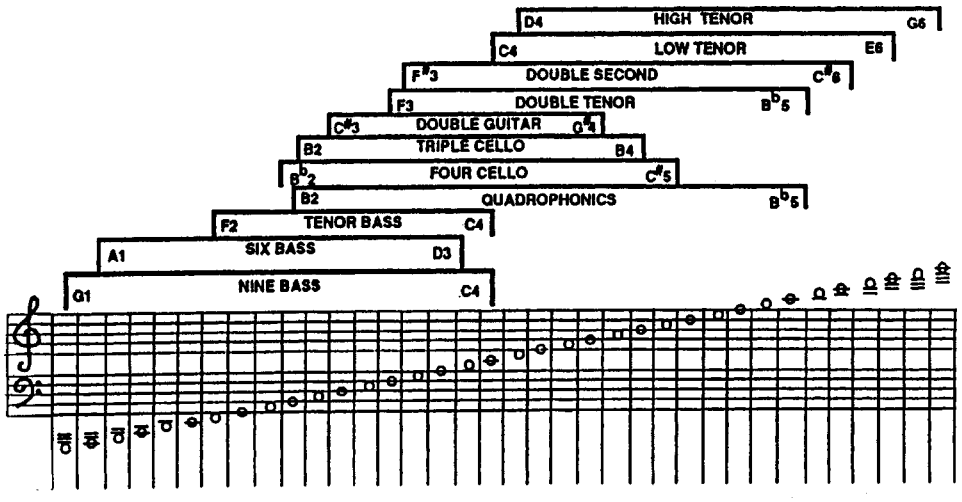


FIG. 1. Typical playing ranges of steelpans (Pichary, 1990).

other hand, which are fabricated from custom-made blanks, have skirts that are 10% to 15% thicker than this, often of stainless steel, and they are sometimes stiffened by pressing in one or two ridges near the bottom of the skirt. His tenor pans have a longer skirt length (22.4 cm) than those of Alexis and Mannette (13 cm).

One of the aims of this paper is to compare the acoustical and mechanical behavior of steelpans constructed by different pan makers, using different practices and designs.

II. NORMAL MODES OF VIBRATION

A normal mode of vibration represents the motion of a linear system at a normal frequency (eigenfrequency). It should be possible to excite a normal mode of vibration at any point in a musical instrument that is not a node and to observe motion at any point that is not a node. A normal mode is a characteristic only of the structure itself independent of the way it is excited or observed.

Normal mode shapes are unique for a structure, whereas the deflection of a structure at a particular frequency, called an operating deflection shape (ODS), may result from the excitation of more than one normal mode. Normal mode testing has traditionally been done using sinusoidal excitation, either mechanical or acoustical. Detection of motion may be accomplished by attaching a small accelerometer, although optical methods are less intrusive.

Normal modes of a complex structure, such as a steelpan, are functions of the entire structure. A normal mode shape should describe how every point on the pan moves when the instrument is excited at any point. As a practical matter, at any given frequency, some parts of the pan move much more than other parts. For example, we often speak of the normal modes of vibration of a particular note area, but we must always be aware of the fact that when one note area vibrates, all other parts of the pan vibrate at the same frequency, although their motion may be very small—too small to be observed, in many cases.

Holographic interferometry (Powell and Stetson, 1965) offers the best spatial resolution of operating deflection shapes (and hence of normal modes). Recording holograms on film tends to be rather time consuming, however. Electronic TV holography, on the other hand, offers one the opportunity to observe vibration motion nearly in real time and a fast, convenient way to record operational deflection shapes and to determine the normal modes (Jansson *et al.*, 1994).

III. EXPERIMENTAL METHOD

The optical system used for TV holography is shown in Fig. 2. A beam splitter (BS) divides the laser light to produce a reference and an object beam. The reference beam illuminates the charge coupled device (CCD) camera via a phase-stepping mirror (PS) (which introduces four 90° steps) and an optical fiber, while the object beam is reflected by mirror PM so that it illuminates the object to be studied. Reflected light from the object reaches the CCD camera, where it interferes with the reference beam to produce the holographic image. The time-averaged interferogram is calculated by the image processor from the four phase-stepped frames. The speckle-averaging mechanism (SAM) in the object beam alters the illumination angle in small steps in order to reduce laser speckle noise in the interferograms.

For holographic observation of vibration, the pans were supported at three points on their rim on an optical table. A sinusoidal driving force was applied to the pan either by attaching a small NdFeB magnet between note areas (so as not to load a note area) in a sinusoidal magnetic field from a solenoid driven by an audio amplifier or by applying a sinusoidal force directly to the steel surface by means of the

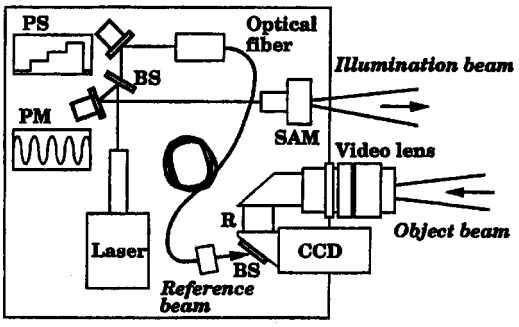


FIG. 2. Optical system for TV holography. BS=beam splitter; PS=phase-stepping mirror; SAM=speckle-averaging mechanism; CCD=video camera.

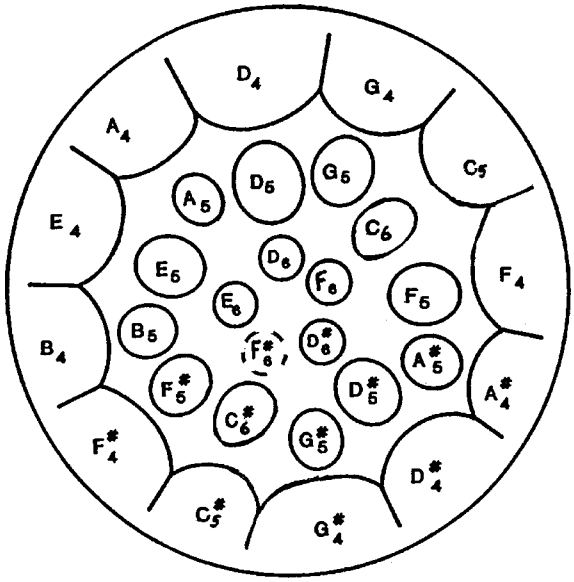


FIG. 3. Layout of note areas on the Alexis tenor pan.

solenoid. (In the latter case the force had twice the frequency of the current in the solenoid since the magnetic force was attractive for current flow in both directions.) Mode shapes were photographed from the TV screen.

The thickness was measured by using two dial gauges along with a special anvil, so that the exact height could be determined both with and without the pan in place on the anvil. The thickness can be determined to within 0.05 mm by this method.

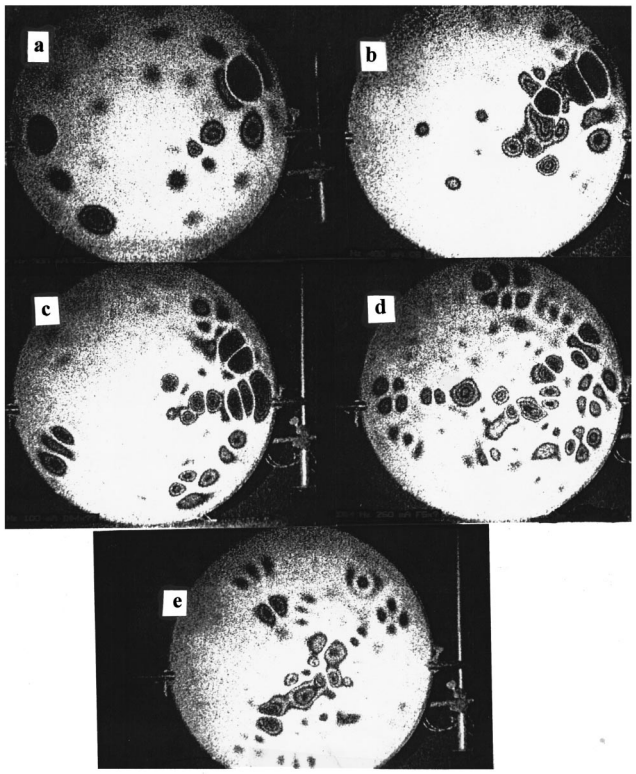


FIG. 4. Vibrations of the Alexis pan at frequencies that excite modes in the C_5 note area: (a) 522; (b) 1050; (c) 1421; (d) 2064; (e) 2184 Hz.

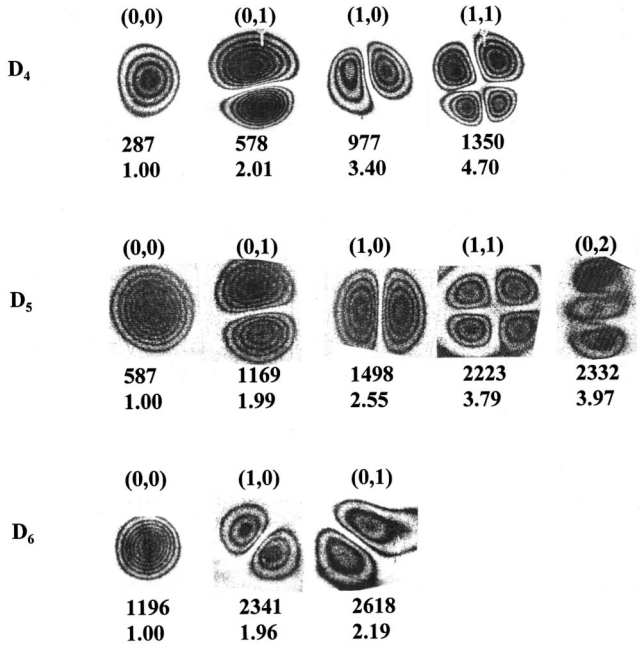


FIG. 5. Holographic interferograms of several modes in the D_4, D_5, and D_6 note areas of the Alexis tenor pan. Modes are designated by (m,n), where m and n are the numbers of radial and circumferential nodes, respectively. Frequency ratios to the fundamentals are given.

IV. RESULTS

A. Results on the Alexis tenor pan

The layout of the notes on the Alexis tenor pan is shown in Fig. 3. The outer ring has 12 more or less trapezoidal note areas tuned from D_4 (294 Hz) to $C_5^{\#}$ (554 Hz). The middle ring has 12 more or less elliptical note areas tuned from D_5 to $C_6^{\#}$ and the inner ring has five nearly circular note areas tuned from D_6 to $F_6^{\#}$. Note that the notes in the outer and middle rings are arranged in ''circles of fifths'': moving counterclockwise, one goes up a perfect fifth or down a perfect fourth (which is equivalent to going up a fifth and down an octave) on the musical scale. Each note in the middle ring is an octave higher than the corresponding note on the outside ring. The inside ring adds a third octave to five notes (six notes when an optional $F_6^{\#}$ is included). Thus each note has several harmonics in common with its nearest neighbors, which leads to the strong interaction between notes characteristic of steelpans.

Holographic interferograms of the Alexis tenor pan driven at several frequencies that excite modes of vibration in the C_5 note area (at two o'clock, viewing the pan as a clock face) are shown in Fig. 4. Note that several other note areas show appreciable vibration, especially at the higher frequencies. At the lowest frequency the active (roughly elliptical) portion of the note area moves in a single phase, while at the frequency of the second harmonic, there is a nodal line parallel to the rim dividing the note area into halves. At 1421 Hz $(2.72f_1)$ there is a radial node; at 2064 Hz $(3.95f_1)$ there are two nodal lines parallel to the rim, and at 2184 Hz, two nodal lines perpendicular to each other. Besides five modes of the C_5 note area, many sympathetic vibrations of other note areas are apparent in Fig. 4. The

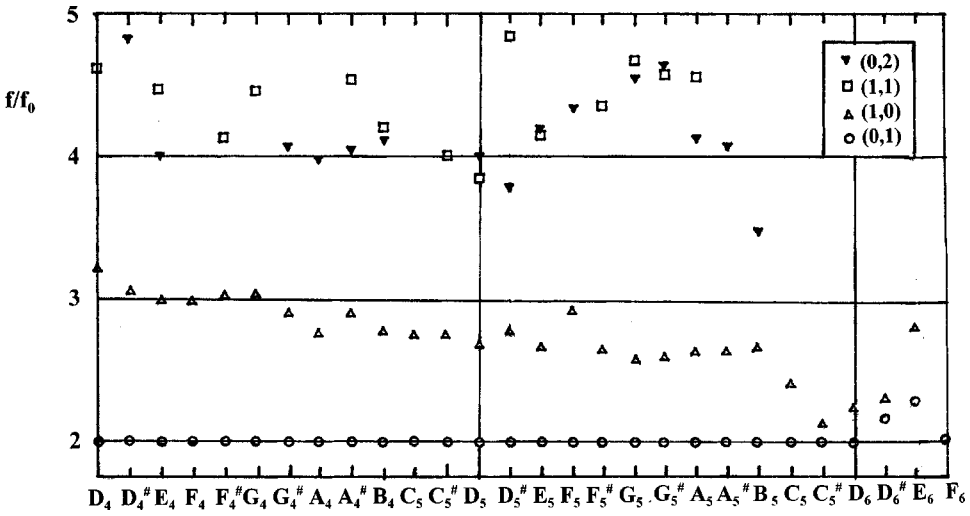


FIG. 6. Relative frequencies of the lowest modes in the Alexis tenor pan.

various note areas can be identified by comparing Fig. 4 with Fig. 3.

Holographic interferograms of several modes observed in the D_4, D_5, and D_6 note areas of the Alexis tenor pan are shown in Fig. 5. Following the mode designation used for elliptical or rectangular plates, modes are designated by (m,n), where m is the number of radial nodal lines and n is the number of nodal lines parallel to the circumference on the pan. In other note areas, (2,0) and (0,2) nodes were also observed.

Relative frequencies of the lowest modes observed in all 28 note areas are shown in Fig. 6. Note that the (0,1) is tuned to the second harmonic of the fundamental in all the notes. In the lowest six or seven notes, the (1,0) has been tuned to the third harmonic, with varying degrees of accuracy, although in the highest notes of the outer ring and throughout the middle ring it is closer to 2.5 times the fundamental frequency (an interval of a major tenth or an octave plus a major third above the fundamental). In a few notes, the (0,2) mode is tuned to the double octave, and in three notes ($F_5^{\#}$, $C_5^{\#}$, and E_5) the (1,1) mode fulfills that role.

Modal shapes in the skirt of the Alexis tenor pan are shown in Fig. 7. The modes correspond to standing bending waves propagating around the ring. At very low frequency (below 200 Hz), the modes have a nodal line at the center and antinodes at the top and bottom edge, as shown in Fig. 7(a), in spite of the rim at the top. From 200 to about 1000 Hz, the modes have a node at the top rim and n antinodes along the bottom edge of the skirt, as shown in Figs. 7(b) and 7(c). Above 1000 Hz, the modes have a second nodal ring half a wavelength above the lower edge, as shown in Fig. 7(d).

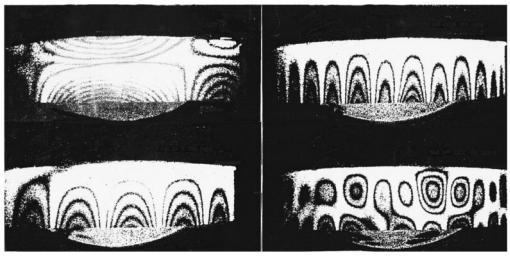


FIG. 7. Vibration modes of the skirt on the Alexis tenor pan: (a) 107; (b) 294; (c) 667; (d) 1400 Hz.

From holographic interferograms of the skirt, such as those shown in Fig. 7, it is possible to determine the wavelength and thus the speed of the bending waves. A graph of wave speed vs the square root of frequency is shown in Fig. 8. The bending wave speed in a thin plate is given by: $v=(1.8fhc_L)^{1/2}$, where $c_L=(E/\rho[1-v^2])^{1/2}$ is the speed of longitudinal waves (Fletcher and Rossing, 1998). For a steel plate with a thickness $h=0.9$ mm and $c_L=5050$ m/s, this predicts $v=2.86f^{1/2}$. The slope of the regression line in Fig. 8 is 2.54 m/s, which is reasonably close to the value predicted by the simple theory for a flat steel plate of the same thickness.

B. Results on the Rohner tenor pan

The layout of the notes on the Rohner tenor pan, which is slightly different from that of the Alexis pan, is shown in Fig. 9. Twenty-nine note areas span the range C_4 (262 Hz) to E_6. (1319 Hz). The notes in this pan also follow circles of fifths in the outer two rings.

The holographic interferograms in Fig. 10 show how the pan vibrates at several frequencies. We have selected frequencies at which resonances occur in the $D_4^{\#}$ note area (at ''five o'clock,'' if we think of the pan as a clock). Note, however, that the rest of the drum moves considerably at each of these frequencies as well. At the lowest frequency (we could call this the ''fundamental'' mode of the $D_4^{\#}$ note area), the entire note area moves in the same direction and thus, as we will see, radiates sound efficiently. At 618 Hz, two halves of the note area move in opposite directions, in the manner of a ''see-saw.'' The vibration pattern becomes increasingly complicated as the frequency increases.

As an aid in interpreting the holographic interferograms in Fig. 10, sketches of several modes of vibration (operating deflection shapes, strictly speaking, but they closely resemble the normal modes of vibration) in the $D_4^{\#}$ note area are shown in Fig. 11. The frequencies and ratios to the nominal fundamental frequency (311 Hz) are given below the sketches. In this note area, the second two modes have intervals of an octave (2:1) and a twelfth (3:1) above the funda-

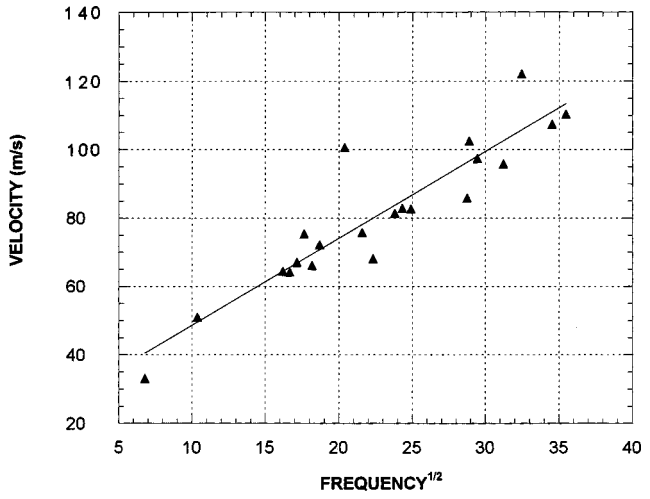


FIG. 8. Bending wave velocity in the skirt of the Alexis tenor pan (determined from modes of vibration such as those in Fig. 7).

mental, which is true in all the larger note areas in the outside ring. In the smaller note areas in the middle and innermost rings, only the octave is tuned harmonically.

Relative frequencies of the lowest modes observed in all 29 note areas are shown in Fig. 12. Note that the (1,0) is tuned to the second harmonic of the fundamental in all the notes. The (0,1) has been tuned to 3 times the fundamental frequency in the larger notes and 2.5 times the fundamental frequency in the others, although the transition from twelfth (third harmonic) to major tenth occurs at a higher frequency than in the Alexis tenor pan. In a few notes, the (2,0) mode is tuned to the double octave, as in the Alexis pan.

Modal shapes in the skirt of the Rohner tenor pan are shown in Fig. 13. The modes correspond to standing bending waves propagating around the ring. At very low frequency (below 200 Hz), the modes show antinodes at the top rim and bottom edge, as shown in Fig. 13(a), although the nodal circle is quite near the top rim rather than at the center, as in the pan shown in Fig. 7(a). From 200 to about 400 Hz, the modes have a node at the top rim and *n* antinodes along the bottom edge of the skirt, as shown in Fig. 13(b), but then the

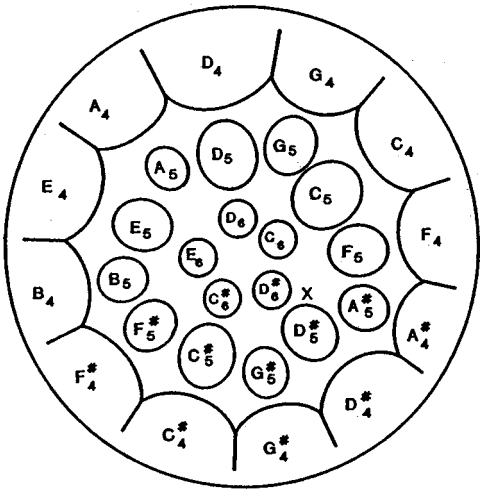


FIG. 9. Layout of note areas on the Rohner tenor pan. The point of application of the driving force is marked by X.

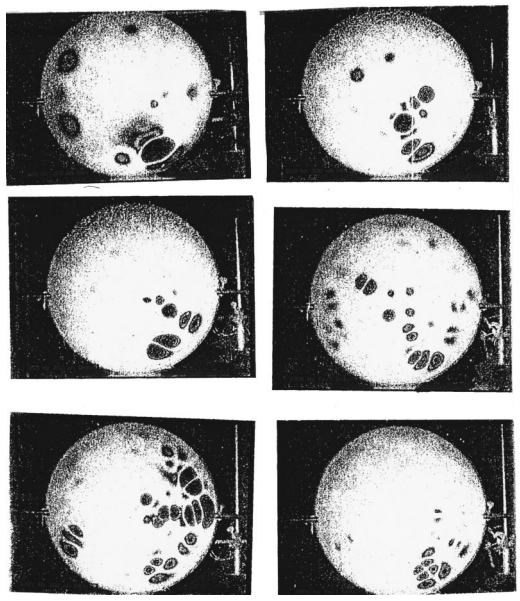


FIG. 10. Modes of vibration in the Rohner tenor pan at frequencies that excite resonances in the $D_4^{\#}$ note are: (a) 310; (b) 618; (c) 952; (d) 1302; (e) 1421; (f) 2005 Hz.

antinode moves up on the skirt, as shown in Fig. 13(c). Figure 13(d) at 728 Hz shows nodal rings at the top rim and also a third of the way above the bottom edge, while Fig. 13(e) at 1238 Hz shows three nodal rings and Fig. 13(f) at 1768 Hz shows four.

A graph of bending wave velocity vs square root of frequency in this skirt is shown in Fig. 14. Note that the slope of the regression line is more than twice that of Fig. 8, although the steel is only 10% thicker. Only at the very highest frequency does the speed of the bending waves reach the speed of sound in air (the frequency at which radiation efficiency shows a marked increase).

C. Results on the Mannette tenor pan

The layout of the notes on the Mannette pan is the same as on the Rohner pan in Fig. 9 except that the center ring has six notes with a seventh at the center. Thus, the Mannette pan has 31 notes spanning the range C_4 to G_6, compared to

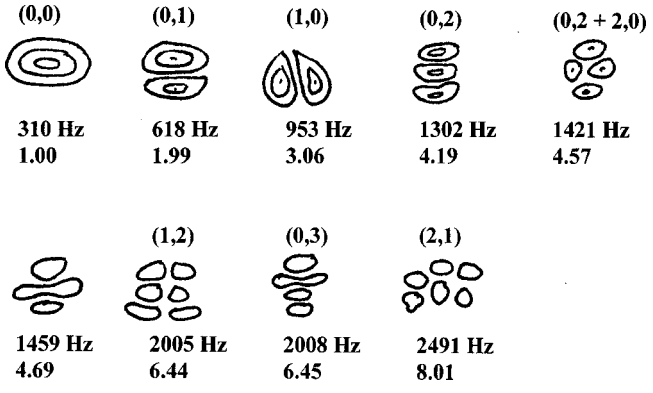


FIG. 11. Sketches of vibration modes (operating deflection shapes) in the $D_4^{\#}$ note area of the tenor pan shown in Fig. 10. Modal frequencies are given in the top line with ratios to the nominal fundamental frequency below.

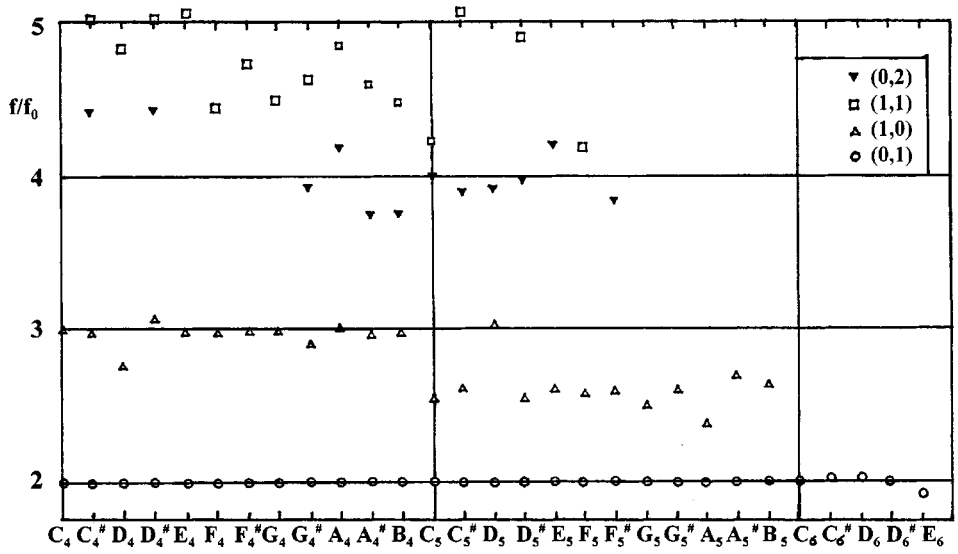


FIG. 12. Relative frequencies of the note areas in the Rohner tenor pan.

28 notes on the Alexis pan and 29 on the Rohner pan. The notes in this pan also follow circles of fifths in the outer two rings.

Holographic interferograms of four modes in the $C_4^{\#}$ note area are shown in Fig. 15. The (0,1) mode is tuned an octave above the fundamental (0,0), and the (1,0) mode is tuned a twelfth (octave plus a fifth) above the fundamental. The (1,1) mode has a frequency near the 5th harmonic of the fundamental, but that mode was probably not tuned. In all the notes in the outer ring, the (0,1) mode is tuned to the second harmonic, but in some notes the (1,0) mode is tuned to the fourth harmonic rather than the third harmonic, as can be seen in Fig. 16, which gives observed frequency ratios for all the notes.

In the middle row, the (0,1) mode is tuned to the second harmonic in all the note areas, but the (1,0) mode is sometimes tuned to the third harmonic, sometimes not; the frequencies range from 2.33 to 3.03 times the fundamental. Apparently the tuner tunes this mode harmonically if possible. In the six notes of the inner row and the one note at the center, only a second harmonic is tuned.

D. Results on the Alexis double-second pan

The double-second steelpan plays in the alto range. It consists of two pans, each having 14 to 18 notes. Thus the note areas are considerably larger than those of the tenor pan. The note layouts in two different double-second pans, a 15-note pan and a 14-note double-grooved pan, are shown in Fig. 17.

Modal shapes observed in three different $B^{\flat}$ note areas in a 15-note double-second pan by Clifford Alexis are shown in Fig. 18 (Hampton, 1995). The lines in these sketches are lines of constant amplitude; regions on either side of a nodal line move in opposite directions. Because the note areas are larger in a double-second pan, it is possible to tune both second and third harmonics in all the note areas, and sometimes a fourth harmonic is tuned as well. In general the second- and third-harmonic modes are ''orthogonal''; that is, the nodal lines are perpendicular to each other.

Harmonic tunings of modes 2–5 in Fig. 18 are shown in Fig. 19 for all 15 notes in the drum. Note that the (0,1) mode is always tuned to the 2nd harmonic, and the (1,0) mode is

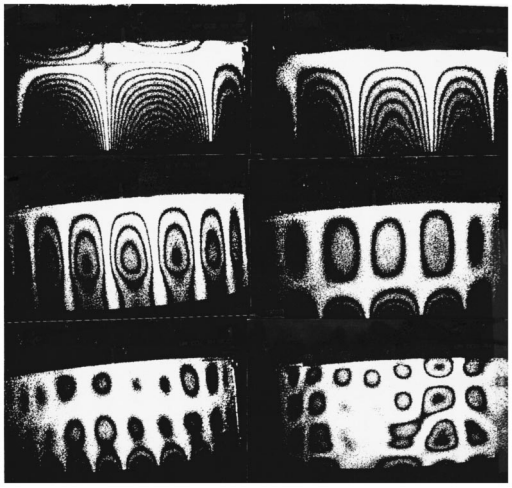


FIG. 13. Vibration modes in the skirt of the Rohner tenor pan: (a) 154; (b) 245; (c) 561; (d) 728; (e) 1238; (f) 1786 Hz.

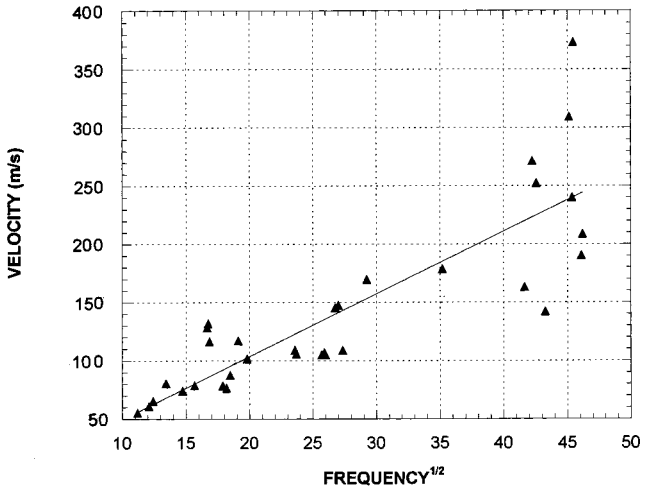


FIG. 14. Bending wave velocities in the skirt of the Rohner tenor pan (determined from modes of vibration such as those in Fig. 13).

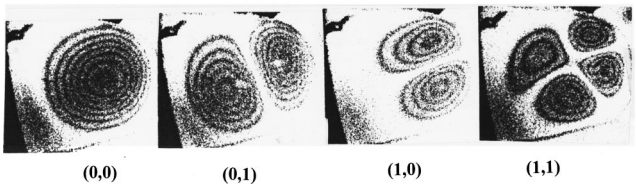


FIG. 15. First four modes in the $C_4^{\#}$ note area in the Mannette tenor pan.

generally tuned to the 3rd harmonic. However, either the (1,1) mode or the (0,2) mode may be tuned to the fourth harmonic [it is the (2,0) mode in the $B_3^{\flat}$ note area]. Tuning the (1,1) mode to the 4th harmonic is sometimes referred by pan tuners as ''diagonal'' tuning for obvious reasons.

The double-second pan whose mode shapes are shown in Fig. 18 is a ''single-grooved'' pan; a single groove separates the larger note areas from each other. Most pan makers these days prefer a ''double-grooved'' design, such as the one shown in Fig. 17(b). Mode shapes in a double-grooved double-second pan are shown in Fig. 20.

Harmonic tunings of modes 2–5 in Fig. 20 are shown in Fig. 21 for all 14 notes in the pan. Note that the (1,0) mode is tuned to the 2nd harmonic, and the (0,1) mode is generally tuned to the 3rd harmonic except in the two lowest notes where it is tuned to the 4th harmonic. In three notes the (1,1) mode is tuned close to the 4th harmonic, while in two notes (including $B_3^{\flat}$) it is the (2,0) mode that is tuned close to this harmonic.

E. Results on the Mannette double-second pan

The Mannette double-second pan consists of two pans, each with 18 notes, as shown in Fig. 22, covering the range of E_3 to $D_6^{\#}$. A double-groove design is used in the outer ring, as in the pan shown in Fig. 17(b), although each pan has four more notes than the Alexis pan in Fig. 21. The thickness of the playing surfaces varies from 1.0 mm in the outer notes to 0.5 mm near the center. The skirt lengths are 22 cm, the same as those in the Alexis double-second pan.

Modal shapes in the F_3 note area are shown in Fig. 23. The (1,0) mode is tuned to an octave above the fundamental

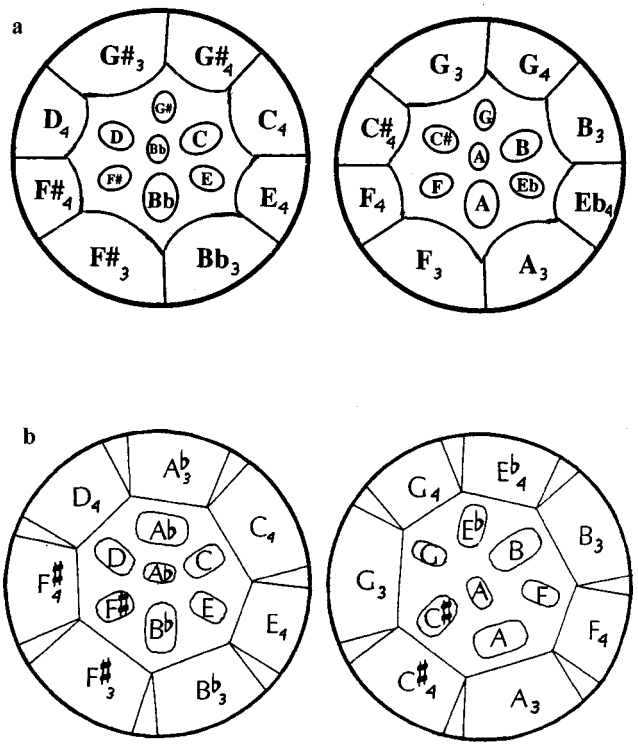


FIG. 17. Note layout on two Alexis double-second pans: (a) 15-note; (b) 14-note, double-grooved in the outer ring.

(0,0) mode, and the (0,1) mode is tuned to two octaves above the fundamental. Other modes occur at 4.72, 5.09, 6.71, and 8.77 times the fundamental.

Tuning of the lower modes in all 18 note areas in one pan is shown in Fig. 24. Note that the (1,0) mode is generally tuned to the octave, but in some notes it is the (0,1) mode that is tuned to the octave instead. In the outer ring, the (0,1) mode is tuned to the double octave only in the two largest notes, while it is tuned to the twelfth (3 times the fundamental frequency) in three notes, and not harmonically tuned in the remaining two note areas. ''Diagonal tuning'' of the (1,1) mode to the double octave appears to have been done in one or two notes. In the middle ring, where the note areas are nearly circular, the (1,0) and (0,1) modes have nearly the

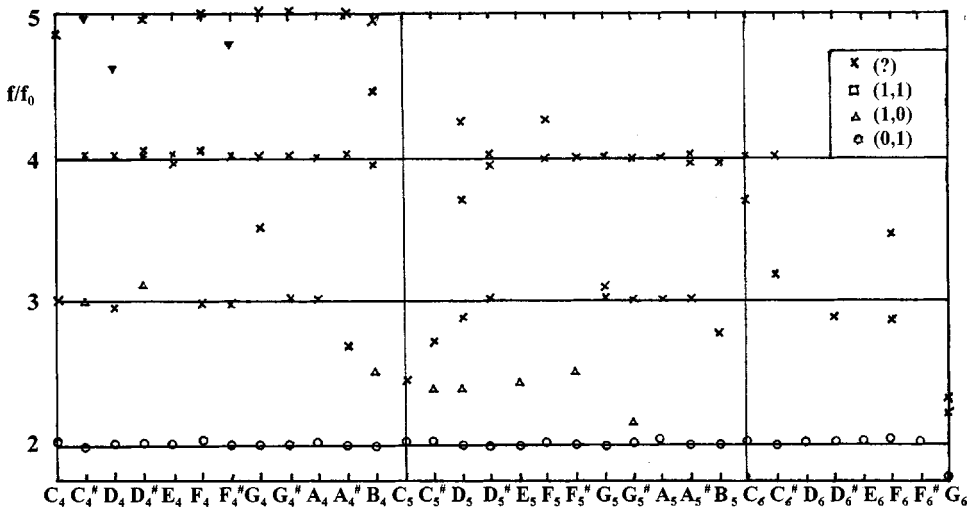


FIG. 16. Relative frequencies of the note areas in the Mannette tenor pan.

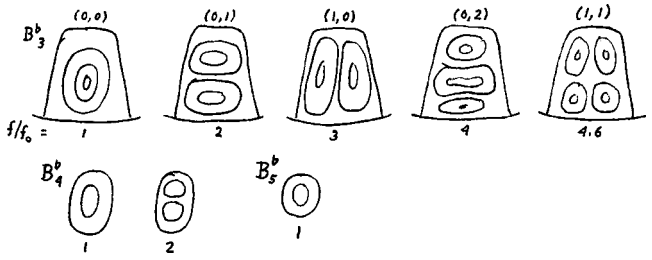


FIG. 18. Mode shapes in three different B^b note areas in the Alexis 15-note double-second pan (Hampton, 1995).

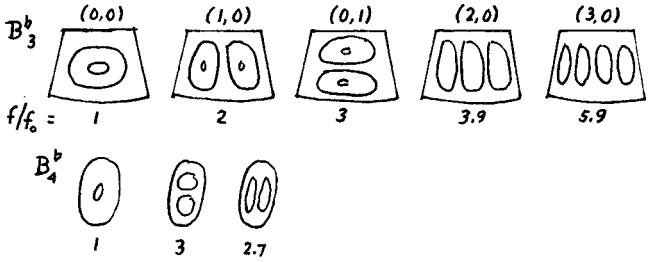


FIG. 20. Mode shapes in three different B^b note areas in the Alexis 14-note double-second pan with a ''double-grooved'' design.

same frequency. The same appears to be true in the inner circle, although holographic modal analysis was not made.

V. DISCUSSION OF THE RESULTS

The note areas in the various steelpans differ widely in size and shape. Some are nearly trapezoidal, some are oval, and some are nearly circular. When struck with a soft beater, a note area will vibrate in a complex manner which can be analyzed in terms of normal modes of vibration. The pan tuner has several options at his/her disposal for tuning these modes of vibration to harmonics of the fundamental. The (0,1) mode is nearly always tuned to the second harmonic, and in many notes, the (1,0) mode is tuned to the third harmonic (a musical twelfth above the fundamental) or to a frequency about 2.5 times the fundamental (a major tenth above the fundamental). Although most tuners use electronic tuners, considerable reliance is placed on their experienced ears. The exact tuning is determined partly by how a particular note responds; when it is not possible to tune modes to the third or fourth harmonics, for example, the tuner adjusts the note to give the ''best'' sound.

The rather daunting task of designing the layout of notes on various pans has been largely a matter of trial and error. Although tuners still customize pans for certain players, there has been a movement toward standardizing note layout (Roach, 1992). Likewise, there has been a trend toward standardization of note size and shape. In a previous study we showed that the longest dimension L of note areas in a large variety of pans appears to follow a scaling law $L=Kf^{-2/3}$, where f is the fundamental frequency (Rossing *et al.*, 1996). There is little known scientific basis for this empirical formula, however, and it may or may not be optimum.

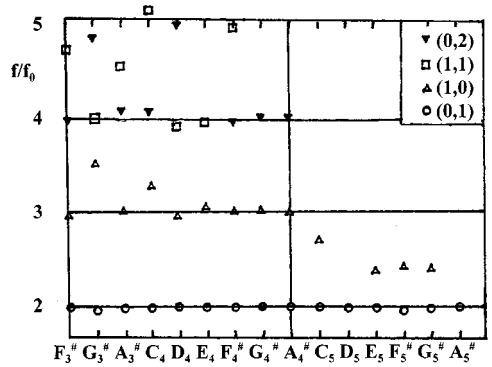


FIG. 19. Harmonic tuning of the (0,1), (1,0), (1,1), and (0,2) modes (modes 2-5 in Fig. 15) in the Alexis 15-note double-second pan (Hampton, 1995).

The only known attempt to model note areas is that of Hampton (1995), who modeled them as shallow rectangular shells with length L, width W, thickness h, and radii of curvature R_L and R_W along the length and width, respectively. Using a computer, Hampton tested a total of 143 520 physical configurations for various mode hierarchies. About 1.2% of the configurations tested produced a second harmonic (1,0) mode. About 4% of those that produced a second harmonic also produced a third-harmonic (0,1) mode; about 9% of them produced a fourth-harmonic (2,0) mode. In most of the harmonic cases, the profile parameter $R=R_W/R_L$ was negative and the curvature parameter R_W/W was small (less than 10).

Profiles of note areas in pans tuned by various tuners reveal the rather surprising result that most of them are nearly flat. Although the notes appear convex, when viewed against the concave surface of the pan, they are, in fact, nearly flat, with local areas of both positive and negative curvature. The mode frequencies of a flat rectangular plate with simply supported (hinged) edges are fairly easy to calculate. In a plate with $L/S=\sqrt{2}$, for example, the lowest three modes are harmonically related, and the (1,1) mode has 4 times the frequency of the fundamental for all values of L/W. However, the boundary condition at the edge of a note area might be expected to be somewhere between simply supported and clamped.

The edges of note areas are quite difficult to model. Most notes are bounded by a groove, some 2–3-mm wide, made with a punch. Narrow curved regions are found on both sides of the groove, and these may actually contribute as much to formation of standing waves as the grooves them-

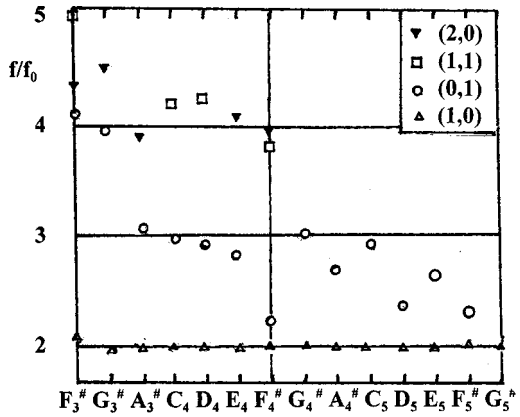


FIG. 21. Harmonic tuning of the (1,0), (0,1), (1,1), and (2,0) modes (modes 2–5 in Fig. 20) in the 14-note double-second pan (Hampton, 1995).

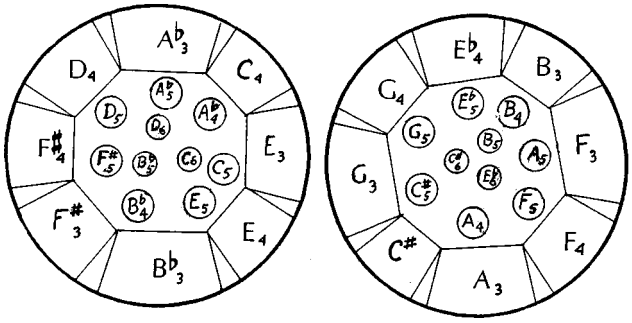


FIG. 22. Note arrangement in the Mannette double-second pan.

selves. Some pan makers in Trinidad have replaced the groove with a series of bored holes, and the one example of a ''bored'' pan that we have studied behaved pretty much the same as a ''grooved'' pan. Some tuners have experimented with grooveless pans. Some vibrational modes are observed to extend slightly outside the note boundaries.

The results of the modal study presented in this paper should be of value to a pan designer who wishes to be guided by scientific principles in improving the sound of pans. Tuning notes with a hammer presumably changes the local thickness and the local stresses. The effect of thickness change on a given mode of vibration should be most effective in an area of greatest deflection curvature, which will ordinarily be at an antinode (the ''bullseye'' in the holographic interferogram of that mode). Thinning a plate reduces the mass but it also reduces the stiffness and the latter effect dominates, so modal frequencies are decreased by thinning.

Glancing blows in which the hammer has a large component of motion parallel to the pan surface are used to change internal stress. The effect is sometimes described as ''tightening'' a note. Murr *et al.* (1999) found that hammering reduces grain size and that harmonic tuning may involve a complex interaction between plastic and elastic properties of the metal. Areas between notes (outside the active note area) are also hammered to ''tighten'' the note. Some tuners have adopted the technique of using an oxyfuel torch to ''tighten'' or ''slacken'' individual notes.

Holographic interferograms such as those in Figs. 4 and 10 reveal the complexity of vibrational modes due to the strong mechanical coupling between note areas. The sound of a note changes markedly when neighboring note areas are damped (Leung and Rossing, 1988). This coupling is found to be quite nonlinear; it increases as the vibration amplitude of the struck (or driven) note increases (Hampton *et al.*, 1987). Enhancing the mechanical coupling between note areas is a key consideration in planning the layout of a pan. Arranging the notes in a circle of fifths with octave intervals between successive rings of notes maximizes this coupling.

The sounds of steelpan notes are known to be rich in harmonics. In addition to the harmonic tuning of modes and the strong coupling between note areas, there is another effect that contributes to harmonic generation. That is nonsinusoidal vibrational motion resulting from asymmetry in the stiffness of a note area due to its shape which leads to harmonic generation. In studying this effect several years ago, we found that when a note area is driven at its fundamental frequency, the amplitude of the second harmonic increased as the square of the fundamental amplitude, the third harmonic as the cube of the fundamental amplitude, and the fourth harmonic as the fourth power of the fundamental amplitude (Rossing and Defrance, 1991). Beyond the fourth harmonic, it was difficult to establish a quantitative relationship between fundamental amplitude and harmonic amplitude, but harmonics out to the eighth were observed at large fundamental amplitude. These measurements, incidentally, were made with the other note areas damped with sandbags so that sympathetic vibrations of harmonically tuned note areas were discounted.

Sound radiation from the skirt is important, although the total amount of sound radiated by the skirt has not yet been determined. In a previous study, we found that in a tenor pan the vibrational amplitude of the skirt on the side nearest the struck note averaged 19 ± 9 dB less than the note amplitude (averaging over all the notes), while it was 23 ± 11 dB less than the average note amplitude on the side of the drum opposite the note (Rossing *et al.*, 1996). In a double-second pan, the corresponding differences in amplitude averaged 27 ± 12 dB on the same side and 37 ± 14 dB on the opposite side. These represent amplitude ratios from about 10:1 to 100:1, suggesting that total radiation from the skirt, in spite of its large radiating area, is considerably less than from the playing surface. However, the vibrational spectrum of the skirt is sufficiently different from that of the note area to influence the timbre of the pan. Damping the skirt markedly changes the timbre.

VI. FUTURE RESEARCH

The Caribbean steelpan is a folk instrument that has found its way into the world's concert halls. This is partly due to the skill of young performers, now trained in universities and music conservatories, and partly due to the skill of craftsmen who have developed the fine art of pan making. Like makers of fine violins, these pan makers have learned their art by observing and serving as apprentices to other pan makers and by considerable trial and error. Crafting fine instruments, however, takes a great deal of time and skill, and as the demand for quality instruments begins to exceed the supply, fine instruments are becoming expensive. As in the

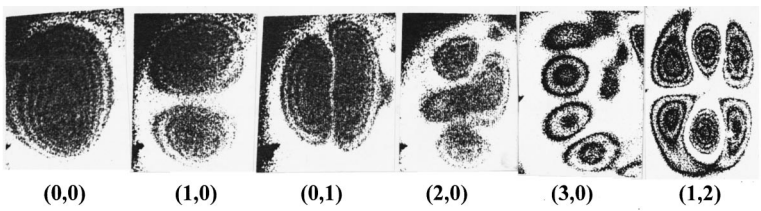


FIG. 23. Note shapes in the F_3 note area of the Mannette double-second pan: (0,0), (1,0), (0.1), (2,0), (3,0), and (1,2).

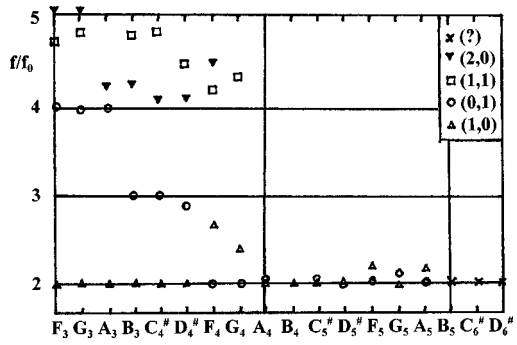


FIG. 24. Tuning of the modes in an 18-note double-second pan by Ellie Mannette.

field of violin making, the questions arises as to how less experienced makers, guided by scientific testing and measurements, can build and tune fine instruments. This appears to be possible as the acoustics of steelpans becomes better understood through scientific research.

Steelpans were originally constructed from used oil drums. The pan maker had little control over the quality of the raw material, which was a constant source of frustration. Many pan makers now purchase steel barrels directly from the manufacturer, and specify standards of quality. Still, steel barrels vary widely in quality, and it is difficult for the pan maker to test them and to avoid investing many hours in shaping an inferior barrel which will result in a steelpan of disappointing quality. Recent studies of metallurgy (Murr *et al.*, 1999) and heat treating (Ferreyra *et al.*, 1999) of steelpan materials are a step in the right direction. We have reported measurements of a steelpan constructed from custom-made blanks that include a tuning surface cold drawn (rather than stretch formed manually by hammering) and hardened by nitriding. We hope that other pan makers will experiment with such materials, and that careful studies of their mechanical and metallurgical properties will continue.

The importance of materials in musical instruments, including the steelpan, is being recognized. In the past four years, the Materials Research Society, in cooperation with the Musical Acoustics Technical Committee of the Acoustical Society of America, has sponsored three special symposia on Materials in Musical Instruments (papers from the first symposium were published in the March 1995 issue of MRS Bulletin). A chapter on ''Materials for Musical Instruments'' has been included in the 2nd edition of *The Physics of Musical Instruments* (Fletcher and Rossing, 1998).

The effect of skirt size on both the total sound radiation and the timbre should be determined. In custom blanks, the thickness of the skirt can also be varied to determine the effect on timbre.

This paper included results from only two members of the steelpan family, the tenor and the double second, and from only three pan makers. In the second paper, acoustical properties of other members of the family will be reported. These studies should be expanded to include pans by other makers, including those in Trinidad, as well.

VII. CONCLUSION

The Caribbean steelpan, an important new musical instrument, is a very complex vibrating system. Striking one note area excites vibration in other note areas and in the skirt as well, all of which radiate sound. Understanding the vibrational mode shapes is the key to understanding the acoustical behavior of the instrument. Using a hammer, the pan maker carefully tunes two or three modes in each note area to have harmonic frequencies in order to obtain a pleasing musical timbre. Although the tuning techniques have developed mainly by trial and error, a scientific understanding of the vibrational modes enables the tuning to proceed with greater accuracy and less effort.

ACKNOWLEDGMENTS

The authors thank Clifford Alexis, Felix Rohner, Jimmie Finnie, and Susie Jackson for the loan of steelpans used in the investigation. Conversations with Allan O'Connor, Liam Teague, and with the three pan makers and their associates were very enlightening.

Achong, A. (**1996**). ''The steelpan as a system of non-linear mode-localized oscillators, I: Theory, simulations, experiments and bifurcations,'' J. Sound Vib. **197**, 471–487.

Achong, A., and Sinanan-Singh, K. A. (**1997**). ''The steelpan as a system of mode-localized oscillators, Part II: Coupled sub-systems, simulations, and experiments,'' J. Sound Vib. **203**, 547–561.

Achong, A. (**1998**). ''The steelpan as a system of non-linear mode-localized oscillators, Part III: The inverse problem-Parameter estimation,'' J. Sound Vib. **212**, 623–635.

Ferreyra, E., Maldonado, J. G., Murr, L. E., Pappu, S., Trillo, E. A., Kennedy, C., Posada, M., De Alba, J., Chitre, R., and Russell, D. P. (**1999**). ''Materials science and metallurgy of the Caribbean steel drum,'' J. Mater. Sci. **34**, 981–996.

Fletcher, N. H., and Rossing, T. D. (**1998**). *The Physics of Musical Instruments*, 2nd ed. (Springer-Verlag, New York).

Hampton, D. S. (**1995**). Investigation of the vibrational modes of steel drums by holographic interferometry, M.S. thesis, Northern Illinois University, DeKalb, Illinois.

Hampton, D. S., Alexis, C., and Rossing, T. D. (**1987**). ''Note coupling in Caribbean steel drums,'' J. Acoust. Soc. Am. **82**, S68 (abstract).

Hansen, U. J., Rossing, T. D., Mannette, E., and George, K. (**1995**). ''The Caribbean steelpan: Tuning and mode studies,'' MRS Bull. **20**, 44–46.

Jansson, E. V., Molin, N.-E., and Saldner, H. O. (**1994**). ''On eigenmodes of the violin—electronic holography and admittance measurements,'' J. Acoust. Soc. Am. **95**, 1100–1105.

Kronman, U. (**1991**). *Steel Pan Tuning* (Musikmuseet, Stockholm).

Leung, K. K., and Rossing, T. D. (**1988**). ''Sound spectra of Caribbean steel drums,'' paper presented at joint annual meeting of APS and AAPT, Arlington, Va.

Murr, L. E., Ferreyra, E., Maldonado, J. G., Trillo, E. A., Pappu, S., Kennedy, C., DeAlba, J., Posada, M., Russell, D. P., and White, J. L. (**1999**). ''Materials science and metallurgy of the Caribbean steel drum,'' J. Mater. Sci. **34**, 967–979.

Pichary, L. (**1990**). *Results of the Steel Pan Survey 90* (Trinidad and Tobago Bureau of Standards, Tunapuna).

Powell, R. L., and Stetson, K. A. (**1965**). ''Interferometric vibration analysis by wavefront reconstruction,'' J. Opt. Soc. Am. **55**, 1593–1598.

Roach, K. (**1992**). *The Imperatives for Standardization* (Trinidad and Tobago Bureau of Standards, Tunapuna).

Rossing, T. D., and Defrance, J. (**1991**). ''Nonlinear vibrations in Caribbean steelpans,'' Illinois Academy of Science, Chicago.

Rossing, T. D., Hampton, D. S., and Hansen, U. J. (**1996**). ''Music from oil drums: The acoustics of the steelpan,'' Phys. Today **49**, 24–29.

Broadband time-domain reflectometry measurement of attenuation and phase velocity in highly attenuating suspensions with application to the ultrasound contrast medium Albunex®

Michael S. Hughes and Alexander L. Klibanov
Mallinckrodt Inc., St. Louis, Missouri 63042

Jon N. Marsh
Washington University School of Medicine, St. Louis, Missouri 63110

James G. Miller
Department of Physics, Washington University in St. Louis, Missouri 63130

Gary H. Brandenburger
Mallinckrodt Inc., St. Louis, Missouri 63042

(Received 10 September 1999; revised 16 January 2000; accepted 12 April 2000)

We describe a technique for broadband measurements of the attenuation coefficient and phase velocity of highly attenuating liquid suspensions. To validate the technique we apply it to the ultrasound contrast agent Albunex® at concentrations ranging from 0.69×10^6 particles/mL to 364×10^6 particles/mL. These longitudinal wave measurements were performed on Albunex suspensions maintained at 37 °C in a special time-domain reflectometer designed and constructed in our laboratory. The frequency-dependent attenuation coefficients and phase velocities obtained in the reflectometer are compared to broadband through-transmission measurements of these same quantities, which were also performed in our laboratory. Although comparison data between the two techniques are only available at lower concentrations, the agreement is quite good and serves to validate the methods described in this paper. © *2000 Acoustical Society of America.* [S0001-4966(00)05507-7]

PACS numbers: 43.80.Ev, 43.80.Qf [FD]

INTRODUCTION

Previous publications have described the development of a broadband, low-noise ultrasonic spectrometer in our laboratory.[1–3] That work resulted in measurement of the attenuation coefficient, phase velocity, and backscatter coefficient of the ultrasound contrast agent, Albunex, for a range of concentrations between 0.09×10^6 particles/mL to 21.0×10^6 particles/mL. Those results represent the broadest-bandwidth single-transducer measurements published to date for Albunex. One outcome of these measurements was the experimental verification of the shelled microbubble theories of de Jong and of Church.[4–9]

A key feature of both theories is that the behavior of the attenuation, phase velocity, and backscatter are based on the behavior of single bubbles at low concentrations. Thus in their seminal work both de Jong and Church anticipated that, at higher concentrations, a different theory might be required to describe the acoustic behavior of suspensions of shelled microbubbles. One purpose of this study is to determine the concentration at which such acoustic behavior becomes measurable. Our previously published through-transmission measurements have shown that the low concentration shelled microbubble theories hold over the entire range of microbubble concentrations that we are able to measure in through-transmission mode. This is a much wider range of concentrations than may be expected theoretically.[4–9] Although these data might therefore also be useful for development of new theories, our aim is primarily the description of methods and experimental results.

We do hope, however, to realize certain secondary benefits from the development of an instrument capable of measuring acoustic properties of highly concentrated suspensions. Such a device would enable us to search for new acoustic behavior, perhaps based on multiple scattering or other collective processes, e.g., radiation reaction or Bjerknes forces. Potentially, these processes might have clinical application (analogous to the discovery of second harmonic imaging by characterizing contrast media behavior at moderate-to-high insonifying power levels). The present work also suggests the possibility of a continuous, real-time, in-line process monitor, capable of measuring the acoustic properties of ultrasound contrast media (for example, Albunex or Optison) during manufacture. In-process concentrations can be comparable to those of the final product in the vial and are much higher than those we can measure with our through-transmission apparatus. Such a sensor could also be applied outside the production environment, for example, for quality control assessment of contrast media in a hospital pharmacy prior to clinical use, which is the current practice for application of nuclear contrast agents.

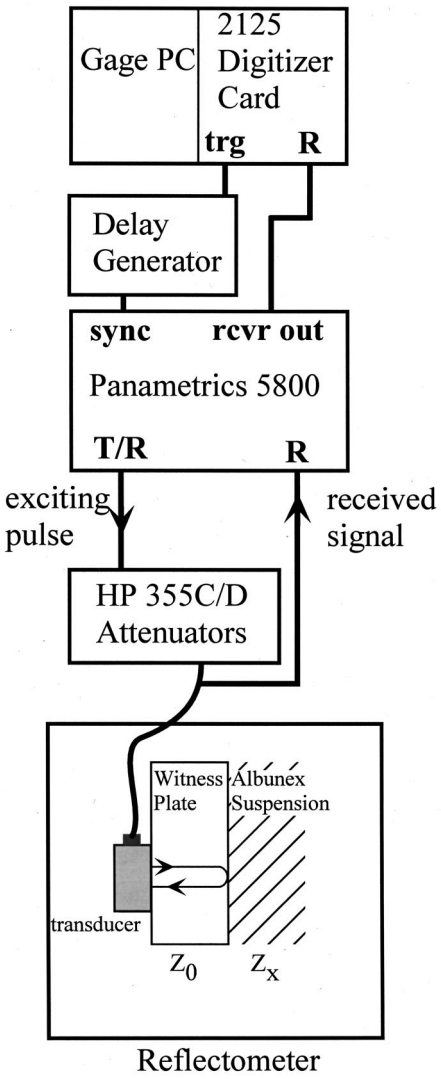


FIG. 1. Control instrumentation and geometry of the broadband time-domain reflectometer. The properties of the witness plate are assumed known. The reflected pulse is used to measure the attenuation coefficient and phase velocity of the Albunex suspension.

I. EXPERIMENTAL METHODS

A. Data acquisition and reduction

The basic design for the broadband reflectometer described in this paper is shown at the bottom of Fig. 1. A commercial transducer emits a broadband (1–10 MHz) longitudinal pulse into a homogeneous plate, with flat and parallel faces, hereafter referred to as the ''witness plate,'' of known complex acoustic impedance Z_0. This wave travels a short distance to the opposite side of the plate, which is in contact with the suspension of Albunex, whose acoustic properties are unknown. The pulse is partially reflected from this interface back to the emitting transducer. Analysis of this reflected pulse yields information about the complex acoustic impedance, Z_x, of the Albunex suspension. If the Albunex suspension behaves linearly, the attenuation coefficient and phase velocity of the suspension can be computed from this quantity, as we will shortly describe [Eqs. (7), (8)].

The magnitude of the reflected pulse is determined by the pressure reflection coefficient, R. R depends on Z_0, which is the complex acoustic impedance in the witness plate, and Z_x, which is the complex acoustic impedance in the sample (in this study the Albunex suspension), which is assumed to be unknown. From these two quantities the complex pressure reflection coefficient may be computed:

$$R(Z_0,Z_x)=\frac{Z_x-Z_0}{Z_x+Z_0}. \tag{1}$$

This equation is the basis of data reduction for the reflectometry method. A third complex acoustic impedance is required to calibrate the measurement. This is obtained by digitizing the reflection from the witness plate when it is in contact with a reference material of known acoustic impedance—in this case, Isoton®. We will indicate acoustic quantities associated with this measurement by the subscript ''ref.''

We will discuss the data reduction equations in approximately the order in which they are defined and used in the data reduction program. The first quantities of interest are the frequency-dependent complex wave numbers:

$$k_j=\frac{\omega}{c_j}, \tag{2}$$

where $j=$''x,'' ''0,'' or ''ref,'' defined above, and, $\omega=2\pi f$, with frequency f. Given these complex wave numbers, we may use the attenuation coefficients: α_{ref}, α_0, α_x, and their corresponding phase velocities c_{ref}, c_0, c_x, to write the expressions:

$$Z_j=\rho_j c_j\left(\frac{1+i\alpha_j/k_j}{1+\alpha_j^2/k_j^2}\right). \tag{3}$$

These are the general definitions of the complex characteristic impedances of the reference standard (Isoton), the witness plate, and the sample (Albunex suspension), respectively.[10]

Next we rewrite Eq. (1), solving for Z_x in terms of $R(Z_0,Z_x),Z_0$, to obtain:

$$Z_x=\frac{1+R(Z_0,Z_x)}{1-R(Z_0,Z_x)}Z_0. \tag{4}$$

It can be shown that the quantity $R(Z_0,Z_x)$ can be written in terms of the complex spectra of the time-domain waveforms, obtained when the witness plate is in contact with the specimen and calibration medium, and of the reflection coefficient R_{ref} (see Ref. 11, Chap. 6):

$$R(Z_0,Z_x)$$

$$R_{\mathrm{ref}}\frac{\text{Fourier transform [specimen waveform]}}{\text{Fourier transform [reference waveform]}}, \tag{5}$$

where R_{ref} is computed using the properties of the reference standard (Isoton) and the witness plate following the form of Eq. (1):

$$R_{\mathrm{ref}}=\frac{Z_{\mathrm{ref}}-Z_0}{Z_{\mathrm{ref}}+Z_0}. \tag{6}$$

We now have enough information to express c_x and α_x as functions of Z_0 and Z_x as follows:

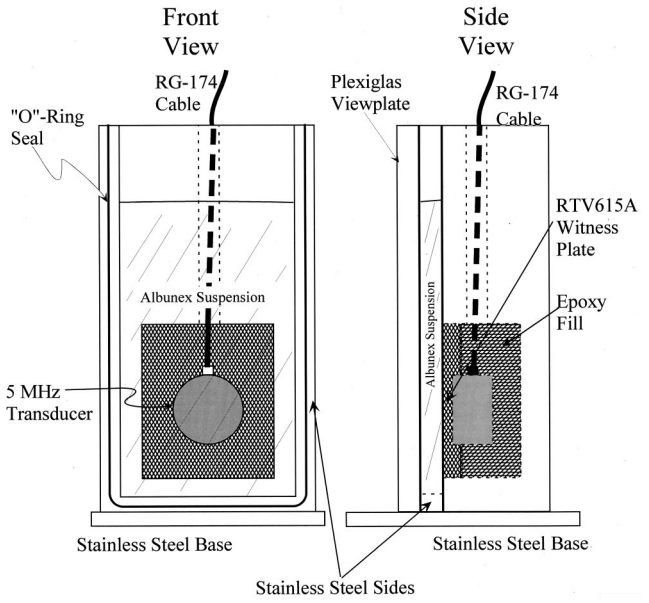


FIG. 2. A drawing of the broadband time-domain reflectometer. Front and side views are shown.

$$c_x = \frac{\mathrm{Re}(Z_x)^2 + \mathrm{Im}(Z_x)^2}{\rho_x \,\mathrm{Re}(Z_x)} = \frac{|Z_x|^2}{\rho_x \,\mathrm{Re}(Z_x)}, \tag{7}$$

$$\alpha_x = \frac{\rho_x \omega \,\mathrm{Im}(Z_x)}{|Z_x|^2}. \tag{8}$$

These are easily obtained by taking the real and imaginary parts of Eq. (6) and inverting.

In Sec. IV we describe the effect of the frequency-dependent linear term, which appears in Eq. (8) for the attenuation coefficient of the sample. This term has the property of amplifying small experimental errors at higher frequencies. We also note that comparison of Eq. (7) with Eq. (8) shows that the phase velocity does not suffer from this effect.

B. Apparatus

All data described in this study were obtained using the apparatus shown in Fig. 1. The broadband pulses used to excite the reflectometer were generated using a Panametrics 5800 pulser/receiver. The exciting pulse was fed into a pair of Hewlett Packard 355 C/D attenuators, which were set to 20 dB attenuation for all results presented in this paper. Use of the attenuators permitted greater control over the output pulse height than was possible using only the controls in the Panametrics 5800. On the receive cycle, the return echo passed directly into the receive input of the Panametrics 5800, where it was amplified, and then into a PC-based Gage 2125 digitizer. A Stanford Research Systems DG535 delay generator was used to control the delay between transmission of the excitation pulse and beginning of the acquisition interval. This device has a rms trigger stability of 60 ps for the data presented in this study.

Figure 2 shows the front and side views of the reflectometer. There are three major components: a stainless steel base supporting a ''U'' shaped member with ''O''-ring seals on both of its sides, a Plexiglas plate, and the transducer/witness-plate assembly. The Plexiglas plate and the transducer/witness-plate assembly are clamped onto the ''U'' shaped support using two welder's clamps (not shown). In this way, a watertight rectangular volume 2 mm thick is created in which the Albunex suspension is contained during measurement. The details of the transducer/witness-plate assembly are shown in the figure as well. It consists of a 4 in.$\times$2 in.$\times$1.5 in. Delrin™ block. A rectangular hole, having dimensions 2.5 in.$\times$1.5 in.$\times$1 in., has been machined into this block. Also shown is the electrical feedthrough hole, which permits connection of a commercial transducer (Panametrics V107, 5.0 MHz, 0.5 in. diameter, planar, contact style transducer) to the driving electronics using a coaxial cable. The active element is the Panametrics V107 transducer described above. It is partially plotted in an epoxy matrix to fix its position 3 mm below the face of the Delrin block. A 3.67 mm thick plate of pre-cured RTV615 silicone rubber is glued to the face of this transducer using RTV615. The entire assembly was heated for 72 h at 57 °C to ensure that the RTV plate and bond had cured. This was necessary to ensure the stability of the material properties over the course of the study.

The choice of RTV615 as the witness plate material was made after evaluation of several reflectometers having witness plates made of silica glass, Lexan™, Plexiglas, and agarose gel. All of these materials were found to be unsuitable, due to either lack of thermal stability (agarose), or lack of reflectometer sensitivity (glass, Lexan™, Plexiglas). The sensitivity loss accompanying the use of glass, Lexan™, Plexiglas arises from the impedance mismatch between the Albunex suspension and the witness plate material.[12,13]

The successful application of Eqs. (7) and (8) depends on the accurate determination of the material properties of the witness plate. This requires measurement of three quantities: ρ_0, α_0, c_0. Our measured value of 1.005 g/cc for the density of RTV615 agrees to within 0.5% with the published value.[14] However, the physical properties of RTV615 exhibit a slight, but reversible, dependence on temperature and on the duration of immersion in water. For instance, the thickness of one of the witness plates is 0.699 cm after sitting in air at room temperature for 24 h. Under these conditions the material is optically transparent. If the specimen is subsequently immersed in 37 °C water for 24 h its appearance will become somewhat cloudy, and its thickness will *decrease* to 0.673 cm. Consistent with this decrease is an increase in density to 1.01 g/cc.[14] This thermal behavior is typical of many rubberlike materials. As a result of this behavior it was necessary to measure the properties of RTV615 under the same conditions that were present in the reflectometer during measurement. We chose these conditions to be immersion in water for 24 h at 37 °C.

Both the attenuation coefficient, α_0, and the phase velocity, c_0, of the RTV witness plate were measured under these conditions in an immersion tank using the standard substitution technique. The resulting attenuation coefficient is shown in the top panel of Fig. 3. We have plotted the attenuation coefficient in dB/cm versus frequency (in MHz) between 1 and 10 MHz. At every fourth point of the mea-

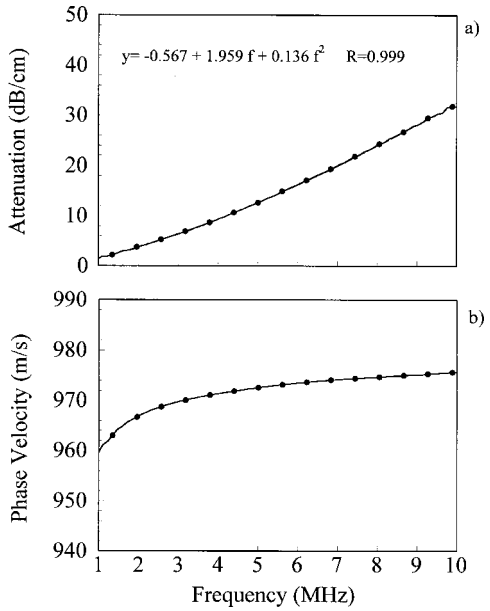


FIG. 3. The attenuation coefficient (top) and phase velocity (bottom) for the RTV witness plate used in our reflectometer. The attenuation coefficient is weakly, but definitely, quadratic, a behavior typical of many rubberlike materials.

sured data a solid black circle has been drawn. We have also plotted the best-fit quadratic curve, which has the equation:

$$\alpha(f) = -0.567 + 1.959f + 0.136\,f^2, \tag{9}$$

with a correlation coefficient of $r = 0.999$. This best fit function was used in our analysis program, which was implemented using a *Mathematica*® notebook, with a loadable *MathLink*® module for reading the raw binary data stored by the experimental apparatus. The best-fit quadratic was used due to its obvious accuracy as an approximating function, and because there is a large body of experimental evidence indicating that rubberlike materials, such as RTV615A, are well described by quadratic attenuation coefficients.[15,16]

The phase velocity, c_0, of the witness plate is shown in the bottom panel of Fig. 3. The plot extends over the same frequency range as the plot of the attenuation coefficient. The values for ρ_{Isoton}, α_{Isoton}, c_{Isoton} at 37 °C were also required for data reduction. We used a salinity of 11 parts per thousand to obtain the following properties for Isoton at 37 °C: $\rho_{\text{Isoton}} = 1.0 \times 10^{-9}$ kg/m^3, $\alpha_{\text{Isoton}} = 0.0$ dB/m, $c_{\text{Isoton}} = 1484$ m/s.[16]

II. METHODS

The reflectometer assembly was clamped together and tested for leaks. The surface of the reflectometer was then pretreated with a solution of 5% human serum albumin to prevent nonspecific binding of the Albunex microspheres to the RTV witness plate of the reflectometer. Failure to execute this step leads to inaccurate results since Albunex bubbles tend to stick to clean RTV. This results in an accumulation of bubbles on the witness plate. Because these bubbles are more closely coupled to the reflection than the

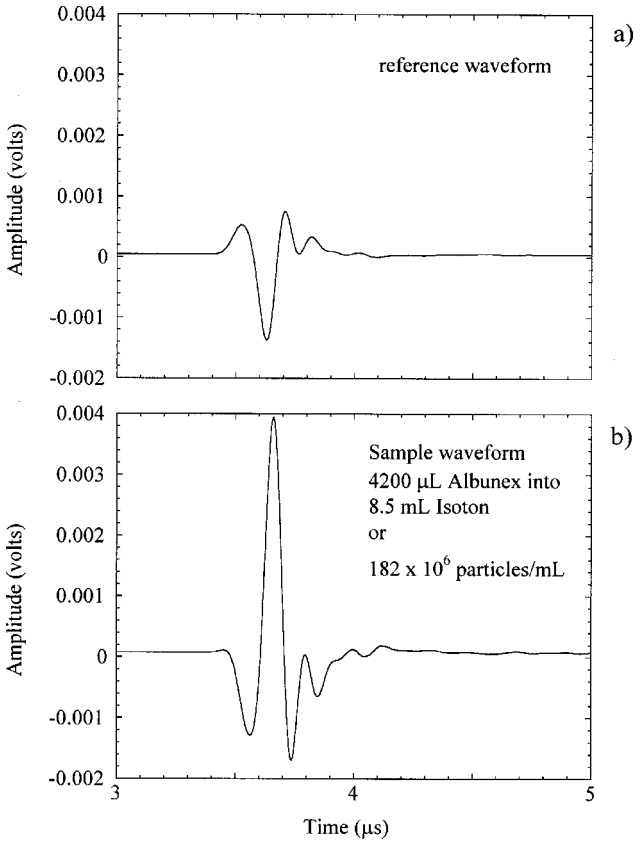


FIG. 4. A plot of typical waveforms obtained from the reflectometer. A reference (Isoton only) waveform is shown in the top panel. A sample waveform is shown in the bottom panel. The baseline must be flat for the technique described in this paper to yield accurate results. Panel (a) shows a typical reference (i.e., Isoton only) waveform. Panel (b) shows a waveform obtained after a total dose of 4.2 mL of Isoton has been introduced into the reflectometer.

remaining microbubbles in the Albunex suspension, they dominate the measurement process, leading to incorrect results for attenuation and phase velocity.

The reflectometer was then placed in a temperature-controlled water bath for the duration of the experiment together with all fixtures and samples used for this study. This was found to be necessary due to the extreme temperature sensitivity of the apparatus.

Next, a calibration or reference trace was acquired using the reflection from the RTV/Isoton interface. This was accomplished by digitizing five thousand, 2048 point waveforms at a sampling rate of 250 MHz, using a Gage 2125 8-bit, single-shot digitizer card. These were averaged together in the time domain and stored in the Gage PC's hard disk for off-line analysis. The pulse repetition rate was 1 kHz for all data described in this paper. The same digitization and averaging procedures were used for all traces acquired in this study. Figure 4(a) shows a typical reference waveform obtained from our apparatus. The horizontal axis represents time in microseconds. Its origin is placed four microseconds after the onset of the exciting pulse. The vertical axis represents the received amplitude in volts.

After acquisition of the calibration trace, the specimen traces were acquired. To obtain the first of these, 8 μL of 37 °C Albunex (equivalent to 0.69×10^6 particles/mL) was

introduced into the Isoton, using a micropipette whose tip had been preheated. A preheated mixing fixture (a nylon cable tie) was used to slowly mix the Albunex during the course of the measurement (which took under 5 s). It was found to be necessary to mix during the measurement in order to obtain accurate results. However, while mixing was underway, a stable waveform was observed on the display of the digitizer.

After the 8 μL data were acquired, an additional 8 μL of Albunex were added to the chamber in order to measure a total dose of 16 μL of Albunex. Subsequent to this measurement, an additional 16 μL of Albunex were added to the chamber to double the concentrations of microspheres. This doubling of concentrations was continued until the final dose of 4.2 mL was added to the chamber, leading to a final total dose of 8.4 mL of Albunex into 8.5 mL of Isoton (equivalent to 364×10^6 particles/mL). Between doses, all mixers and pipettes were maintained at 37 °C. This was done to minimize temperature variations over the course of the measurements. Figure 4(b) shows the average waveform obtained for the total dose of 4.2 mL of Albunex into 8.5 mL of Isoton. We note that the waveform exhibits both a time shift and a significant change in shape relative to the reference waveform. One of the main hurdles to obtaining accurate results using broadband reflectometry is precise determination of this time shift. (More will be said about this in the discussion of Fig. 8.)

The peak positive pressure at the surface of the transducer/witness plate assembly was measured using a calibrated Specialty Engineering Associates 0.4 mm diameter hydrophone and found to be a uniform 1 kPa over a circular 0.5 in. diameter region directly above the aperture of the transducer.

III. RESULTS

The attenuation coefficients obtained over the full range of concentrations are shown in the top panel of Fig. 5. The four lowest concentrations are difficult to visualize; these data are also presented in Fig. 6. The 11×10^6 particles/mL curve exhibits a peak value for the attenuation coefficient at approximately 2.5 MHz, the same value as at the lower concentrations and these data are also presented in Fig. 6. However, the peak in the attenuation coefficient shifts monotonically toward higher frequencies with increasing concentration. The theoretical models[4–9] tested in our previous publications[1–3] predict the same location for the peak attenuation coefficient regardless of dose. Hence, we interpret the data shown in Fig. 5 as an indication that at high concentrations one or more of the assumptions of these models are violated.

The measured phase velocities for all concentrations between 0.69 and 364×10^6 particles/mL are plotted in the bottom panel of Fig. 5, with the data for the four lowest concentrations again given in Fig. 6. As concentration is increased, the location of this maximum shifts toward higher frequencies. It is also interesting that at the higher concentrations, the phase velocity at 1 MHz appears to approach approximately 300 m/s; this is very close to the speed of sound in air at room temperature. Finally, we note that the

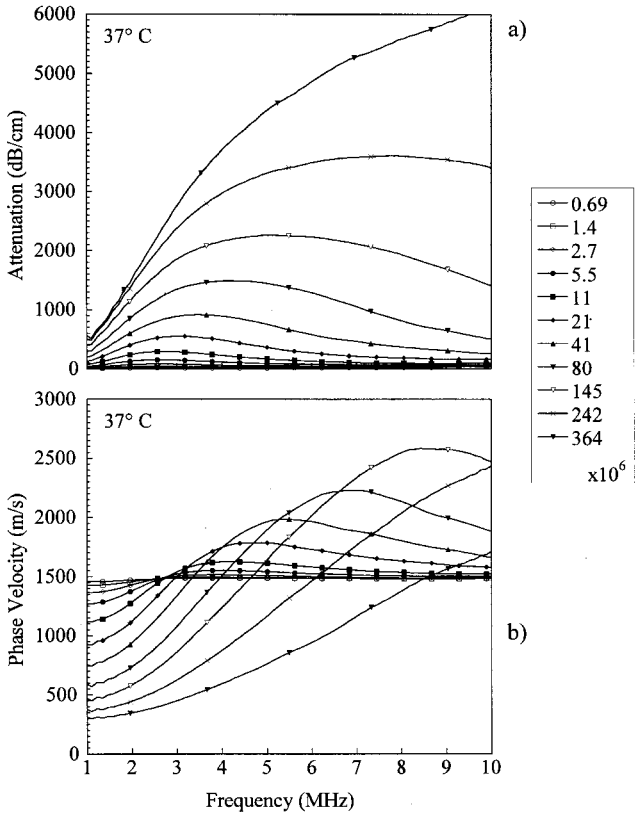


FIG. 5. The attenuation coefficient (top panel) and phase velocity (bottom panel) for 11 concentrations of Albunex. The legend shows the particles/mL for each curve in the plot.

velocity dispersion over the interval between 1 and 10 MHz, increases with increasing concentration with the exception of the two highest concentrations. At the two highest concentrations, we were unable to determine the maximum due to limited bandwidth.

For the lowest four doses, 0.69, 1.4, 2.7, and 5.5×10^6, full bandwidth through-transmission measurements were also available for comparison. These were obtained using our through-transmission, water-tank apparatus, which has been described in previous publications.[1–3] We note that these through-transmission data were acquired at a peak positive pressure of 38 kPa. This pressure was chosen since it is the lowest useable output pressure obtainable with the through-transmission apparatus, and hence provides the closest comparison that we may make with the 1 kPa pressure output of the reflectometer. In addition, previous work from our lab has shown that Albunex suspensions behave linearly with increasing acoustic pressures up to 100 kPa (Ref. 3, p. 1664). A comparison between the attenuation coefficients and phase velocities obtained in through-transmission versus reflectometry is shown in Fig. 6. In all plots, the horizontal axis extends from 1 to 10 MHz, the usable bandwidth of the reflectometer. The attenuation coefficient plots are shown in the left column and are plotted in units of dB/cm. The phase velocities are shown in the right-hand column. Through-transmission data are represented using solid triangles at every fourth point in all plots. The solid black squares represent the reflectometry data at every fourth point and are so indicated by the legend. For the attenuation data, we see that

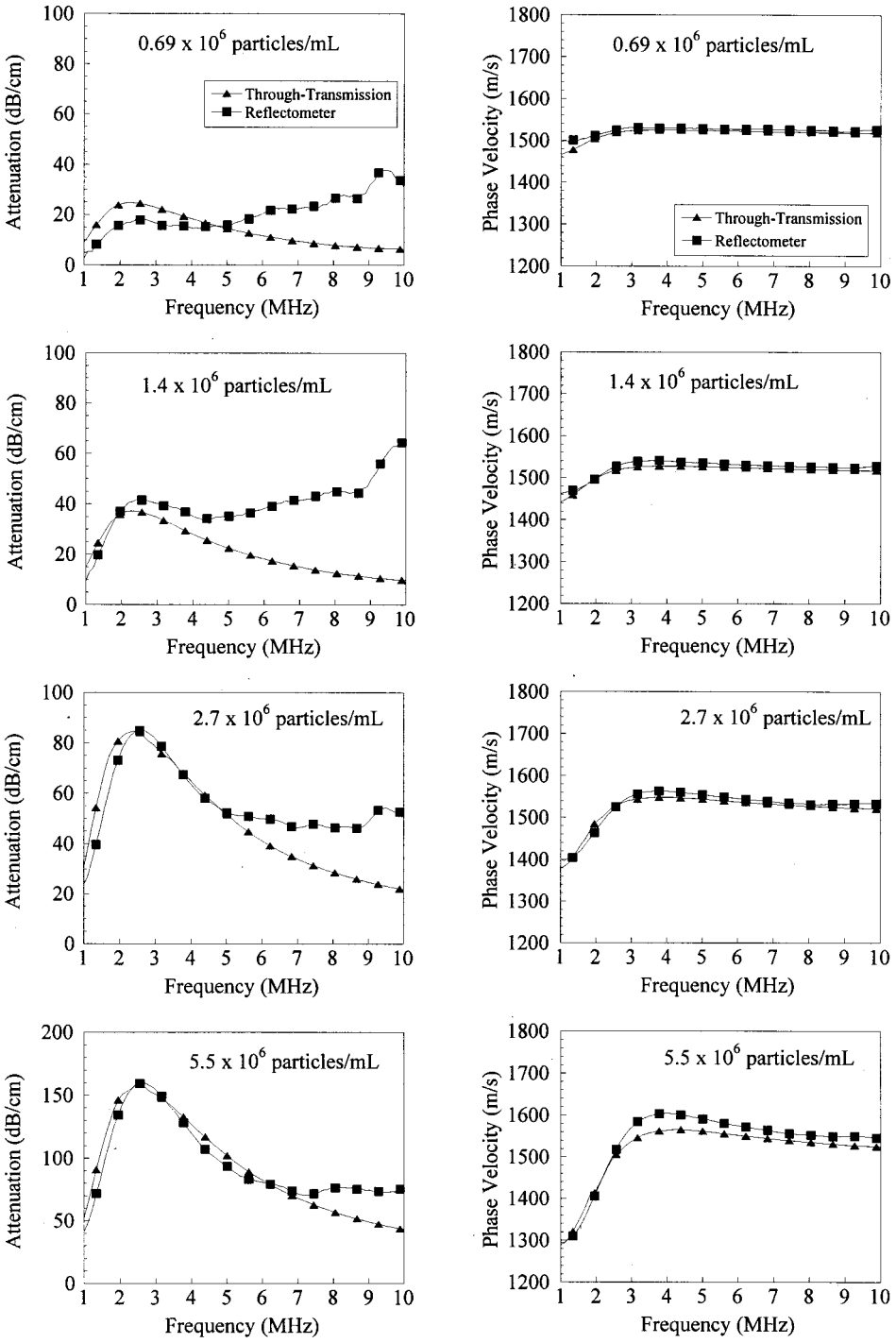


FIG. 6. A comparison of the attenuation coefficients (left column) and phase velocities (right column) of Albunex obtained via broadband reflectometry and through-transmission measurement. The solid triangles were obtained using the through-transmission technique, the solid squares using broadband time-domain reflectometry. All data were obtained at 37 °C.

both techniques exhibit a resonant peak at roughly the same location and with approximately the same height in dB/cm. However, the reflectometry data for attenuation at the two lowest concentrations exhibits a second peak between 9 and 10 MHz. These secondary peaks are an artifact, which will be discussed below.

IV. DISCUSSION

Figure 7 shows the locations of the attenuation peaks as a function of concentration with the values at the lowest concentrations taken as the positions of the local maxima agreeing with those determined using the through-transmission technique. The first four points represent the range of concentrations where the attenuation peak may be determined successfully using through-transmission techniques.[1–3] At higher concentrations, the attenuation is so great that no measurable energy is transmitted through the specimen in the frequency range near the peak. Hence, location of the peak at these concentrations is impossible to determine using through-transmission measurements. The first five data points plotted have essentially the same value. However, above the concentration 11×10^6 microspheres/mL, the location of the attenuation peak shifts toward higher frequencies with increasing dose. To the best of our knowledge, this is a new observation for Albunex, and shows that

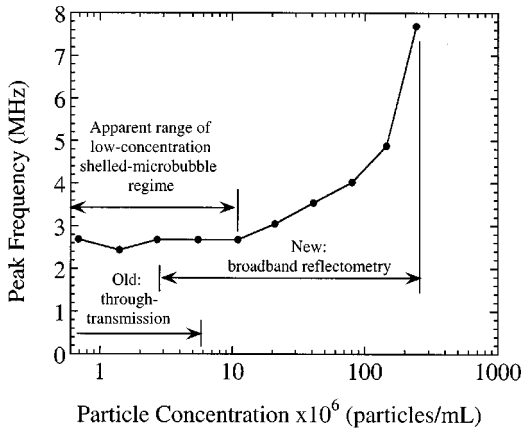


FIG. 7. A plot showing where the low concentration shelled microbubble models break down. The limit of previous through-transmission measurements (labeled as Old) is compared with the new range available using broadband reflectometry. However, all data plotted in this graph were obtained using broadband reflectometry.

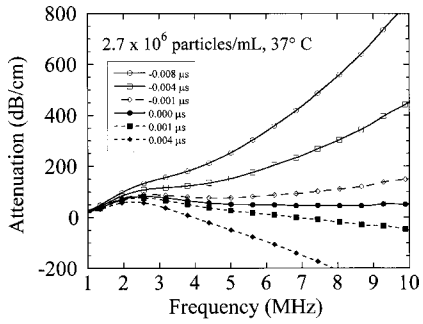


FIG. 8. These plots show the effect of time-shift errors on the reflectometry estimate of attenuation coefficient. The heavy solid black line shows the unaltered attenuation coefficient.

these measurements fall outside of the experimental regime covered by previously published models for shelled microbubble behavior.[4–9]

Figure 6 shows that quantitative agreement between the two techniques improves with increasing concentration. Furthermore, the reflectometer data exhibit a tendency toward increasing attenuation at the upper end of the useable bandwidth. This trend is more pronounced at lower concentrations. Both of these trends are to be expected based on the reflectometer data analysis and the effective medium model discussed below. Furthermore, this trend increases our confidence that the higher concentration results shown in Fig. 5 are correct.

Our experience with the reflectometer leads us to believe that very slight timing errors arising from the extreme thermal sensitivity of the apparatus are the source of the discrepancy seen in the attenuation data of Fig. 6. Early in the investigation, the experimental procedure involved the temporary removal of the reflectometer from the water bath for each re-dosing. This had the effect of permitting slight cooling of the witness plate between rinsing and re-dosing. The attenuation and phase velocities obtained in this manner exhibited large errors in comparison to through-transmission data.

The mechanisms for this error appear to be thermally induced changes in the thickness or phase velocity, or both, of the witness plate. These variations have been experimentally observed to introduce small relative time shifts between the calibration and specimen data traces. We have modeled the errors that would be introduced by these types of shifts and found that they are characteristic of the errors observed when the temperature during sample measurements is slightly below its value during the reference measurement. This modeling was performed by numerically shifting the specimen traces by nanosecond increments and computing the resulting attenuation coefficient. The shifting was accomplished by: Fourier transforming the time domain data, multiplying by a complex phase shift $e^{i\omega\tau}$, and then inverse Fourier transforming, to obtain a new data set which has been shifted by τ nanoseconds. In this way, shifts smaller than the experimental sample interval (4 ns) were obtained. The effect on the calculated attenuation coefficient resulting from time shifting by $-8,-4,-1,0,1$, and 4 ns (i.e., shifting by $-2\Delta t$, $-1\Delta t$, $-0.25\Delta t$, $0\Delta t$, $0.25\Delta t$, and $1\Delta t$, where $\Delta t=4$ ns) is shown in Fig. 8. This figure makes use of the raw data set acquired from a suspension having a concentration of 2.7×10^6 particles/mL. The heavy solid line is the unaltered attenuation coefficient resulting from not shifting the specimen data prior to analysis. For positive time shifts, equivalent to adding a small delay to the actual specimen trace, an apparent attenuation coefficient that becomes negative at lower and lower frequencies is obtained. The greater the shift, the sooner the apparent attenuation coefficient becomes negative. For negative time shifts, there is a significant overestimation of the attenuation coefficient, which increases as frequency increases. This overestimate worsens as the size of the shift is increased. In fact, this particular type of error can sometimes lead to a predicted attenuation that is nearly linear with frequency. For our through-transmission measurements of low concentration microbubble suspensions, we know that the attenuation should exhibit a resonant peak between 2 and 3 MHz. Hence, the error is immediately obvious. However, for measurement of materials having a linear attenuation coefficient, such as plastic, metal, ceramic, or composite plates, or for measurement of adhesive bonds, this effect may represent an especially insidious source of error.

We modeled the impact of timing errors on the attenuation coefficient because this quantity appears to be much more sensitive to measurement errors than the phase velocity. This is due to the presence of the frequency-dependent term in Eq. (8). This term has the effect of magnifying any experimental errors present in the data, such as the time-shift errors modeled in Fig. 8. The importance of time-shift errors can be expected to decrease as concentration is increased; at higher concentrations the magnitude of the mismatch in acoustic impedance increases, thus decreasing the relative impact of small but ever-present errors in signal timing. This reasoning is consistent with the data shown in Fig. 6. At the two highest concentrations shown, agreement between the two measurement techniques appears to be good for the attenuation coefficient at all frequencies below 5 MHz (for 2.7×10^6 particles/mL) and 7 MHz (for 5.5×10^6 particles/mL). For these frequency ranges the small discrepancies observed in the comparison plots are of the same order of mag-

nitude as we have observed in through-transmission measurements made in our laboratory.[17] In contrast, for phase velocities the reflectometry and through-transmission techniques show agreement to within 3.5% at all frequencies; see Fig. 6. This is might also be expected based on Eq. (7), which does not require multiplication by an error-magnifying term and hence, is much less sensitive to experimental errors.

The width of the pulses shown in Fig. 4 are also important to understanding the limitations of the measurement technique described in this study. We postulate that the pulse width places an upper bound on the depth of the Albunex suspension that may be probed by the reflectometry measurement. The size of this boundary layer may be estimated by dividing the duration of the pulse, approximately 1 μs, by the speed of sound in the Albunex suspension. *A priori*, we may only estimate this speed, which we do by using the speed of sound in water at 37 °C (1494 m/s).[16] At best, this can only be a crude estimate since, as we have shown in Figs. 5(b) and 6, Albunex suspensions have a large velocity dispersion. If, however, we assume that the velocity is different from water by no more than a factor of 5 over the useful bandwidth (and the data in Fig. 5 bear this assumption out), the boundary layer is at least 300 m/s$\times 10^{-6}$ s$=$300 microns thick. This thickness is at least 60 bubble-radii for Albunex, which has a maximum bubble diameter of 10 microns. Thus during the time period of the ultrasonic pulse, the witness plate generates mechanical effects that interrogate a boundary layer of the Albunex suspension 60 bubble radii thick. This is important since the formalism underlying our method assumes homogeneous media on both sides of the interface. To apply this formalism, we must think of the Albunex suspension as an effective medium having a compressibility which is representative of the average properties of the microbubble/liquid mixture. At least two conditions must be met for this approximation to be credible: the size of the microbubbles must be small relative to the interrogating wavelength, and there must be a large number of microbubbles in the boundary layer. The extent to which these criteria have been met for our experimental system can be assessed from Fig. 6. Thinking of the experiment in this way also makes it obvious that the suspension must be well mixed during the course of the experiment. Indeed, implementation of the appropriate mixing protocol and reflectometer configuration was one of the major hurdles we had to overcome in order to obtain accurate data. In addition, the accuracy of the technique can be expected to increase with increasing microbubble concentration. This supposition is born out in Fig. 6, which shows increasing agreement between the two techniques as microbubble concentration is increased.

V. CONCLUSION

We have measured the attenuation coefficients and phase velocities for a range of Albunex concentrations between 0.69×10^6 particles/mL and 364×10^6 particles/mL using the technique of broadband reflectometry. These concentrations correspond to dilutions ranging from 0.1% to 50% and to nominal gas volumes of 0.0025% to 1.4%. At lower concentrations, where through-transmission measurements are also available, the techniques agree well, thus validating the reflectometer method. The basis for this measurement technique is an effective medium approximation, in which the suspension of microbubbles in a fluid medium is assumed to behave as a homogenous medium. The conditions necessary to make this approximation were shown to be valid in our apparatus. These conditions also dictate the need to continuously mix the suspension during data acquisition. Several other conditions must also be met. Temperature stability must be maintained over the course of the measurement, and good matching of acoustic impedance between witness plate and specimen is required. Although the technique described here has been applied to Albunex suspensions, it should be equally applicable to other liquid suspensions having high attenuation.

[1] J. N. Marsh, C. S. Hall, M. S. Hughes, J. Mobley, J. G. Miller, and G. H. Brandenburger, ''Broadband through-transmission signal loss measurements of Albunex suspensions at concentrations approaching in vivo does,'' J. Acoust. Soc. Am. **101**, 1155–1161 (1997).

[2] J. Mobley, J. N. Marsh, C. S. Hall, M. S. Hughes, G. H. Brandenburger, and J. G. Miller, ''Broadband measurements of phase velocity in Albunex,'' J. Acoust. Soc. Am. **103**, 2145–2153 (1998).

[3] J. N. Marsh, M. S. Hughes, C. S. Hall, S. H. Lewis, R. L. Trousil, G. H. Brandenburger, H. Levene, and J. G. Miller, ''Frequency and concentration dependence of the backscatter coefficient of the ultrasound contrast agent Albunex,'' J. Acoust. Soc. Am. **104**, 1654–1666 (1998).

[4] N. de Jong, ''Acoustic properties of contrast agents,'' Ph.D. Thesis, Rotterdam (1993).

[5] N. de Jong, N. Ten Cate, F. J. Lancee, C. T. Roelandt, J. R. C. T. , and N. Bom, ''Principles and recent developments in ultrasound contrast agents,'' Ultrasonics **29**, 324–330 (1991).

[6] N. de Jong, L. Hoff, T. Skotland, and N. Bom, ''Absorption and scatter of encapsulated gas filled microspheres, theoretical considerations and some measurements,'' Ultrasonics **15**, 95–103 (1992).

[7] N. de Jong and L. Hoff, ''Ultrasound scatter properties of albunex microspheres,'' Ultrasonics **31**, 175–181 (1992).

[8] T. G. Leighton, *The Acoustic Bubble* (Academic, New York, 1994).

[9] C. C. Church, ''The effects of an elastic solid surface layer on the radial pulsations of gas bubbles,'' J. Acoust. Soc. Am. **97**, 1510–1521 (1995).

[10] J. G. Miller and D. I. Bolef, ''Technique for the study of highly attenuated ultrasonic waves in gases,'' Rev. Sci. Instrum. **41**, 1305–1306 (1970).

[11] J. N. Marsh, ''Qualitative acoustic characterization of ultrasonic contrast agents,'' Ph. D. Thesis, Washington University in St. Louis (1998).

[12] P. O. Lipovko and V. M. Lube, ''Acoustical testing of the properties of a medium by measurement of its characteristic impedance,'' Sov. Phys. Acoust. **17**, 135–137 (1971).

[13] J. J. Brown, V. E. Stubblefield, and J. G. Miller, ''Optimizing signal-to-error ratio in standing wave ultrasonic measurements,'' Proceedings of the 1974 IEEE Symposium in Ultrasonics, 74CH0896-1SU, pp. 549–554.

[14] A. Selfridge, ''Approximate material properties in isotropic materials,'' IEEE Trans. Sonics Ultrason. **SU-32**, 381–394 (1985).

[15] R. D. Corsaro and J. D. Klunder, ''A filled silicone rubber materials system with selectable acoustic properties for molding and coating applications at ultrasonic frequencies,'' NRL Report 8301 (1979).

[16] J. R. Lovett, ''Comments concerning the determination of absolute sound speeds in distilled and seawater and Pacific Sofar speeds,'' J. Acoust. Soc. Am. **45**, 1051–1052 (1969).

[17] M. S. Hughes, J. N. Marsh, N. A. Schmich, A. L. Klibanov, G. L. Cantrell, J. G. Miller, and G. H. Brandenburger, ''Behavior of resonant peak of attenuation of albunex at varying power levels and durations,'' Ultrason. Imaging **20**, 544–545 (1998).

A class of chaotic bird calls?

N. H. Fletcher[a)]
School of Aerospace and Mechanical Engineering, Australian Defence Force Academy, Canberra 2600, Australia

(Received 9 February 2000; accepted for publication 16 May 2000)

Evidence is presented that the basic vocalized sound produced by some cockatoos, specifically the Australian sulfur-crested cockatoo (*Cacatua galerita*) and the gang-gang cockatoo (*Callocephalon fimbriatum*), has a chaotic acoustic structure rather than the harmonic structure characteristic of most birdsongs. These findings support those of Fee *et al.* [Nature (London) **395**(3), 67–71 (1999)] on nonlinear period-doubling transitions in the song of the zebra finch (*Taeniopygia guttata*). It is suggested that syllables with chaotic structure may be a feature of the songs of many birds. © *2000 Acoustical Society of America.* [S0001-4966(00)04108-4]

PACS numbers: 43.80.Ka [WA]

INTRODUCTION

Birdsong has been the subject of study for many years, from physiological, acoustical, and behavioral points of view. A comprehensive account is given in the classic book by Greenewalt[1] and there are many more recent papers to which we do not need to refer here. An excellent recent account of the relation between physiology and birdsong has been given by Suthers *et al.*[2]

Until recently, only three types of birdsong have been recognized, which we now discuss in turn. They might be termed simple voiced song, double-voiced song, and whistled song, respectively.

Simple voiced song is characterized by an acoustic spectrum consisting of a series of exact harmonics of a fundamental frequency F_0 (to use a notation common in human phonetics). The envelope of the spectrum of this song at any instant is characterized by a set of formant bands at frequencies $F_1, F_2, F_3, \ldots$ in which the amplitude of the harmonics of F_0 rises above the general declining trend with increasing frequency. The frequencies of the fundamental F_0 and of the formant bands F_n may vary with time under voluntary control by the bird, the extent of the variation being very different for different species of bird.

This simple voiced song is closely analogous to human speech and song generally, and to human vowel sounds in particular. The physiological interpretation is given in terms of a ''source-filter model,'' as for human speech. The avian syrinx acts as a self-excited oscillator, driven by air pressure from the lungs to vibrate at a characteristic frequency F_0 that is determined by the mass and tension of the syringeal membranes. This vibration in turn modulates air flow into the upper vocal tract at the characteristic frequency F_0. Because the flow equations for the syringeal valve are nonlinear, this flow U contains a complete spectrum of exact harmonics nF_0 of the fundamental, with amplitudes U_n generally decreasing steadily with increasing frequency. The acoustical properties of the upper vocal tract can be expressed in terms of the frequency-dependent acoustic impedance $Z(f)$ that it presents to air flow from the syrinx, and the formant bands occur around the resonances of the vocal tract for which this impedance is a maximum, the acoustic pressure associated with the flow component U_n being $p_n = Z(nF_0)U_n$. Integration of these concepts into a calculable model gives good agreement with experiment.[3] The frequency F_0 of the fundamental can be varied by varying the muscle tension on the syringeal membranes, while the frequencies F_n of the vocal tract resonances can be controlled by changing tongue position and beak opening.[4] Details of the mathematical approach to calculations such as these have been given elsewhere.[5]

In the case of song-birds, account must be taken of the fact that they possess two independent syringeal valves, one located in each bronchus. There are, in fact, two possibilities that are exploited differently by different species. In one case, the bird simply closes off one syringeal valve during song and uses the other, perhaps permanently and perhaps on an alternating basis. In the second case, both syringeal valves operate at the same time, but their oscillations are locked into synchronism by the common oscillating pressure at the base of the trachea. Such frequency and phase locking is common in many types of oscillators, and requires only that their natural frequencies be not too far apart and that there exist a nonlinear physical coupling mechanism between them.[6] These conditions are easily met in the avian vocal system, where the two syringeal valves are at least nominally identical in structure, the flow through them is a nonlinear function of pressure, and the tracheal pressure oscillation provides an effective coupling mechanism.

The second type of song, which might be called double-voiced song, is one in which the frequencies of the two syringeal valves are controlled to be so different that locking is impossible. When this happens, each bronchus feeds a flow signal to the trachea at its characteristic oscillation frequency $F_0^{(1)}$ or $F_0^{(2)}$. Each of these flows contains harmonic components at integer multiples of its fundamental frequency but, in addition, there will be nonlinearly generated components at multiple sum and difference frequencies $nF_0^{(1)} \pm mF_0^{(2)}$, where n and m are positive integers. The amplitudes of these mixture terms decline approximately as x^{n+m}, where x is a

[a)]Permanent address: Research School of Physical Sciences and Engineering, Australian National University, Canberra 0200, Australia.

quantity less than unity that depends upon the exact nature of the flow nonlinearity and is proportional to the amplitude of the fundamental of the oscillation. Songs of this type sound rather like pairs of notes played on musical instruments, provided the nonlinearity is not too great.

The third generally recognized type of song is termed whistled song. Analysis shows that it consists of an almost pure sine wave with no upper harmonics, and the frequency can often be changed rapidly in a sweep over a range of about a factor two. The physiological mechanism for production of such song has not been established. The fact that whistles are often interpolated within a voiced song suggests a common generation mechanism, and it has been suggested that this might be a retraction of the syringeal membranes while in oscillation so that they no longer completely close, leading to a great reduction in the harmonic content of the flow. Flow calculations, however, indicate that, while such a procedure would reduce the relative amplitude of higher partials in the flow, it would not eliminate them entirely, as appears to be the case in whistled song. It should also be noted that some human languages, notably those from central Africa, involve the interpolation of aerodynamically produced whistles amidst voiced speech.

An alternative explanation of whistled song is that it is produced by pure aerodynamic means without any vibrating surfaces, in much the same way as sound in flutes, organ pipes and whistles. In all these cases the source of sound is the interaction of an unstable air jet with a resonator. If the resonator has many modes, as in an organ pipe, then the sound has many harmonic overtones. If, however, it has but one mode, as in the case of the Helmholtz resonator of an occarina, which is a simple cavity vented by finger holes, then the sound output will approximate a pure sine wave. A simple biological example is human whistling, in which the resonator is the mouth cavity and the jet issuing through the aperture between the lips is subject to a varicose instability (or change in diameter) which feeds back to influence flow through the aperture and thus internal pressure. It is not clear what structures in the avian vocal system might be responsible for sound production by this mechanism. Possible candidates are structures at the base of the trachea, but another possibility involves the larynx, a raised tongue, and a further aperture between tongue and beak.

It is the purpose of the present note to present further evidence for a modified type of voiced utterance that might be called chaotic song. The possibility that some bird calls might be chaotic was raised by the present author some time ago[7] in relation to the Australian sulfur-crested cockatoo *Cacatua galerita*, and since then a pioneering study of a transition from normal to period-doubled and perhaps chaotic song in the case of the zebra finch *Taeniopygia guttata* has been published by Fee *et al.*[8,9] The difference between the calls reported here and those of the zebra finch will be discussed briefly in the conclusion to the present paper.

I. ANALYSIS OF CHAOTIC CALLS

To the ear, the calls of Australian cockatoos have the sound of a steady raucous screech, quite unlike the more melodious sounds of other birds. This observation was the stimulus of the present investigation. A good collection of the songs of Australian birds of the region near Canberra, recorded in their native habitat, is available,[10] and this was used for the required sound samples. The two cockatoos selected for study were the sulfur-crested cockatoo (*Cacatua galerita*) and the gang-gang cockatoo (*Callocephalon fimbriatum*). Each selected call was digitized at 32 kb/s and its waveform examined. A section of nearly constant amplitude and duration about 0.7 sec was then selected for detailed analysis. From an auditory point of view these sections of the call were constant in loudness and sound quality, and this constancy was confirmed by the nature of the oscillograph records of the waveforms, suggesting strongly that they are the result of processes that are ''stationary'' in a statistical sense. Two similar segments from the songs of birds with simple sound, namely the boobook owl (*Ninox novaeseelandiae*) and the Eastern rosella parrot (*Platycercus eximius*), were similarly studied for comparison purposes.

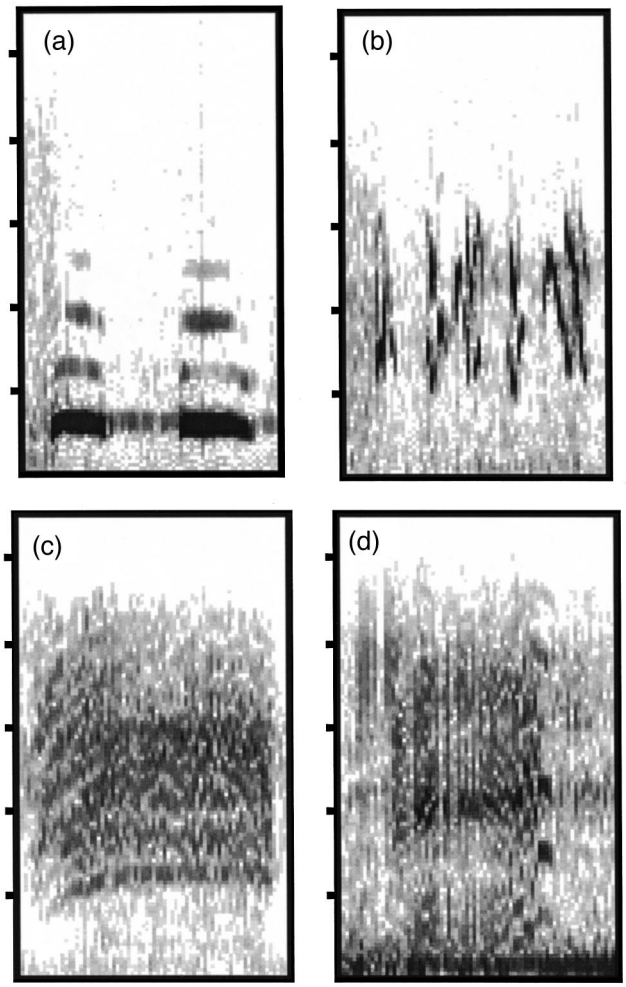


FIG. 1. Time-resolved spectrograms of the calls of the four birds in the study: (a) the boobook owl, (b) the Eastern rosella parrot, (c) the sulfur-crested cockatoo, and (d) the gang-gang cockatoo. In each case the duration of the sample is about 0.7 sec and the frequency range is 0 to 5.5 kHz.

As an initial study, sound spectrograms of the four bird calls were made, and the results are shown in Fig. 1. The call of the owl, in Fig. 1(a), has a simply harmonic structure and no pitch change, while that of the parrot, in Fig. 1(b), consists essentially of a single simple tone with rapid frequency

variations. In contrast, the sulfur-crested cockatoo call in Fig. 1(c) has a very broad-band spectrum, though with traces of two independent voices and vestiges of quasiharmonic structure. The call of the gang-gang in Fig. 1(d) is broadly similar, but with a higher pitch and less obvious substructure. It appears reasonable to treat the major part of these latter two songs as being stationary in a statistical sense.

It was the assumption at the beginning of the study that the calls in question might have a chaotic structure, and the digitized records were therefore subjected to analysis on this basis. Such an analysis of time-series data to detect chaos and other interesting phenomena has been the subject of much detailed study,[11,12] and a set of computer programs[13] with the title ''Chaos Data Analyzer'' is available to carry out the analysis. In addition, a newly available suite of programs[14] for ''Visual Recurrence Analysis'' provides an independent and rather different approach. These programs were applied to the observational records with the results detailed below.

The essential feature of normal nonchaotic song is its predictability. As discussed above, the short-term spectrum consists of an assembly of pure sinusoidal tones that are generally in harmonic relationship, giving an exactly repeating waveform. When two syringeal sources produce sounds at unrelated frequencies the overall harmonic relationship is complicated by multiple sum and difference tones and the waveform no longer repeats. In either case, however, once the spectral composition of the sound is known, its future waveform can be predicted exactly. This statement applies, of course, only to short segments of song in which the bird does not deliberately change the defining parameters such as pitch and loudness.

At the other end of the scale comes random noise. Here the future waveform is entirely unpredictable and the signal can be described only in statistical terms.

Between these two extremes lies chaotic behavior, in which the oscillation is governed by well-defined and often simple laws but, because of nonlinearity in the basic vibration mechanism, the future waveform is unpredictable in the absence of knowledge of the precise initial conditions—a very small change in initial conditions makes an immense change in the exact future course of the oscillation.[12] Chaos is often studied by examining the behavior of the system when the parameters of the underlying differential equations are progressively changed. In a natural biological system such as birdsong such an approach is not possible; rather we must attempt to find something of the nature of the underlying equations by examining the sound output.[11]

A. Data analysis

The syrinx is, of course, a complex vibrating system, since the pressure-controlled vibrating valve, whether it be the syringeal membrane or some other structure, can support a large number of possible oscillation modes. It therefore turns out not to be possible to discover much about the exact nature of its vibration from the sound output, but a demonstration that it is indeed chaotic is relatively simple. The approach is to calculate the so-called Lyapunov exponents that describe the waveform. In essence this exponent measures the degree of sensitivity of the oscillation to its initial conditions and its deviation from predictability. First the waveform is mapped onto a phase space which, in the simplest case, plots the slope of the waveform at each point against the magnitude of the waveform at that point. For a wave consisting of exactly harmonic components, this phase map consists of a simple closed curve, while for more complex waveforms the map is correspondingly more involved. The Lyapunov exponent essentially measures the rate at which the distance between two points on the curve (related by the sampling time between them) increases as the waveform evolves. For a simple harmonic or multifrequency deterministic system, the distance between the points settles down to a constant value, on average, while for a chaotic system the distance increases steadily. The Lyapunov exponent measures the exponential rate of this increase in separation. Chaotic systems are characterized by moderate positive exponents and simple systems by negative or zero exponents. Random systems have large positive exponents.

TABLE I. Parameters computed from CDA program.

	Lyapunov exponent	Capacity dimension	Correlation dimension
Two nonlinearly coupled signals	0.0	1.4	2.2
Henon attractor	0.6	1.4	1.3
Lorenz attractor	0.1	1.7	2.0
Random noise	0.15	2→4	1→6
Boobook owl	0.06	1.5	1.9
Rosella parrot	0.02	2.5	2.1
Sulfur-crested cockatoo	0.3	2.2	3.2
Gang-gang cockatoo	0.2	2.3	3.8

The Chaos Data Analyzer programs[13] provide facilities for plotting phase-space maps, calculating the largest Lyapunov exponent, and otherwise examining time-series data. As a control, short segments of song from the boobook owl (*Ninox novaeseelandiae*) and from the Eastern rosella parrot (*Platycercus eximius*) were also analyzed. The boobook song was nearly sinusoidal, with a single dominant frequency, giving a simple closed-loop phase map and a largest Lyapunov exponent of 0.06 ± 0.03. The rosella song contained two dominant frequencies, had a more complex but still generally ring-shaped phase map, and a largest exponent of 0.02 ± 0.02. Both these exponents are essentially zero, as would be expected for deterministic signals. Further details of the analysis are given in Table I, which also shows the capacity dimension (Hausdorff dimension) and correlation dimension calculated from the data. For both of these songs, the correlation function had the form of a cosine function and did not decay in amplitude over the sample length, indicating a closely predictable behavior. The phase-space plot in both cases had the form of a broad elliptic ring, as would be expected for a basically sinusoidal signal modulated by other frequencies. Comparison with computed data for a signal consisting of two nonlinearly interacting unrelated harmonic oscillations, given in the first line of Table I, indicates close similarity and confirms the predictable nature of these songs, despite their irregular waveform. We return to discuss such songs in a later section.

The contrast with the calls of the sulfur-crested cockatoo

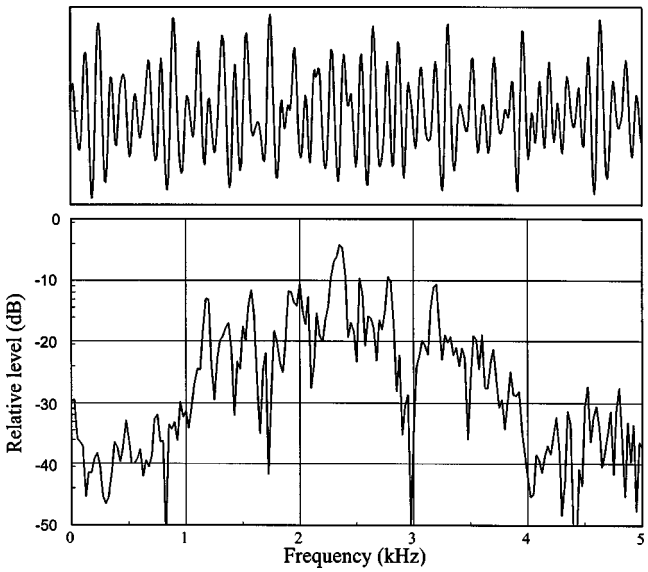


FIG. 2. (a) Waveform and (b) spectrum of a short sample of the song of the sulfur-crested cockatoo. Plots for the gang-gang cockatoo are very similar.

and the gang-gang cockatoo is marked. While the waveform of the song of the boobook owl and rosella parrot appears ''smooth'' in each case, that of the two cockatoos has a ''rough'' appearance, as shown for the case of the sulfur-crested cockatoo in Fig. 2(a). Each cockatoo call had a broad power spectrum peaked at about 2.5 kHz and with many subsidiary peaks, as shown for a short segment of the sulfur-crested cockatoo call in Fig. 2(b). This confirms the analysis provided by the spectrograms of Fig. 1. The quasiperiodic components are, however, of large amplitude. The largest Lyapunov exponent for the sulfur-crested cockatoo call was 0.28 ± 0.06 and that for the gang-gang 0.23 ± 0.05. As shown in Table I, these values are comparable with those computed for typical chaotic signals, such as those associated with the Lorenz and Henon systems, and indicate a lack of predictability. For each bird the correlation function has the form of a decaying cosine wave, as shown for the sulfur-crested cockatoo in Fig. 3, the decay rate being rather more rapid in the case of the gang-gang. In neither case, however, is the decay as rapid as for the Lorenz or Henon systems, and there appears to be significant sinusoidal residue.

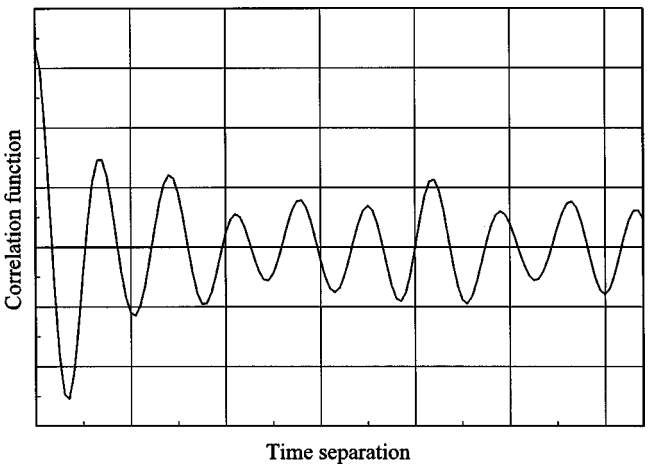


FIG. 3. Computed correlation function for the song of the sulfur-crested cockatoo.

From the data given in Table I it is clear that there is a close resemblance between the parameters for these two cockatoo calls and those for the well-known Lorenz and Henon chaotic attractors. These two cases, however, derive from analysis of well-defined and simple nonlinear differential equations and lead to phase-space maps that are fractal ''strange attractors.'' In the case of the cockatoo calls, the phase-space maps essentially filled uniformly an elliptical area, and no further information could be gained from them, even if the embedding dimension was raised. The explanations for this are probably that the call is not truly stationary in a statistical sense, and that it also contains noise, both of which features tend to blur any phase-space pattern, although it could also result from the dimensionality of the system being higher than 3. The further fact, however, that the computed correlation dimension, regarded as a function of the embedding dimension, rises to a gently sloping plateau value for an embedding dimension greater than about 3, distinguishes the call structure from simple band-limited noise, for which the correlation dimension rises smoothly to large values with increasing embedding dimension.

B. Visual recurrence analysis

As a second approach to understanding the nature of the cockatoo call, the technique known as visual recurrence analysis[14] was used. In this approach, the signal is sampled at a set of equally spaced times t_i and at each time the following N sample values are used to define an N-dimensional vector Y_i associated with the time t_i. A color-coded matrix plot is then made of the Euclidean distance between all pairs of vectors, such that the (i,j)th element of matrix is the distance between the vectors Y_i and Y_j. Although it is difficult to draw any quantitative conclusions from such a plot, it does show up patterns in the data in a very clear manner—a well-correlated signal gives a repetitive pattern, a chaotic signal has an irregular but definite pattern, and simple noise has a pattern that is irregular at all scales.

Figure 4(a) shows a recurrence plot for a signal consisting of two incommensurate nonlinearly interacting harmonic signals, as discussed in the next section, and Fig. 4(b) a plot to a similar scale for a random noise signal. These figures set the range of variation to be expected. Figure 5(a) then shows the recurrence plot for the Lorenz attractor, while Fig. 5(b) shows a plot to a similar scale for the cockatoo call. The similarity between Figs. 5(a) and 5(b) is clear, as is their difference from the plots of Fig. 4. We again conclude that the call of the cockatoo likely has a chaotic nature.

II. PHYSICAL MODEL

It is important to seek the reasons why the cockatoo calls might be chaotic, or at least the mechanism that might lead to this result. For the present this is largely speculation, but as such is necessary as a guide to further studies. In particular it is important to distinguish chaotic song from the type of nonlinearly mixed two-voice song described in the Introduction.

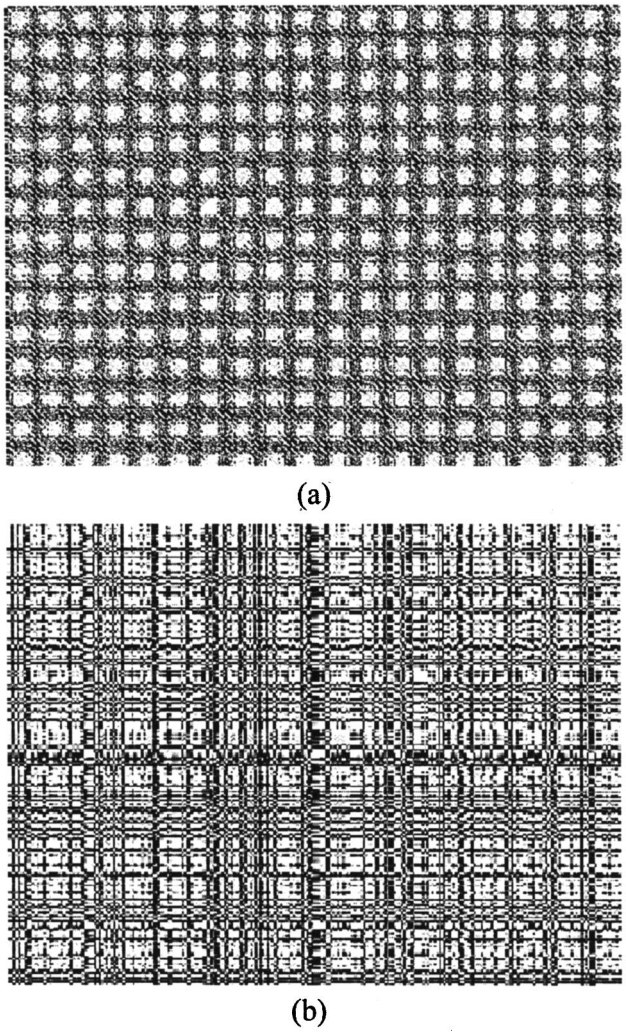


FIG. 4. Visual recurrence analysis plot of (a) two nonlinearly coupled harmonic signals, as defined by Eqs. (1)–(3), and (b) a random noise signal. (The original plots were in color.)

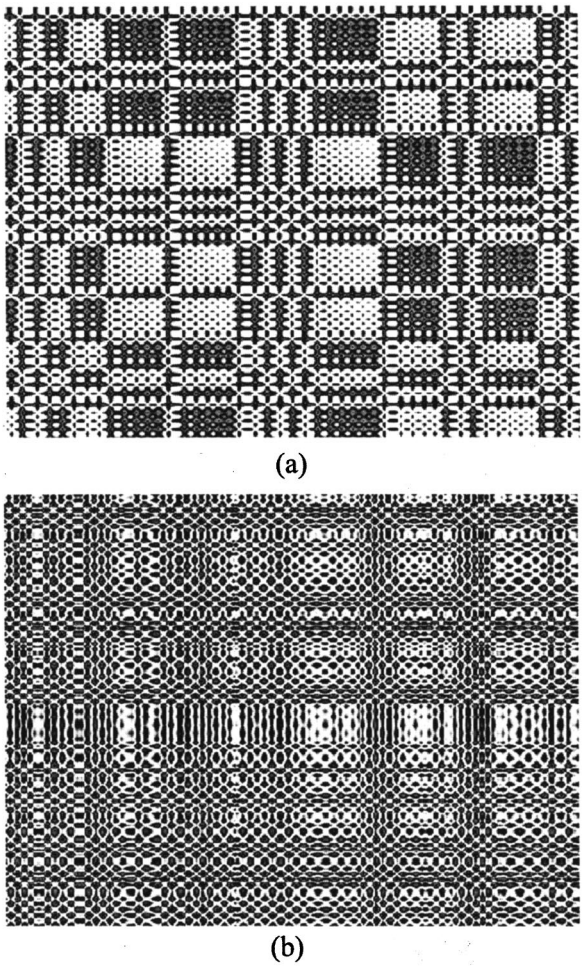


FIG. 5. Visual recurrence analysis plot of (a) the Lorenz attractor time series, and (b) the song of the sulfur-crested cockatoo. The recurrence plot for the Henon attractor is qualitatively very similar to that for the Lorenz attractor. (The original plots were in color.)

Although the largest Lyapunov exponent is significantly different from zero, the waveform in Fig. 2(a) has at least one strong quasiperiodic component, which is also visible on the spectrogram in Fig. 2(b). In addition, the correlation function of Fig. 3 also has a periodic structure, although one that decays toward zero with time. These facts suggest that the mechanism might be one in which the syringeal valves in the two bronchi of the bird oscillate simultaneously and are coupled by a strong interaction through the sound pressure at the base of the trachea. The fact that the bird has a very loud cry, with an estimated radiated power of at least 100 mW, implies an acoustic pressure at the base of the trachea that is perhaps as high as 1 kPa (154 dB). Such an acoustic pressure is comparable to the blowing pressure below the syrinx and can therefore strongly influence the vibration of the syringeal valves. Furthermore, the airflow through these valves is a highly nonlinear function of pressure drop across them, and the membrane vibrations themselves are probably quite nonlinear because they are likely to contact the wall once in each cycle.

With two nonlinear vibrating sources and a strong nonlinear coupling between them, there is an adequate number of degrees of freedom for chaotic oscillation, rather than simply for the generation of multiple sum and difference frequencies. At the same time, the natural frequencies of the two syringeal valves may well contribute quasiperiodic components to the overall oscillation, as observed.

On the other hand, it is possible that the two nonlinear membrane vibrations simply couple nonlinearly to produce a multitude of multiple sum and difference frequencies that gives the appearance of chaos. To test this, a synthetic signal $y(t)$ was created from two harmonic signals $x_1(t)$ and $x_2(t)$ combined nonlinearly. Specifically,

$$x_1(t)=\sin(\omega_1 t)+a_1\sin(2\omega_1 t+\phi_1)+a_2\sin(3\omega_1 t+\phi_2), \quad (1)$$

$$x_2(t)=\sin(\omega_2 t)+a_1\sin(2\omega_2 t+\phi_3)+a_2\sin(3\omega_2 t+\phi_4), \quad (2)$$

$$y(t)=[x_1(t)+x_2(t)]+b_1[x_1(t)+x_2(t)]^2+b_2[x_1(t)+x_2(t)]^3, \quad (3)$$

where the fundamental frequencies ω_1 and ω_2 have no simple integer relationship ($\omega_2=1.3824\omega_1$), and the amplitudes a_n and b_n are given as $a_1=0.3$, $a_2=0.1$ and $b_1=b_2$

$=0.1$. The waveform and spectrum of this signal showed patterns broadly similar to those recorded for the cockatoo cries. As shown in Table I, however, the largest Lyapunov exponent in this case is essentially zero, the actual value given by the CDA program being -0.01 ± 0.03. The correlation function is irregular but shows no decline in amplitude with time, while the visual recurrence plot, shown in Fig. 4(a), has an obvious regularly repeating structure. We conclude that a simple nonlinear mixing of two periodic signals, as might indeed be achieved in the syrinx, cannot explain the nature of the observed sound signal.

The final possibility to be investigated is that the signal consists of a nonlinearly combined pair of harmonic signals, as just discussed, with a significant admixture of random noise. A signal with this property was constructed, the ratio of peak noise to peak signal being either 0.1:1, 0.3:1 or 1:1. The CDA program was then used to determine the largest Lyapunov exponent for the signal and also the behavior of capacity and correlation dimensions with increasing embedding dimension. For noise ratios of 0.1:1 and 0.3:1, the largest Lyapunov exponent was less than 0.01, indicating nonchaotic behavior. For a noise ratio of 1:1, the calculated Lyapunov exponent was 0.2, but the capacity and correlation dimensions increased steadily with increasing embedding dimension, as for an essentially random-noise behavior. It can be concluded that the bird cry is not adequately simulated by such signal.

III. CONCLUSIONS

The conclusions to be drawn from this brief examination of the calls of these particular birds appear clear. The songs of birds such as the boobook owl and the rosella parrot are basically sinusoidal, but with rapid modulations in the case of the rosella. Such songs are quite normal variations of basic voiced or whistled song. The basic repetitive ''carrier'' is provided by vibration of the syringeal membranes, and its amplitude and frequency are modulated at a slower rate by oscillatory changes in the tension of the supporting muscles.[2] These songs can therefore be classified as ''normal'' in the present context. This conclusion is supported by the facts that the largest Lyapunov exponent is approximately zero, that the phase-space plot is essentially an elliptical ring, which is perhaps the projection of a torus in a space of higher dimension, and that the correlation function extends with undiminished amplitude over a long time period.

For cockatoos, specifically the sulfur-crested and the gang-gang cockatoos, on the other hand, we conclude that the call structure is very different, and in fact chaotic. This conclusion is supported by the fact that the largest Lyapunov exponent is in each case in the range 0.2–0.3, a range typical of chaotic behavior rather than random noise or nonlinearly coupled periodic signals, and by the visual appearance of the recurrence mappings. The power spectrum does have several dominant frequencies, which presumably represent major oscillation modes of the syringeal structures, but the amplitudes and phases of these modes vary in a chaotic fashion. The chaotic nature of these calls is audibly recognizable in their harsh screeching character.

The investigation of chaotic oscillations in simple systems, by which is generally meant systems with a small number of degrees of freedom or dynamic variables, is now a well-established subject.[12] The case of birdsong is likely to be more complex because the primary oscillating system, the syrinx, has many degrees of freedom corresponding to wavelike motions on the membranes or other vibrating structures, and because there may often be two vocal sources involved. This makes detailed analysis much more difficult.

This study thus supports the detailed work of Fee and colleagues[8] on the song of the zebra finch. Their analysis showed the existence of abrupt period-doubling transitions in the song, a feature that is characteristic of one major route to chaos.[15] Their numerical analysis showed that such a transition could be derived for the case of a simple model of airflow through a compliant constriction, without requiring two independent oscillatory sources. It is also interesting to note that the double reeds of woodwind instruments such as oboes or bassoons show a period-doubling transition when blown vigorously and not fitted to the instrument — a test carried out when the reed is being adjusted by the player.

While the vocal utterances of these Australian cockatoos appear to be entirely chaotic in structure, and thus an extreme case, it seems likely that shorter syllables with highly nonlinear or even chaotic structure may be part of the vocal repertoire of many birds.

ACKNOWLEDGMENT

This work is part of a program supported by a grant from the Australian Research Council.

[1] C. H. Greenewalt, *Birdsong: Acoustics and Physiology* (Smithsonian Institution, Washington, 1968).
[2] R. A. Suthers, F. Goller, and C. Pytte, ''The neuromuscular control of birdsong,'' Philos. Trans. R. Soc. London, Ser. B **354**, 927–939 (1999).
[3] N. H. Fletcher, ''Birdsong—A quantitative acoustic model,'' J. Theor. Biol. **135**, 455–481 (1988).
[4] N. H. Fletcher and A. Tarnopolsky, ''Acoustics of the avian vocal tract,'' J. Acoust. Soc. Am. **105**, 35–49 (1998).
[5] N. H. Fletcher, *Acoustic Systems in Biology* (Oxford University Press, New York, 1992).
[6] N. H. Fletcher, ''Mode locking in nonlinearly excited inharmonic musical oscillators,'' J. Acoust. Soc. Am. **64**, 1566–1569 (1978).
[7] N. H. Fletcher, ''Nonlinearity, complexity, and control in vocal systems,'' in *Vocal Fold Physiology: Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular, San Diego, 1996), pp. 3–16.
[8] M. S. Fee, B. Shraiman, B. Pesaran, and P. P. Mitra, ''The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird,'' Nature (London) **395**, 67–71 (1999).
[9] F. Goller, ''Vocal gymnastics and the bird brain,'' Nature (London) **395**(3), 11–12 (1999).
[10] Canberra Ornithologists Group, *Birdsongs of Canberra* (audio cassette), 1988.
[11] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, ''Geometry from a time series,'' Phys. Rev. Lett. **45**, 712–716 (1980).
[12] G. L. Baker and J. P. Gollub, *Chaotic Dynamics*, 2nd ed. (Cambridge University Press, Cambridge, 1996).
[13] J. C. Sprott and G. Rowlands, *Chaos Data Analyzer*, Physics Academic Software (American Institute of Physics, New York, 1992).
[14] E. Kononov, *Visual Recurrence Analysis*, v 4.0, http://pwl.netcom.com/~eugenek (1999).
[15] M. J. Feigenbaum, ''Universal behavior in nonlinear systems,'' Los Alamos Sci. **1**, 4–27 (1980).

Simulation of circular array ultrasound transducers for intravascular applications

Jerome M. G. Borsboom
Thoraxcentre, Erasmus University Rotterdam, Rotterdam, The Netherlands and Circuits and Systems Laboratory, Department of Electrical Engineering, Delft University of Technology, Delft, The Netherlands

E. Ignacio Céspedes
Thoraxcentre, Erasmus University Rotterdam, Rotterdam, The Netherlands and Endosonics Corporation, Rancho Cordova, California 95670

Antonius F. W. van der Steen[a)]
Thoraxcentre, Erasmus University Rotterdam, Rotterdam, The Netherlands and Interuniversity Cardiology Institute Netherlands (ICIN), Utrecht, The Netherlands

Charles T. Lancée
Thoraxcentre, Erasmus University Rotterdam, Rotterdam, The Netherlands

Ed F. Deprettere
Circuits and Systems Laboratory, Department of Electrical Engineering, Delft University of Technology, Delft, The Netherlands

(Received 26 December 1998; accepted for publication 3 April 2000)

The beam shape of a circular array transducer that is commonly used in intravascular ultrasound catheters was investigated in linear mode of operation. For this use, a simulation program which can simulate the radio frequency (rf)-response of a number of scatterers has been developed. The program is based on the impulse response method, which is implemented in the frequency domain. Due to the unusual geometry of the transducer, the far field gets peculiarly shaped for large apertures. Instead of having a far field with its maximum intensity in a single lobe on the acoustical axis, the far field splits up into a dual-lobe far field with maximum intensity in two lobes off the acoustical axis. A formula is derived that predicts the occurrence of these beam shapes. © *2000 Acoustical Society of America.* [S0001-4966(00)00108-9]

PACS numbers: 43.80.Vj [FD]

INTRODUCTION

Intravascular ultrasound (IVUS) imaging is the only currently available clinical method to directly visualize atherosclerosis and other pathologic conditions within the wall of blood vessels.[1–4] The transducers that are used for IVUS are mounted at the tip of a catheter, which is small enough (typically around 1 mm in diameter) to fit inside a coronary artery. The commonly used modalities generate a tomographic image. Currently, two kinds of transducers are available for this use:[5] a rotating single-element transducer[6] and a circular array transducer.[7–9] Both modalities have proven to be capable of producing high-quality images. Current developments in rf data analysis allow a functional extension of IVUS beyond imaging.[10–12] Two areas of interest are the assessment of the local elasticity of the vascular wall[13,14] and intravascular flow estimation.[15] For example, the decorrelation of repetitively acquired rf signals is directly related to the local blood velocity, which can be used for flow assessment. This method has been evaluated in a computer model,[16] and validated for a number of flow conditions *in vitro*,[17,18] in animal models,[19,20] and in patients.[21,22] Although IVUS flow assessment has been demonstrated using a rotating single-element transducer, array catheters are in principle better suited as they can fire multiple times at a fixed beam position. In this paper we focus on some specific properties of the acoustic beam of the circular array transducer as it relates to flow imaging.

We investigate the acoustic field of a circular array transducer consisting of 64 elements wrapped around the surface of a 1.2-mm diameter catheter and operating at a center frequency of 20 MHz. The length of the array elements is 0.7 mm and the pitch is 59 μm. Due to the limited space in the catheter, it is not possible to connect each element with a separate, outgoing wire. Instead, an integrated circuit is included in the tip in order to select and multiplex elements onto the few wires possible in the catheter. In this setup, transmission and reception is limited to only one signal with one group of elements at a time.

The array transducer has two modes of operation: the *phased mode* operation, in which synthetic beamforming is used to obtain a narrow beam at the imaging location, and the *linear mode* operation, in which a number of elements are electrically tied together and used as a single element for both transmission and reception. High-quality imaging is realized in *phased mode* operation as it requires a narrow and well-focused beam and the ability to synthesize more than 64 beams to form the image.[8,9] However, synthetic beam forma-

[a)]Address for correspondence: Room Ee2302, Erasmus University Rotterdam, P.O. Box 1738, 3000DR, Rotterdam, The Netherlands. Electronic mail: vandersteen@tch.fgg.eur.nl

tion is computationally demanding and the entire process takes place in the order of a millisecond.[23] By concurrent synthesis of adjacent beams, current systems perform this operation in real time, achieving frame rates up to 30 s^{-1}. For other purposes, *linear mode* of operation may be better. When a high pulse repetition rate is required for multipulse sequences, as in flow imaging,[15,16] time constraints exclude the use of synthetic beam formation. Real-time electronic focusing using delay lines is also impossible as only one group of elements can be used at a time. In this case a lower and fixed number of beams (N for an N-element array) can be obtained, which might be sufficient for flow imaging.

Implementation and use of either mode of operation requires detailed knowledge of the beam shape of the array in various configurations. The array we have been investigating has an unusual circular geometry for which no analytical expression exists to accurately predict three-dimensional beam shapes.[24] Its small size makes field measurements cumbersome. To our knowledge, no hydrophones are available to measure the minute pressure distribution produced by this transducer, especially in the near field. In order to predict the performance of the array, a simulation approach has been taken. To validate this approach, we first simulated continuous wave (cw) beam shapes for square and rectangular transducers for which theory is available. After validation, it is assumed that the simulation program is functionally correct and can be used for the simulation of curved transducers.

In this paper, a wideband simulation program for a circular array transducer in linear mode is presented. The aim is to develop a program which is capable of simulating beam shapes for various element configurations as well as rf signals in response to scatterers in the ultrasound beam. As this study is designed for the linear mode of operation, phasing of separate elements was not taken into account. Beam shapes used for validation of the program and beam shapes of the array in various configurations are presented and discussed. We show that due to the high radius of curvature of the transducer, the beam shape under certain conditions becomes peculiar and unfit for imaging. We derive a formula that predicts the occurrence of these beam shapes.

I. THEORY

In general, the pressure field along the acoustical axis of an ultrasound transducer exhibits rapid fluctuations in the near field and a gradual decay in a single lobe in the far field.[25–27] The transition from near field to far field can be defined to occur at the last axial maximum of the pressure response. For a disc-shaped transducer in cw mode, the last axial maximum, and hence the transition from near field to far field, is approximately located at

$$z=\frac{D^2}{4\lambda}, \tag{1}$$

where z is the distance to the transducer surface, D is the diameter of the transducer, and λ is the wavelength at the center frequency of the radiated ultrasound. This approximate relation is derived by considering the axial point P that satisfies

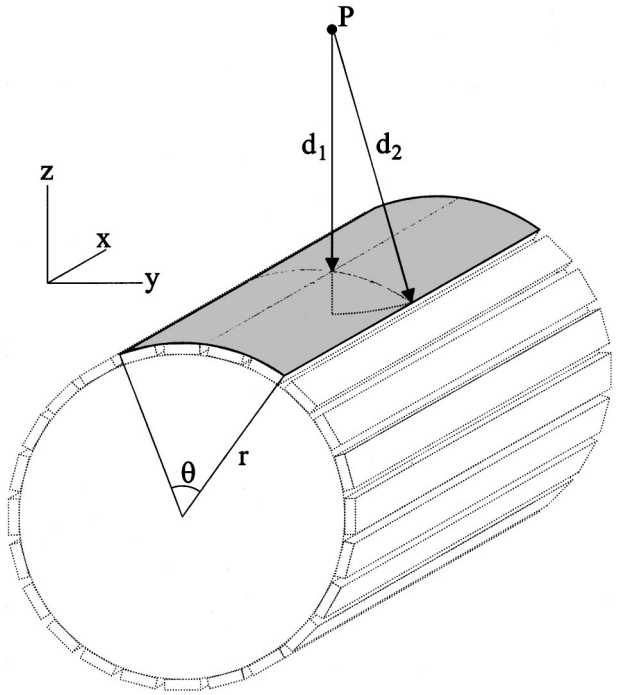


FIG. 1. Geometrical setup for derivation of the 'limit aperture angle' θ_L, that gives the largest aperture angle for which a conventional single-lobe far field exists.

$$d_2-d_1=\frac{\lambda}{2}, \tag{2}$$

where d_1 is the distance from P to the surface of the transducer, and d_2 is the distance from P to the edge of the transducer.[25] Beyond this point, no destructive interference can occur as the maximum path difference is less than half a wavelength and hence the axial pressure field decays monotonically. For circular array transducers with a convex surface, however, the existence of a transition from near field to far field is not trivial. We show that for certain combinations of curvature radius, wavelength, and aperture, there is no such axial transition point.

Consider a circular array transducer with an axial point P at a distance d_1 from the transducer surface (see Fig. 1). The distance from P to the straight edge of the surface is indicated by d_2, and can be written as

$$d_2=\sqrt{r^2+(d_1+r)^2-2r(d_1+r)\cos\left(\frac{\theta}{2}\right)}, \tag{3}$$

where r is the radius of the catheter and θ is the aperture of the active transducer surface. By choosing P so that Eq. (2) is satisfied, a lower boundary for the transition from near field to far field can be obtained. Substituting Eq. (2) into Eq. (3) and solving for d_1, this bound is obtained as

$$d_1=\frac{-\left(\frac{\lambda}{2}\right)^2+2r^2-2r^2\cos\left(\frac{\theta}{2}\right)}{\lambda-2r+2r\cos\left(\frac{\theta}{2}\right)}. \tag{4}$$

From Fig. 1 it can be seen that when the aperture θ increases for fixed r and λ, both d_1 and d_2 have to increase in order for

Eq. (2) to hold. In other words, for larger apertures, the lower bound moves away from the transducer and hence so does the transition from near field to far field. Then, for certain combinations of r and λ, a ''limit aperture'' can be defined as the aperture that requires P to be at infinity in order to satisfy Eq. (2). When P is at infinity, the denominator in Eq. (4) is zero. Solving for θ, the limit aperture θ_L is given as

$$\theta_L = 2\cos^{-1}\left(1-\frac{\lambda}{2r}\right), \tag{5}$$

or in number of elements N_e

$$N_e = \frac{2N_t}{2\pi}\cos^{-1}\left(1-\frac{\lambda}{2r}\right), \tag{6}$$

where N_t is the total number of elements in the array. Only for apertures smaller than the limit aperture, $\theta < \theta_L$, will there be axial points for which there is no destructive interference and hence a single-lobed decaying far field. For apertures larger than the limit aperture, $\theta > \theta_L$, the conventional definition of a decaying, single-lobe far field is no longer valid.

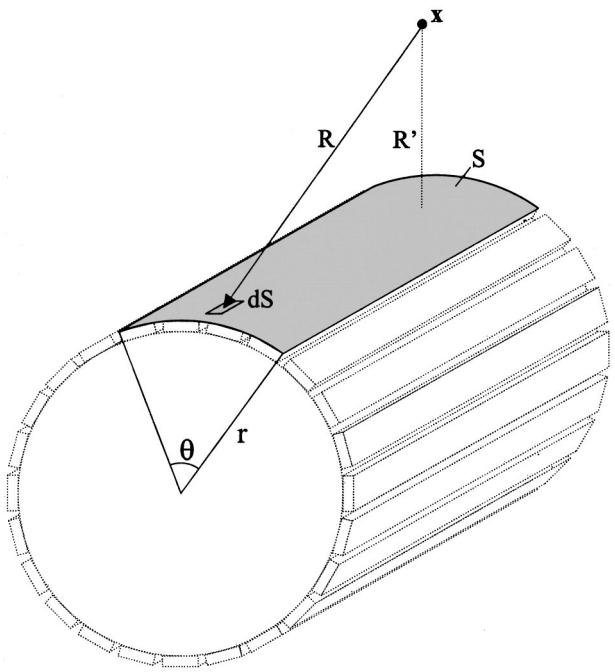


FIG. 2. Geometrical setup for calculation of the ultrasound rf signal in response to a scatterer at position **x** in the sound field generated by surface *S*.

II. SIMULATION APPROACH

A. Program development

A finite-element simulation program was developed using the impulse-response method, a common approach in ultrasound studies.[16,28] With this method, it is possible to calculate the rf response of any number of scatterers. In this method, each infinitesimal area of the sound-emitting part of the transducer transmits a predefined sound pulse. At a given scattering point, its response of the transmitted pulses is reflected back to the transducer. The reflected pulse is received by the entire transducer and eventually summed with the responses from other scatterers.

The setup used for calculation of the rf response is shown in Fig. 2. The one-way pressure response at point **x** due to the vibrating surface S can be expressed as[28]

$$p_1(\mathbf{x},t) = C\int_S\int_{-\infty}^{\infty}\psi(\tau)\frac{\delta\left(t-\tau-\frac{R}{c}\right)}{R}d\tau\,dS, \tag{7}$$

where τ is a convolution time variable, C is a conversion constant, c is the speed of sound, and R is the distance between point **x** and the surface element dS. The function $\psi(t)$ holds the impulse response of the transducer, and $\delta(t)$ is a Dirac pulse. As C is a constant converting velocity into pressure, it is arbitrarily set to one. Since it contains a convolution integral, Eq. (7) is computationally advantageously transformed into the frequency domain

$$P_1(\mathbf{x},k) = \Psi(k)\int_S\frac{e^{-jkR}}{R}dS = \Psi(k)H(\mathbf{x},k). \tag{8}$$

In order to implement this equation, the integral and the frequency are discretized. For the integral, this is done by subdividing S into a regular, two-dimensional grid of small integration elements. The size of these elements is chosen to be much smaller than half the wavelength corresponding to the largest frequency in the radiated sound pulse. The responses from the small integration elements are summed together to obtain an approximation for the integral in Eq. (8)

$$P_1(\mathbf{x},k) = \Psi(k)dX\,dY\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\frac{e^{-jkR_{i,j}}}{R_{i,j}}, \tag{9}$$

where N_1 and N_2 are the number of integration steps in each direction, and dX and dY are the sizes of these steps.

Discretization of the frequency has to be done with care. In order to prevent the occurrence of aliasing, we must make sure that the time $R_{i,j}/c$, which models the traveling time of a pulse from an integration element $dX\,dY$ to the point **x**, never exceeds the length of the Fourier transform window minus the length of the pulse, i.e., the pulse must remain within the Fourier window. This is accomplished by calculating a time-offset R'/c, where R' is the distance between **x** and the transducer surface. Inclusion of the offset in Eq. (9) yields

$$P_1^*(\mathbf{x},k) = \Psi(k)dX\,dY\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\frac{e^{+jkR'}e^{-jkR_{i,j}}}{R_{i,j}}. \tag{10}$$

The length of the fast Fourier transform is chosen so that all possible responses can be accommodated, thus preventing aliasing.

Calculation of an rf signal in response to a cloud of scatterers is based on summation of the two-way response for each scatterer. Assuming the absence of multiple scattering, the two-way response of a scatterer at position **x** can be found using the reciprocity theorem.[26] In the time domain, calculation of the two-way response requires the convolution of the transmission response at the scatterer position with itself. In the frequency domain, the convolution is simply

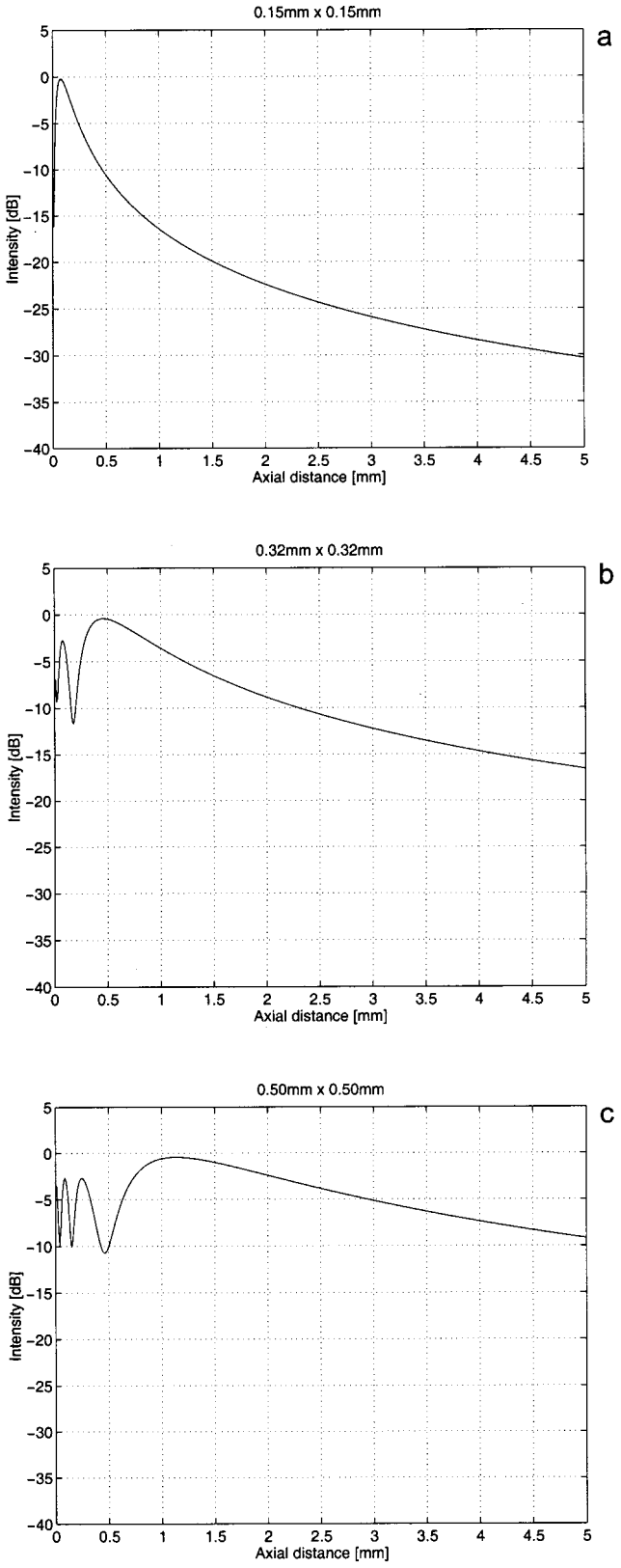


FIG. 3. Simulated one-way axial intensity beam profiles at 20-MHz cw for flat, square transducers with sides of (a) 0.15 mm, (b) 0.32 mm, and (c) 0.50 mm.

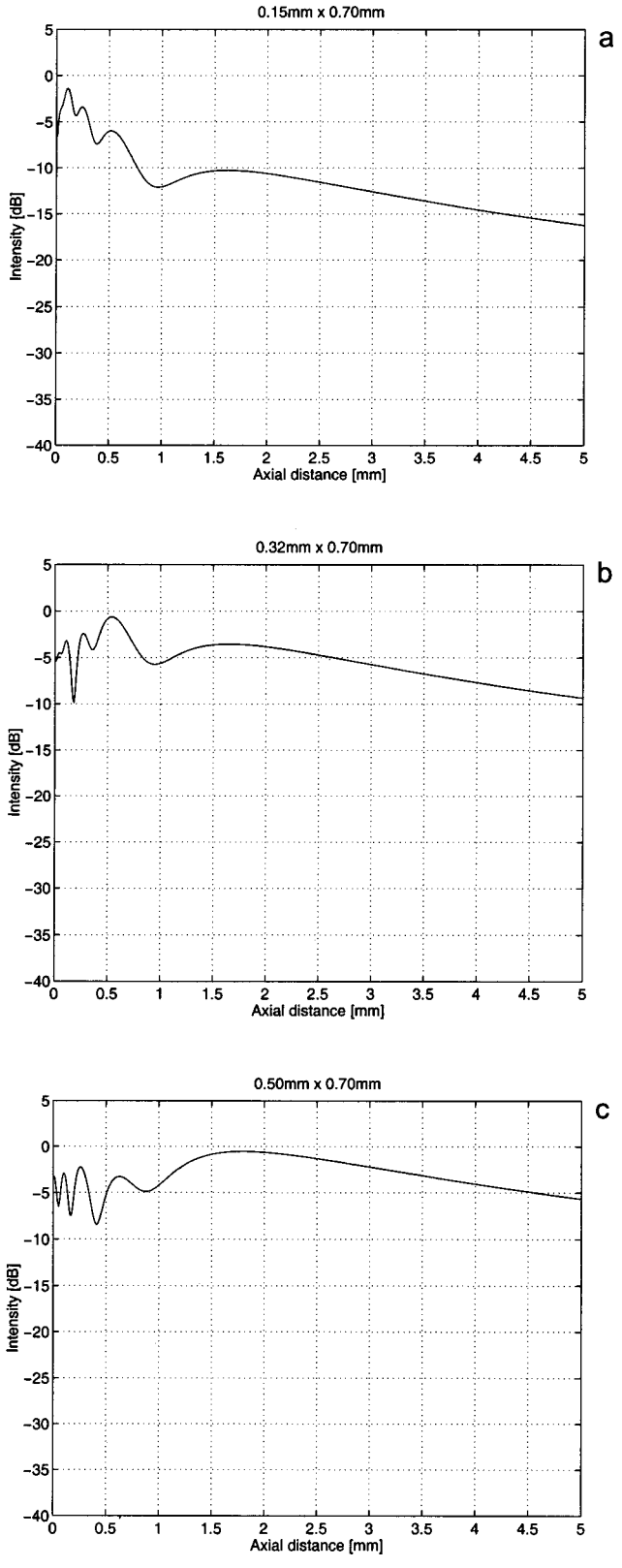


FIG. 4. Simulated one-way axial intensity beam profiles at 20-MHz cw for flat, rectangular transducers with short sides of (a) 0.15 mm, (b) 0.32 mm, and (c) 0.50 mm and a long side of 0.70 mm.

expressed as the product of two identical one-way terms. Squaring the one-way response given by Eq. (10) gives the two-way frequency response

$$P_2(\mathbf{x},k) = P_1^*(\mathbf{x},k)^2. \tag{11}$$

Finally, the two-way time response $p_2(\mathbf{x},t)$ for a single scatterer is found by taking the inverse Fourier transform of $P_2(\mathbf{x},k)$. After correction for the time-offset R'/c, this response is summed with the responses of other scatterers to give the total rf response

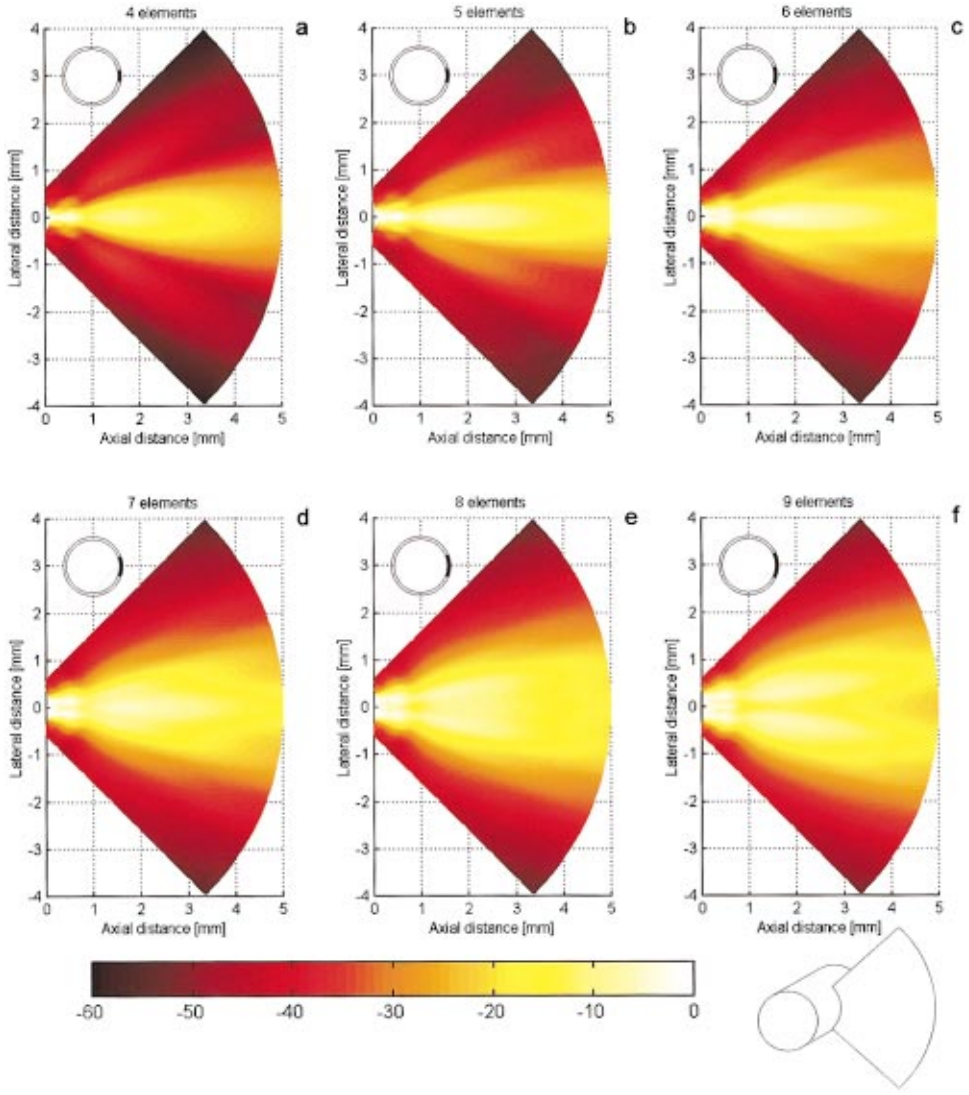


FIG. 5. Simulated two-way intensity broadband beam profiles in the $y-z$ plane for apertures corresponding to four to nine elements. The aperture is indicated by the black area in the upper-left ring icon.

$$p_{\mathrm{rf}}=\sum_{j=1}^{M} p_2(\mathbf{x},t+2R_j'/c), \tag{12}$$

where M is the number of scatterers.

B. Echosignal simulation

With the simulated rf responses, properties of the array transducer can be investigated. However, computation time increases linearly with the number of scatterers for which the response has to be calculated. Although this is not a problem when simulating beam profiles, computation time may become very large for flow simulations, for which clouds with many scatterers are required and for which the rf response must be calculated at multiple displacement steps. In order to keep computation time within practical limits, the surface of the active elements of the array was approximated by a single, curved surface by ignoring the gap between elements. The aperture of this surface equals the aperture formed by the active elements and approximates the curved setup of the elements. Although the actual elements on the transducer are flat, the width of the individual elements is small enough, approximately half a wavelength, to neglect the curvature that is introduced into the elements by this approximation.

In order to compare beam profiles of the array in linear mode for different apertures, simulations were performed. Beam profiles were obtained by stepping a single point target through the ultrasound beam and calculating at each point the peak amplitude of the received rf response. By stepping the point target along a line or in a two-dimensional grid, one- and two-dimensional beam profiles could be calculated. The step size used to create the beam profiles was 100 μm for the axial or z direction (see Fig. 1), 28 μm for the elevational or x direction, and, as cylindrical coordinates were used in the calculations, 1.8° for the lateral direction. Apertures corresponding to one to nine consecutive array elements with an element size of 0.7×0.036 mm and an array diameter of 1.2 mm were simulated.

III. EVALUATION AND VALIDATION

The first validation was performed by simulating 20 MHz cw one-way, axial intensity profiles for flat *square* transducers. Figures 3(a)–3(c) show these profiles. The sides of the transducers measure 0.15, 0.32, and 0.50 mm, which approximately corresponds to the length of the aperture of 3,

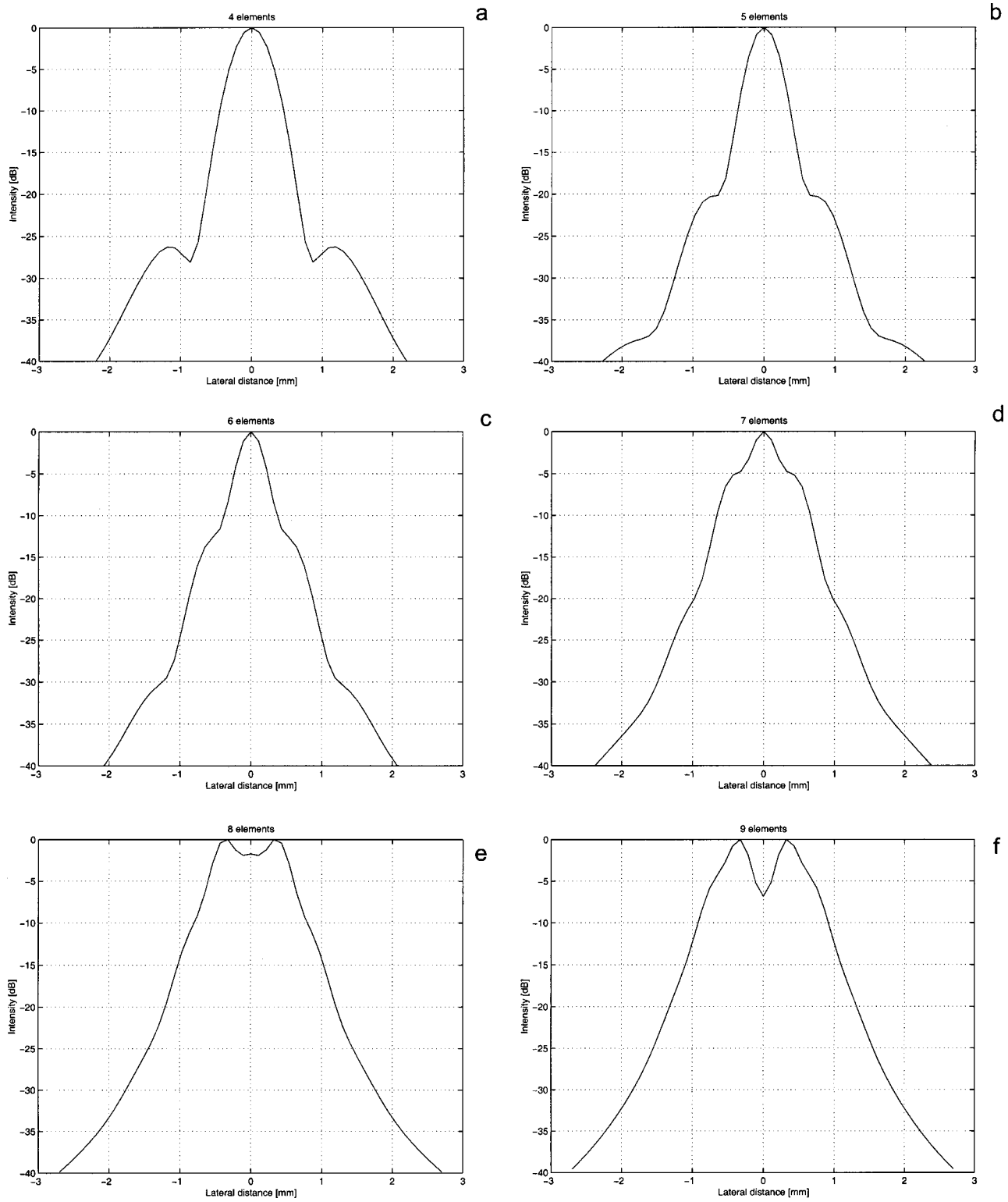


FIG. 6. Simulated two-way intensity broadband beam cuts in the *y* direction for apertures corresponding to four to nine elements at 2.0 mm from the transducer.

6, and 9 elements, respectively, of the circular array transducer. These profiles show a near field with a regularly fluctuating intensity, a last axial maximum, and a far field with its typical slow decay. There are no significant differences compared to theoretical intensity profiles.[29]

The same simulations were performed for flat *rectangular* transducers. One side of the transducer was fixed at 0.7 mm, which corresponds to the length of an element of the array, while the other side had the same lengths as for the square transducers. Thus, this simulation represents a flattened version of the curved array transducer. The simulated axial intensity profiles are shown in Figs. 4(a)–4(c). Compared to the profiles of square transducers, these profiles show more irregular fluctuations in the near field, which are

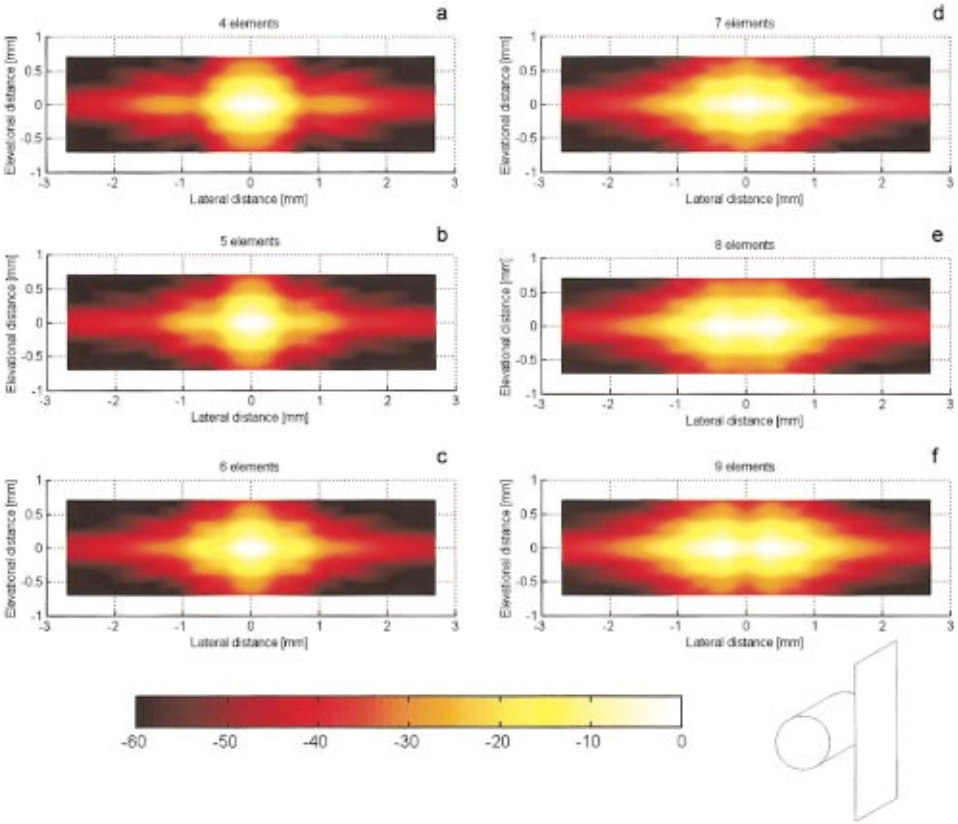


FIG. 7. Simulated two-way intensity broadband beam profiles in the $x-y$ plane for apertures corresponding to four to nine elements at 2.0 mm from the transducer.

due to the multiplicative nature of the beam profile of rectangular transducers.[25,26]

All validation simulations show good agreement between theoretically expected and simulated beam profiles, thus showing that cw beam profiles are simulated with adequate accuracy. Therefore, it is assumed that the program simulates cw and wideband beam profiles of the cylindrically curved surface of the transducer as well.

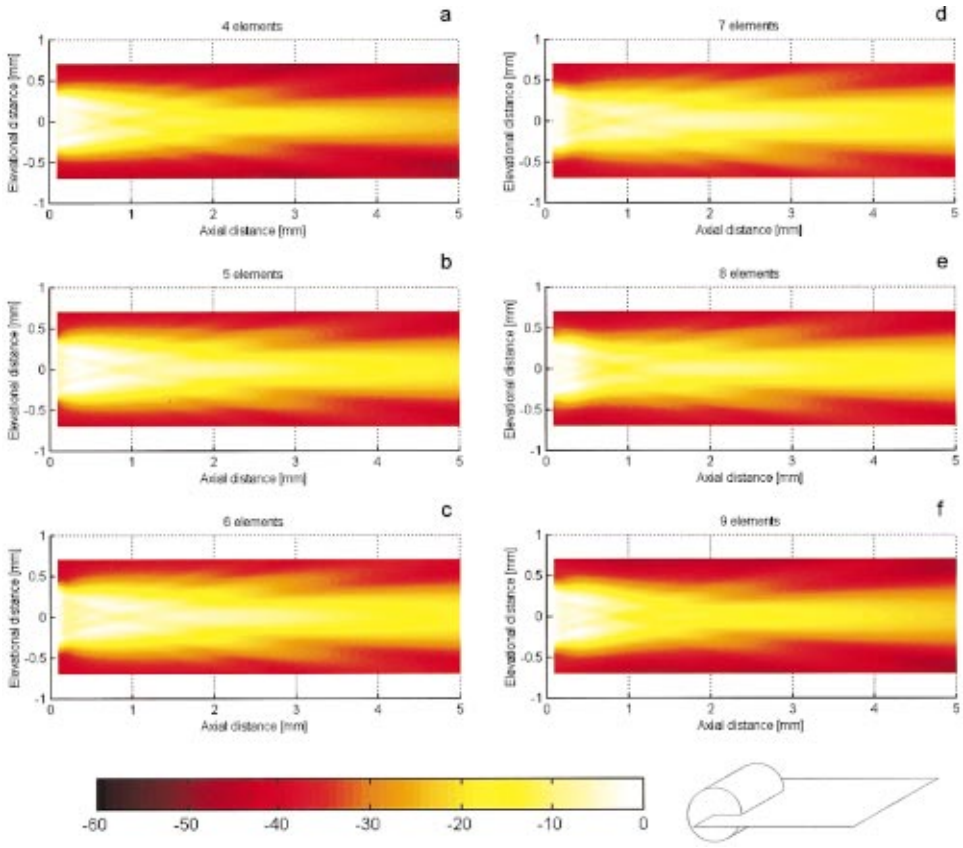


FIG. 8. Simulated two-way intensity broadband beam profiles in the $x-z$ plane for apertures corresponding to four to nine elements.

IV. RESULTS

Several two-way, intensity, broadband beam profiles were simulated in three perpendicular planes for various apertures to illustrate the 3D ultrasound beam. Directional conventions are indicated in Fig. 1, with each of the three simulated planes defined by a pair of axes. The $x-y$ plane corresponds to a C plane, the $y-z$ plane to the scan plane. The broadband pulse has a center frequency of 20 MHz and 50% (-6 dB) fractional bandwidth. The dimensions of the simulated array are 0.7 mm length, 1.2 mm, diameter and apertures corresponding to one to nine elements with an element size of 0.7×0.036 mm.

Two-way intensity beam profiles in the $y-z$ plane are shown in Figs. 5(a)–5(f). This figure contains six beam profiles corresponding to four to nine element apertures. Each profile is a 90° sector of the beam ranging over 5-mm depth. The two-way intensity is coded in color ranging from yellow to black and displayed over a range from 0 to -60 dB with yellow corresponding to higher intensities. The small ring icon in the upper-left corner of each picture indicates in black the aperture formed by the elements. In order to enable quantitative judgments, each beam profile in Fig. 5 is cut at 2-mm axial distance and plotted as a one-dimensional intensity profile in Figs. 6(a)–6(f).

Beam profiles for the $x-y$ plane are shown in Figs. 7(a)–7(f). Each picture is a 5.0×1.4-mm cut through the beam at 2.0 mm from the transducer surface and shows the two-way intensity beam profile. The horizontal and vertical directions correspond to the curved and the straight dimensions of the transducer, respectively. Finally, the $x-z$ plane beam profiles are shown in Figs. 8(a)–8(f). Here, each picture is a 5×1.4-mm cut through the ultrasound beam along the acoustical axis and the straight direction of the transducer. In each figure, the orientation of the beam profiles relative to the transducer is indicated by a small pictogram.

V. DISCUSSION AND CONCLUSION

Looking at the beam profiles for apertures ranging from four to nine elements, it can be seen that the single-lobe far field which is present for small apertures turns into a dual-lobe far field for larger apertures. In the dual-lobe far-field condition, the maximum intensity is no longer found on the acoustical axis, but is located at each of the two off-axis main lobes. This indicates that there is destructive interference on the acoustical axis in the far field which is caused by the curvature of the surface.

The limit aperture for this transducer is found with Eq. (5) to be $\theta_L=0.71$, or equivalently, 7.2 elements. The apertures for seven and eight elements are $\theta_7=0.64$ and $\theta_8=0.73$. Thus, from the theory described in this paper, the aperture for eight elements exceeds the limit aperture. This is also demonstrated in Fig. 5 by the beam split for eight elements which is not present for the seven-element aperture. This result confirms Eq. (5), which calculates the aperture beyond which a dual-lobe far field occurs. As a dual-lobe far field severely decreases lateral resolution, this limits the number of consecutive elements to be used in linear mode. Equation (6) determines the maximum number of elements to be used in order to prevent a dual-lobe far field with its decreased lateral resolution.

Flow assessment is most accurately done with a narrow beam, but is also dependent on a high pulse repetition rate. Blood is flowing across the transducer and the same blood particles must be sampled several times to accurately estimate the flow velocity. Therefore, flow assessment uses the linear mode to obtain a high pulse repetition frequency. In addition, beam width is optimized for estimation accuracy, given pulse repetition frequency. From Fig. 5 we see that the narrowest beam is obtained with five elements, which therefore, in principle, is best for flow assessment. However, as the same blood particles have to be sampled multiple times, the beam may be too small to get accurate velocity estimates, thus requiring a broader beam. Using more elements causes the beam to widen, but the number of elements is limited by Eq. (6), which indicates when the beam starts to split. The use of Eq. (6) is in defining an upper limit on the number of elements usable for velocity estimation.

We have developed a simulation program for the circular array transducer. It is a useful tool to investigate beam profiles for this transducer for which there is no theory and for which accurate field measurements are practically impossible. Beam profiles are easily simulated and flow simulations with highly controlled flow conditions are within reach. This provides a tool for designing array transducers and investigating processing algorithms based on simulated rf signals. For further research, the simulation program can be modified to represent other array configurations, such as an increase of the number of elements or a change in catheter diameter. Currently, the rf simulations are utilized to investigate decorrelation patterns of simulated blood, which will be reported separately.

ACKNOWLEDGMENT

This work was financially supported by the Dutch Technology Foundation. (STW RGN 44.3462)

[1] *Intravascular Ultrasound*, edited by J. Roelandt, E. J. Gussenhoven, and N. Bom (Kluwer Academic, Dordrecht, 1993).
[2] *Intravascular Ultrasound*, edited by R. Erbel, J. Roelandt, J. Ge, and G. Görge (Martin Dunitz, London, 1998).
[3] *Intravascular Ultrasound Imaging in Coronary Artery Disease*, edited by R. J. Siegel (Dekker, New York, 1998).
[4] P. G. Yock, P. J. Fitzgerald, and Y. Honda, ''Intravascular Ultrasound,'' in *Textbook of Interventional Cardiology*, edited by E. J. Topol (Saunders, Philadelphia, 1999).
[5] N. Bom, H. ten Hoff, C. T. Lancée, W. J. Gussenhoven, and J. G. Bosch, ''Early and recent intraluminal ultrasound devices,'' Int. J. Card. Imaging **4**, 79–88 (1989).
[6] H. ten Hoff, ''Scanning Mechanisms for Intravascular Ultrasound Imaging,'' Ph.D. thesis (Erasmus University Rotterdam, Rotterdam, 1993).
[7] N. Bom, C. T. Lancée, and F. C. van Egmond, ''An ultrasonic intracardiac scanner,'' Ultrasonics **10**, 72–77 (1972).
[8] M. O'Donnell and L. J. Thomas, ''Efficient synthetic aperture imaging from a circular aperture with possible application to catheter-based imaging,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 366–380 (1992).
[9] M. O'Donnell, M. J. Eberle, D. N. Stephens, J. L. Litzza, K. San Vincente, and B. M. Shapo, ''Synthetic phased arrays for intraluminal imaging of coronary arteries,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 714–721 (1997).
[10] M. O'Donnell, M. J. Eberle, D. N. Stephens, J. L. Litzza, B. M. Shapo, J.

R. Crowe, C. D. Choi, J. J. Chen, D. M. W. Muller, J. A. Kovach, R. L. Lederman, R. C. Ziegenbein, C. C. Wu, K. SanVincente, and D. Bleam, ''Catheter arrays: Can intravascular ultrasound make a difference in managing coronary artery disease,'' Proceedings of the 1997 IEEE Ultrasonics Symposium, pp. 1447–1456.
[11] A. F. W. van der Steen, E. I. Céspedes, C. L. de Korte, S. G. Carlier, W. Li, F. Mastik, C. T. Lancée, J. Borsboom, F. Lupotti, R. Krams, P. W. Serruys, and N. Bom, ''Novel developments in intravascular imaging,'' Proceedings of the 1998 IEEE Ultrasonics Symposium, pp. 1733–1742.
[12] N. Bom, W. Li, A. F. W. van der Steen, C. T. Lancée, E. I. Céspedes, C. J. Slager, and C. L. de Korte, ''New developments in intravascular ultrasound imaging,'' Eur. J. Ultrasound **7**, 9–14 (1998).
[13] B. M. Shapo, J. R. Crowe, A. R. Skovoroda, M. Eberle, N. A. Cohn, and M. O'Donnell, ''Displacement and strain imaging of coronary arteries with intraluminal ultrasound,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 234–246 (1996).
[14] C. L. de Korte, A. F. W. van der Steen, E. I. Céspedes, and G. Pasterkamp, ''Intravascular ultrasound elastography of human arteries: Initial experience *in vitro*,'' Ultrasound Med. Biol. **24**, 401–408 (1998).
[15] W. Li, A. F. W. van der Steen, C. T. Lancée, E. I. Céspedes, and N. Bom, ''Blood flow imaging and volume flow quantitation with intravascular ultrasound,'' Ultrasound Med. Biol. **24**, 203–214 (1998).
[16] W. Li, C. T. Lancée, E. I. Cespedes, A. F. W. van der Steen, and N. Bom, ''Decorrelation of intravascular echo signals: Potentials for blood velocity estimation,'' J. Acoust. Soc. Am. **102**, 3785–3794 (1997).
[17] W. Li, A. F. W. van der Steen, C. T. Lancée, E. I. Céspedes, S. G. Carlier, E. J. Gussenhoven, and N. Bom, ''Potentials of volumetric blood-flow measurement,'' Semin. Interv. Cardiol. **2**, 49–54 (1997).
[18] W. Li, C. T. Lancée, A. F. W. van der Steen, E. J. Gussenhoven, and N. Bom, ''Blood velocity estimation with high frequency intravascular ultrasound,'' Proceedings of the 1996 IEEE Ultrasonics Symposium, pp. 1485–1488.
[19] E. I. Céspedes, S. Carlier, W. Li, F. Mastik, A. F. W. van der Steen, N. Bom, P. Verdouw, and P. W. Serruys, ''Blood flow assessment using a mechanical intravascular ultrasound catheter: Initial evaluation *in vivo*,'' J. Vasc. Invest. **4**, 39–44 (1998).
[20] A. F. W. van der Steen, W. Li, E. I. Céspedes, S. Carlier, M. Eberle, P. D. Verdouw, P. W. Serruys, and N. Bom, ''*In vivo* validation of blood flow estimation using the decorrelation of radiofrequency intravascular echo signals,'' Proceedings of the 1997 IEEE Ultrasonics Symposium, pp. 1247–1250.
[21] S. G. Carlier, W. Li, E. I. Céspedes, A. F. W. van der Steen, J. N. Hamburger, N. Bom, and P. W. Serruys, ''Simultaneous morphological and functional assessment of a renal artery stent intervention with intravascular ultrasound,'' Circulation **97**, 2575–2576 (1998).
[22] S. G. Carlier, E. I. Céspedes, A. F. W. van der Steen, N. Bom, and P. W. Serruys, ''Intracoronary blood flow assessment with IVUS,'' Cardiovasc. Intervent. CVI Online **2**, 22–25 (1998).
[23] M. J. Eberle, ''The latest in electronic imaging,'' Semin. Interv. Cardiol. **2**, 19–23 (1997).
[24] H. F. Olson, *Acoustical Engineering* (Van Nostrand, Princeton, NJ, 1957).
[25] D. A. Christensen, *Ultrasonic Bioinstrumentation* (Wiley, New York, 1988), Chap. 8.
[26] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[27] J. Zemanek, ''Beam behavior within the nearfield of a vibrating piston,'' J. Acoust. Soc. Am. **49**, 181–191 (1971).
[28] J. C. 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).
[29] A. Weyns, ''Studie en optimalisatie van elementaire diagnostische ultrasoonzenders en ultrasoonontvangers'' [Study and optimization of basic diagnostic ultrasound transmitters and receivers], Ph.D. thesis (K. U. Leuven, Leuven, 1982).

Foliage echoes: A probe into the ecological acoustics of bat echolocation

Rolf Müller[a] and Roman Kuc
Intelligent Sensors Laboratory, Department of Electrical Engineering, Yale University, New Haven, Connecticut 06520-8284

(Received 13 October 1999; revised 2 April 2000; accepted 1 May 2000)

The research reported here aims at understanding the biosonar system of bats based on the properties of its natural inputs (ecological acoustics). Echoes from foliages are studied as examples of ubiquitous, natural targets. The echo properties and their qualitative relationship to plant architecture are described. The echoes were found to be profoundly stochastic and in general neither Gaussian nor stationary. Consequently, features useful for discrimination of such target classes will be confined to estimated random process parameters. Several such statistical signal features which are sufficiently invariant to allow a classification of the used example plants were identified: the characteristic exponent and the dispersion of an α-stable model for the amplitude distribution, a crest factor defined as the ratio of maximum squared amplitude and signal energy, the dispersion of the first threshold passage distribution, the structure of the correlation matrix, and a nonstationarity in sound channel gain. Discrimination error probability could be reduced by combining features pairwise. The best combination was the crest factor and the correlation coefficient of a log-linear model of the time-variant sound channel gain; it yielded an estimated Bayes risk of 6.9% for data pooled from different views. [S0001-4966(00)01808-7]

PACS numbers: 43.80.Ka [WA]

INTRODUCTION

Any mobile, active mammal should benefit considerably from a detailed knowledge of its surroundings. In order to meet the informational requirements of the various biological tasks crucial to survival, its sensory systems need to perform all three types of estimation tasks: detection, localization, and classification. Biosonar constitutes the dominating far sense in bats. As such it can be expected to supply the bulk of the sensory information about any object which the animals encounter. In order to understand how this is accomplished, it is indispensable to determine the salient features of echoes generated in natural environments. This is a topic of ecological acoustics (in the sense of Gibson, 1979).

Since the actual tracking of prey is usually completed within a short time interval, a significant portion of the echoes a bat receives originates from the surfaces which make up the environment of the respective species (assuming there are surfaces within the detection distance). The ability to classify these surfaces is at the core of fundamental biological tasks like habitat choice and navigation: In order to assess the fitness of a surrounding for a given behavioral goal (e.g., hunting for a certain type of prey, seeking shelter), it will suffice most of the time to assign a fairly broad class label to the encountered surfaces. For navigation, however, identification of a landmark has to be performed. While identification is in general more demanding than assigning objects to broader classes, it should be pointed out that the only requirement for landmark identification is uniqueness within the current context. The following example illustrates the notion of uniqueness within context: If a bat has flown about 10 min since leaving its roost and it encounters an elm tree, landmark identification is feasible by determining the tree species correctly and knowing that there are no other elm trees within any distance that could have been possibly bridged within 10 min. Therefore, even a fairly broad categorization of natural surfaces may aid the animals considerably in navigating through their nightly environments. In the present work, the properties of echoes from examples of natural surfaces will be presented, which are readily commensurate with classification beyond the coarsest possible categories, but do not attempt to trace the very limits to the resolution of classification.

Sonar-based target classification is by no means a new research topic. For the special case of fluttering insect prey, it has been convincingly demonstrated that certain bat species can use the pattern of transient Doppler shifts and specular reflections associated with the insect's wing movement for prey classification (von der Emde and Schnitzler, 1990; Kober and Schnitzler, 1990). It remains to be seen, however, whether target classification constitutes a more universal function of biosonar in bats, allowing the animals to attach labels to most of the echoes they receive. Using small deterministic objects as examples, Kuc (1997) showed in an experimental approach that classification of such targets is readily feasible within a framework of biosonar functional principles including adaptive mobility, focusing, and a lossy signal representation.

The basic auditory signal representation is generally assumed to be spectrogramlike (e.g., Lyon, 1982; Patterson *et al.*, 1992). Altes (1984) investigated likelihood ratio classification of zero-mean nonstationary random processes and found the maximum likelihood classifier for this problem to be a spectrogram correlator. While the work of Altes (1984)

[a]Electronic mail: rolf.mueller@yale.edu

opens a window on the usage of the signal representation generated by the auditory periphery for the texture classification problem, the present work focuses on achieving an understanding of how echo properties are linked to properties of natural textures in the first place. For this purpose *ad hoc* methods of feature extraction suffice and we will postpone the evaluation of systematic approaches to feature selection in the auditory basis representation to a later stage of our research.

In the present work, classification of foliages as a ubiquitous example of natural structures in the habitats of many bat species will be considered. Previous work by McKerrow and Harper (1999) has already demonstrated an ability to classify plants using continuous wave frequency modulated (CWFM-) sonar. Here the echo properties which are relevant to classifying different foliage types (e.g., of different plant species) will be highlighted, in order to lay the foundations for establishing whether the sensory world of bats contains such class information. A clear picture of the echo properties relevant to classification is crucial to developing realistic hypotheses for the extraction of biologically meaningful parameters from these signals by the bats' auditory system. Two architecturally distinct plant species will be considered as examples of how echo properties relate to the structural layout of a foliage. These findings should generalize well to any foliage displaying the relevant structural properties.

I. GENERAL METHODS

Two plant species, each with a pronouncedly different architecture, were chosen as examples for studying the echo generating process in foliages: The weeping fig (*Ficus benjamini*) has flat, oval leaves with a length of 85.2 ± 8.5 mm and a width of 36.4 ± 4 mm in mature leaves ($N=100$ leaves). Its foliage is sparse. The spreading yew (*Taxus media*) has small, needlelike leaves (length$=19.5\pm2.5$ mm, width$=2.4\pm0.3$ mm, $N=100$ leaves), which are much more densely packed than the leaves of the fig. One potted plant of each species was used in the experiments. The fig had a diameter of 0.9 m and a height of 1.1 m; for the yew the respective values were 55 and 40 cm. In the experiments reported in Sec. II C the fig was replaced with branches of holly (*Ilex sp.*), which is similar to the fig in size, shape, and distribution of the leaves. The reasons for this replacement are given in Sec. II C 1.

A sonar head was mounted on a five degree-of-freedom robot arm (x, y, z, bearing and elevation). It comprised three transducers (Polaroid 7000 Series; 2.5-cm-diam circular aperture; frequency response was maximal at ~50–60 kHz; the first null in the directivity pattern for 50 kHz occurred at $\sim\pm20$ degrees). The housing of the transducers was removed and their centers were arranged on the corners of an upside-down equilateral triangle with ~6-cm sidelength. All three transducers were tilted inward by ~15 degrees. The bottom transducer (''mouth'') served as an emitter, the two top transducers (''ears'') as receivers. Excitation of the emitter was impulsive, leading to a Gabor-wavelet-like waveform of ~60-μs duration. Most signal energy is contained in the frequency band ~30–100 kHz. This is similar to the frequency range of operation known for bat sonar.

The receiver output was digitized at a sampling rate of 2 MHz and with 12-bit resolution (signal-to-noise ratio ~65 dB). Pre- and posttrigger intervals were chosen to ensure the complete echo waveform was contained within any datavector. Echoes were collected according to a B-scan protocol, with both plants scanned a total of four times: from opposite sides of the plant to obtain two independent views of it and with two different bearing angles: normal (0 degrees) and oblique incidence (45 degrees). In all cases, the sonar head passed the foliage at a distance of about 25 cm, ensuring a good signal-to-noise ratio. Echoes were recorded binaurally every 1 mm along a cross-range interval of 30 cm.

The data were processed to extract various echo features, namely impulsiveness, total signal energy, maximum amplitudes, variance of range estimates, time dependence of sound channel gain as well as range of visibility and uniformity of individual reflections. The specific methods used are stated later together with the results (Sec. II). For all echo features as well as their pairwise combinations, probability density functions were estimated using a Gaussian kernel estimator in conjunction with biased cross-validation (Sain *et al.*, 1994).

Probability density function estimates obtained for normal and oblique sound incidence separately are based on 2400 echoes each. The sample size is twice as large in cases where data was pooled regardless of orientation. Since correlation distances of up to ~1 cm were found, not all the echoes contained in a B-scan with 1-mm cross-range between data acquisitions constitute independent samples. Thus the effective sample size is somewhat smaller than the numbers given earlier. Binaural data obtained at each position show little correlation beyond what is introduced by the pulse, which is the same for both ears (mean binaural correlation coefficient $r=0.48\pm0.09$, $N=2400$); therefore, pooling data from both ears should not result in a noticeable reduction of the effective sample size.

II. EXPERIMENTS

A. Impulsiveness of the echoes

Since the echo trains from all foliages were pronouncedly random in nature, features linked to statistical properties of the random process (and to the echo-generating mechanism) have to be estimated from each individual realization in order to achieve the necessary invariance. An example of such a property is the probability density function of the signal amplitudes, a suitable description of which is provided in this section.

1. Method

To select data for analysis an *ad hoc* method was employed to crop the leading and trailing stretches of noise in each recording: Blocks of 200 samples (0.1 ms) at the beginning and at the end of the recordings were assumed to contain noise only. Moving inwards from the edges of the datavector, the 75%-percentiles of the sample values in subsequent blocks were compared to those of the assumed ''noise-only blocks.'' Blocks were discarded as long as their 75%-percentiles did not exceed twice the percentile of the marginal blocks. Two techniques were applied to reduce the

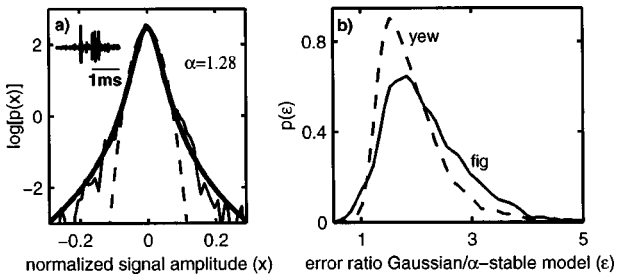


FIG. 1. (a) Example of an amplitude distribution in a single echo obtained from a fig foliage (waveform comprises $N=4000$ samples). The echo waveform is shown in the inset. Thin line: Kernel density estimate (estimated asymptotic mean integrated squared error 0.016); thick line: symmetric α-stable probability density function, with parameters estimated from the waveform; dashed line: Gaussian probability density function which minimizes the mean squared error to the kernel density estimate. Amplitude values were normalized by the maximum value of the entire sample. (b) Estimates of the probability density functions for the ratio of rms errors obtained for the Gaussian and the α-stable model with respect to the kernel density estimates. Solid line: fig; dashed line: yew. Estimated asymptotic mean integral squared errors: 1.2×10^{-3} (fig) and 1.6×10^{-3} (yew).

nonstationarity of the echo sequences: First, the effect of geometric attenuation (inverse square law) was corrected for. Second, if some confidence in a log-linear model of the channel gain was obtained (metric correlation coefficient $r\geqslant0.6$, see Sec. II E), this influence was compensated for as well by multiplying with the inverse of the estimated gain value.

From the resulting time series of amplitude values the parameters of an α-stable distribution were estimated using the quantile method proposed by McCulloch (1986). The characteristic function $\Phi(t)$ of an α-stable distribution is given by

$$\Phi(t)=e^{i\delta t-|ct|^{\alpha}[1+i\beta\ \mathrm{sgn}(t)\omega(t,\alpha)]}, \tag{1}$$

where $\omega(t,\alpha)$ is defined to be $(2/\pi)\log|t|$ for $\alpha=1$ and $\tan(\pi\alpha/2)$ otherwise. The parameter α is the characteristic exponent; β, c, and δ quantify symmetry, dispersion and location, respectively. A review of α-stable distributions may be found in Nikias and Shao (1995); here the characteristic exponent $0<\alpha\leqslant2$ [Eq. (1)] is used as a convenient measure for the heaviness of the amplitude distribution's tail and hence for the impulsiveness of the echoes. Landmarks for interpreting α in terms of a curve shape are provided by the Gaussian probability density with $\alpha=2$ and the more heavy-tailed Cauchy density with $\alpha=1$. Apart from the heaviness of tails, the dispersion c (c^2 is twice the variance in case of a Gaussian probability density) is of interest. The location parameter δ was found to be close to zero for all echoes; the same applies to the skewness parameter β, indicating a symmetric α-stable distribution. The latter two parameters are therefore ignored in the subsequent analysis.

2. Results

In general the α-stable model provided a better fit to the distribution of amplitudes found in the echoes (Fig. 1). This is true for both plants, but more pronounced for the fig [Fig. 1(b)], which generally produces more heavy-tailed amplitude distributions than the yew under normal sound incidence [Fig. 2(a)]. Almost all echoes are more impulsive than would

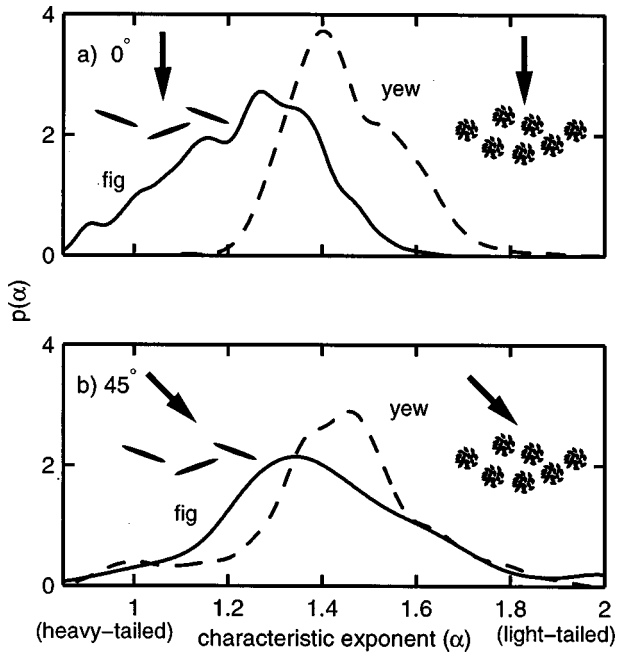


FIG. 2. Distributions of the characteristic exponent α. Solid lines: fig; dashed lines: yew. (a) Normal sound incidence. (b) Oblique sound incidence. The cartoons in the insets represent the orientation of the sonar beam in the horizontal plane with respect to the foliage in the fig (left) and in the yew (right). Leaves are depicted schematically by their cross section. Estimated asymptotic mean-integrated squared errors range between 0.011 and 0.015.

be expected if the amplitude distribution were Gaussian (Fig. 2). For the fig, the degree of impulsiveness is dependent on the orientation of the sonar beam with respect to the foliage: Normal sound incidence [Fig. 2(a)] produces more impulsive echoes than oblique sound incidence [Fig. 2(b)]. For the yew such a strong influence of view angle on the heaviness of the amplitudes distribution's tails is not evident. There is, however, a strong effect of view angle on the dispersion c in both plants (Fig. 3). Separability based on the dispersion c is increased for the oblique condition (Fig. 3). This is in contrast to the results for the characteristic exponent α (Fig. 2).

3. Discussion

The observed echo properties are tentatively explained as follows: The echoes are formed by a superposition of impulselike contributions from the reflecting facets in the foliage. The impulsiveness of the entire echo train is a measure of the visibility of these individual contributions in the superposition result. The comparatively large, planar leaves of the fig act like little mirrors, contributing strong coherent reflections clearly visible in the echoes [see inset in Fig. 1(a)]. In the yew the echoes are a superposition of many low-amplitude contributions from small facets, which are less mirrorlike. The planar fig leaves tend to be oriented uniformly towards the outside, thus strong coherent reflections are more likely to be seen at normal sound incidence on this foliage. The yew foliage has no large facets capable of generating strong mirrorlike reflections, hence the absence of this effect (see insets in Fig. 2 for cartoons depicting leave shape and orientation). This does not imply that echoes from

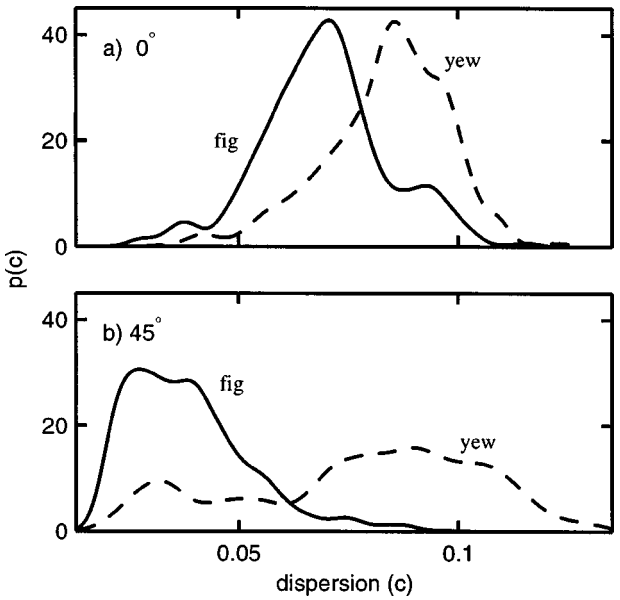


FIG. 3. Distributions of the dispersion parameter c. Solid lines: fig; dashed lines: yew. (a) Normal sound incidence. (b) Oblique sound incidence. Estimated asymptotic mean integrated squared errors range between 0.08 and 0.31.

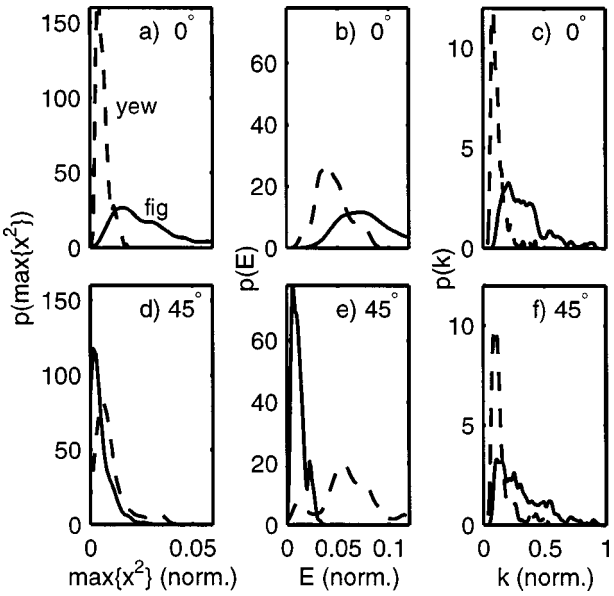


FIG. 4. Distributions estimates for maximum squared amplitude [(a) and (d)], energy [(b) and (e)] and crest factor [(c) and (f)] under normal (a)–(c) and oblique sound incidence (d)–(f). Solid lines: fig; dashed line: yew. Estimated asymptotic mean integrated squared errors (fig/yew): (a) 0.10/0.75, (b) 0.27/0.28, (c) 0.01/0.04, (d) 10.62/1.37, (e) 0.88/0.24, and (f) 0.02/0.06.

the yew are entirely view independent; the dramatic change in the distribution of the dispersion parameter (Fig. 3) provides evidence to the contrary. However, other than in the case of the characteristic exponent α, there is no obvious architectural correlate in the yew foliage for this effect. For the fig, α and c show a similar dependence on viewing angle. This can be interpreted as evidence for coherent reflections producing both, the few large peaks responsible for the tails of the amplitude distribution and the more numerous smaller peaks, which determine the value of c. Specular, coherent reflections can contribute to the recorded echo waveform with large or small amplitudes dependent on how well the directivity of the mirror source and the receiver are aligned.

The finding that most of the recorded foliages echoes have overtly non-Gaussian amplitude distributions may have consequences for the usefulness of estimation techniques based on assumptions of Gaussianity and finite second-order moments, but it should be pointed out that α-stable distributions do not provide an entirely satisfying description of any aspect of a signal [see Middleton (1998) and Sec. II B].

B. Energy and maximum amplitude of the echoes

For a given noise level, the energy of an echo determines the best achievable signal-to-noise ratio, which places fundamental limits on any sonar estimation problem. At the same time, estimation of the signal energy is straightforward. It is therefore worthwhile to compare the echo energy for the foliage types under consideration.

1. Method

The maximum value $\hat{x}_{\max}$ in each sampled waveform $x[t]$, $t=1,...,N$, was measured and an energy estimate $\hat{E}$ was computed by

$$\hat{E}=T\sum_{N} x^2[t], \tag{2}$$

where T is the sampling period. The crest factor is another measure for the impulsiveness of a waveform. Here a crest factor k was computed as the ratio of the maximum squared amplitude and the energy. The common definition of the crest factor ($\hat{x}_{\max}/x_{\mathrm{rms}}$) is related to k by a squaring and a factor of $1/(NT)$. It should be noted that the notion of finite signal power or energy may in principle not be reconciled with the amplitudes being α-stable distributed with $\alpha<2$. However, physical echoes doubtless have finite energy, thus, while α-stable distributions are employed here as a convenient method to quantify the heaviness of the amplitude distribution's tails, they are not adopted as a closed framework for any estimation carried out on the signals. Concerns about the variance of the estimator given by Eq. (2) can be addressed by considering the distribution of the estimates obtained.

2. Results

The distributions obtained for energy and maximum squared amplitude show a pronounced view dependence [Figs. 4(a) and 4(d), 4(b) and 4(e)]: For normal sound incidence the echoes from the fig tend to have larger values than echoes from the yew [Figs. 4(a) and 4(d)]. The dispersion is also greater in case of the fig; this is particularly true for the maximum squared amplitude [Fig. 4(a)]. The similar view dependence of both signal properties is the result of a simple linear relationship between them (Fig. 5). For the fig this relationship is particularly strong ($r=0.945$, $0.938\leqslant r\leqslant 0.952$ with error probability $\alpha=0.001$); for the yew it is a visible tendency ($r=0.731$, $0.698\leqslant r\leqslant 0.761$ with error probability $\alpha=0.001$). The energy content of yew echoes having a given

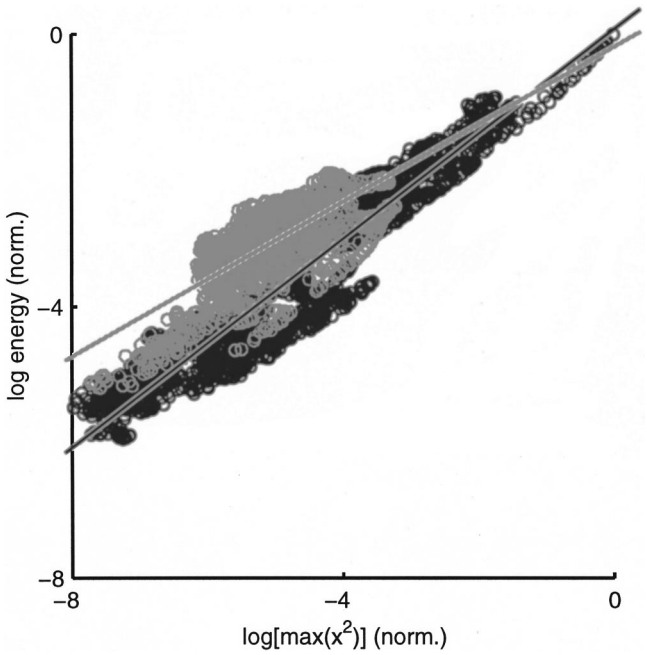


FIG. 5. Scatter plot of signal energy against maximum squared amplitude. Both quantities are normalized with respect to the maximum value obtained for the entire data set. Black symbols and regression line: fig; gray symbols and regression line: yew.

maximum amplitude tends to be larger than for echoes from the fig with the same maximum amplitude value (Fig. 5). This is mirrored by the crest factor which tends to be larger for the fig at both viewing angles [Figs. 4(c) and 4(f)].

3. Discussion

The data on maximum amplitude and signal energy shows two remarkable properties: the relationship between the two variables and their pronounced view dependence. The simple linear relationship between maximum amplitude and energy can be readily explained by considering the impulsive nature of the echoes: The energy is a linear function of $x^2[t]$ for any t [Eq. (2)]. If $x[t_{\max}] \gg x[t \neq t_{\max}]$, $\hat{E} \approx T x^2[t_{\max}]$. The same approximately linear dependence on the maximum value may also be found for impulselike peaks in the waveform which extend beyond a single amplitude sample, as long as different maximum values just scale the respective peak; in this case

$$\hat{E} \approx x^2[t_{\max}] T \sum_{t_{\max}-P/2}^{t_{\max}+P/2} k^2[t], \tag{3}$$

where P is the number of samples within the duration of the peak and $k[t]=x[t]/x[t_{\max}]$. Here the energy of the peak with normal maximum amplitude $T\Sigma_{t_{\max}-P/2}^{t_{\max}+P/2} k^2[t]$ acts as a proportionality constant. Since large peaks are a more prominent element in the echo waveforms obtained for the fig, the linear relationship is much better developed than for the yew (Fig. 5). Peaks from strong specular reflections likewise explain the pronounced view dependence seen in the fig: It is due to the limited extent of the solid angle from which the specular reflection originating from a single facet may be observed in conjunction with a preferred direction of leaf-normals in this foliage. The greater dispersion in the fig echoes at normal incidence [Fig. 4(a)] indicates that a head-on hit of a planar reflecting surface does not always occur in the fig. Thus, while coherent reflections of very large amplitudes are a feature seen exclusively in the fig, they are also frequently absent and hence may not be useful for classification of any echo. It should be noted that the view dependence of the energy seen in both the fig and the yew [Figs. 4(b) and 4(e)] mirrors the behavior of the dispersion c (Fig. 3), which equals the signal power in the case of a Gaussian amplitude distribution. The strong view dependence renders energy and maximum amplitude an unreliable substrate for classification based on single echoes, as long as no independent knowledge of the grazing angle is available. This shortcoming can be eliminated to a large extent by formation of the crest factor k [Figs. 4(c) and 4(f)].

C. Penetration of foliage

The durations of the echo trains recorded for fig and yew foliages frequently corresponded to range intervals of up to ~30 cm [see inset in Fig. 1(a) for an example]. The experiments reported in this section were designed to clarify, whether these ranges corresponded to an actual penetration of the foliage by the sonar beam or whether they resulted from multi-path reflections in the peripheral layer of leaves.

1. Method

In order to determine the locations from which echoes originate, experiments with two single branches stacked in range were performed. In these experiments, the fig was replaced with holly, which has planar leaves similar in size to the fig, but offers the additional advantage that the leaves of a single branch tend to be arranged in a plane. B-scans were performed of the following three arrangements: a single branch at a range of ~20 cm, the same branch with a second branch placed immediately behind it, the second branch alone (first branch removed). The sonar head was aligned for normal sound incidence on the foliage.

2. Results

Reflections from both first branches (holly and yew) covered the entire cross-range interval examined; the yew appeared to be more dense than the holly [Figs. 6(a) and 6(d)]. Nevertheless, the second yew branch was clearly visible [Fig. 6(e)], whereas the second holly branch contributed to the B-scan image only at the fringe of the first branch, where the leaf layer of the first branch became patchy [Fig. 6(b)]. Removal of the first branch shows that both second branches were capable of generating strong echoes [Figs. 6(c) and 6(f)]. Finally, comparison of the different conditions for one plant emphasizes the reproducibility of the B-scan images for each individual branch.

3. Discussion

The results demonstrate that echoes can originate in the depth of all studied foliages; multi-path phenomena—if present—do not contribute noticeably to the observed echo

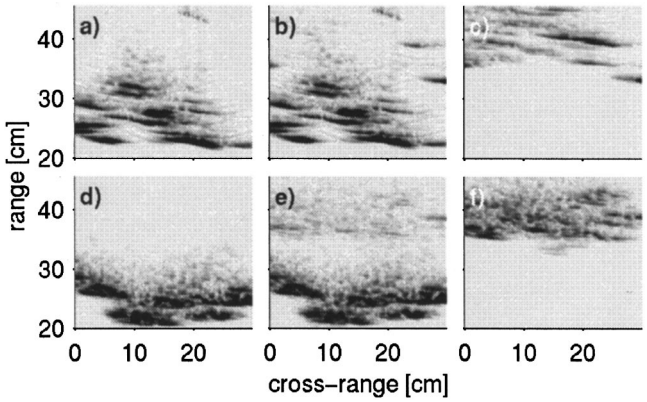


FIG. 6. B-scans obtained from two stacked branches of holly (a)–(c) and yew (d)–(f). (a) and (d) first branch alone; (b) and (e) both branches; (c) and (f) second branch alone. The dynamic range of the images has been restricted to 23 dB.

durations. The mechanisms which allow echoes to originate from inside the foliages differ between the yew and the sparser foliages of the holly and fig: The yew foliage paradoxically combines a dense packing of reflecting facets in the outer foliage layer with good visibility of the inside. The larger, planar leaves of the holly (and the fig) may be effective sound barriers themselves, but their sparser distribution allows the sound to travel through the larger voids between the leaves, again giving access to reflecting facets distributed in a volume rather than on a surface. In Sec. II E we deal with signal features which are most likely attributable to the different ways by which sound reaches inner layers of the foliages.

The finding that the example foliages studied are volume scatterers has implications with regard to the sensory world of bats: In many of their natural environments (e.g., in forests), the animals should perceive their action spaces as delimited not by a collection of two-dimensional surfaces (Bozma and Kuc, 1991), but by three-dimensional ''clouds'' of reflecting facets. These ''clouds'' can differ in properties of the individual facets as well as in their distribution. A sonar-based characterization of the leaf distribution is presented in the next section (Sec. II D); properties of the individual reflectors are dealt with in Sec. II F.

D. Leaf distribution

Visual inspection gives rise to the impression that leaves are distributed much more sparsely in the foliages of leafy trees like the fig or the holly than in the yew. This is also evident from B-scan images (Fig. 6). In this section it is shown that a simple sonar ranging paradigm, which is well within the capabilities of bat biosonar, can readily demonstrate these differences.

1. Method

The employed ranging paradigm is based on incoherent reception and crossing of an amplitude threshold: After correcting for the geometric attenuation, the envelopes were computed for each echo in a B-scan as the magnitude of the analytic signal formed by the Hilbert transform. Threshold values were linearly spaced between zero and the maximum envelope amplitude. The first time at which a given threshold was exceeded was noted for each echo.

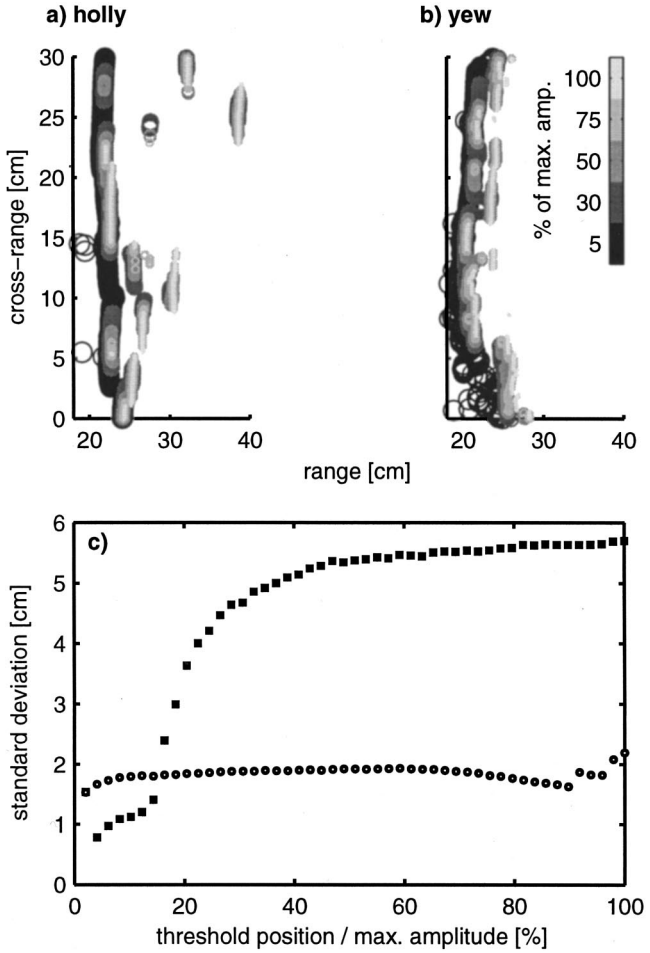


FIG. 7. First-passage ranges in example B-scans for (a) holly and (b) yew. (c) shows the standard deviation of the first passage distributions as a function of threshold position in relation to the maximum amplitude of the waveforms. Filled squares: holly; open circles: yew.

2. Results

In both plants the ranges of first-time level crossings varied little along the cross-range axis for low-amplitude thresholds [Figs. 7(a) and 7(b)]. In the yew this behavior was hardly affected by raising the threshold level, whereas in the holly the distribution of first-passage range started spreading out with rising thresholds. This difference in dependence on threshold level is clearly visible if the standard deviation of the first-passage distribution is plotted as a function of threshold position (in relation to the maximum amplitude value in the waveform): For the yew, an essentially horizontal line results, whereas for the holly the curve has a sigmoid shape with a saturation level almost three times above the values obtained for the yew [Fig. 7(c)].

3. Discussion

When using sonar to characterize the leaf distribution, this distribution is observed through the directivity function of the sonar system. Setting low threshold levels makes the

system sensitive to echoes from a wide range of bearing angles. The sonar beam's diameter is large in relation to gaps between the reflecting facets in the foliage at low threshold. Consequently, the roughness in the initial layer of reflecting facets is not visible. Raising the threshold does not change this in the case of the yew, since the beam cannot be narrowed to fit the small crevices between the needles. In the case of the holly, choosing a higher threshold level is sufficient for the sonar beam to pass through the voids between the leaves and thereby reveal the variability in the range of the individual reflectors. It should be noted that other sources of variability in the amplitude of a particular scatter's contribution to an echo exist (these influence the scattering coefficient). These factors may well assign a leaf with unfavorable weighting by the directivity a higher echo amplitude than one in the center of the beam. But since these factors are independent of beam position, on average the effect of the beam-weighting function will be seen, when comparing first passage distributions obtained for different amplitude thresholds.

E. Time dependence

Sound propagates deeply in both types of foliage studied here (Sec. II C). However, the channel it uses for doing so is pronouncedly different in each case: In the fig, comparatively large voids allow access to reflecting facets inside the foliage, whereas in the yew, only small crevices are open to sound propagation. It may be hypothesized that these obvious differences in the sound channels are mirrored in likewise distinct echo properties.

1. Method

A log-linear model was fitted to the data: After correction for spreading losses, the echo envelopes were extracted (Sec. II D). Each envelope was logarithmically compressed and the time determined where the echo amplitude exceeded 4.2 times the standard deviation of the amplitudes in this time series (echo together with short leading and trailing stretches of noise). The factor of 4.2 was chosen because it maximized separation between the two foliage types (see next section) and observed the condition that the chosen amplitude threshold is exceeded in any of the recorded echoes. A linear regression model was then fitted to all sample values from the maximum amplitude to the end of the recorded waveform.

2. Results

In the yew there is a visible negative correlation between time and echo amplitude, i.e., there is a time-dependent, decreasing sound channel gain. For the fig, no such relationship could be demonstrated. Thus, there was little overlap between distributions for the regression coefficient r (Fig. 8). This finding holds for both normal and oblique sound incidence, although at oblique incidence the difference was reduced. The first moments of the regression coefficient's distribution were -0.072 in the fig and -0.623 in the yew at normal incidence as opposed to -0.253 and -0.572, respectively at oblique incidence.

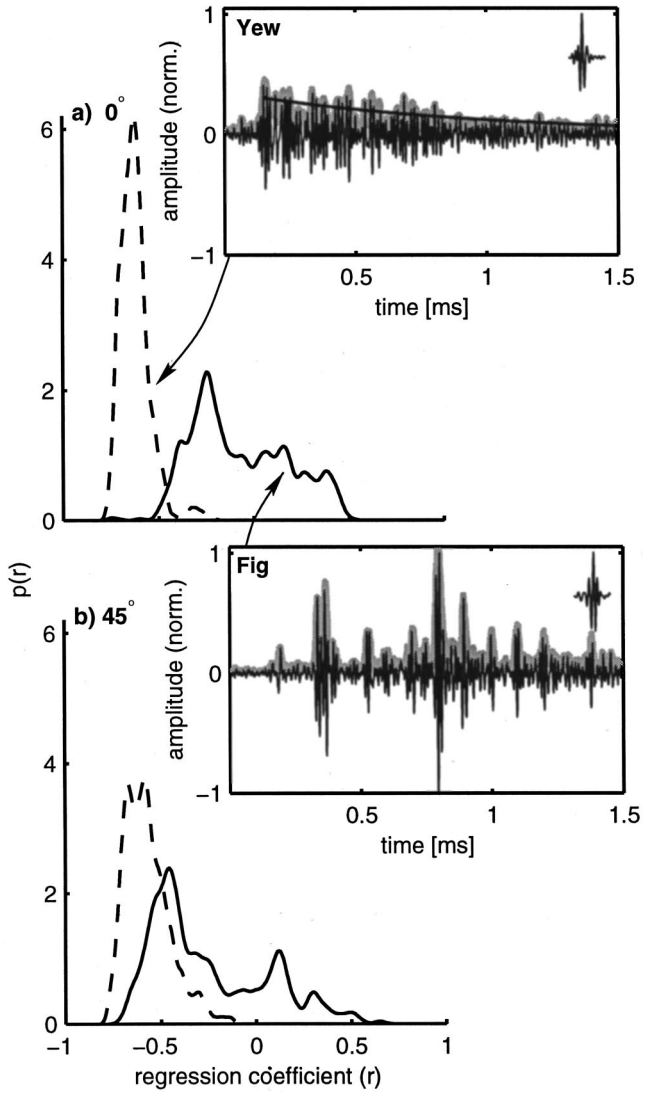


FIG. 8. Time dependence of sound channel gain: probability density function estimates for the regression coefficient (r) in the fig (solid lines) and the yew (dashed lines): (a) normal sound incidence; (b) oblique sound incidence. The insets show typical waveforms obtained at normal sound incidence for both plants. For the yew, the fit by the log-linear regression model is shown ($r=-0.72$), for the fit it is omitted, because the model does not provide a good fit to the data ($r=-0.19$). In the upper right corner of each inset echoes from an individual leaf of the respective plant are shown for comparison. Estimated asymptotic mean integrated squared errors (fig/yew): (a) 0.013/0.018; (b) 0.014/0.015.

3. Discussion

The results provide evidence that after an initial buildup of echo amplitude in the yew, the channel gain tends to decrease as a power function. The regression coefficients obtained still indicate a large variability in echo amplitude at any given time inside the echo train. Hence, the results demonstrate a general tendency which is superposed on this variability (see yew inset in Fig. 8 for an example). A decrease of sound channel gain according to a power function is compatible with assuming the foliage to transmit a fixed fraction of the signal amplitude in each layer. Since the sound propagates into the fig foliage through comparatively large voids, no visible attenuation (beyond geometrical attenuation and

absorption in the medium) is associated with this process and no trend in the echo amplitudes can be discerned (see fig inset in Fig. 8 for an example).

The log-linear regression coefficient constitutes a remarkably good classification feature for the two plants: The overlap between the distributions is small at normal incidence, moderate at oblique incidence and both distributions show little view independence. This should allow us to tell the two plant species apart with a low probability of an error from any aspect angle.

F. Individual reflector properties

The echo-generating process differs between the fig and the yew because the reflecting facets are more sparsely distributed in the fig and each facet is capable of generating strong coherent reflections, if viewed under a favorable angle. In this section the dominance of strong, coherent components in the echo trains produced by the fig foliage will be demonstrated.

1. Method

Echoes in all eight B-scans (normal and oblique sound incidence) were compared pairwise by means of a cross correlation. The cross correlation was normalized to a unity value of the autocorrelation at zero lag. The maximum value obtained in each cross correlation was stored in a correlation matrix. The correlation matrix modifies the concept of the covariance matrix by using the cross correlation instead of just the expectation of the products at zero lag. This ensures that similarities in the compared waveforms are found irrespective of time shifts introduced by an eventual interdependence of cross range and range in a given reflector.

2. Results

The correlation matrix for the yew B-scans has a very simple structure, for $|x-y|>0$ (off the main diagonal) $\max[R_{xy}(\tau)]$ decreases rapidly (within differences in cross range of about 1 cm) to a pedestal value of ~ 0.4 [Fig. 9(b)]. In the results for the fig B-scan the same pattern of rapidly decaying correlation with cross-range distance is discernible, but it is augmented by blunt peaks of higher correlation located on the main diagonal as well as away from it. The differences shown in the examples (Fig. 9) can be easily quantified by considering, for example, the average value of $\max[R_{xy}(\tau)]$ off the main diagonal. Pooling oblique and normal sound incidence, for the yew this average value ranged between 0.42 and 0.47 and between 0.45 and 0.85 for the fig ($N=8$ B-scans; each correlation matrix was evaluated at $\sim 10^4$ points on one side of the main diagonal).

3. Discussion

The observed structure of the correlation matrices is explained by the following hypothetical scenario: In the yew, the matrix reflects the fact that each echo train constitutes a superposition of many small components. Over only a small cross-range increment, the weights of the components are so thoroughly changed that hardly any similarity remains. The pedestal correlation value seen in both foliage types is due to usage of the same pulse in eliciting all echoes. The broadenings in high correlation values along the main diagonal occurring in the fig indicate that some strong specular reflections remain visible over a larger cross-range interval. The areas of high correlation off the diagonal can hardly be due to the same reflecting facet; they stem from similarities between the echoes generated by different facets. These similarities are not surprising, since the large reflecting facets in the fig are mirrorlike and hence generate an echo waveform closely resembling the pulse.

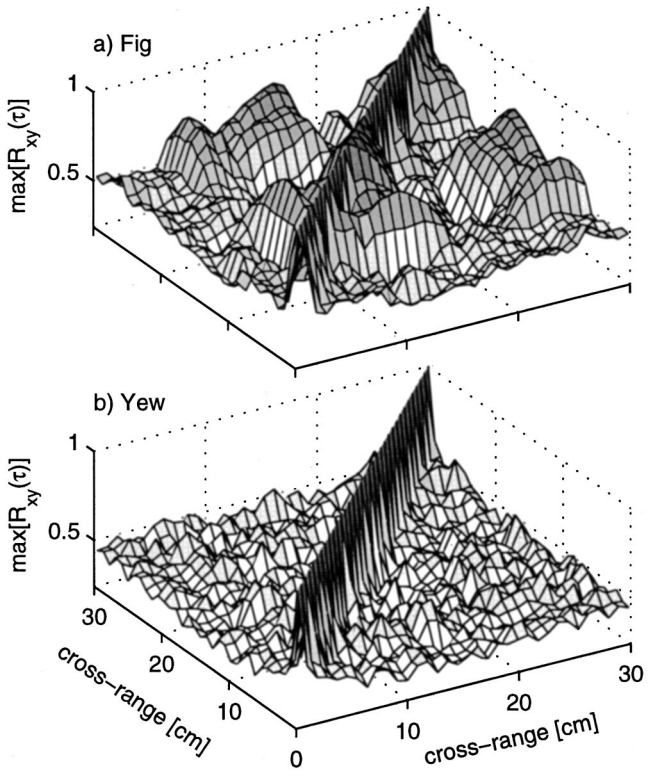


FIG. 9. Examples of correlation matrices computed for B-scans in the fig (a) and the yew (b).

G. Estimated Bayes risks

The probability density function estimates given in the previous sections can be used to assess the separability of the two foliage classes. This was done here by estimating the Bayes risk, a measure of error in a discrimination task, assuming equal *a priori* probabilities of the target classes. A summary of the results shows that classification based on single features estimated from a single echo is possible with a moderate error probability (Table I): The most powerful single feature for normal incidence and pooled data is a nonstationarity in the time domain (correlation coefficient r, Bayes error $\sim 2.4\%$ for normal, $\sim 13\%$ for pooled data). It is followed by the crest factor k and the dispersion c in case of the pooled data, by squared maximum amplitude $\max^2$ and k for the normal incidence data. For oblique incidence, the energy E is the best individual feature followed by c. Using pairwise combinations of the features, the Bayes risk can be reduced even further: The best combinations are r and $\max^2$

TABLE I. Estimates of Bayes errors (in %) for single features and their pairwise combinations. Feature pairs are evaluated by their joint (2-D) probability density function. Combining a feature with itself means that the (1-D) probability function of this feature is evaluated (a) shows normal incidence, (b) shows oblique incidence, and (c) shows normal and oblique incidence. The feature name abbreviations are α: characteristic exponent of the α-stable distribution; c: distribution parameter of the α-stable distribution; E: signal energy; $\max^2$: maximum squared signal amplitude; k: crest factor; and r: metric correlation coefficient of the log-linear regression.

α	c	E	$\max^2$	k	r	
(a)						
22.1	20.0	16.6	7.1	9.2	0.9	α
	26.7	10.9	6.0	13.1	2.0	c
		20.1	8.6	10.3	1.9	E
			9.4	7.9	0.7	$\max^2$
				14.2	1.6	k
					2.5	r
(b)						
40.5	15.4	7.0	26.1	18.5	19.5	α
	16.9	6.1	14.6	14.0	14.5	c
		7.4	5.5	3.9	6.3	E
			32.4	7.8	20.5	$\max^2$
				24.1	10.7	k
					23.1	r
(c)						
31.7	20.0	17.4	18.2	16.1	10.7	α
	21.9	13.3	10.2	14.1	8.8	c
		23.8	13.2	15.7	9.9	E
			22.7	14.5	9.1	$\max^2$
				19.1	6.9	k
					13.0	r

for normal incidence, E and k for oblique incidence, and r and k for the pooled data. Estimated Bayes risks are ~0.7%, ~3.9%, and ~6.9%, respectively.

III. GENERAL DISCUSSION

The experiments conducted in this work have demonstrated signal features in the impulse responses of two architecturally distinct plant species, which are suitable for classifying these example plants with a reasonably low error probability. This classification is possible based on just a single echo, and is irrespective of whether sound incidence is normal or oblique. How fine is the discrimination of different plant architectures using sonar? The fig and the yew are fairly distinct, but not extremes in terms of leaf size, orientation, and distribution. The distance between them in the studied feature space leaves little room for reliable discrimination of intermediate foliage types. However, many possible ways to increase the resolving power of foliage classification have not yet been explored. Notable examples are comparisons between several echoes and adaptive mobility, which may allow the animals to actively descend a gradient in some signal feature towards a better viewing position. It can therefore be concluded with some confidence that the impulse responses of foliages contain enough classification-relevant features in order to assign useful labels to many of these natural surfaces.

The methods used to obtain the reported results deviated from the functional characteristics of bat sonar in important ways: (1) Pulse durations were shorter than those typically seen in bats. (2) Echoes were collected according to a B-scan paradigm with precisely known positions of the sonar head. (3) Some of the feature extraction methods are most likely not reasonable models of the auditory signal processing in bats (e.g., coherent reception for computing the correlation matrices of the B-scans). These non-biomimetic experimental designs were deliberately chosen because they facilitate understanding the properties of the echo generating process and its relation to the echo features. Realistic auditory classifier models can hardly be developed without this knowledge.

A fundamental insight into the nature of foliage echoes is their profound stochasticity and non-Gaussianity. In short, the same echo from a foliage is never seen twice. Salient, sufficiently invariant features may therefore be found only in the realm of statistical expectations. Likewise, the placement of the reflecting facets in a foliage is random to the observer. This renders, for example, the position of an individual leaf meaningless (tomorrow it may be in a different position due to the influence of wind). Informative for classification are only the parameters of the spatial leaf distribution and the distribution of reflector properties. Hence recognition of echoes from natural textures is fundamentally different from recognizing simple deterministic objects. The notion of a statistical approach to impulse responses is particularly well known in the processing of geophysical signals (e.g., Robinson and Treitel, 1980), but novel to the study of biosonar. Based on psychoacoustical work, a ''two-wavefront interference'' paradigm, which requires bats to discriminate different delays between two echoes, has been suggested as a model for the perception of textures in the bats' natural habitats (Schmidt, 1988). This model is unable to cope with the stochasticity of foliages and the echoes to which they give rise: Discrimination of different delays between two identical replicas of the input can be based on the invertible relationship between the deterministic spacing of two reflectors and the likewise deterministic spacing of spectral notches. Such deterministic comb-filter transfer functions can be readily evaluated using a library of spectral templates (Schmidt, 1999). However, any such library will fail on the echoes produced by the foliages studied here, because they combine a large variability with an essentially white expectation of the transfer function.

The finding that ubiquitous natural textures give rise to pronouncedly random signals is highly recommended for consideration as a starting point for experimental work in bats. What needs to be investigated is the kinds of tasks bats can actually perform on random inputs, e.g., can they be trained to classify different types of foliage? Likewise the neural responses elicited by the convolution sum of bat echolocation and non-Gaussian random sequences deserve attention.

What would be the most advantageous (auditory) representation for the classification of foliage echoes? A satisfying answer to this question has to await the results of a systematic approach to feature selection for this estimation problem. Nevertheless, it is obvious from the work reported here that very powerful features (e.g., the nonstationary gain

seen in the yew, Sec. II E) require a representation which includes a temporal dimension. Given the surprisingly long duration of the foliage echoes (~2 ms and even longer), this is not beyond what the auditory system could accomplish. In order to make progress towards a hypothetical auditory representation of these foliage impulse responses, batlike chirps will be employed and a biomimetic representation of the resulting echoes will be evaluated in future work. Important elements of this representation would be a bandpass filter bank, demodulation, and a sparse spike-code.

ACKNOWLEDGMENTS

This work was supported by a postdoctoral fellowship granted by DAAD and NSF Grant No. IRI-9504079. The authors wish to thank Dave Mellinger for his suggestions on an earlier draft of the manuscript.

Altes, R. A. (**1984**). ''Texture analysis with spetrograms,'' IEEE Trans. Sonics Ultrason. **31**, 407–417.

Bozma, Ö., and Kuc, R. (**1991**). ''Characterizing pulses reflected from rough surfaces using ultrasound,'' J. Acoust. Soc. Am. **89**, 2519–2531.

Gibson, J. J. (**1979**). *The Ecological Approach to Visual Perception* (Houghton Mifflin, Boston).

Kober, R., and Schnitzler, H.-U. (**1990**). ''Information in sonar echoes of fluttering insects available for echolocating bats,'' J. Acoust. Soc. Am. **87**, 882–896.

Kuc, R. (**1997**). ''Biomimetic sonar recognizes objects using binaural information,'' J. Acoust. Soc. Am. **102**, 689–696.

Lyon, R. F. (**1982**). ''A computational model of filtering, detection and compression in the cochlea,'' *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*.

McCulloch, J. H. (**1986**). ''Simple consistent estimators of stable distribution parameters,'' Commun. Stat.-Simul. Comput. **15**(4), 1109–1136.

McKerrow, P. J., and Harper, N. L. (**1999**). ''Recognizing leafy plants with in-air sonar,'' Sensor Rev. **19**, 202–206.

Middleton, D. (**1998**). ''Fundamental limitations on the use of α-stable probability models for physical noise processes,'' in *Proceedings of the 1998 Conference on Information Sciences and Systems*, CISS '98 (Princeton University, Princeton, NJ, 1998), Vol. II, pp. 856–863.

Nikias, C. L., and Shao, M. (**1995**). *Signal Processing with Alpha-stable Distributions and Applications* (Wiley, New York).

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1992**). ''Complex sounds and auditory images,'' in *Auditory Physiology and Perception*, edited by L. Cazals and Y. Demany (Pergamon, Oxford), pp. 429–446.

Robinson, E. A., and Treitel, S. (**1980**). *Geophysical Signal Analysis* (Prentice–Hall, Englewood Cliffs, NJ).

Sain, S., Baggerly, K., and Scott, D. (**1994**). ''Cross-validation of multivariate densities,'' J. Am. Stat. Assoc. **89**, 807–817.

Schmidt, S. (**1988**). ''Evidence for a spectral basis of texture perception in bat sonar,'' Nature (London) **331**, 617–619.

Schmidt, S. (**1999**). ''Auditory categories for the perception of echo spectra in bats,'' J. Acoust. Soc. Am. **105**, 1204.

von der Emde, G., and Schnitzler, H.-U. (**1990**). ''Classification of insects by echolocating greater horseshoe bats,'' J. Comp. Physiol. A **167**, 423–430.

LETTERS TO THE EDITOR

This Letters section is for publishing (a) brief acoustical research or applied acoustical reports, (b) comments on articles or letters previously published in this Journal, and (c) a reply by the article author to criticism by the Letter author in (b). Extensive reports should be submitted as articles, not in a letter series. Letters are peer-reviewed on the same basis as articles, but usually require less review time before acceptance. Letters cannot exceed four printed pages (approximately 3000–4000 words) including figures, tables, references, and a required abstract of about 100 words.

About a double-body immersion horn system to be used for quantitative sonochemical studies

J. Reisse,[a] K. Bartik, O. Fabre, and J. Vandercammen
Université Libre de Bruxelles, CP 165/64, 50 av. F. D. Roosevelt, 1050 Bruxelles, Belgium

(Received 26 July 1999; accepted for publication 24 April 2000)

The majority of quantitative sonochemical studies in the 20–100-kHz frequency range are performed by using an immersion horn system. The new system described in this letter consists of a double immersion horn acted on by a single pair of piezoelectric ceramics. This instrument is well adapted to the quantitative measure of effects, i.e., variations in the rate constant of a specific reaction associated to the change of an experimental parameter. The gas effect, obtained by comparing the rate of a reaction under air and under argon, illustrates the efficiency of the system. © *2000 Acoustical Society of America.* [S0001-4966(00)00908-5]

PACS numbers: 43.35.Vz [HEB]

Many, if not all, devices used to perform sonochemical experiments in the 20-kHz frequency range are systems derived from instruments initially developed for other purposes. The so-called sonication bath is nothing else than a cleaning bath, while the immersion horn system is directly derived from the ultrasonic cell disrupter, which itself is similar to an ultrasonic tool commonly used for working plastic. In this letter a new instrument specifically designed for quantitative sonochemical studies is described.

Frequently, if not always, chemists performing sonochemical reactions in pure liquids or homogeneous solutions (homogeneous sonochemistry) are interested in the change of the reaction rate, or the change of the relative concentration of the reaction products, associated with the change of one experimental parameter (temperature, dissolved gas, pressure, ...). In simpler words, these chemists are more interested in the temperature effect, the gas effect, the pressure effect, than in the absolute value of the rate constants.

The study of these various effects, when performed in the 20-kHz range, is generally undertaken with an immersion horn system because sonication baths are not adapted to quantitative studies. An immersion horn consists of piezoelectric ceramics compressed between the active part of the horn and a counter mass, which corresponds to the passive part of the horn.

The new system being developed in our laboratory is shown in Fig. 1. We like to call it the Janus system. It is a double-body system with two similar active horns acted on by a unique pair of piezoelectric ceramics. Each active horn acts as counter mass for the other one. The system is well adapted for measuring effects since the comparison between two different experimental conditions can be performed directly. Another advantage of this system is that any accidental change in the transduction factor of the piezoelectric ceramics or in the electrical power will affect the two reactions. If one of these reactions is a test reaction, the accidental change will immediately be detected. As a criticism of the double-body system, it could be argued that any change in the characteristics of cell 1 will have consequences on the ultrasound intensity in cell 2, and vice versa. Strictly speaking this is true, but in many cases, these effects are second-order effects which can be neglected. An example will be given below. The 20-kHz generator and the amplifier used with this system are home-built instruments. The active surface of both horns is 0.785 cm^2. The internal volume of each double-walled cell is 150 cm^3. The temperature can be controlled. During the series of experiments described below,

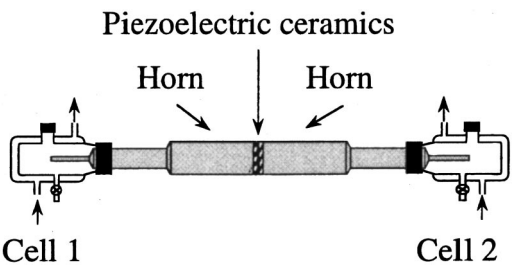


FIG. 1. Schematic representation of the double-body system.

[a]Electronic mail: jreisse@ulb.ac.be

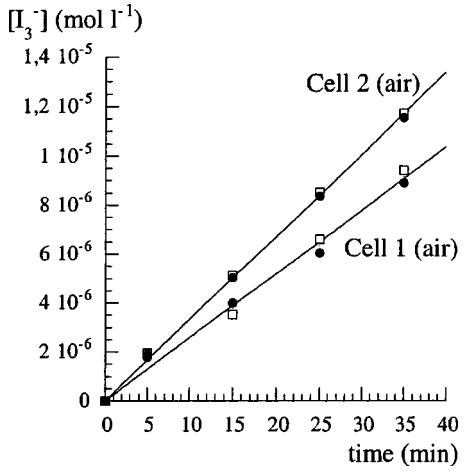


FIG. 2. Evolution of $[I_3^-]$ as a function of time. The solution in both cells is saturated with air. The two different symbols correspond to two successive and independent set of reactions. The lines correspond to a linear regression through the average values.

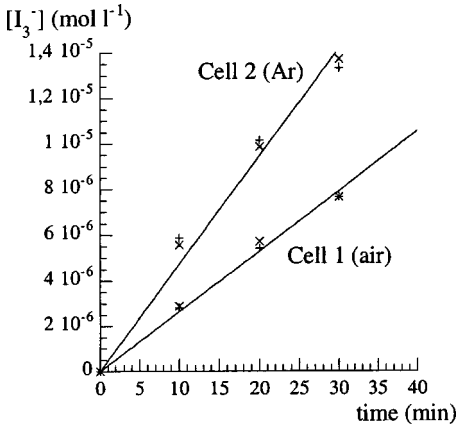


FIG. 3. Evolution of $[I_3^-]$ as a function of time. The solution in reference cell 1 is saturated with air and the solution in the working cell 2 is saturated with argon. Each symbol corresponds to an independent set of reactions. The lines correspond to a linear regression through the average values.

the temperature was $20\,°\pm1\,°$C, and the electrical intensity was set to 300 mA.

Before demonstrating the advantages that the double-body reactor offers for studying a gas effect, we performed the oxidation of I^- (initial concentration: 0.1 M) into I_2 in both cells, under air, using water as solvent and in the presence of a small arnount of CBr_4 $(10^{-4}\,\mathrm{M})$[1,2] in order to test the reproducibility of the system. The results of two experiments performed are shown in Fig. 2 [concentration of I_3^- as a function of time as measured by UV absorption spectroscopy at 352 nm ($\varepsilon = 26.200\,\mathrm{l\,mol^{-1}\,cm^{-1}}$)]. The agreement between the two experiments is fully satisfactory. The difference observed between cell 1 and cell 2 is due to the lack of perfect symmetry of the double-body system: the two horns, or the shape of the two cells, are probably slightly different. Obviously, this difference is of no importance since it is reproducible. It simply leads to the definition of an intrinsic cell correction factor, $k(\mathrm{cell\,2})/k(\mathrm{cell\,1})$, which is equal to 1.29 for this system and which will be applied in the study of a gas effect.

To study a gas effect with this system, the oxidation rate of I^- into I_2 under air and under argon were compared. The results are reported in Fig. 3. Comparing Figs. 2 and 3, it is possible to see that the reaction kinetics in the reference cell (cell 1, reaction performed under air) is the same if the dissolved gas in cell 2 is air or argon and therefore, as suggested earlier, changing the gas in cell 2 truly has a negligible effect on cell 1. The ratio of the two slopes in Fig. 3 is equal to 1.78, the reaction being faster in cell 2. Taking into account the intrinsic correction factor (1.29), the $k(\mathrm{Ar})/k(\mathrm{air})$ ratio becomes $1.78/1.29 = 1.38$. A control experiment was also performed where cell 2 was used as the reference cell (reaction under air) and cell 1 was the working cell (reaction under argon). Taking into account the cell correction factor, the $k(\mathrm{Ar})/k(\mathrm{air})$ ratio is equal to 1.44. The good agreement between this last value and the value obtained with the other configuration (1.38) shows that the double-body system is well adapted to the quantitative measure of effects in sonochemistry. A 4% difference on a rate constants ratio is much below the limit of what can be interpreted in chemical kinetics. Gas effects on sonochemical reactions were, and are still, a subject of great interest.[1–5] Nevertheless, we will not discuss these issues here because the aim of this Letter is simply to demonstrate how the double-body system leads easily to the measure of an effect, and in this particular case, to the measure of a gas effect. The interpretation of the effect is far from trivial because it includes not only intrinsic gas properties, like the polytropic ratio or the heat conductivity, but also concentration terms. The stationary concentration of air and argon under sonication are obviously not the same.[6] Last but not least, N_2 and O_2 are reactive gas while Ar is not.

The system used to perform these experiments is a prototype with two aluminum horns. The final version of the instrument will include two titanium horns in order to avoid excessive erosion of the horn tips.

[1] A. Weissler, J. Acoust. Soc. Am. **25**, 651–657 (1953).
[2] A. Henglein, ''Contributions to Various Aspects of Cavitation Chemistry,'' in *Advances in Sonochemistry*, edited by T. Mason (Jai, London, 1993), Vol. 3, pp. 17–83.
[3] E. L. Mead, R. G. Sutherland, and R. E. Verrall, Can. J. Chem. **54**, 1114–1120 (1976).
[4] P. Riesz, ''Free Radical Generation by Ultrasound in Aqueous Solutions of Volatile and Non Volatile Solutes,'' in *Advances in Sonochemistry*, edited by T. Mason (Jai, London, 1991), Vol. 2, pp. 23–64.
[5] J. Reisse, T. Caulier, C. Dekerckheer, Y. Kegelaers, N. Segebarth, and K. Bartik, ''Some physico-chemical aspects of so-called homogeneous sonochemistry,'' in *Sonochemistry and Sonoluminescence*, edited by L. A. Crum, T. J. Mason, J. Reisse, and K. S. Suslick, Nato ASI Series C, Vol. 254 (Kluwer Academic, Dordrecht, 1999), pp. 205–224.
[6] L. Broeckaert, T. Caulier, O. Fabre, C. Maerschalk, J. Reisse, J. Vandercammen, D. H. Yang, Th. Lepoint, and F. Mullie, ''Quantitative Homogeneous Sonochemistry,'' in *Current Trends in Sonochemistry*, edited by G. J. Price (The Royal Society of Chemistry, 1992), pp. 8–25.

Generalized thermoelastic waves in homogeneous isotropic plates

J. N. Sharma
Department of Mathematics, Regional Engineering College, Hamirpur-177 005 H.P., India

Devinder Singh
Department of Mathematics, SBCMS Institute of Engineering, Mukerian-144 001 Pb, India

Rajneesh Kumar
Department of Mathematics, Kurukshetra University, Kurukshetra-136 179 Haryana, India

(Received 2 April 1999; accepted for publication 25 April 2000)

The propagation of thermoelastic waves in homogeneous isotropic plate subjected to stress-free and rigid insulated and isothermal conditions is investigated in the context of conventional coupled thermoelasticity (CT), Lord–Shulman (LS), Green–Lindsay (GL), and Green–Nagdhi (GN) theories of thermoelasticity. Secular equations for the plate in closed form and isolated mathematical conditions for symmetric and skew-symmetric wave mode propagation in completely separate terms are derived. It is shown that the motion for SH modes gets decoupled from the rest of the motion and remains unaffected due to thermo-mechanical coupling and thermal relaxation effects. The phase velocities for SH modes have also been obtained. The results for coupled and uncoupled theories of thermoelasticity have been obtained as particular cases from the derived secular equations. At short wavelength limits the secular equations for symmetric and skew-symmetric waves in a stress-free insulated and isothermal plate reduce to Rayleigh surface waves frequency equations. Finally, the numerical solution is carried out for aluminum–epoxy composite material and the dispersion curves for symmetric and skew-symmetric wave modes are presented to illustrate and compare the theoretical results.
[S0001-4966(00)01108-5]

PACS numbers: 43.40.At [PJR]

INTRODUCTION

The theory of thermoelasticity is well established.[1] The effect of heat conduction upon the propagation of Rayleigh waves in the semi-infinite elastic solid having isothermal or thermally insulated boundaries has already been discussed in detail.[2,3] To explain and remove the paradox of infinite velocity of heat propagation, some researchers[4–8] derived and formulated generalized theories of thermoelasticity. Some researchers[9–11] investigated the influence of temperature, strain fields, and heat flux on the static and dynamic behavior of coupled thermoelastic plates. Banerjee and Pao[12] investigated the propagation of plane harmonic waves in homogeneous anisotropic solids, taking into account the thermal relaxation time. The propagation of plane harmonic waves in homogeneous transversely isotropic materials has also been studied in generalized theories of thermoelasticity.[13] Chandrashekhariah[14] brought out a comprehensive review of literature on the subject. In the present paper we have discussed the propagation of plane waves in an infinite homogeneous, isotropic thermoelastic plate of thickness $2d$ in the context of various theories of thermoelasticity.

I. FORMULATION OF THE PROBLEM AND ITS SOLUTION

We consider an infinite homogeneous isotropic thermally conducting elastic plate of thickness $2d$ initially at uniform temperature T_0. We take origin of the coordinate system (x_1,x_2,x_3) on the middle surface of the plate. The x_1-x_2 plane is chosen to coincide with the middle surface and the x_3 axis normal to it along the thickness as illustrated in Fig. 1. The surfaces $x_3=\pm d$ are assumed to be (i) stress-free insulated or isothermal, and (ii) rigid insulated or isothermal, boundaries.

The basic governing equations of generalized thermoelasticity[4,5] in the absence of heat sources and body forces are

$$(\lambda+\mu)\nabla(\nabla\cdot\mathbf{u})+\mu\nabla^2\mathbf{u}-\beta\nabla(T+\delta_{2k}t_1\dot{T})=\rho\ddot{\mathbf{u}}, \tag{1}$$

$$K\nabla^2T-\rho C_e(\dot{T}+t_0\ddot{T})=\beta T_0(\dot{e}+\delta_{1k}t_0\ddot{e}), \tag{2}$$

where $\mathbf{u}=(u_1,u_2,u_3)$ is the displacement vector, $T(x_1,x_2,x_3,t)$ is the temperature change; λ, μ are Lame's parameters; K is thermal conductivity; ρ and C_e are, respectively, the density and specific heat at constant strain, and $\beta=(3\lambda+2\mu)\alpha_t$, α_t is the linear thermal expansion and e is the dilatation. The dot notation denotes time differentiation and δ_{ij} is the Kronecker's delta. Here, $k=1$ for Lord–Shulman (LS) theory and $k=2$ for Green–Lindsay (GL) theory. The thermal relaxation times t_0 and t_1 satisfy the inequalities[15]

$$t_0\geqslant t_1\geqslant 0, \tag{3}$$

for GL theory only. We take the x_1-x_3 plane as the plane of incidence and we assume that the solutions are explicitly independent of x_2, but implicit dependence is there so that the transverse component u_2 of displacement is nonvanishing in Eqs. (1) and (2). In view of this, the governing equations (1) and (2) in the nondimensional form can be rewritten as

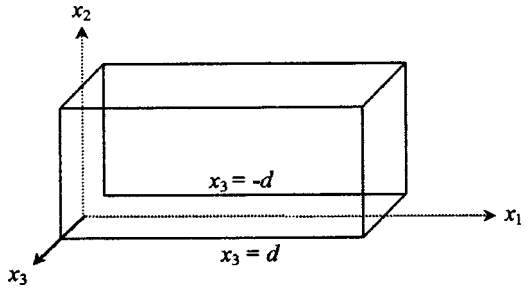


FIG. 1. Geometry of the problem.

$$u_{1,11}+(1-\delta^2)u_{3,13}+\delta^2 u_{1,33}-(T+\delta_{2k}t_1\dot{T})_{,1}=\ddot{u}_1, \tag{4}$$

$$(1-\delta^2)u_{1,13}+\delta^2 u_{3,11}+u_{3,33}-(T+\delta_{2k}t_1\dot{T})_{,3}=\ddot{u}_3, \tag{5}$$

$$\delta^2(u_{2,11}+u_{2,33})=\ddot{u}_2, \tag{6}$$

$$T_{,11}+T_{,33}-(\dot{T}+t_0\ddot{T})=\epsilon[u_{1,1}+\dot{u}_{3,3}+\delta_{1k}t_0(\ddot{u}_{1,1}+\ddot{u}_{3,3})], \tag{7}$$

where comma notation is used for spatial derivatives and we have defined the quantities

$$\begin{aligned}
&x_i'=\omega^* x_i/c_1, \quad t'=\omega^* t, \quad u_i'=\rho\omega^* c_1 u_i/\beta T_0,\\
&T'=T/T_0, \quad t_1'=\omega^* t_1, \quad t_0'=\omega^* t_0,\\
&\omega^*=C_e(\lambda+2\mu)/K, \quad \epsilon=T_0\beta^2/\rho C_e(\lambda+2\mu),\\
&\delta^2=\mu/(\lambda+2\mu), \quad c_1^2=(\lambda+2\mu)/\rho, \quad c_2^2=\mu/\rho,
\end{aligned} \tag{8}$$

The primes have been suppressed for convenience. The nondimensional mechanical boundary conditions at $x_3=\pm d$ are given by

$$\begin{aligned}
&\sigma_{33}=u_{3,3}+(1-2\delta^2)u_{1,1}-(T+t_1\delta_{2k}\dot{T})=0\\
&\sigma_{13}=u_{1,3}+u_{3,1}=0, \quad \sigma_{23}=u_{2,3}=0
\end{aligned} \tag{9}$$

for stress-free boundary, and

$$u_1=u_2=u_3=0 \tag{10}$$

for rigidly fixed boundaries. The nondimensional thermal boundary conditions at $x_3=\pm d$ are given by

$$T_{,3}+hT=0, \tag{11}$$

where h is the surface heat transfer coefficient. Here, $h\to 0$ corresponds to thermally insulated boundaries and $h\to\infty$ refers to the isothermal one. We introduce the potential functions ϕ, ψ through the relations

$$u_1=\phi_{,1}+\psi_{,3}, \quad u_3=\phi_{,3}-\psi_{,1} \tag{12}$$

in Eqs. (4), (5), and (7), and assume the solutions of the form

$$(\phi,u_2,\psi,T)=[f(z),u_2(z),g(z),h(z)]\exp[i\xi(x_1-ct)], \tag{13}$$

where $c=\omega/\xi$ is the phase velocity, ω is frequency, and ξ is the wave number. Upon using solutions (13) in Eqs. (4) to (7) via (12) and solving the resulting differential equations, the expressions for ϕ, T, u_1, u_2, and u_3 are obtained as

$$\begin{aligned}
T=&\,i\tau_1^{-1}\omega^{-1}\{(\alpha^2-m_1^2)(Ar_1+Bc_1)\\
&+(\alpha^2-m_3^2)(Cr_3+Dc_3)\}\exp[i\xi(x_1-ct)],
\end{aligned} \tag{14}$$

$$\begin{aligned}
u_1=&\{i\xi[Ar_1+Bc_1+Cr_3+Dc_3]\\
&+\beta[Ec-Fr]\}\exp[i\xi(x_1-ct)],
\end{aligned} \tag{15}$$

$$u_2=[Gr+Hc]\exp[i\xi(x_1-ct)], \tag{16}$$

$$\begin{aligned}
u_3=&\{m_1[Ac_1-Br_1]+m_3[Cc_3-Dr_3]\\
&-i\xi[Er+Fc]\}\exp[i\xi(x_1-ct)],
\end{aligned} \tag{17}$$

where $r=\sin\beta x_3$, $c=\cos\beta x_3$, $r_i=\sin m_i x_3$, $c_i=\cos m_i x_3$, $i=1,3$,

$$\alpha^2=\xi^2(c^2-1), \quad \beta^2=\xi^2[(c^2/\delta^2)-1], \tag{18}$$

$$m_1^2=\xi^2(a_1^2c^2-1), \quad m_3^2=\xi^2(a_3^2c^2-1), \tag{19}$$

$$\begin{aligned}
a_1^2,a_3^2=&\tfrac{1}{2}\{(1+\tau_0-i\omega\epsilon\tau'\tau_1)\\
&\pm[(1-\tau_0-i\omega\epsilon\tau_0'\tau_1)^2-4i\omega\epsilon\tau_0'\tau_1]^{1/2}\},
\end{aligned} \tag{20}$$

$$\begin{aligned}
&\tau_0=t_0+i\omega^{-1}, \quad \tau_0'=t_0\delta_{1k}+i\omega^{-1},\\
&\tau_1=t_1\delta_{2k}+i\omega^{-1}.
\end{aligned} \tag{21}$$

II. SECULAR EQUATIONS

Invoking the boundary conditions (9)–(11) at the surfaces $x_3=\pm d$ of the plate and using Eqs. (14)–(17), we obtain a system of eight simultaneous linear equations in each case which has nontrivial solution if the determinant of the coefficients of amplitudes $[A,B,C,D,E,F,G,H]^T$ vanishes. This, after applying lengthy algebraic reductions and manipulations, leads to the secular equations for a plate with thermally insulated and isothermal boundaries, respectively, as follows:

$$\begin{aligned}
&\left[\frac{\tan m_1 d}{\tan\beta d}\right]^{\pm 1}-\frac{m_1(\alpha^2-m_1^2)}{m_3(\alpha^2-m_3^2)}\left[\frac{\tan m_3 d}{\tan\beta d}\right]^{\pm 1}\\
&=\begin{cases}\dfrac{4\xi^2\beta m_1(m_3^2-m_1^2)}{(\xi^2-\beta^2)^2(\alpha^2-m_3^2)}, & \text{for stress free,}\\[2ex] \dfrac{\beta m_1(m_3^2-m_1^2)}{\xi^2(\alpha^2-m_3^2)}, & \text{for rigidly fixed.}\end{cases}
\end{aligned} \tag{22}$$

$$\begin{aligned}
&\left[\frac{\tan m_1 d}{\tan\beta d}\right]^{\pm 1}-\frac{m_3(\alpha^2-m_1^2)}{m_1(\alpha^2-m_3^2)}\left[\frac{\tan m_3 d}{\tan\beta d}\right]^{\pm 1}\\
&=\begin{cases}\dfrac{(\xi^2-\beta^2)^2(m_3^2-m_1^2)}{4\xi^2\beta m_1(\alpha^2-m_3^2)}, & \text{for stress free,}\\[2ex] \dfrac{\xi^2(m_3^2-m_1^2)}{\beta m_1(\alpha^2-m_3^2)}, & \text{for rigidly fixed.}\end{cases}
\end{aligned} \tag{23}$$

and

$$\sin(2\beta d)=0, \tag{24}$$

for decoupled SH modes of wave propagation in the plate.

Here, the superscript $+1$ corresponds to skew symmetric and -1 refers to symmetric modes. Equations (22), (23), and (24) are the secular equations for the propagation of modified guided thermoelastic waves in the plate. We refer to such

waves as thermoelastic plate waves rather than Lamb waves, whose properties were derived by Lamb in 1917 for isotropic solids in elastokinetics.

III. COUPLED AND UNCOUPLED THERMOELASTICITIES

In the case of the coupled theory of thermoelasticity, the thermal relaxation times vanish, i.e., $t_0=0=t_1$ so that $\tau_0=\tau_0'=\tau_1=i\omega^{-1}$, and consequently, we have

$$a_1^2,a_3^2 = \frac{[1+i\omega^{-1}(1+\epsilon)\pm\{[1-i\omega^{-1}(1-\epsilon)]^2+4\epsilon\omega^{-1}\}^{1/2}]}{2}. \tag{25}$$

The wave propagation in the plate is governed by the secular equations (22) to (24) with reduced/changed values of the characteristic roots m_i, $i=1,3$ in these equations. In the case of uncoupled thermoelasticity, the coupling constant $\epsilon=0$, which leads to $a_1^2=1$, $a_3^2=\tau_0$ so that $m_1^2=\alpha^2$, $m_3^2=\xi^2(\tau_0 c^2-1)$. Consequently, the secular equations (22) and (23) reduce to

$$\frac{\tan\alpha d}{\tan\beta d}=\begin{cases} -\left[\dfrac{4\xi^2\alpha\beta}{(\xi^2-\beta^2)^2}\right]^{\pm 1}, & \text{for stress free,} \\ -\left[\dfrac{\alpha\beta}{\xi^2}\right]^{\pm 1}, & \text{for rigidly fixed.} \end{cases} \tag{26}$$

The transcendental equation (26) is the Rayleigh Lamb frequency equation and has already been discussed in detail by Graff[16] for various possibilities and situations. The frequency equation for plate waves with rigidly fixed boundaries can be discussed on similar lines as Eq. (26).

Using the definition of β in Eq. (24), we get

$$c=\pm\delta\sqrt{\frac{n^2\pi^2}{4\xi^2 d^2}+1}, \quad n=0,1,2,3,\ldots. \tag{27}$$

This gives the phase velocities of the decoupled SH waves. These modes of propagation are independent of the thermal variations and nature of the boundaries and hence can be similarly discussed as was done in Graff.[16]

IV. WAVES OF SHORT WAVELENGTH

Some information on the asymptotic behavior is obtainable by setting $\xi\to\infty$. If we take $\xi>\omega/\delta$, it follows that $\xi>\omega$ and that $c<\delta$, 1; we then replace α, β, m_1, and m_3 in the frequency equations by $i\alpha'$, $i\beta'$, $i\alpha_1$. For $\xi\to\infty$, $\tanh(\alpha_1' d)/\tanh(\beta' d)\to 1$, $i=1,3$ so that the resulting frequency equations (22) and (23) for both symmetric and skew-symmetric cases reduce to

$$(2-c^2/\delta^2)^2[\alpha_1^2+\alpha_1\alpha_3+\alpha_3^2-1+c^2]=4\beta'\alpha_1\alpha_3(\alpha_1+\alpha_3), \tag{28}$$

and

$$(2-c^2/\delta^2)^2(\alpha_1+\alpha_3)=4\beta'(\alpha_1\alpha_3+1-c^2), \tag{29}$$

respectively. These are merely Rayleigh surface wave equations.[2,3,7] The Rayleigh results enter here since, for such small wavelengths, the finite-thickness plate appears as a semi-infinite medium. Hence, vibrational energy is transmitted mainly along the surface of the plate.

V. THERMOELASTICITY WITHOUT ENERGY DISSIPATION

The nondimensional basic governing equations of thermoelasticity without energy dissipation[8] in the absence of heat sources and body forces for a homogeneous isotropic plate are given by[17]

$$u_{1,11}+(1-\delta_1^2)u_{3,13}+\delta_1^2 u_{1,33}-T_{,1}=(1/c_p^2)\,\ddot{u}_1, \tag{30}$$

$$(1-\delta_1^2)u_{1,33}+\delta_1^2 u_{3,11}+u_{3,33}-T_{,3}=(1/c_p^2)\,\ddot{u}_3, \tag{31}$$

$$c_s^2(u_{2,11}+u_{2,33})=\ddot{u}_2, \tag{32}$$

$$c_T^2(T_{,11}+T_{,33})-\dot{T}=\epsilon^*(\ddot{u}_{1,1}+\ddot{u}_{3,3}). \tag{33}$$

Here,

$$\epsilon^*=\gamma^2 T_0/\rho C_e(\lambda+2\mu), \quad \delta_1^2=c_s^2/c_p^2,$$
$$\gamma=(3\lambda+2\mu)\beta^*, \quad c_p^2=(\lambda+2\mu)/\rho v^2, \tag{34}$$
$$c_s^2=\mu/\rho v^2, \quad c_T^2=K^*/\rho C_e v^2,$$

where v is the standard speed, β^* the coefficient of volume expansion, and K^* is the material constant characteristic of the theory. Following the procedure of the previous sections, the secular equations for stress-free and rigidly fixed boundaries are again given by Eqs. (22) and (23), respectively. But, here the roots m_1, m_2, and α,β are obtained as

$$\alpha^2=\xi(c^2/c_p^2-1), \quad \beta^2=\xi^2(c^2/c_s^2-1), \tag{35}$$

$$m_1^2=\xi(p_1^2c^2-1), \quad m_3^2=\xi^2(p_2^2c^2-1), \tag{36}$$

where

$$p_1^2,p_2^2=\frac{1}{2}\left\{\frac{1}{c_p^2}+\frac{\epsilon^*}{2c_T^2}\pm\frac{1}{c_T}\left[\epsilon^*\left(\frac{1}{c_p^2}+\frac{\epsilon^*}{4c_T^2}\right)\right]^{1/2}\right\}. \tag{37}$$

For the uncoupled theory of thermoelasticity, $\epsilon^*=0$ and we have $m_1^2=\alpha^2=m_2^2$. We again arrive at Eqs. (26) in the case of stress-free and rigidly fixed boundaries of the plate. Equation (32) leads to the dispersion relation of decoupled shear wave modes of the type (24), where phase velocities in this case are given by

$$c=\pm c_s\sqrt{n^2\pi^2/4\xi^2 d^2+1}, \quad n=0,1,2,3\ldots.$$

VI. NUMERICAL RESULTS AND DISCUSSION

With the view of illustrating the theoretical results obtained in the preceding sections and comparing these in the context of various theories of thermoelasticity, we now present some numerical results. The material chosen for this purpose is aluminum–epoxy composite, the physical data for which is given as[18] $\epsilon=0.073$, $\lambda=7.59\times10^{10}$ Nm^{-2}, $\mu=1.89\times10^{10}$ Nm^{-2}, $\rho=2.19\times10^3$ kg m^{-3}, $K=2.508$ J/ms °C, $C_e=961.4$ J kg^{-1}/°C, $T_0=23$ °C, $t_0=6.131\times10^{-13}$ s, $t_1=8.765\times10^{-13}$ s, $d=1.0$.

In the case of GN theory, K^* is not the thermal conductivity but a material constant characteristic of the theory. In

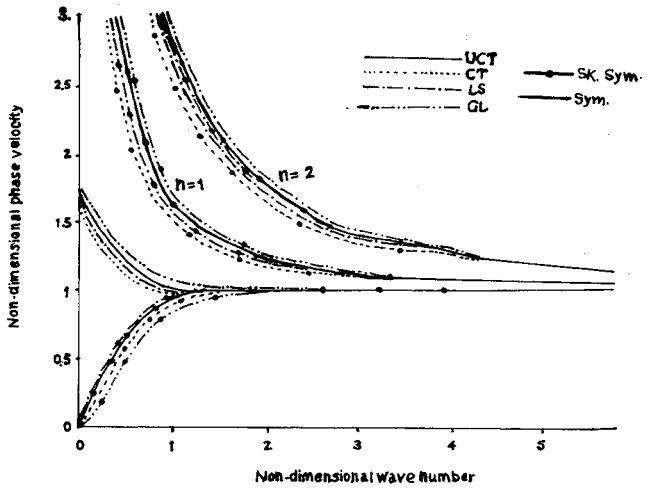


FIG. 2. Dispersion curves for skew symmetric and symmetric modes.

this theory the thermal wave speed (C_T) depends upon K^* instead of thermal relaxation times as in LS and GL theories of thermoelasticity. To analyze the theoretical results in this theory, we have chosen hypothetical values[17] of material parameters as $c_p=1.0$, $c_T=0.5$ so that $v^2=(\lambda+2\mu)/\rho=c_1^2$, $c_s^2=\mu/(\lambda+2\mu)=\delta^2$, and $K^*=[C_e(\lambda+2\mu)]/4$, for aluminum–epoxy composite-like material. We have taken $\epsilon^*=0.073=\epsilon$ here. The phase velocities of symmetric and skew-symmetric modes of wave propagation have been computed for various values of wave number (ξ) from the dispersion relations (22) for stress-free insulated boundary conditions. The corresponding classical and modified dispersion curves for Rayleigh–Lamb-type modes are presented in Figs. 2 and 3 in the context of various theories of thermoelasticity. The phase velocity of the lowest skew-symmetric mode is observed to increase from zero value at vanishing wave number to become closer to the Rayleigh wave velocity at higher values of the wave number, where as in the case of the lowest symmetric mode it decreases from a value greater than

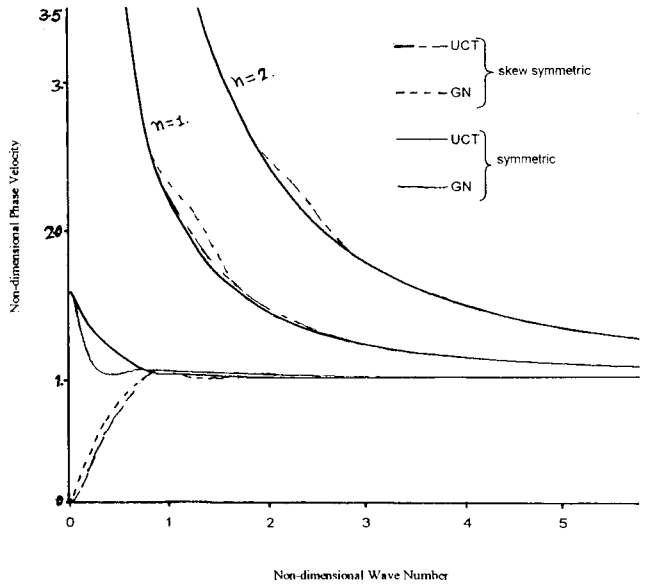


FIG. 3. Dispersion curves for skew symmetric and symmetric modes in UCT and GN theories of thermoelasticity.

unity towards the Rayleigh velocity asymptotically with an increase in wave number. The phase velocities of higher modes of propagation, symmetric and skew-symmetric, attain quite large values at vanishing wave number, which sharply slash down to become steady with increasing values of the wave number. The magnitude of velocities of higher symmetric and skew-symmetric modes is observed to develop at a rate which is approximately n times the magnitude of the velocity of the first mode ($n=1$). These numerically computed results are found to be quite in agreement with the corresponding analytical result and their trends are similar to those reported by Graff[16] in the case of homogeneous isotropic elastic plate, except for the modification due to the thermo-mechanical coupling and thermal relaxation. The wave propagation modes are also observed to behave alike for plates with stress-free isothermal boundaries.

The computations for a plate having rigid insulated boundaries show that the lowest skew-symmetric modes have nonzero velocity and the lowest symmetric modes have zero velocity at vanishing wave numbers. But, the phase velocities of these modes also become asymptotically close to the corresponding surface wave velocity with increasing value of the wave number. The behavior of higher modes of propagation is observed to be similar to other cases. The modes of propagation also behave alike in a plate with rigidly fixed isothermal boundaries. The effect of thermal relaxation time is observed to be negligibly small.

[1] W. Nowacki, *Dynamic Problems of Thermoelasticity* (Noordhoff, Leyden, The Netherlands, 1975).
[2] P. Chadwick and D. W. Windle, ''Propagation of Rayleigh waves along isothermal and insulated boundaries,'' Proc. R. Soc. Am. **280**, 47 (1964).
[3] R. J. Atkin and P. Chadwick, ''Surface waves in heat conduction elastic body correction and extension of a paper of Chadwick and Windle,'' J. Therm. Stresses, **4**, 509 (1981).
[4] H. W. Lord and Y. Shulman, ''Generalized dynamical theory of thermoelasticity,'' J. Mech. Phys. Solids **15**, 229–309 (1967).
[5] A. E. Green and K. A. Lindsay, ''Thermoelasticity,'' J. Elast. **2**, 1–7 (1972).
[6] R. S. Dhaliwal and H. H. Sherief, ''Generalized thermoelasticity for anisotropic media,'' Quarterly Appl. Math. **38**, 1–8 (1980).
[7] A. H. Nayfeh and S. N. Nasser, ''Thermoelastic waves in solids with and thermal relaxation,'' Acta Mech. **12**, 53–69 (1971).
[8] A. E. Green and P. M. Nagdhi, ''On thermoelasticity without energy dissipation,'' J. Elast. **31**, 189–208 (1993).
[9] V. Kozlov, ''Thermoelastic vibrations of rectangular plates,'' Prikl. Mekh. **8**, 123–127 (1972).
[10] C. Massalas, ''Symmetric non-linear response of a circular plate due to heat flux at $z=\pm h/2$,'' Rev. Roum. Sci. Tech., Ser. Mec. Appl. **28**, 643–650 (1983).
[11] A. B. Kumar, ''Thermoelastic waves from a suddenly punched hole in stretched plate,'' Indian J. Pure Appl. Math. **20**, 181 (1989).
[12] D. K. Banerjee and Y. H. Sherief, ''Thermoelastic waves in anisotropic solids,'' J. Acoust. Soc. Am. **56**, 1444–1453 (1974).
[13] J. N. Sharma, ''On the low and high frequency behaviour of generalized thermoelastic waves,'' Arch. Mech. **38**, 665–673 (1986).
[14] D. S. Chandrasekhariah, ''Thermoelasticity with second sound—A review, Appl. Mech. Rev.'' Appl. Mech. Rev. **39**, 355–376 (1986).
[15] A. E. Green, ''A note on linear thermoelasticity,'' Mathematika **19**, 69–75 (1972).
[16] K. F. Graff, *Wave Motion in Elastic Solids* (Dover, New York, 1991).
[17] D. S. Chandrasekharaiah, ''One-dimensional wave propagation in the linear theory of thermoelasticity energy dissipation,'' J. Therm. Stresses **19**, 695–710 (1996).
[18] R. Kumar and B. Singh, ''Reflection of plane waves from the flat boundary of micropolar generalised thermoelastic half space,'' Int. J. Eng. Sci. **36**, 865–890 (1998).